Volume 4 Issue 4, October 2024

Association for Computing Machinery

Advancing Computing as a Science & Profession

EIR

Newsletter

Energy Informatics Review

Sponsored by:

ACM SIGENERGY

Part of:

DACH+ Conference on Energy Informatics Special Issue

SIGEnergy

Energy Informatics Review

a publication of the ACM Special Interest Group on Energy Informatics

Contents

Preface **1**
Michael Moser (*Swiss Federal Office of Energy*)

Modeling protection systems in smart active distribution grids for a performance analysis **2**
Amit Dilip Patil (*University of Passau*); Hermann de Meer (*University of Passau*); Poul Heegaard (*NTNU Trondheim*); Bjarne Helvik (*NTNU Trondheim*)

Don't Touch the Power Line - A Proof-of-Concept for Aligned LLM-based Assistance Systems to Support the Maintenance in the Electricity Distribution System **17**
Sascha Kaltenpoth (*Paderborn University*); Oliver Müller (*Paderborn University*)

Practical Application of Energy-related Use Case Methodology and SGAM Framework: A Study of User Perspectives and Challenges **24**
René Kuchenbuch (*OFFIS e.V.*); Pranit Nale (*OFFIS e.V.*); Laura Niemann (*OFFIS e.V.*); Oliver Werth (*OFFIS e.V.*); Jürgen Sauer (*University of Oldenburg*)

Choosing the right Ontology to describe Research Data in the Energy Domain **34**
Alexandro Steinert (*OFFIS e.V., Carl von Ossietzky Universität of Oldenburg*); Stephan Ferenz (*OFFIS e.V., Carl von Ossietzky Universität of Oldenburg*); Astrid Nieße (*OFFIS e.V., Carl von Ossietzky Universität of Oldenburg*)

Utilizing autonomous airport parking facilities as virtual power plants to provide frequency containment reserve **50**
Dominik Vereno (*Josef Ressel Centre for Dependable System-of-Systems Engineering*); René Schütz (*Josef Ressel Centre for Dependable System-of-Systems Engineering*); Katharina Polanec (*Josef Ressel Centre for Dependable System-of-Systems Engineering*); Francesco Maldonato (*CARIAD SE / Technical University of Berlin*); Florian Mielke-Sulz (*Vienna University of Technology*); Damien Kögler (*Robert Bosch GmbH*); Rolf Nicodemus (*Robert Bosch GmbH*); Johanna Weber (*Robert Bosch GmbH*); Jounes-Alexander Gross (*Josef Ressel Centre for Dependable System-of-Systems Engineering*); Christian Neureiter (*Josef Ressel Centre for Dependable System-of-Systems Engineering*); Amin Khodaei (*University of Denver*); Sebastian Lehnhoff (*OFFIS - Institut für Informatik*)

How Satisfactory Can Deep Reinforcement Learning Methods Simulate Electricity Market Dynamics? Bechmarking via Bi-level Optimization **66**
Nick Harder (*Albert Ludwigs University of Freiburg*); Lesia Mitridati (*Technical University of Denmark*); Farzaneh Pourahmadi (*Technical University of Denmark*); Anke Weidlich (*University of Freiburg*); Jalal Kazempour (*Technical University of Denmark*)

Power-dependent price profiles - defining grid- and market-oriented incentives for building energy management systems **80**
Tobias Riedel (*FZI Research Center for Information Technology*); Carl Hauschke (*Research Center for Information Technology*); Hartmut Schmeck (*Karlsruhe Institute of Technology*)

Monitoring Germany's Core Energy System Dataset: A Data Quality Analysis of the Marktstammdatenregister **90**
Florian Kotthoff (*OFFIS Institute for Information Technology*); Christoph Muschner (*Reiner Lemoine Institut gGmbH*); Deniz Tepe (*fortiss, Research Institute of the Free State of Bavaria*); Esther Vogt (*Institute*

for Enterprise Systems, University of Mannheim); Ludwig Hülk (*Reiner Lemoine Institut gGmbH*); Florian Kotthoff (*OFFIS Institute for Information Technology*)

Decentralized Privacy-preserving Power Flow Forecast on Smart Meters **102**
Severin Nowak (*Lucerne Univeristy of Applied Sciences and Arts*); Fabian Widmer (*Lucerne Univeristy of Applied Sciences and Arts*); Jeremy Taylor (*Via Science*); Antonios Papaemmanouil (*Lucerne Univeristy of Applied Sciences and Arts*); Daniel Raimundo (*Lucerne Univeristy of Applied Sciences and Arts*)

The plan4.energy Approach for Planning Support for Positive Energy Districts **111**
Yannick Wimmer (*AIT Austrian Institute of Technology*); Daniel Schwabeneder (*AIT Austrian Institute of Technology*); Michael Niederkofler (*Innovation Lab act4.energy*); Patrizia-Ilda Valentine (*Mobilize Financial Services*)

Enabling Moral Agency in Distributed Energy Management: An Ethics Score for Negotiations in Multi-Agent Systems **118**
Malte Stomberg (*OFFIS e. V.*); Martin Tröschel (*OFFIS e. V.*)

The impact of redispatch on grid and storage expansion planning in the German energy system **131**
Clara Büttner (*Flensburg University of Applied Sciences*); Ulf Müller (*Flensburg University of Applied Sciences*)

Towards a State Explanation Framework in Cyber-Physical Systems **144**
Katrin Schreiberhuber (*Wirtschaftsuniversität Wien*); Fajar J. Ekaputra (*Wirtschaftsuniversität Wien*); Marta Sabou (*Wirtschaftsuniversität Wien*); Daniel Hauer (*Siemens AG Österreich*); Konrad Diwold (*Siemens AG Österreich*); Thomas Frühwirth (*TU Wien*); Gernot Steindl (*TU Wien*); Tobias Schwarzinger (*TU Wien*)

Analysing and Predicting Extreme Frequency Deviations: A Case Study in the Balearic Power Grid **157**
Zakariea Sharfeddine (*Karlsruhe Institute of Technology*); Sebastian Pütz (*Karlsruhe Institute of Technology*); Vedang Tamhane (*Karlsruhe Institute of Technology*); Veit Hagenmeyer (*Karlsruhe Institute of Technology*); Benjamin Schäfer (*Karlsruhe Institute of Technology*)

An Algorithm for Modelling Rolling Intrinsic Battery Trading on the Continuous Intraday Market **165**
Leo Semmelmann (*Karlsruher Institut für Technologie (KIT)*); Jannik Dresselhaus (*Karlsruher Institut für Technologie (KIT)*); Kim K. Miskiw (*Karlsruher Institut für Technologie (KIT)*); Jan Ludwig (*Karlsruher Institut für Technologie (KIT)*); Christof Weinhardt (*Karlsruher Institut für Technologie (KIT)*)

Probabilistic energy forecasting through quantile regression in reproducing kernel Hilbert spaces **177**
Luca Pernigo (*Università della Svizzera italiana*); Rohan Sen (*Università della Svizzera italiana*); Davide Baroli (*Università della Svizzera italiana*); Davide Baroli (*Università della Svizzera italiana*)

Forecasting Renewable Energy at European Markets **189**
Hun Rim (*Università della Svizzera italiana*); Juraj Kardos (*Università della Svizzera italiana*); Olaf Schenk (*Università della Svizzera italiana*)

A Flexible and Holistic Multi-Agent Framework for Local Energy Aggregation Approaches **192**

Michael Bettermann (*University of Passau*); Hermann de Meer (*University of Passau*)

Field Survey of Wireless M-Bus Encryption for Energy Metering Applications in Residential Buildings **195**
Friedrich Hiller v. Gaertringen (*Karlsruhe Institute of Technology*); Johannes Galenzowski (*Karlsruhe Institute of Technology*); Kaibin Bao (*Karlsruhe Institute of Technology*); Simon Waczowicz (*Karlsruhe Institute of Technology*); Veit Hagenmeyer (*Karlsruhe Institute of Technology*)

A Distributed Game Theoretic Approach for Optimal Battery Use in an Energy Community **209**
Nicolas Kirsch (*École Polytechnique Fédérale de Lausanne*); Giulio Salizzoni (*École Polytechnique Fédérale de Lausanne*); Tomasz Gorecki (*Sustainable Energy Center, CSEM*); Maryam Kamgarpour (*École Polytechnique Fédérale de Lausanne*)

Data-Driven Identification and Operational Optimization of Energy-Flexible Thermal Supply Systems **216**
Jan Zangenberg (*Institute of Production Management, Technology and Machine Tools Technical University of Darmstadt*); Jonas Wendt (*Institute of Production Management, Technology and Machine Tools Technical University of Darmstadt*); Tobias Koch (*Institute of Production Management, Technology and Machine Tools Technical University of Darmstadt*); Tobias Kapser (*Technical University of Darmstadt*); Matthias Weigold (*Institute of Production Management, Technology and Machine Tools Technical University of Darmstadt*)

A Platform Ecosystem Providing New Data For The Energy Transition **228**
Markus Duchon (*fortiss GmbH*); Patrick Buchenberg (*Technische Universität München*); Jessy Matar (*fortiss GmbH*); Mahsa Faraji Shoyari (*fortiss GmbH*); Ingmar Kessler (*fortiss GmbH*); Alexander Perzylo (*fortiss GmbH*); Philipp Kuhn (*Technische Universität München*); Thomas Hamacher (*Technische Universität München*); Ghayathri Suriyamoorthy (*Karlsruhe Institute of Technology*); Thorsten Schlachter (*Karlsruhe Institute of Technology*); Wolfgang Süß (*Karlsruhe Institute of Technology*); Nguyen Xuan Thinh (*Technische Universität Dortmund*); Haniyeh Ebrahimi Salari (*Technische Universität Dortmund*); Jasmin Latko (*Technische Universität Dortmund*); Minsheng Xu (*RWTH Aachen University*); Maxim Shamovich (*RWTH Aachen University*); Dominik Schlütter (*RWTH Aachen University*); Jérôme Frisch (*RWTH Aachen University*); Markus Kraft (*Computational Modelling Pirmasens GmbH*); Kushagar Rustagi (*Computational Modelling Pirmasens GmbH*); Carolin Ayasse (*Technische Universität Darmstadt*); Florian Steinke (*Technische Universität Darmstadt*); Michael Metzger (*Siemens AG, Technology*); Laura Kuper (*Siemens AG, Technology*)

Empowering Energy Communities and P2P Energy Sharing: A Novel End-to-End Ecosystem for Planning, Deployment, and Operation **240**
Shievam Kashyap (*University of Applied Sciences Upper Austria*); Christoph Schaffer (*University of Applied Sciences Upper Austria*); Tobias Fischer (*University of Applied Sciences Upper Austria*); Michael Zauner (*University of Applied Sciences Upper Austria*); Franz Fischer (*University of Applied Sciences Upper Austria*); Marc Kurz (*University of Applied Sciences Upper Austria*); Georg Hartner (*University of Applied Sciences Upper Austria*); Stefan Grünberger (*University of Applied Sciences Upper Austria*); Oliver Hödl (*University of Applied Sciences Upper Austria*)

Poster Abstract: A Digital Twin Platform Applied to Hydrogen Electrolyzers **247**
Amit Kumar Singh (*OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany*); Jelke Wibbeke (*OFFIS – Institute for Information Technology, Energy Division, Escherweg 2,*

Oldenburg, Germany); Amin Raeiszadeh (*OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany*); Nils Huxoll (*OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany*); Michael Brand (*OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany*)

Poster Abstract: Peer-to-Peer Communication Using Enhanced German Smart Meter Gateway Infrastructure . . . 251
Eike Niehs (*TU Braunschweig*); Julien Essers (*TU Braunschweig*); Johannes Rothert (*TU Braunschweig*); Bernd Engel (*TU Braunschweig*)

Poster Abstract: Forecasting and Optimization as a Service for Energy Management Applications with the Energy Service Generics Framework . . . 254
David Wölfle (*FZI Research Center for Information Technology*); Hartmut Schmeck (*Karlsruhe Institute of Technology*)

Poster Abstract: Interactive Simulation of Swiss Energy and Environment Designed for Gaming and Choice Experiments . . . 258
Toby Simpson (*Università della Svizzera italiana*); Michael Multerer (*Università della Svizzera italiana*); Rolf Krause (*Università della Svizzera italiana*)

Poster Abstract: Controllability Architecture for Energy Devices . . . 261
Alexander Hill (*OFFIS e.V.*); Jan-Henrik Bruhn (*OFFIS e.V.*); Christian Pieper (*OFFIS e.V.*)

Poster Abstract: Hardware-in-the-Loop Simulation Environment for Validating Distribution System Control Applications . . . 264
Marcel Otte (*OFFIS – Institute for Information Technology*); Carsten Krüger (*OFFIS – Institute for Information Technology*); Pratyush Das (*OFFIS - Institute for Information Technology*); Sebastian Rohjans (*Jade University of Applied Sciences & OFFIS - Institute for Information Technology*); Sebastian Lehnhoff (*OFFIS – Institute for Information Technology*)

Poster Abstract: Optimizing the feature space for power system modeling . . . 267
Aleksandr Berezin (*OFFIS – e.V., Germany*)

Poster Abstract: learning techno-economic parameters of electricity generation units from their operational data . . . 269
Chiara Fusar Bassini (*Hertie School, Germany*)

Poster Abstract: Social Inclusion by Design - Policies for Bridging Electricity Prosumerism to Low-Income Households . . . 272
Christina Speck (*Karlsruhe Institute of Technology, Germany*)

Poster Abstract: Optimizing the provision of distributed energy resources operational load flexibility on local market platforms for grid stability . . . 275
Daniel Bull (*Karlsruhe University of Applied Sciences, Institute of Refrigeration, Air-Conditioning and Environmental Engineering (IKKU), Germany and University of Freiburg, Department of Sustainable Systems Engineering (INATECH), Germany*)

Poster Abstract: Novel Privacy Attacks on Load Data: Understanding their Impact and Mitigation Strategies . . . 278

Dejan Radovanovic (*Center for Secure Energy Informatics at the Salzburg University of Applied Sciences, Austria*)

Poster Abstract: Load Forecasting in Energy Communities for Flexibility Services Provision **281**
Dimitrios Papadopoulos (*INATECH-University of Freiburg, Germany*)

Poster Abstract: Towards Online Meta-Modeling of Communication Networks in Energy Systems **284**
Malin Radtke (*University of Oldenburg, Germany*)

Poster Abstract: Coordinating the Heterogeneity of Aggregators in Digitalised Power Systems **287**
Marcel Otte (*OFFIS - Institute for Information Technology, Germany*)

Poster Abstract: Procedural Generation of Communication Networks in Power Systems **290**
Xavier Weiss, Lars Nordström, Patrik Hilber, and Emre Süren (*KTH, Sweden*)

Preface

Dear Reader,

This Special Issue of ACM Energy Informatics Review is a collection of works presented at the 13th DACH+ Conference on Energy Informatics, Lugano, Switzerland, October 9-11, 2024. As Swiss Federal Office of Energy, we are delighted to have been supporting this year's conference.

The objective of the DACH+ conference series on Energy Informatics is to promote the research, development, and implementation of information and communication technologies in the energy domain and to foster the exchange between academia, industry, and service providers in the German-Austrian-Swiss region and its neighbouring countries (DACH+). With its focus on this region, the conference is supported by the ACM Special Interest Group ACM SIGEnergy since 2023.

In this special issue you will find original work and poster abstracts from the conference, as well as abstracts from the Doctoral Workshop.

The DACH+ Energy Informatics conference series is a joint initiative of the German Federal Ministry for Economic Affairs and Climate Action, the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology, and the Swiss Federal Office of Energy. Since the beginning of the conference series, this important community meeting has alternated between the participating countries - each time with the support of the respective ministry.

The transformation of energy systems towards a sustainable energy supply cannot be achieved in time when seen as an isolated national concern. Research shows that the exchange of current findings, the discussion on new technologies and the critical examination of relevant developments is most effective as a joint effort. The DACH+ Energy Informatics conference series has been demonstrating this in an outstanding way for many years now.

We are happy to be one of the three main supporters for the Energy Informatics community, fostering the transformation of energy systems as digitalized systems, and to host this year's conference in Switzerland once again.

Michael Moser
Swiss Federal Office of Energy

Modeling protection systems in smart active distribution grids for a performance analysis

AMIT DILIP PATIL and HERMANN DE MEER, University of Passau, Germany
POUL EINAR HEEGAARD and BJARNE EMIL HELVIK, Norwegian University of Science and Technology, Norway

Distributed renewable energy sources increase the fault level in distribution grids, impacting protection systems and potentially causing misoperations like blinding, which increases the fault clearing time. High impedance faults result in low fault currents, further delaying protection tripping. Communication-based adaptive protection systems can improve protection sensitivity but are prone to delays, affecting fault clearing times. Therefore, this research addresses how the impact of distributed renewable energy sources, fault impedance and communication delay on protection systems can be quantified in terms of operation time. A discrete event simulation model is proposed to study protection performance in active distribution grids. Time distribution assumptions are investigated to capture the circuit breaker trip time accurately. The relative fault level of distribution and external grid is formalized. The impact of the relative fault level and communication delay on the protection system is determined by measuring fault clearing times using discrete event simulation. This impact is further quantified using a new protection performance index. Results show that for the system studied, protection blinding is critical for low impedance faults in grids with high fault levels, while high impedance faults are critical in grids with low fault levels. Moreover, sympathetic tripping is seen at increased distribution grid fault levels and fault impedance. Furthermore, while communication systems reduce fault clearing times, increased delays harm protection systems.

CCS Concepts: • **Networks → Network performance modeling**; **Network performance analysis**; **Network simulations**; • **Computing methodologies → Modeling and simulation**.

Additional Key Words and Phrases: Adaptive protection, communication, distributed generation, discrete event simulation

1 INTRODUCTION

Traditionally, the external grid has provided fault current in distribution grids. Consequently, protection relays have had fixed *sensitivity*, i.e., the ability of the protection system to detect faults, configured to operate without the possibility to modify settings. Integrating Distributed Renewable Energy Sources (DRESs) using power electronics-based converters into distribution grids decentralizes the fault current source and could result in bidirectional power flow. Therefore, the assumptions of *traditional* protection systems may not hold in modern systems, as these DRESs are also expected to supply fault current. Prominent issues that arise due to DRESs are that of *protection blinding*, e.g., a DRES in the same feeder as a fault supplies fault current, reducing current through an upstream Circuit Breaker (CB) resulting in increased Fault Clearing Times (FCTs) and *sympathetic tripping*, e.g., when a relay in fault-free feeder operates due to fault current from a DRES in the same feeder [15]. Another issue in distribution grids is that of *High Impedance Faults* (HIFs). HIFs also increase FCTs due to the low fault current drawn by them [35]. As faults pose a system safety and stability risk, these faults should be isolated as soon as possible, i.e., in the order of a hundred milliseconds. Higher FCTs could lead to loss of synchronicity.

Information and Communication Technology (ICT) may be used to *adapt* the *sensitivity*, i.e., the protection system's ability to detect fault currents, e.g., communicating new settings to a relay in real-time to sense low fault currents [25, 35]. Hence, an online *adaptive protection* system is preferable to a manual on-site reconfiguration of relays [17, 20, 28]. However, the ICT system is susceptible to variable communication delays. This could cause delayed adaptation of relay sensitivity, resulting in long FCTs. Long FCTs could impact the stability of the power system since the fault and the resulting fault current persist during this time. This could lead to line overloading, equipment damage and even cascading failures. Therefore, structural changes due to DRESs and the integration of ICT into power systems impact the performance of the protection system. Hence, the impact of delay, DRES and fault impedance on protection systems is an important problem. However, these impacts have been studied in isolation [10, 14, 15]. Therefore, a new method is required to assess the protection system performance considering these changes together in ICT and power systems.

The main objective of this research is *to develop a model that quantifies the impact of communication delay, fault impedance and DRESs on adaptive protection systems.* This model can determine the protection system performance for various system design and planning scenarios. The impact of communication delay and DRESs on an adaptive protection system, especially during protection misoperation, is investigated using Discrete Event Simulation (DES). DES is chosen to assess the impact of communication delay, DRES fault current contribution and fault impedance on FCTs as DES can capture the time elapsed after events, e.g., relay operation time. The system studied is a radial distribution grid with a traditional unidirectional overcurrent protection system and time-dependent relays. The fault impedance relative to bus impedance and DRES fault current contribution relative to the external grid fault current contribution is formalized. This formalization is considered in the analysis of the impact of communication delay. The analysis determines the conditions under which protection misoperation occurs. These conditions are essential to understand the risk the protection system carries in distribution grids as DRESs are integrated. The results indicate how this risk is exacerbated by communication delay for various DRES integration levels for different fault types. This risk is quantified in a new protection performance index. To model the protection operation accurately, various time distributions are investigated. Thus, the proposed model allows an assessment of adaptive protection systems while considering ICT and power system interdependencies. The main contributions of this research are as follows:

Authors' addresses: Amit Dilip Patil, amit.patil@uni-passau.de; Hermann de Meer, demeer@uni-passau.de, University of Passau, Innstrasse 41, Passau, Bavaria, Germany, 94032; Poul Einar Heegaard, poul.heegaard@ntnu.no; Bjarne Emil Helvik, bjarne.helvik@ntnu.no, Norwegian University of Science and Technology, Høgskoleringen 1, Trondheim, Norway.

- A DES model to assess the impact of communication delay, fault impedance and DRESs on protection systems,
- Determination of DRES penetration level and fault impedance beyond which protection misoperation occurs using the proposed DES model,
- Formulation of a new protection performance index that quantifies protection system performance for a range of communication delays and fault levels,
- Investigation of the model's sensitivity to time distribution assumptions for protection operation modeling in DES.

The rest of the paper is structured as follows: Section 2 presents the relevant research in this domain while highlighting the research gaps and stating the research question. Section 3 describes the system considered and its assumptions. Section 4 presents the formal model and its various elements. Section 5 describes the simulation setup and the practical implementation. Section 6 presents the evaluation and results obtained from simulations using the model. Section 8 concludes the research while highlighting future work.

2 RELATED WORK

The topic of ICT-dependent power systems is subject to continuous research. In this section, the relevant research is presented. Table 1 compares this research with related research regarding system properties considered in modeling *and* assessing protection systems. A ✓ mark indicates the parameter mentioned in the column is considered in the research mentioned in the row, while a blank cell indicates the parameter is not considered.

Impact of DRESs and fault impedance on traditional protection systems. Protection system evaluation requires a suitable power system fault and protection system model. Faults are typically modeled as an impedance between two points, e.g., a phase and the ground for single-phase faults [4, 32]. This research adopts this fault modeling approach. Traditional protection systems have had pre-defined settings typically defined for scenarios with high fault current, i.e., Low Impedance Faults (LIFs) with no possibility of adaptation. In this case, faults such as HIFs may not be detected and cleared. Furthermore, due to the integration of DRESs, a new set of challenges emerges, e.g., protection blinding [10, 15]. The issue of protection misoperation due to DRESs has been tackled using offline approaches [15, 16]. However, the aforementioned approaches do not determine when such misoperations occur, i.e., in terms of grid fault level and fault impedance. In contrast, fault impedance and fault level are considered in this research. Majambere et al. [23] and Dubey et al. [13] considered the impact of impedance on differential and distance protection but did not investigate the impact of communication delay or DRESs. *This research explicitly considers fault impedance, DRESs and communication delay.*

Adaptive protection systems and impact of ICT performance. *Adaptive* protection is a potential solution to challenges such as misoperation, e.g., blinding. Adaptive protection systems have been studied rigorously in research [8]. Jain et al. studied the benefits of adaptive protection using ICT, especially when the grid operation mode is considered [20]. Farraj et al. studied the impact of communication delay on stability time for faults but did not consider DRESs [14]. Ali et al. also studied the impact of delay on stability time while considering various fault locations, but they did not consider DRESs [2, 7]. In contrast, this research considers the impact of communication delay and DRESs on adaptive protection in a single context. Therefore, *the impact of communication delay and DRESs together on protection systems has yet to be investigated*, especially considering protection misoperation such as *blinding* and is tackled in this research.

Protection system modeling for Power-ICT system interdependencies. ICT-based protection systems have been modeled using various methods. Zerihun et al. used stochastic activity nets to model communication-based protection systems to assess their dependability, i.e., component or link failures but did not model or investigate the impact of communication delay [5]. Furthermore, the impact of power system parameters has not been studied in conjunction with ICT parameters in protection systems. Aleixo et al. mentioned communication delay requirements for adaptive protection based on IEC 61850 [19]. However, they did not investigate the impact of power system parameters, e.g., fault impedance, on the protection system performance [1]. Coffele et al. investigated the impact of LIFs on protection, especially considering the location of DRESs [10].

Reference	DRES	DRES fault level	Fault impedance	Fault type	Communication delay	Time distribution	Protection misoperation
[15]	✓	✓					✓
[10]	✓	✓		✓			✓
[23]	✓						
[20]	✓	✓	✓	✓			
[14]				✓	✓		
[2]				✓	✓		
[9]						✓	
[5]						✓	
[1]					✓		
[24]	✓		✓				✓
This research	✓	✓	✓	✓	✓	✓	✓

Table 1. Research gap in existing literature on protection system modeling.

However, they did not consider the impact of ICT on protection systems or protection blinding. Modeling of protection trip time in DES requires a time distribution assumption. The negative exponential distribution has been used for this purpose [5, 9]. However, this assumption is required for numerical solutions and not simulations. *This research determines an appropriate time distribution assumption for protection modeling such that the impact of DRESs, fault impedance and delay is captured.* Furthermore, an IEC 61850-based communication network for the protection system is assumed to assess ICT *performance*.

In summary, an approach encapsulating power-ICT system interdependencies in protection systems requires attention. This research proposes a DES-based modeling approach to tackle the gaps mentioned in this section. The proposed model explicitly considers the impact of communication delay, the consequences of DRESs and fault parameters such as impedance on the performance of a protection system while investigating suitable modeling assumptions for protection systems. Therefore, the main research question tackled in this paper is: *How can the protection system be modeled to capture the impact of fault impedance, communication delay and DRESs?*

3 SYSTEM DESCRIPTION AND ASSUMPTIONS

This section describes the interdependent power and ICT systems studied in this research. The integrated power-ICT system is defined explicitly to study interdependencies. An exemplary system is depicted in Fig. 1 and is described in this section.

3.1 Power system

The power system consists of a radial low-voltage distribution grid comprising $\mathtt{m}$ buses, i.e., $\{\mathtt{b}_1, \ldots, \mathtt{b}_\mathtt{m}\} \in \mathcal{B}$, where $\mathcal{B}$ is the set of buses and $\mathtt{n}$ lines, i.e., $\{\mathtt{l}_1, \mathtt{l}_2, \ldots, \mathtt{l}_\mathtt{p}\} \in \mathcal{L}$, where is $\mathcal{L}$ is the set of lines that connect the buses. Each bus $\mathtt{j}$ has an impedance $\mathtt{z}_{\mathtt{jj}}$ between the busbar and the ground. This research is focused on solidly grounded radial distribution networks. Faults are assumed to occur at buses, a commonly used assumption that allows the analysis of fault impedance relative to the bus impedance [4, 32]. These faults are classified based on the fault current they produce.

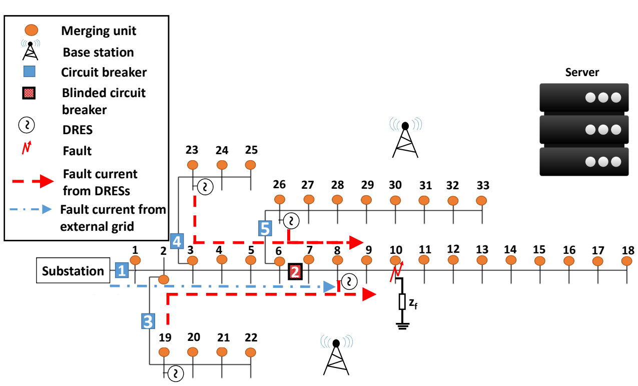


Fig. 1. Example power system depicting protection blinding with circuit breakers, DRESs, and a communication network.

LIFs are detected by their high fault current, while HIFs are detected based on voltage $\mathtt{v}$ and current $\mathtt{i}$ measurements. HIFs cause a fault current that is hard to distinguish from the load current, hence the need for adaptive protection. Furthermore, they cause non-linear voltage and current characteristics and distortion in measurement waveforms [18]. Single-phase to-ground faults may cause such characteristics and are the basis of HIF detection [25]. For this reason, single-phase-to-ground faults are studied in this research. However, other fault types are considered in the evaluation. Due to faults, a fault current is drawn from the external grid and DRESs. The external grid and DRESs have a fault level $\mathtt{S}_{\mathtt{eg}}$ and $\mathtt{S}_{\mathtt{dres}}$, respectively. Fault level is defined as the potential maximum current that flows in a network during a fault.

3.1.1 Protection system. Converter-coupled DRESs and protection devices such as overcurrent relays that operate CBs are integrated for the protection scenario (see Fig. 1). One unidirectional overcurrent protection device operating the respective CB is placed at the beginning of each feeder. The overcurrent protection may have different characteristics, e.g., standard inverse, very inverse, etc, based on IEC 60255 [29]. Standard inverse characteristics are most commonly used in distribution grids. This type only operates when the fault current exceeds its pickup current value. Very inverse characteristics are used when the fault current capacity of a network is low and as a backup to standard inverse types. Overcurrent relays with very inverse characteristics are suitable for faults that produce low and those that produce high fault currents. It is assumed that the relays are initialised with pickup-current settings, which determine the current at which the relay begins to operate, and the time multiplier setting, which determines the time delay to trip, corresponding to the fault current produced by commonly occurring faults such as LIFs. The relay operation is based on a numerical overcurrent relay, e.g., wavelet transform sampling, which issues a trip command instantaneously. After this, a trip delay may be induced [29]. The general protection philosophy is based on the approach of Gkavanoudis et al. [15], with the difference that ICT-based protection is used in this research.

3.1.2 Protection misoperation. Fault currents to trip the protection are provided by DRESs and the external grid. The fault current contribution is based on IEC 60909 [3, 22]. They supply fault current immediately after the fault occurs. If a DRES exists between a fault and a relay, e.g., at bus 8 in Fig. 1), the current through the protection contributed from the external grid will be reduced. This increases the trip time of the relay immediately upstream, i.e., the relay operating CB 2 in Fig. 1. This is known as *protection blinding* [15]. Note that since the upstream fault current reduces, the current through the secondary protection, i.e., CB 1, also reduces. Thus, the overcurrent protection relay operating CB 1 is also blinded. An increase in trip time could threaten grid stability in some circumstances, e.g., prolonged faults can cause frequency deviations, which, if not corrected, can lead to generator trips destabilizing the grid. Therefore, the settings of a protection device must be adapted after the fault occurs so that the protection device can sense low fault currents (due to blinding) and decrease the trip time. The FCT is calculated as the time from when a fault occurs to when a protection device clears a fault by isolating the faulty part of the network. This FCT is denoted by Θ in this research. The difference in FCTs with and without DRESs is called the *blinding time*, denoted by Θ_d.

Another type of misoperation is *sympathetic tripping*. This misoperation type occurs when the protection device in a healthy feeder (a feeder without a fault) trips before the relevant primary protection [15]. The difference between the protection trip time in a healthy feeder and the primary protection is denoted by Φ in this research. Therefore, a negative value of Φ indicates sympathetic tripping has occurred. Interlocking schemes can be used to resolve sympathetic tripping where, based on communication, relays are blocked from operation until the primary relay has cleared the fault. However, this could lead to long FCTs as the primary relay may take a long time to detect and clear a fault.

Since both fault impedance and DRESs impact the protection operation, this research investigates the impact of both these parameters on the protection system. Note that protection blinding and sympathetic tripping are issues specific to overcurrent protection devices, which are modeled in this research. Furthermore, this research focuses on the impact of DRESs on primary protection devices with respect to a fault.

3.2 ICT system

The ICT system contains a server, communication links, and merging units (see Fig. 1). The main functions of the ICT system are described in this section.

3.2.1 Communication to detect faults. The IEC 61850 standard defines an intelligent electronic device with measurement, protection and control functionality. This intelligent electronic device, hosted at the centralized server, is responsible for fault detection, location, and calculation of the new relay settings. The measurements for detection and location are sent with a communication delay $\mathtt{t}_{\text{delay}}$ using an IEC 61850-based communication network from a merging unit at each bus to the server. Merging units send measurements as Sampled Value (SV) messages defined in the IEC 61850 standard. Assuming a 4 kHz sample rate defined in this standard, the SV messages are transmitted every 250 microseconds. There is at least one merging unit on each bus. The time to detect a fault after reception of measurements is assumed to be constant. Specific detection schemes may have different detection times and can be explored in future work.

3.2.2 Communication to adapt protection device. Upon reception of the relevant measurements at the server and a successful detection process, appropriate control actions are taken. Generic Object Oriented Substation Event (GOOSE) messages are used to send new protection settings after a fault occurs. The GOOSE message contains the new pickup current settings. These settings allow the relay to sense the fault current and open in an appropriate time frame. A detailed overview of the protection adaptation is described alongside the model in Section 4. SV and GOOSE messages are assumed to be sent with a communication delay $\mathtt{t}_{\text{delay}}$, which impacts the FCT Θ. The communication delay is the end-to-end delay from the sender (e.g., server) to the receiver (e.g., relay). This end-to-end delay is also called transfer time in IEC 61850 [19]. All components in power and ICT systems are susceptible to failure. In the event of a component failure, the component does not deliver its service until a repair or recovery action is completed.

3.3 Power-ICT system interdependencies in adaptive protection

Using an IEC 61850-based ICT system introduces power-ICT system interdependencies. The main interdependency is the reliance of the protection system on ICT for fault detection and adapted settings. Fault detection relies on merging units transmitting SV messages consisting of voltage and current measurements. Furthermore, the protection settings are adapted using GOOSE messages. If an ICT link fails, a fault remains undetected, or the relay settings are assumed to remain as initially set for a fault-free scenario. However, protection misoperation, e.g., blinding, may increase the protection trip time due to low fault current. In this case, the relay must wait for fault detection and to receive new settings. Since faults must be cleared in the order of milliseconds, the communication delay involved in detecting faults and adapting the relay sensitivity should be as low as possible to maintain a low FCT. An increased delay will increase the time to detect and time to adapt a relay. Therefore, the FCT Θ increases. Consequently, an increased communication delay may not improve the FCT Θ in ICT-based protection systems compared to protection systems without ICT.

Considering the system description, the objective is to determine the influence of DRESs and communication delay on protection systems using a suitable methodology. This methodology is described in Section 4.

4 INTERDEPENDENT POWER-ICT SYSTEM MODEL

This section presents the performance model proposed to analyse the performance of ICT-based adaptive protection systems described in Section 3. An overview of the events and parameters involved in the modeling methodology is shown in Fig. 2. The horizontal arrows indicate events, while the vertical arrows indicate the time required for these events. This figure is described in this section.

4.1 Power system model

A fault is modeled as a single phase-to-ground connection, i.e., an impedance between a phase and ground parallel to bus $\mathtt{j}$ impedance $\mathtt{z}_{\mathtt{jj}}$. This fault is characterized by a fault impedance $\mathtt{z}_{\mathtt{f}}$ and exists if the corresponding fault impedance is non-negative [4, 32], as shown in Equation (1). The occurrence of a fault is the first event in the model (*fault occurs at bus* in Fig. 2):

$$|\mathtt{z}_{\mathtt{jj}}| \leftarrow \frac{|\mathtt{z}_{\mathtt{jj}}| \cdot |\mathtt{z}_{\mathtt{f}}|}{|\mathtt{z}_{\mathtt{jj}}| + |\mathtt{z}_{\mathtt{f}}|}, \text{where } \mathtt{z}_{\mathtt{f}} = \{\, \mathtt{x} + \mathtt{iy} \mid \mathtt{x}, \mathtt{y} \in \mathbb{R}\} \qquad (1)$$

The ratio of the fault to bus impedance, called the relative impedance $\mathtt{z}_\rho$, is given by Equation (2):

$$\mathtt{z}_\rho = \frac{|\mathtt{z}_{\mathtt{f}}|}{|\mathtt{z}_{\mathtt{jj}}|} \times 100 \qquad (2)$$

Note that the relative impedance is distinct from the source impedance ratio. The source impedance ratio is the ratio of source impedance to the line impedance. This ratio is the preferred method to study

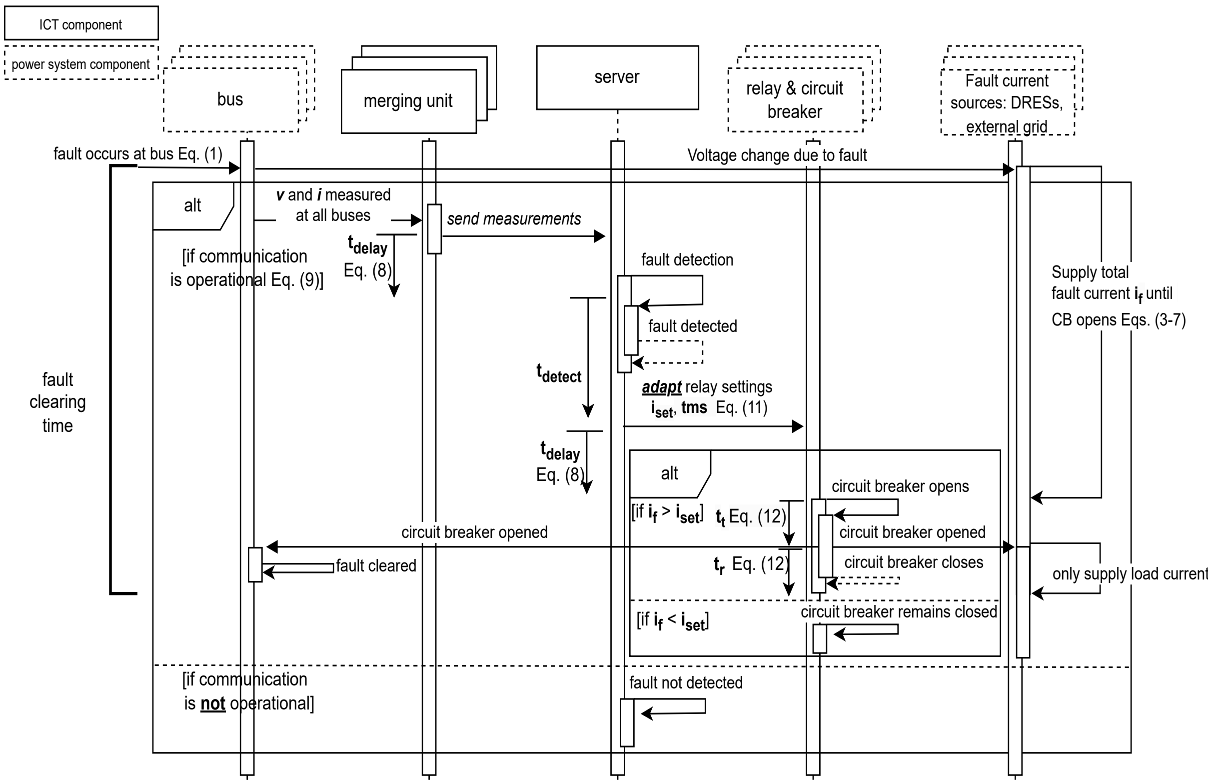


Fig. 2. Sequence of power and ICT system events in adaptive protection and their respective time delays.

the performance of distance relays, e.g., if the relay under or overreaches. In contrast, the relative impedance depicts the ratio of fault impedance to the bus impedance. The bus j impedance z_{jj} at which a fault occurs is impacted by the fault. The resulting bus impedance is the equivalent parallel impedance between z_{jj} and z_f, (see Equation (1)). It can be deduced from Equation (1) that low z_f values reduce z_{jj} to a low equivalent impedance. This is the case of LIFs, where a large fault current flows due to the low fault impedance. For HIFs, the equivalent impedance does not drop drastically, which results in a low fault current. In this research, the fault impedance is considered to be purely resistive. In future work, a fault bus, i.e., a ground connection, can be inserted on a line to model a line fault. Each DRES at bus m detects the voltage change due to the fault and injects fault current $i_{m,max}$ after the fault occurs, determined by Equation (3) based on IEC 60909 considering DRESs [3, 6, 22].

$$i_{m,max} = \frac{S_{dres}}{\sqrt{3}\, V_n}, \text{for each bus } m \tag{3}$$

where S_{dres} is the fault level of the DRES and V_n is the network voltage. Note that Equation (3) describes the fault current contribution for 3-phase faults. For 2-phase and 1-phase faults, the fault current contribution should be accordingly adjusted. Equation (4) determines the total fault current i_{dres}, considering each $i_{m,max}$ from DRESs, as defined in IEC 60909 details of which can be found in reference [3].

$$i_{dres} = \frac{1}{z_{jj}} \cdot \sum_{m=1}^{n} z_{jm} \cdot i_{m,max}, \text{for a fault at bus } j \tag{4}$$

where z_{jj} is the bus j impedance, z_{jm} is the impedance between bus j and m, $i_{m,max}$ is the maximum current contribution from the converter-coupled DRES at bus m. The relative fault level of the distribution grid, i.e., all DRESs in the distribution grid to the external grid, is given by δ defined in Equation (5):

$$\delta = \frac{\sum S_{dres}}{S_{eg}} \tag{5}$$

Fault current is also drawn from the external grid, denoted by i_{ext}. This is described by Equation (6) [6]:

$$i_{ext} = \frac{V_n}{\sqrt{3}\, Z_{jj}}, \text{for a fault at bus } j \tag{6}$$

However, due to i_{dres}, the magnitude of i_{ext} reduces to i'_{ext}, i.e., blinding (see Fig. 1). Therefore, the total fault current i_f:

$$i_f = i_{dres} + i'_{ext} \tag{7}$$

Equation (7) describes the total fault current i_f drawn by the fault *with blinding* where the relay operating CB 1 is blinded. Furthermore, The DRES at bus 8 may inject $i_{8,max}$, which will cause any upstream i_f to reduce. Therefore, Equation (7) may change with respect to the system topology and DRES location. In the absence of DRESs, $i_f = i_{ext}$. Consequently, for HIFs, z_{jj} remains high and i_f is low.

4.2 ICT system model

The presence of a fault does impact the current flow and voltage in the power system. Each merging unit samples current and voltage measurements at its respective bus after the fault occurs (v *and* i *measured at all buses* in Fig. 2). This communication of measurements as SV messages to the server is the next event. Each SV message is communicated with a *mean* end-to-end delay t_{delay} if the communication link is operational (*send measurements* in Fig. 2).

$$t_{delay} = \begin{cases} d & d > 0 \text{ if communication link} \\ & \text{is operational} \\ \text{not defined} & \text{otherwise} \end{cases} \tag{8}$$

This delay shown in Equation (8), or transfer time as called in IEC 61850 standard, consists of network stack processing t_a and network transfer delays t_b. Therefore, $d = t_a + t_b$. Moreover, communication devices such as routers in the communication path contribute to the network transfer time. In the presence of losses, the transfer time can vary [19]. Hence, in this research, a mean communication delay is used. Furthermore, failure and recovery events can impact a communication link's state from a reliability perspective. This research depicts a communication link as a wireless radio link. Each radio link communication capable device is being connected through n wireless links simultaneously. Each wireless link is assumed capable of multipath propagation leading to a Rayleigh fading channel. Therefore, in this research, the radio transmission is modeled by a failure and recovery action where fading is considered the cause of transmission failure [37]. The radio link state may also be considered failed if the transmitting device fails, e.g., merging unit failure. The impact of the communication state is shown in Fig. 2, where the fault is not detected if the communication is not operational. Equation (9) indicates the state of a radio link.

$$radio_{state} = \begin{cases} 1 & \text{if link is fully operational} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The states of other components, such as merging units, servers, etc., are similarly determined using failure and repair or recovery events. After the server receives the measurements, these measurements are processed to detect faults. Several methodologies have been developed to identify the type of fault in the system [18]. The low voltage and high fault current caused by a LIF is used to detect this type of fault. HIFs cause non-linear voltage v and current i characteristics due to the ground contact impedance and arcing ($v \not\propto i$) [18]. A HIF detection process is triggered when such characteristics are observed (*fault detection* event in Fig. 2). Although these methodologies vary in implementation, they rely on voltage and current measurements for detection. The communication requirements of these methodologies are formulated as constraints, i.e., SV messages from the faulty buses must be received with delay t_{delay} within a threshold t_{lim} (*send measurements* in Fig. 2). The detection process is successful if necessary SV messages from the faulty buses are obtained. The fault is further classified as a LIF or HIF as shown in Equation (10).

$$fault_{type} = \begin{cases} LIF & \text{if } i_f \geq i_{set} \\ HIF & \text{if } v \not\propto i \text{ and } i_f < i_{set} \end{cases} \tag{10}$$

The fault may be cleared by adapting the settings of the primary relay to trip at a lower pickup current subject to a further communication delay (see Section 3.1.2 and Fig. 2).

4.3 Power - ICT system interdependency model

As relays are assumed to be initialised with settings for LIFs, these settings may need to be adapted using ICT to sense the low fault current due to protection blinding or HIFs. Equation (11) describes the adaptation of relay settings. In the context of protection blinding, the current through the relay i_f is reduced (Fig. 1). If the pickup current setting i_{set} remains as initially set, the relay may issue a trip command after an extended time. In the case of sympathetic tripping, the fault current i_f may be too high in a healthy feeder. This could cause the protection in a healthy feeder to trip if the pickup current setting i_{set} remains as initially set. Therefore, in both cases, the relay settings must be adapted, i.e., adapt pickup setting i_{set} or time multiplier setting tms within a time threshold t_{lim} such that the appropriate CB trips promptly to avoid disastrous consequences. Hence, there is a delay constraint on ICT due to the fault. This encapsulates a dependency of power system on ICT described in Equation (11). Therefore, once a fault is detected with delay t_{detect}, new settings are sent to relays using GOOSE messages with a mean communication delay t_{delay} (*adapt circuit breaker settings* in Fig. 2).

$$(i_{set}, tms) \leftarrow \{\, i_{set,new}\,,\ tms_{new} \mid t_{delay} < t_{lim} \} \tag{11}$$

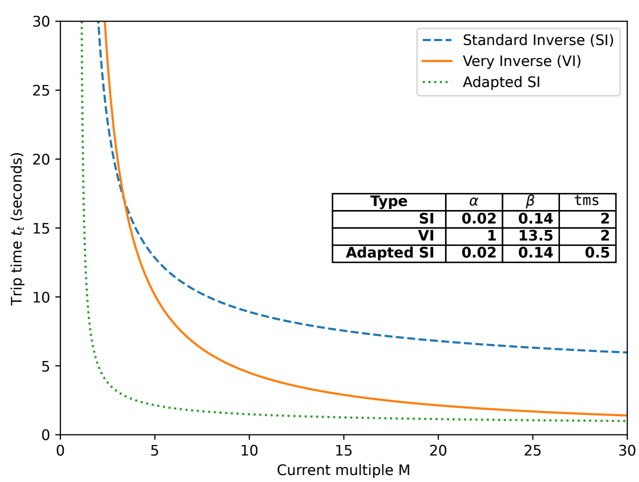


Fig. 3. Current vs time graph for protection relays. Standard inverse and very inverse characteristics are shown. An adapted standard inverse is depicted with a different time multiplier setting.

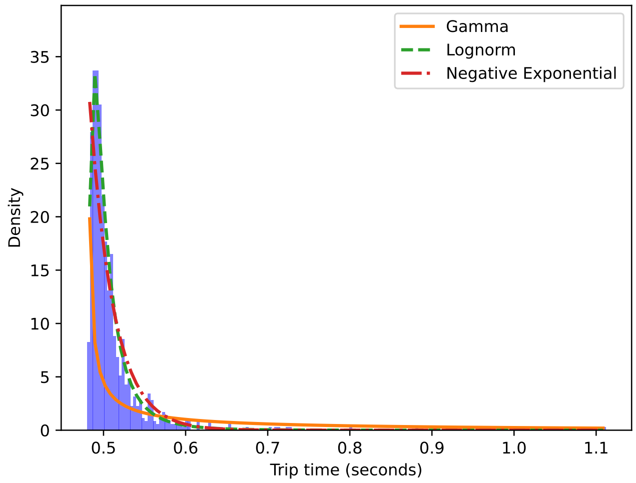


Fig. 4. Example fitting of candidate distributions to trip time data.

Only after $\mathtt{t_{delay}}$, a relay may wait a certain time to issue a trip signal, as determined by the new settings. This is the trip time $\mathtt{t_t}$ described by Equation (12) (*circuit breaker opens* in Fig. 2). Once the CB has tripped, it remains open until it is reset (*circuit breaker closes* in Fig. 2). The reset time $\mathtt{t_r}$ is based on the *adapted* settings (Equation (12)) based on IEC 60255.

$$\mathtt{t_t} = \mathtt{tms} * \frac{\beta}{\mathtt{M}^{\alpha} - 1}, \mathtt{t_r} = \mathtt{tms} * \frac{\beta_r}{1 - \mathtt{M}^{\alpha_r}} \tag{12}$$

where $\mathtt{M} = \mathtt{i_f}/\mathtt{i_{set}}$ is the current multiple, $\mathtt{i_f}$ is the fault current flowing through the relay, tms is the time multiplier setting, and $\mathtt{i_{set}}$ is the pickup current setting at which the relay begins to operate. The M value determines the sensitivity of the relay. A low M value implies that the relay has a low sensitivity (a high pickup current setting $\mathtt{i_{set}}$), i.e., the relay will take longer to issue a trip signal. Thus, the M value is low in a normal, fault-free scenario. The values α, β, α_r and β_r are constants that determine the type of relay characteristics (see Section 3.1) [29]. The protection characteristics are visualised in Fig. 3. The figure shows standard inverse, very inverse and an adapted standard inverse curve [29]. The relays in this research are initialised with the standard inverse curve, as shown in the figure. The relay settings are adapted using new tms and $\mathtt{i_{set}}$ values (see Equation (11)). In this example, the tms value is reduced, which produces the *adapted standard inverse* curve from the figure. In doing so, the relay is more sensitive than the initial setting, i.e., for the same M value, the adapted setting has a lower trip time.

4.3.1 Determining time distribution of circuit breaker events. In this research, to model the performance of the ICT-dependent protection devices in DES, i.e., a relay, the time distribution of trip $\mathtt{t_t}$ and reset time $\mathtt{t_r}$ for DES is investigated. Time distributions are an important aspect of performance modeling as the time delays of an event are sampled based on the time distribution of the event. The negative exponential distribution is commonly assumed as the time distribution for numerical solutions. However, this is a strong assumption, mainly due to its memoryless property, since the history of current flow influences trip time. Furthermore, as DES is used, this distribution is not essential. The possible candidates for time distribution of events, particularly the trip time $\mathtt{t_t}$ of a relay, is approximated as follows [26]. The negative exponential distribution is a special case of the gamma distribution. Hence, the first step is to generalize the negative exponential to a gamma distribution as a potential candidate. Since trip delay measures time, values can only be non-negative, while the distribution tends to be positively skewed with a tail. The positive skewness results from trip and reset times tending to be low, while larger trip times caused by low fault current contribute to a long tail. The lognorm distribution fits these characteristics and is considered a candidate [31].

4.3.2 Determining time distribution parameters. The time distribution parameters for the protection events in DES are determined using the maximum likelihood estimation technique. This technique is based on the log-likelihood function on the trip and reset time samples obtained from Equation (12) [31]. These samples are obtained using Monte Carlo sampling. A uniform distribution with bounds $\mathcal{U}[0, 1000]$ is assumed to sample various fault impedances such that faults with low and high impedances are captured [11, 21]. This assumption allows the model to capture the trip times for the entire range of fault impedance, preventing a bias towards a specific fault impedance, e.g., in importance sampling. While this assumption seems unrealistic, the transformation from fault impedance (consequently the fault current drawn) to trip time allows for the correction of this assumption based on Equation (12). Further distribution assumptions for fault impedance may be studied in future work. A short-circuit calculation is performed for each fault impedance independently, and the trip time is calculated for each fault impedance using Equation (12). The resulting distribution of trip times is shown in Fig. 4. The figure shows how the candidate distributions fit the trip time samples. It can be seen that the density of low trip times is higher than that of larger trip times rather than a uniform density. This is because the short-circuit and trip time calculation corrects the initial uniform distribution assumption. It should also be noted the trip time samples begin shortly before 0.5 seconds. This is due to standard inverse characteristics' minimum trip time delay (see Fig. 3). This minimum trip time delay is dependent on the relay settings.

4.3.3 Statistical tests for time distribution suitability. The observed data and theoretical distributions are compared using various tests. The Quantile-Quantile (Q-Q) plot test results are shown in Fig. 5. Q-Q plots are graphical tools that assess whether a dataset follows a particular probability distribution. In each plot, the x-axis represents the quantiles of the theoretical distribution, while the y-axis represents the quantiles of the dataset. If the data points fall along a straight line at 45°, it indicates that the dataset closely follows the theoretical distribution. Fig. 5 indicates that gamma and lognorm distributions fit the trip time event shown by the data points coinciding with the 45°line. However, this test only indicates the degree of fit between sampled data and theoretical distributions. Further tests

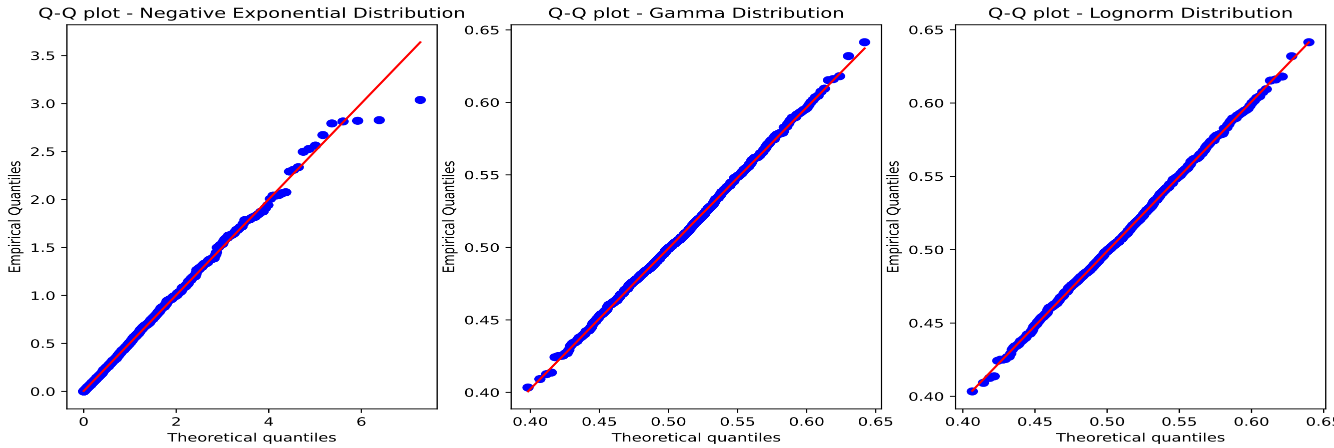


Fig. 5. Quantile-Quantile plot comparing sampled data against theoretical distributions. The plot illustrates the degree of fit between the sampled data distribution and the theoretical distributions, including negative exponential, gamma and lognorm.

are required to determine the fitness of each distribution, e.g., sensitivity to distribution tails. These test results are shown in Table 2. The Kolmogorov-Smirnov Test measures the discrepancy between the observed data's empirical cumulative and theoretical distribution functions. A higher value indicates a better fit. The Cramér-von Mises Criterion test, similar to the Kolmogorov-Smirnov test, compares the observed data's empirical cumulative distribution function with the theoretical distribution function but is more sensitive to differences in the tails. Finally, the Kullback-Leibler Divergence measures the difference between two probability distributions. A smaller value indicates a better fit for both the Cramér-von Mises and Kullback-Leibler tests. These initial distribution comparison tests indicate that the lognorm distribution is a suitable assumption for the protection trip time event. Similarly, the tests can be used to determine distributions for protection reset time $\mathtt{t_r}$. The impact of this distribution assumption on FCT Θ is studied in Section 6.

Test	Negative exponential	Gamma	Lognorm
Kolmogorov-Smirnov	0.0	3.44E-55	0.0038
Cramér-von Mises	97.26	26.34	1.12
Kullback-Leibler Divergence	0.801	0.528	0.124

Table 2. Tests for comparison between theoretical and candidate distributions of trip time $\mathtt{t_t}$.

4.3.4 Quantifying protection system performance. While the model described so far helps determine a system's response in a particular scenario, a metric is required to quantify the impact on the protection system. This metric can be used to compare different grids, e.g., with different fault levels, fault level distribution across DRESs, and communication systems. For each system, the impact of a set of communication delays $\Omega = \{\mathtt{t}_{\mathtt{delay},1}, \ldots, \mathtt{t}_{\mathtt{delay},n}\}$, a set of relative fault levels $\Delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$ and a set of FCTs for those delays and fault levels $T = \{\Theta_{1,1}, \Theta_{1,2}, \ldots, \Theta_{n,n}\}$ should be collected. Then the protection performance index Π to indicate the risk carried by the protection can be calculated by Equation (13):

$$\Pi = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \cdot \Theta_{i,j}}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}}, \forall\, \Theta_{i,j} \in T_{\mathtt{norm}} \tag{13}$$

where the parameter $w_{i,j}$ is a weight defined as $w_{i,j} = \delta_i \cdot \mathtt{t}_{\mathtt{delay},j}$ and $\delta_i \in \Delta_{\mathtt{norm}}, \mathtt{t}_{\mathtt{delay},j} \in \Omega_{\mathtt{norm}}$. The weights are used to consider the impact of delay and fault level on the protection system. Therefore, a higher delay and fault level could lead to a higher Π value due to their impact on the FCT. Consequently, a higher weight implies a higher impact on the system. The sets $\Delta_{\mathtt{norm}}$, $\Omega_{\mathtt{norm}}$ and $T_{\mathtt{norm}}$ are the normalized counterparts of $\Delta, \Omega,$ and T respectively. Due to normalization and weighted average calculation, the value Π lies in the interval $[0, 1]$. A low value of Π indicates a better protection system performance. To determine Π, the sets Ω and Δ should be defined. The space of communication delay is characterized by a continuum of time intervals, ranging from minimal to maximal delays, which include the potential delays encountered during signal propagation. This space is discrete, each value representing a delay value. Similarly, the fault level space covers the range of the fault levels observed within the system, spanning from low to high DRES penetration. This space is also defined by discrete fault levels, each indicative of the extent of DRES penetration in the distribution grid. For the space of communication delay, a progressive sampling strategy is implemented to capture a continuous range of delay values while ensuring an increasing sequence. This involves systematically selecting delay intervals by incrementally increasing time intervals to represent a continuum of delay scenarios. By iteratively sampling delay values from this progressive sequence, a subset is constructed that spans the desired range of communication delay, facilitating a comprehensive analysis of FCT Θ behaviour across increasing delay conditions. Similarly, within the space of fault level, a systematic sampling technique is utilized to define a subset characterized by

increasing fault levels. This entails sequentially selecting fault levels, starting from low and progressively escalating to higher ones. By systematically sampling fault severity levels according to this increasing progression, a continuous subset is constructed that captures the spectrum of fault levels, representing various degrees of DRES penetration.
Considering the model description in this section, as shown in Fig. 2, the FCT Θ consists of several time delays. Two cases can be distinguished: traditional and adaptive protection. In the former case, Θ depends on parameters such as fault impedance, protection sensitivity and DRES penetration. In contrast, the latter depends on communication delay, fault impedance, and DRES penetration [36]. Once the relay settings have been adapted, the relay operates after a delay, clearing the fault while the DRESs cease to supply fault current. This is the end of the fault-clearing process in Fig. 2.

5 MODEL SIMULATION SETUP

In this section, the simulation setup for the model is described. The DES model is described using *Stochastic Activity Nets* (SANs), an extension of stochastic petri nets [30]. SANs enable a comprehensive performance evaluation based on DES. DES allows specific times at which an event occurs to be recorded, enabling the calculation of Θ under varying parameters, e.g., communication delay. The state of power system components is modeled using SANs, and the *pandapower* package in Python is used for short circuit current calculations [33]. *DIGSILENT powerfactory* is used for validation purposes, while Python libraries are used for result processing and plotting. The simulation setup is shown in Fig. 6 and described with arrow labels as reference. The SANs are initialised with power and ICT system topologies based on the system described in Section 3 and parameters shown in Table 3 (arrows 1a and 1b in Fig. 6). The simulation follows the event sequence from Fig. 2. A power system fault event occurs in SANs, indicating a voltage $\mathtt{v}$ and current $\mathtt{i}$ state change (*fault occurs* event in Fig. 2). The resulting values are calculated in *pandapower* and fed back into SANs (arrows 2a and 2b in Fig. 6). The simulation progresses to the next event, i.e., sending measurements with a time delay $\mathtt{t_{delay}}$ if the supporting ICT system is available (send measurements event in Fig. 2). Upon reception of measurements, the next event, fault detection, is triggered, which completes with delay $\mathtt{t_{detect}}$. Then, the relay settings are adapted with a time delay $\mathtt{t_{delay}}$ only if ICT is available (*adapt relay settings* event in Fig. 2). Once the relay settings are adapted, the CB trips after a time delay $\mathtt{t_t}$ to clear the fault. The CB waits for reset time $\mathtt{t_r}$ to close (*CB opening and closing* events in Fig. 2). The simulation then continues until the next fault event. The simulated time of events is stored and used to calculate FCT Θ (arrows 3, 4, 5 in Fig. 6).

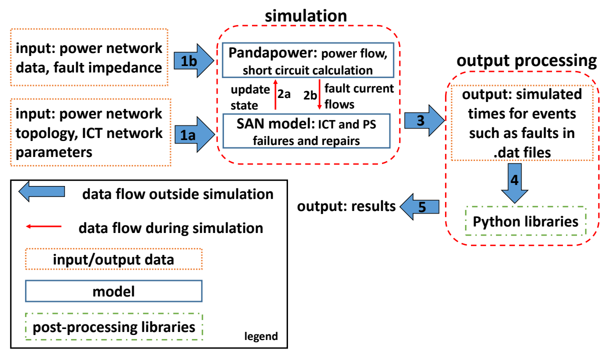
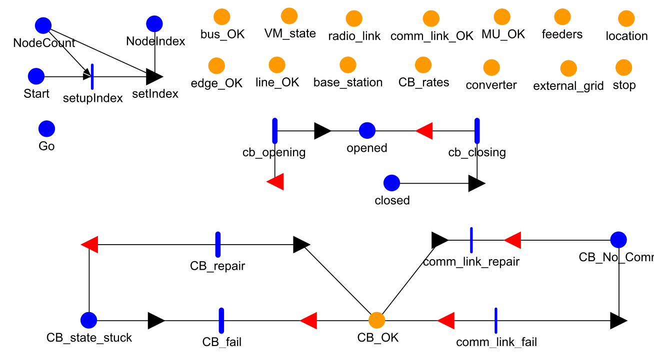
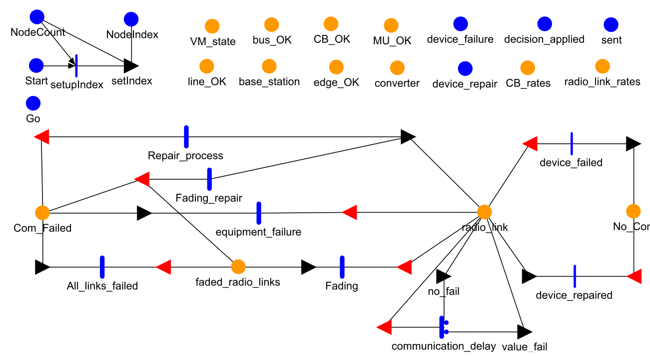


Fig. 6. Simulation setup for the proposed model.



(a) SAN model of a CB



(b) SAN model of a radio link

Fig. 7. SAN models of CBs and radio links.

Figures 7a and 7b show the SAN models of radio links and CBs. State variables are modeled using *places*, visualized by circles [30]. Pre- and post-conditions after events are modeled using *input* and *output* gates, visualized by triangles. Events are modeled using *activities*, visualized by thick and thin bars. The time delays associated with these events (described in Section 4) are captured by these *activities* (transitions in stochastic petri nets). In this research, the time distributions of the protection trip and reset events are of interest. Activity times are often assumed to possess a negative exponential distribution as a modeling simplification for numerical solutions. Therefore, as simulation is used to solve the model, this assumption is relaxed to evaluate its impact on the modeling. This research adopts the approach of da Silva et al. to calculate Θ using SANs [12]. This is accomplished by capturing the simulated time when a fault occurs and when it is cleared. The full SAN model and documentation can be found in [27].

Parameter	Description	Value (hours)
$1/\lambda_{cbf}$	mean time to CB failure	3 636
$1/\mu_{cbr}$	mean time to CB repair	2
$1/\lambda_{ecf}$	mean time to server communication link failure	9 000
$1/\mu_{ecr}$	mean time to server communication link repair	6
$1/\lambda_{escf}$	mean time to server crash failure	4 000
$1/\mu_{escr}$	mean time to server crash repair	4
$1/\lambda_{muf}$	mean time to random merging unit hardware failure	9 000
$1/\mu_{mur}$	mean time to merging unit hardware repair	2
$1/\lambda_{muswf}$	mean time to random merging unit software failure	1 428
$1/\lambda_{rlf}$	mean time to radio link fading	65
$1/\mu_{rlfr}$	mean time to radio link fading recovery	100 ms

Table 3. Fixed simulation parameters

6 EVALUATION

A modified IEEE 33 bus system is used for evaluation (Fig. 1). The fault injection approach is based on that of Gkavanoudis et al. [15]. A single phase-to-ground fault is induced at bus 10 in the power network. The results for this type of fault are shown. Further fault types are considered in the analysis. The validation process follows the three-step validation outlined by Trivedi et al. [34]. The three steps are face validation, input-output transformation, and validation of model assumptions. A simulation run consists of 10 000 hours of simulated time. For all simulation results from the SAN model, a 95% confidence interval is shown.

6.1 Validation

A modified version of the model, representing traditional protection systems without the possibility to adapt relays, is validated. The model is validated by comparing the mean FCT with those obtained from *powerfactory*. A fault level of 25 MVA is considered in this research [10]. The same model is configured in *powerfactory*. FCT Θ is measured from *powerfactory* and SAN models. The results are illustrated in Fig. 8. The FCT Θ obtained from *powerfactory* closely matches those obtained from the model, showing that the model captures the trip times of relays. This result serves as the input-output transformation step of the validation [34]. Note that results where FCTs obtained from the DES are lower than powerfactory do

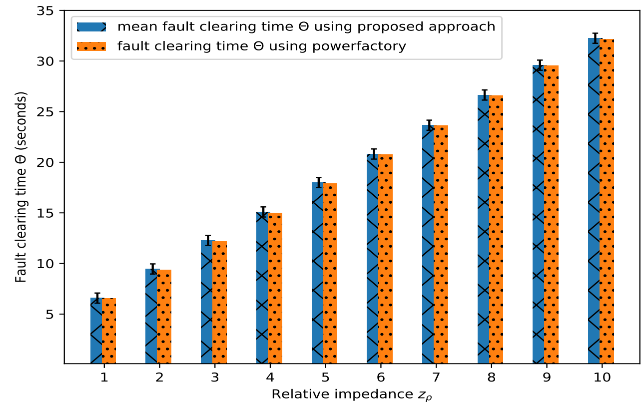


Fig. 8. Comparison of FCT Θ between the proposed approach and powerfactory using standard inverse characteristics for traditional protection systems.

not imply a performance improvement. The DES values are averages with 95% confidence interval.

6.2 Impact of fault impedance

Next, the impact of fault impedance z_f on FCT Θ on traditional protection systems without DRESs is studied. All parameters of the model described in Subsection 6.1 are fixed while the fault impedance is varied. The pickup current setting and the current multiple are fixed as in normal operation mode. Bus 10 in the system described, where the fault is induced, has an equivalent impedance of 3082 Ω. Fig. 8 shows the FCT Θ for this study. The bars depict the mean Θ. The fault impedance z_f is expressed in terms of relative impedance z_ρ. An increasing trend in Θ is observed as z_ρ increases. This is due to the decreased fault current i_f caused by the increased z_ρ (Equation (4)). The decreased fault current decreases the current multiple M. Hence, the trip time t_t increases (Equation (12)). This impact of fault impedance is captured by the DES event that models the protection trip time t_t. Hence, the proposed model captures the impact of fault impedance on the protection system.

6.3 Impact of protection operation distribution choice

For this study, the time distribution of the *cb_opening* and *cb_closing* in traditional protection, i.e., activities that model trip time t_t and reset time t_r respectively, are varied (see Fig. 7). Three distributions

Event	Parameter	Negative exponential	Gamma	Lognorm
CB open	$\hat{\mu}$ (sec)	0.244844	0.244844	0.244844
	σ (sec)	0.244844	0.299871	0.299871
CB close	$\hat{\mu}$ (sec)	0.006279	0.006279	0.006279
	σ (sec)	0.006279	0.00769	0.00769

Table 4. Mean and standard deviation of the distribution of trip time event.

are explored: negative exponential, gamma, and lognorm for reasons discussed in Section 5. The distributions' mean ($\hat{\mu}$) and standard deviation (σ) are shown in Table 4. These distribution values are identical to evaluate the impact of the distribution shape. The negative exponential distribution has an identical $\hat{\mu}$ and σ by definition. A greater σ is used for gamma and lognorm distributions, indicating a larger spread of possible values. A simulation study is performed for each time distribution assumption, producing an FCT Θ distribution. Statistical moments of these Θ distributions are shown in Table 5. Results show that the mean FCT Θ is relatively higher when the gamma and lognorm distributions are assumed. This is because lognorm and gamma distributions are heavy-tailed distributions. Hence, they can capture extreme cases with high FCT Θ, leading to a higher mean. The higher standard deviation for gamma and lognorm distribution also causes the variance to increase in the Θ distribution as there is a greater range of values sampled in these distributions. This increase in variance can be attributed to the heavy-tailedness of the distributions. The skewness moment shows that when the lognorm distribution is assumed, the distribution of Θ is asymmetrical, i.e., the lognorm distribution assumption produces a heavier right tail than negative exponential and gamma distribution. Kurtosis, i.e., the propensity to produce outliers, is higher for lognorm than negative exponential and gamma distribution. This indicates that the lognorm distribution is a suitable modeling choice to model the distribution of Θ, especially when the objective is to capture large Θ values, e.g., in the case of blinding. Thus, it can be seen that appropriate distributions can be used to capture various ranges of Θ. This result demonstrates the validation of model assumptions, i.e., distributions such as lognorm can be used to model protection operation events.

6.4 Impact of distributed renewable energy sources

In this study, the impact of DRESs is investigated. Two cases are distinguished: traditional protection without DRESs and with DRESs. Therefore, the external grid supplies fault current in the former case, while the DRESs *and* the external grid contributes fault current in the latter case. Protection sensitivity remains identical and fixed for both cases. For the blinding scenario, only one DRES is considered in the network in the same feeder as the fault. The presence of a DRES at bus 8 *blinds* the upstream relay, i.e., the fault current $\mathtt{i_f}$ through the relay is reduced. Fig. 9 shows the results of this blinding effect. For a relative impedance $\mathtt{z}_\rho$ (the ratio of fault impedance $\mathtt{z_f}$

Moment	Distribution of CB close and open events		
	Negative exponential	Gamma	Lognorm
Mean (sec)	0.229	0.259	0.275
Variance	0.0001	0.0016	0.0046
Skewness	2.0	6.35	7.14
Kurtosis	6.0	78.98	112.35

Table 5. Statistical moments of fault clearing time distribution for a fault with $\mathtt{z}_\rho = 1$ for bus 10.

to bus impedance $\mathtt{z_{jj}}$), the blinding time Θ_d (difference between the FCTs with and without distributed generation, i.e., blinding time) increases as the relative fault level δ (the ratio of distribution grid fault level to external grid fault level) increases since a higher fault current is drawn, further reducing the current from the external grid. Consequently, the current through the upstream relay is reduced. Hence, by Equation (12), the trip time $\mathtt{t_t}$ increases. The blinding time Θ_d also increases as $\mathtt{z}_\rho$ increases for a low δ since a low fault current is drawn from the external grid due to the higher fault impedance. Interestingly, at high $\mathtt{z}_\rho$ and δ values, Θ_d decreases. This is due to the increased sensitivity of earth fault protection (phase to ground fault) in the presence of DRESs [10]. This result shows that protection blinding is an issue for faults with low $\mathtt{z_f}$ and high δ values and vice versa. *Therefore, for the system studied, LIFs are critical for grids with high fault levels and HIFs are critical for grids with low fault levels for protection blinding.*

Another impact of DRESs is that of sympathetic tripping. This phenomenon is studied by the impact of a single DRES in the network in a feeder parallel to the feeder with the fault. For this study, the impact on the relay operating the CB immediately upstream of a DRES at bus 26 in traditional protection is investigated (CB 5 in Fig. 1). A fault with relative impedance $\mathtt{z}_\rho$ is induced at bus 10. Consequently, the DRES in the parallel feeder contributes to the fault current $\mathtt{i_{dres}}$ along with the external grid. Fig. 10 shows the results for various $\mathtt{z}_\rho$ and δ values. A normalized trip time difference value $|\Phi| = 1$ indicates a scenario where sympathetic tripping does not occur, i.e., primary protection CB 2 trips first. However, CB 5 trips first as the relative fault level δ increases, i.e., sympathetic tripping. The value $|\Phi|$ reduces as the absolute value indicates the CB in the healthy feeder has tripped before the primary protection. This is especially true as fault impedance increases. This is because higher DRES fault levels indicate a higher magnitude of fault current is injected in

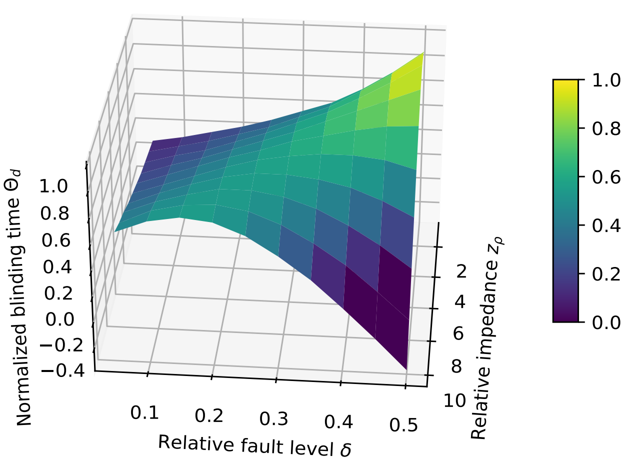


Fig. 9. Protection blinding time due to distributed generation under various fault levels and fault impedances.

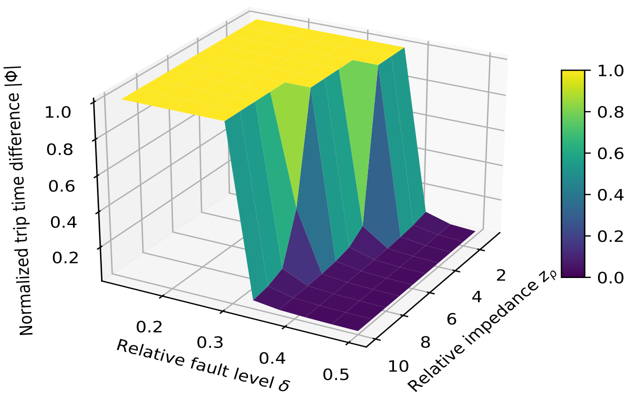


Fig. 10. Trip time difference between relay in healthy feeder and relay in fault feeder indicating sympathetic tripping for various fault levels and fault impedances.

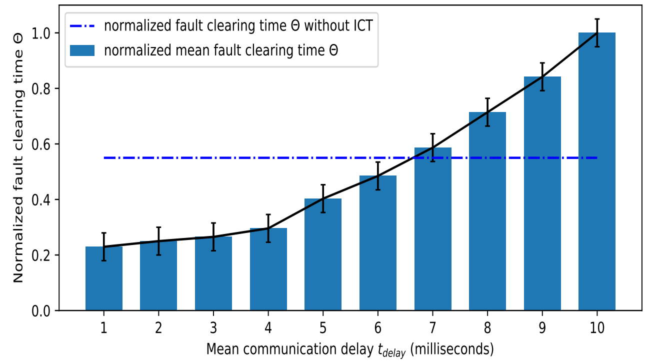


Fig. 11. Impact of communication delay on fault clearing times for protection blinding with $z_\rho = 1$ of bus 10 impedance, $\delta = 0.5$ and standard inverse protection characteristics.

the parallel feeder than the current drawn from the external grid (Equation (3)). Furthermore, a lower fault current is drawn at higher fault impedances (Equation (4)). *Hence, the CB in the healthy feeder is likely to trip first at higher fault impedances and high fault levels for the system studied.*

6.5 Impact of communication delay

In this study, the impact of communication delay $\mathtt{t_{delay}}$ on FCT Θ in adaptive protection systems is investigated (see Equation 11). All parameters are fixed for each simulation, while the mean communication delay of the SV and GOOSE messages is varied (see Section 3) for a sensitivity analysis. The results for the impact of communication delay $\mathtt{t_{delay}}$ on FCT Θ for relative fault level $\delta = 0.5$, a fault with $z_\rho = 1$ and standard inverse protection characteristics are illustrated in Fig. 11. The figure shows the mean Θ and the Θ in traditional protection, i.e., protection without ICT. Fig. 11 shows that using ICT for protection reduces Θ if the delay does not exceed a threshold. The mean Θ for cases with a mean communication delay of up to 6 ms is lower than the mean Θ in traditional protection for the same fault when blinding occurs. Note that the mean Θ for communication delay above 6 ms is larger than without ICT. *Therefore, 6 ms is the communication delay threshold* $\mathtt{t_{lim}}$ *in the sample space studied.* The mean Θ increases as the mean communication delay $\mathtt{t_{delay}}$ increases since there is a larger time delay until the relay is adapted. This impact is captured during DES and serves as face validation, capturing the impact of communication delay.

Fig. 12 shows the impact of communication delay $\mathtt{t_{delay}}$ on Θ for relays using very inverse characteristics. A fault with $z_\rho = 1$ of bus 10 impedance is induced for $\delta = 0.5$. Very inverse characteristics have a lower communication delay threshold, particularly for LIFs, as they have a shorter trip time $\mathtt{t_t}$ (hence, it should be adapted sooner). The model captures this and results in lower mean Θ for both traditional and adaptive protection systems, as illustrated in Fig. 12. The lower mean Θ implies that the communication delay threshold for these characteristics is stricter when compared to standard inverse characteristics. Fig. 12 shows that for a protection system with very inverse characteristics, ICT improves the performance until a mean communication delay of 5 ms. For very inverse characteristics, this serves as a delay threshold beyond which the adaptive protection system performs worse than the traditional protection system.

Fig. 13 shows the results for studies where the impact of communication delay on adaptive protection under various fault impedances without DRESs is analyzed. The relay is adapted to the same sensitivity in all cases. First, it can be noted that the results are consistent with the pattern seen in Figs. 8 and 11, i.e., increasing Θ as $\mathtt{t_{delay}}$ and z_ρ increases. This is because a steadily increasing z_ρ reduces the fault current drawn (see Equation 4) and hence reduces Θ. Similarly, an increasing $\mathtt{t_{delay}}$ causes Θ to increase. This is because of the longer time needed to adapt the relay. Fig. 14 shows the results

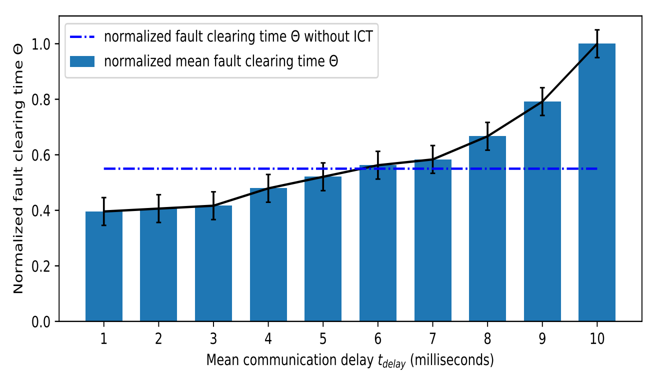


Fig. 12. Impact of communication delay on fault clearing times for protection blinding with $z_\rho = 1$ of bus 10 impedance, $\delta = 0.5$ and very inverse protection characteristics.

for studies where the impact of communication delay on adaptive protection under various DRES penetration levels for a 3-phase fault is analyzed. The relay possesses standard inverse characteristics and is adapted to the same sensitivity in all cases to resolve protection blinding. First, it can be noted that the results are consistent with the pattern seen in Fig. 11, i.e., increasing FCT Θ resulting in increasing blinding time Θ_d (difference between the FCTs with and without distributed generation) as $\mathtt{t_{delay}}$ increases. An important result is that communication delay has a low impact on low fault level grids. This is due to the lower fault current injected, leading to lower Θ_d. Next, Θ_d also increases as δ increases. This is because the DRESs inject a higher fault current. Consequently, the external grid fault current is reduced further, increasing the blinding effect. This effect and a high communication delay drastically increase Θ_d. The combined effect is the reason for the extreme Θ_d at high fault levels and high delay.

Fig. 15 shows the results for studies where the impact of communication delay on adaptive protection under various DRES penetration levels for a 2-phase fault is analyzed. Note that in this result, the impact of $\mathtt{t_{delay}}$ and δ are more prominent than for 3-phase faults. This is because fault current is lower for 2-phase faults than for 3-phase faults, and the relay is adapted to the same sensitivity in both studies. Furthermore, as δ increases, the blinding impact reduces an already low fault current caused by the 2-phase fault. Therefore, a steeper increase in Θ_d is seen. Similar to 3-phase faults, a high $\mathtt{t_{delay}}$ and a high δ results in a longer blinding time.

Fig. 16 shows the results for studies where the impact of communication delay on adaptive protection under various DRES penetration levels for a 1-phase fault. For this fault type, the impact of $\mathtt{t_{delay}}$ and δ are more prominent than for 3-phase and 2-phase faults. This is because fault current is even lower for 1-phase faults compared to 2-phase faults and 3-phase faults, while the relay is adapted to the same sensitivity in both studies. Similar to 2-phase faults, a steep increase in Θ_d is seen due to low fault current. This fault type also shows a high blinding time for high $\mathtt{t_{delay}}$ and δ.

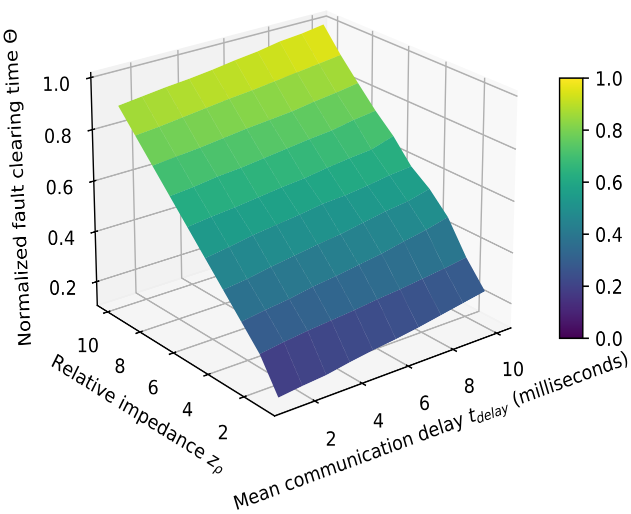


Fig. 13. Impact of communication delay $\mathtt{t_{delay}}$ and fault impedance z_ρ on fault clearing time Θ for $\delta = 0.5$.

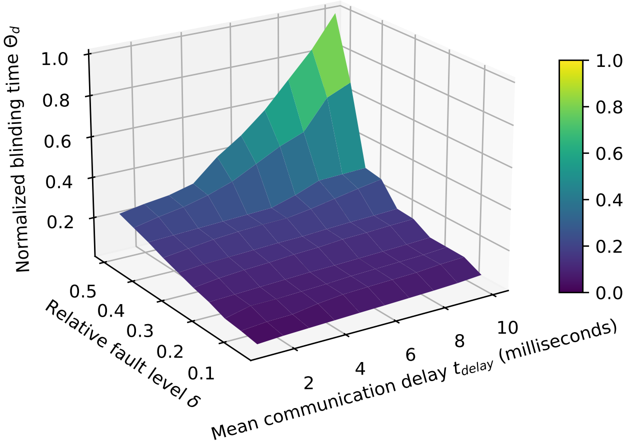


Fig. 14. Impact of communication delay $\mathtt{t_{delay}}$ and fault level δ on fault clearing time Θ for a 3-phase fault with $z_\rho = 1$ in protection blinding.

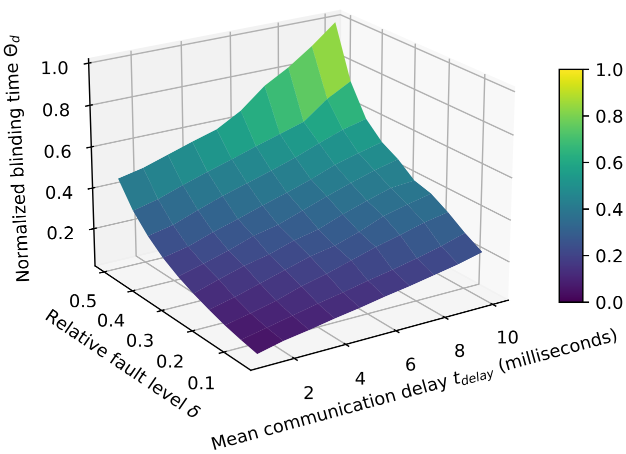


Fig. 15. Impact of communication delay $\mathtt{t_{delay}}$ and fault level δ on fault clearing time Θ for a 2-phase fault with $z_\rho = 1$ in protection blinding.

7 DISCUSSION

This section discusses and contextualises the results from Subsection 6.5 using the proposed protection performance index Π. Table 6 shows the quantification of protection performance index Π for various protection blinding scenarios with fixed fault impedance but varying fault level and communication delay. Results show that Π value increases as communication delay and fault level increases for each fault type. It should be noted that the Π value is high for

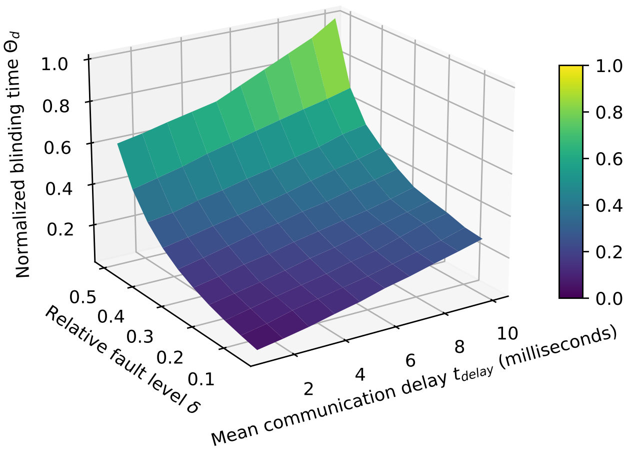


Fig. 16. Impact of communication delay $\mathtt{t_{delay}}$ and fault level δ on fault clearing time Θ for a 1-phase fault with $z_\rho = 1$ in protection blinding.

Case no.	Fault type	$\mathtt{t_{delay}}$ (ms)	δ (25 mva)	Π
1	3-ph	1-10	0.1-0.5	0.36
2	3-ph	10-20	0.1-0.5	0.59
3	3-ph	1-10	0.5-1	0.68
4	3-ph	10-20	0.5-1	0.82
5	2-ph	1-10	0.1-0.5	0.41
6	2-ph	10-20	0.1-0.5	0.67
7	2-ph	1-10	0.5-1	0.82
8	2-ph	10-20	0.5-1	0.9
9	1-ph	1-10	0.1-0.5	0.42
10	1-ph	10-20	0.1-0.5	0.7
11	1-ph	1-10	0.5-1	0.84
12	1-ph	10-20	0.5-1	0.92

Table 6. Protection performance index for the case studies.

grids with high fault levels, even though the communication delay is in a lower range. This is highlighted by cases 2 and 3 (similarly, cases 6 and 7, 10 and 11). This implies that the distribution grid fault level is an important factor that should be considered in protection system performance evaluation, along with communication delay. The results also show that 1-phase faults are more critical than 3-phase faults in protection blinding since the Π value is higher. This is because blinding time Θ_d is higher due to lower fault current for 1-phase faults (see cases 1 and 9). Consequently, the impact of communication delay is severe for 1-phase faults in grids with high fault levels when the relay is adapted to the same sensitivity. *Therefore, in the cases studied, communication delay is more critical in grids with high fault levels than in grids with low fault levels in resolving protection blinding.* The impact of communication delay should be considered with the distribution grid fault level when assessing power system protection. Therefore, the integration of DRESs and the use of ICT in protection systems does impact the protection system adversely. These aspects of a smart distribution grid should be considered in the system design for resilient operation. In particular, the proposed model can be used to assess the impact on the FCT Θ in terms of the time elapsed, e.g., communication delay, CB sensitivity, for various grid fault levels.

8 CONCLUSION

The main contribution of this research is a model to assess the performance of adaptive protection systems in smart active distribution grids. The model considers formal descriptions of power system properties, e.g., fault impedance and DRES fault current contribution, and performance-related properties of ICT, e.g., delay. These formal descriptions are considered to calculate the time delay in discrete-event simulation, which impacts the fault clearing time. Modeling assumptions such as time distribution functions are investigated. Results show that the lognorm distribution is a suitable time distribution assumption to capture extreme fault clearing times, especially in the case of protection blinding due to its heavy-tailedness. Results from a sensitivity analysis show that rising fault impedance increases the fault clearing times. Protection misoperation results show that in the context of protection blinding for the system studied, low-impedance faults are critical for grids with high fault levels, and high-impedance faults are critical for grids with low fault levels. Sympathetic tripping is observed at high fault levels and tends to occur as fault impedance increases. Communication as a solution to protection misoperation is discussed. Results show that the fault clearing time increases when communication delay increases. In particular, as fault impedance and communication delay increase, the fault clearing time increases almost linearly. An important insight is that communication delay has a higher impact in grids with high fault levels compared to lower fault levels. This insight holds true for 3-phase, 2-phase and 1-phase faults. Studies on these fault types show that grid fault level should be considered along with communication delay since grids with high fault levels and a low communication delay can perform worse than grids with low fault levels and high communication delays. The newly formulated protection performance index quantifies this impact. Therefore, the protection system performance should consider communication delay and the fault level of the distribution grid. As more renewables are integrated into the distribution grid, the impact on the protection system becomes more critical. Further time distributions will be investigated in future work to model the operation of protection devices such as reclosers. The impact of time-critical communication failures, such as packet loss, will also be investigated.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation DFG as part of the project “Multi-ResiServD” with the project identification number 360352892 of the priority program DFG SPP 1984 - Hybrid and multimodal energy systems: System theory methods for the transformation and operation of complex networks. The authors acknowledge support from COST (European Cooperation in Science and Technology) Action RECODIS CA15127.

REFERENCES

[1] Ana Aleixo, Rogério D Paulo, Rui D Jorge, Alberto Rodrigues, Carlos Arantes, José Cabaça, and Pedro M Neves. 2020. Resilient 5G technologies optimized for power grid protection solutions using IEC 61850 time-critical communications. In *15th International Conference on Developments in Power System Protection (DPSP 2020)*. IET, 1–6.

[2] Mohd Hasan Ali and Dipankar Dasgupta. 2011. Effects of communication delays in electric grid. In *2011 Future of Instrumentation International Workshop (FIIW) Proceedings*. IEEE, 38–41.

[3] Rafat Aljarrah, Hesamoddin Marzooghi, Vladimir Terzija, and James Yu. 2019. Modifying IEC 60909 standard to consider fault contribution from renewable energy resources utilizing fully-rated converters. In *2019 9th International Conference on Power and Energy Systems (ICPES)*. IEEE, 1–6.

[4] Abdulaziz Aljohani and Ibrahim Habiballah. 2020. High-impedance fault diagnosis: a review. *Energies* 13, 23 (2020), 1–18. 6447.

[5] Tesfaye Amare, Charles M Adrah, and Bjarne E Helvik. 2019. A method for performability study on wide area communication architectures for smart grid. In *2019 7th International Conference on Smart Grid (icSmartGrid)*. IEEE, 64–73.

[6] Gerd Balzer. 2016. Short-circuit calculation with fullsize converters according to iec 60909. In *21 st Conference of Electric Power Supply Industry, CEPSI*.

[7] Dimitris Baros, Nick Rigogiannis, Nick Papanikolaou, and Michael Loupis. 2020. Investigation of communication delay impact on DC microgrids with adaptive droop control. In *2020 International Symposium on Industrial Electronics and Applications (INDEL)*. IEEE, 1–6.

[8] Pedro Henrique Aquino Barra, Denis Vinicius Coury, and Ricardo Augusto Souza Fernandes. 2020. A survey on adaptive protection of microgrids and distribution systems with distributed generators. *Renewable and Sustainable Energy Reviews* 118 (2020), 1–16. 109524.

[9] Silvano Chiaradonna, Paolo Lollini, and Felicita Di Giandomenico. 2007. *Modelling framework of an instance of the electric power system: Functional description and implementation*. Technical Report Technical Report RCL071202. University of Florence, Dip. Sistemi Informatica, RCL group.

[10] Federico Coffele, Campbell Booth, A Dyśko, and Graeme Burt. 2012. Quantitative analysis of network protection blinding for systems incorporating distributed generation. *IET Generation, Transmission & Distribution* 6, 12 (2012), 1218–1224.

[11] Qiushi Cui and Yang Weng. 2019. Enhance high impedance fault detection and location accuracy via μ-PMUs. *IEEE Transactions on Smart Grid* 11, 1 (2019), 797–809.

[12] Leandro Dias da Silva, Paolo Lollini, Diamantea Mongelli, Andrea Bondavalli, and Gianluca Mandò. 2021. A stochastic modeling approach for traffic analysis of a tramway system with virtual tags and local positioning. *Journal of the Brazilian Computer Society* 27 (2021), 1–38.

[13] Kartika Dubey and Premalata Jena. 2020. Impedance angle-based differential protection scheme for microgrid feeders. *IEEE Systems Journal* 15, 3 (2020), 3291–3300.

[14] Abdallah Farraj, Eman Hammad, and Deepa Kundur. 2015. A systematic approach to delay-adaptive control design for smart grids. In *2015 IEEE international conference on smart grid communications (SmartGridComm)*. IEEE, 768–773.

[15] Spyros I Gkavanoudis, Dimitrios Tampakis, Kyriaki-Nefeli D Malamaki, Georgios C Kryonidis, Eleftherios O Kontis, Konstantinos O Oureilidis, José María Maza-Ortega, and Charis S Demoulias. 2020. Protection philosophy in low short-circuit capacity distribution grids with high penetration of converter-interfaced distributed renewable energy sources. *IET Generation, Transmission & Distribution* 14, 22 (2020), 4978–4988.

[16] Felix Glinka, Nicolas Schulte, R Bertram, Armin Schnettler, and Mitja Koprivšek. 2018. Solutions for blinding of protection in today's and future German LV grids with high inverter penetration–simulative and experimental analysis. *The Journal of Engineering* 2018, 15 (2018), 1256–1260.

[17] Hany F Habib, Christopher R Lashway, and Osama A Mohammed. 2017. On the adaptive protection of microgrids: A review on how to mitigate cyber attacks and communication failures. In *2017 IEEE industry applications society annual meeting*. IEEE, 1–8.

[18] Ester Hamatwi, Odunayo Imoru, Matheus M Kanime, and Hitila SA Kanelombe. 2023. Comparative analysis of high impedance fault detection techniques on distribution networks. *IEEE Access* 11 (2023), 25817–25834.

[19] IEC. 2011. IEC 61850-5, Communication networks and systems for power utility automation - Part 5: Communication requirements for functions and device models. , 60–63 pages.

[20] Rishabh Jain, David L Lubkeman, and Srdjan M Lukic. 2018. Dynamic adaptive protection for distribution systems in grid-connected and islanded modes. *IEEE Transactions on Power Delivery* 34, 1 (2018), 281–289.

[21] NSB Jamili, MR Adzman, SRA Rahim, SM Zali, M Isa, and H Hanafi. 2019. Evaluation of earth fault location algorithm in medium voltage distribution network with correction technique. *International Journal of Electrical and Computer Engineering (IJECE)* 9, 3 (2019), 1987–1996.

[22] Ismail Kasikci. 2018. *Short circuits in power systems: A practical guide to IEC 60909-0*. John Wiley & Sons, 14–18.

[23] David Majambere and Goro Fujita. 2017. Impact of fault resistance on impedance relay: Adaptive Mho directional type scheme development using LabVIEW. In *TENCON 2017-2017 IEEE Region 10 Conference*. IEEE, 3006–3011.

[24] Kari Mäki, Sami Repo, and Pertti Järventausta. 2005. Blinding of feeder protection caused by distributed generation in distribution network. In *5th WSEAS Int. Conf. on Power Systems and Electromagnetic Compatibility*. Citeseer, 377–382.

[25] S Hamid Mortazavi, Zahra Moravej, and S Mohammad Shahrtash. 2018. A searching based method for locating high impedance arcing fault in distribution networks. *IEEE Transactions on Power Delivery* 34, 2 (2018), 438–447.

[26] Johnathan Mun. 2015. Understanding and choosing the right probability distributions. *Advanced Analytical Models* (2015), 899–917.

[27] A. D. Patil. 2023. *Stochastic Activity Net Models for Adaptive Protection*. Technical Report. University of Passau. 18 pages. https://www.fim.uni-passau.de/fileadmin/dokumente/fakultaeten/fim/forschung/mip-berichte/MIP-2301.pdf

[28] Binod Poudel, Daniel Ruiz Garcia, Ali Bidram, Matthew J Reno, and Adam Summers. 2021. Circuit topology estimation in an adaptive protection system. In *2020 52nd North American Power Symposium (NAPS)*. IEEE, 1–6.

[29] Hari Prasetijo and Daru Tri Nugroho. 2019. Overcurrent relays coordination: comparison characteristics standard inverse, very inverse and extremely inverse. In *Journal of Physics: Conference Series*, Vol. 1367. IOP Publishing, 1–9. 012051.

[30] William H Sanders and John F Meyer. 2001. Stochastic activity networks: formal definitions and concepts. *Lectures on Formal Methods and Performance Analysis: First EEF/Euro Summer School on Trends in Computer Science Bergen Dal, The Netherlands, July 3–7, 2000 Revised Lectures 1* (2001), 315–343.

[31] Mohammed K Shakhatreh, Artur J Lemonte, and Germán Moreno-Arenas. 2019. The log-normal modified Weibull distribution and its reliability implications. *Reliability engineering & System safety* 188 (2019), 6–22.

[32] Mads Graungaard Taul, Xiongfei Wang, Pooya Davari, and Frede Blaabjerg. 2019. Systematic approach for transient stability evaluation of grid-tied converters during power system faults. In *2019 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 5191–5198.

[33] Leon Thurner, Alexander Scheidler, Florian Schäfer, Jan-Hendrik Menke, Julian Dollichon, Friederike Meier, Steffen Meinecke, and Martin Braun. 2018. pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Transactions on Power Systems* 33, 6 (2018), 6510–6521.

[34] Kishor S Trivedi and Andrea Bobbio. 2017. *Reliability and availability engineering: modeling, analysis, and applications*. Cambridge University Press, 30–31.

[35] Bin Wang, Jianzhao Geng, and Xinzhou Dong. 2016. High-impedance fault detection based on nonlinear voltage-current characteristic profile identification. *IEEE Transactions on Smart Grid* 9, 4 (2016), 3783–3791.

[36] Yunqi Wang, Jayashri Ravishankar, and Toan Phung. 2016. A study on critical clearing time (CCT) of micro-grids under fault conditions. *Renewable Energy* 95 (2016), 381–395.

[37] Tesfaye Amare Zerihun and Bjarne E Helvik. 2019. Dependability of smart distribution grid protection using 5g. In *2019 3rd International Conference on Smart Grid and Smart Cities (ICSGSC)*. IEEE, 51–59.

Don't Touch the Power Line - A Proof-of-Concept for Aligned LLM-based Assistance Systems to Support the Maintenance in the Electricity Distribution System

SASCHA KALTENPOTH and OLIVER MÜLLER, Paderborn University, Germany

As a possible solution to the demographic change and the resulting knowledge loss due to retirements in the Energy sector, this study aimed to develop a generic pipeline to implement and evaluate proof-of-concepts (PoCs) for LLM-based assistance systems in new domains. Our pipeline contains an LLM-based data generation strategy based on documents, a retrieval-augmented generation (RAG) architecture utilizing prompting techniques on existing German LLMs, and an LLM-based automatic evaluation strategy. We leverage our pipeline to evaluate five LLMs using data from a German DSO. We found that the Llama3 and the Mistral model are appropriately aligned for the task. We plan to pilot the RAG architecture in the DSO's infrastructure for future research and continuously research improvements using the generated human demonstrations.

CCS Concepts: • **Information systems → Expert systems**; • **Computing methodologies → Natural language generation**; • **Hardware → Power networks**.

Additional Key Words and Phrases: Electricity Distribution Grid, Switch Gear Maintenance, Large Language Models, Assistance Systems

1 INTRODUCTION

Climate change is one of the biggest threats of our time [World Economic Forum 2024]. To mitigate climate change, the energy system must transform towards renewable energy sources and integrate electric vehicle charging infrastructure [European Parliament and the Council 2023]. The integration of renewable energy sources and electric vehicle charging is accompanied by voltage fluctuations and loading peaks [Basit et al. 2020; Smith et al. 2022], increasing stress on and maintenance effort of critical assets such as switchgears for distribution system operators (DSOs) [Hoffmann et al. 2020; zur Heiden et al. 2024]. Considering the demographic change in the Energy sector [European Commission Joint Research Center 2024], this is a certain issue, especially in the German *Energiewende* [Federal Ministry for Economic Affairs and Energy 2021]: a majority of the DSO's experienced technicians will retire within this decade, while the onboarding of new technicians may require up to three years, leading to loss of knowledge in the service and maintenance of the distribution system [zur Heiden and Kaltenpoth 2024].

A promising solution to knowledge loss emerged with the maturing of large language models (LLMs) like GPT-3 [Brown et al. 2020] and GPT-4 [OpenAI 2023]: assistance systems based on LLMs support employees by integrating with existing knowledge bases (e.g., internal guidelines and best practices) and facilitate the transfer of existing knowledge [Jiang et al. 2024; Kernan Freire et al. 2023]. Some concepts depend on closed-source cloud models such as GPT-3 [Kernan Freire et al. 2023], while others use open-source LLMs on-premies such as LLaMA 2 [Jiang et al. 2024]. Even concepts for assistance systems to support technicians in the service and maintenance of the distribution system have been developed, although no implementation exists yet [zur Heiden and Kaltenpoth 2024]. While several use cases provide proof-of-concept (PoC) of LLM-based assistance systems, including various domains such as medicine [Shi et al. 2023] and law [Cui et al. 2023], no PoC of LLM-based assistance systems in the maintenance of the distribution system in Germany yet exists. Additionally, LLM-based assistance systems can generate incorrect, off-topic, biased, or harmful content [Askell et al. 2021; Kenton et al. 2021], such as recommending technicians to touch an open power line, possibly leading to injuries or system failure in the distribution grid. Thus, the so-called alignment of LLMs, especially their helpfulness and honesty, must be considered in a PoC [Askell et al. 2021].

Therefore, this study first aims to develop an LLM-based assistance system pipeline for maintenance support in the German distribution grid based on existing PoCs, followed by a PoC including an evaluation of helpfulness and honesty in German LLMs. We build a PoC of an LLM-based assistance system for maintenance support in the German distribution Grid based on retrieval-augmented generation (RAG) [Lewis et al. 2020] of existing German DSO documents leveraging different German LLMs. We then generate a test set of 97 questions and answers based on the documents and analyze the honesty of the different German LLMs using automatic evaluation strategies to determine the most aligned model for piloting in a real-world scenario.

We found that two models, namely the LlaMA3-DiscoLM-8B [Harries et al. 2023] and the LeoLM-Mistral-7b [Plüster 2023], provide appropriate answers regarding helpfulness and honesty while obtaining high inference speed. While our general pipeline can support PoCs on LLM-based assistance systems using state-of-the-art (SOTA) aligned German LLMs, the results of our analysis provide insights into the alignment of German LLMs for technical domains such as the electricity distribution grid. We hope to encourage future research on LLM-based assistant systems in general, while our research directs to the collection of human data in a piloting scenario and extending the abilities of the LLM-based assistance system by further fine-tuning the LLM and incorporating human preferences.

The remainder of this paper is structured as follows. First, the background of especially LLMs and their alignment is given, followed by an analysis of the related work. Afterward, the pipeline of our PoC approach is explained, followed by the PoC's helpfulness and honesty analysis results. Finally, the results are discussed, and an outlook for future research is given.

Authors' address: Sascha Kaltenpoth, sascha.kaltenpoth@uni-paderborn.de; Oliver Müller, oliver.mueller@uni-paderborn.de, Paderborn University, Warburger Str. 100, Paderborn, North Rhine-Westfalia, Germany, 33098.

2 BACKGROUND

2.1 Large Language Models

Large language models (LLMs) such as GPT3 [Brown et al. 2020], GPT4 [OpenAI 2023], or LlaMA2 [Touvron et al. 2023] are a specific form of neural networks based on the Transformer architecture [Vaswani et al. 2017] that are highly capable in solving natural language processing (NLP) tasks. Furthermore, they need to be distinguished between encoder- and decoder-based LLMs. While both encoder and decoder are part of the initial transformer architecture, recent models usually use one part, as their special capabilities differ. While encoder-only models perform better at solving natural language understanding [Devlin et al. 2019], decoder-only models perform better at natural language generation [Brown et al. 2020; Touvron et al. 2023]. Decoder-only LLMs are trained on billions of word pieces, referred to as tokens, and contain billions of parameters [Teubner et al. 2023]. These generative models aim to generate meaningful text, given a sequence of tokens as input, and are trained to autoregressively predict the next token t_{n+1} based on the conditional probability distribution $P(t_{n+1}|t_1...t_n)$, where $t_1...t_n$ refers to the input sequence [Shanahan et al. 2023]. Such pre-trained LLMs are often called foundation models due to their capabilities in solving various tasks [Bommasani and et al. 2022].

2.2 LLM Alignment Basics

LLM alignment has its roots in the artificial intelligence (AI) alignment problem [Hadfield-Menell and Hadfield 2019]. The AI alignment problem arises when the specified reward for an AI model and the outcome or value originally intended by relevant humans (e.g., the designer or user) differ. Although this seems not to apply to LLMs, they can generate incorrect, off-topic, biased, or harmful content [Askell et al. 2021; Brown et al. 2020; Kenton et al. 2021; Touvron et al. 2023]. LLM alignment aims to align LLMs with human, organizational, and social values [Askell et al. 2021; Kenton et al. 2021]. This broad goal can be operationalized drawing on the definition of Askell et al. [2021]. They define an aligned agent as helpful, honest, and harmless. While a helpful agent can solve human-given tasks, honesty refers to giving accurate information [Askell et al. 2021]. In contrast, a harmless agent shall mitigate being offensive, discriminatory, biased, and refusing dangerous behavior, even neglecting other factors, e.g., preventing to be helpful building a bomb but staying harmless [OpenAI 2023].

2.3 LLM Alignment Approaches

The alignment and adaptation of LLMs can be applied using various approaches. While the current most prominent paradigm is prompting LLMs [Liu et al. 2023], LLMs can be further adapted to a specific domain or task by supervised fine-tuning it or incorporating human preferences [Ouyang et al. 2022].

Prompting describes the process of formulating unambiguous instructions, so-called prompts, and leveraging LLMs to generate responses without further training them [Liu et al. 2023]. Prompt engineering, in turn, is the optimization of prompts [Liu et al. 2023]. Two common examples are zero-shot and few-shot prompting [Liu et al. 2023]. Zero-shot prompting directly instructs the task without examples, while few-shot prompting refers to passing one or more examples of a downstream task before the actual prompt [Brown et al. 2020]. In domain-specific or knowledge-intensive scenarios, even humans need additional external knowledge to perform tasks. Retrieval-augmented generation (RAG) is a common strategy to address this [Lewis et al. 2020].

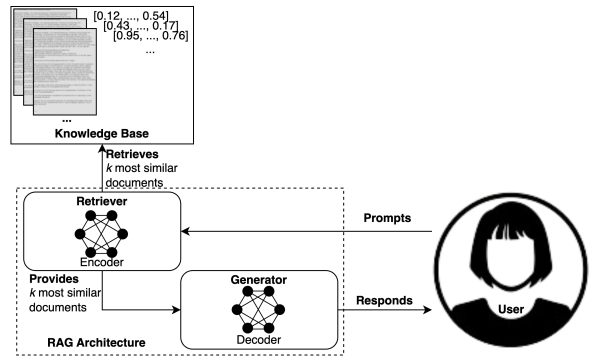


Fig. 1. Retrieval-augmented generation (RAG) (Adapted from [Lewis et al. 2020].)

As illustrated in Figure 1, the knowledge base comprises vector indices, referred to as embeddings. When the user prompts the architecture, the encoder-only retriever determines the k most similar documents using cosine similarity between the database indices and the user prompt embeddings to provide them to the generator. The decoder-only generator responds based on a special prompt that includes the documents provided by the retriever [Lewis et al. 2020]. This prompt is usually in an instruction style [Zhou et al. 2023]. The first part is the instruction, e.g., "Answer the question based on the context below," followed by the context, and finally, the question and an answer trigger phrase [Zhou et al. 2023]. Additionally, RAG architectures have been shown to make models more honest [Lewis et al. 2020].

Supervised fine-tuning (SFT) is simply fine-tuning an LLM based on high-quality human demonstrations [Ouyang et al. 2022]. Therefore, a prompt is derived from a dataset, and a human demonstrator generates the desired LLM output for the prompt [Ouyang et al. 2022]. This strategy has increased the quality, diversity, and alignment of generated texts [Ouyang et al. 2022; Touvron et al. 2023].

Besides SFT, another training and alignment strategy is utilizing human preferences. One approach is reinforcement learning from human feedback (RLHF) [Ouyang et al. 2022]. In this method, a pre-trained LLM produces several responses to a prompt, which human evaluators then rank based on quality [Ouyang et al. 2022]. Based on this ranking, a reward model is trained to automatically assess the responses, which is then used alongside a reference model (a static version of the LLM) to train a reinforcement learning-based LLM [Schulman et al. 2017]. RLHF-trained LLMs usually show higher alignment [Ouyang et al. 2022; Touvron et al. 2023].

2.4 LLM Evaluation

Evaluating LLMs can broadly be divided into human evaluation and automatic evaluation [Chang and et al. 2024]. As a specific automatic evaluation form, benchmark datasets can be used to evaluate LLMs

Table 1. Related Work on LLM-based assistance systems case studies

Domain	LLM	Size B param	Integration	On-Prem	Prompt	CoT	SFT	PoC	Helpful	Honest	German	Evaluation	Source
N/A	LLaMA	7	RAG	X	X	X	X	X	-	X	-	Automatic	[Jiang et al. 2024]
Textile	GPT-3	-	Other	-	X	X	-	X	-	-	-	Manual	[Kernan Freire et al. 2023]
Code	LLaMA2	7, 13, 70	RAG	X	X	-	X	X	-	X	-	Manual	[Liu et al. 2024]
Legal	LLaMA	13	RAG	X	X	X	X	X	-	X	-	Automatic	[Cui et al. 2023]
Cooking	GPT-3.5	-	RAG	-	X	-	-	-	-	-	-	Manual	[Chan et al. 2023]
Medicine	Vicuna	13	RAG	X	X	-	-	X	-	X	-	Automatic	[Shi et al. 2023]
DSO*	-	-	RAG	X	-	-	-	-	X	X	X	Manual	[zur Heiden and Kaltenpoth 2024]
DSO	**Table 2**	**Table 2**	**RAG**	**X**	**X**	**X**	**-**	**X**	**X**	**X**	**X**	**Automatic**	**ours**

*This study proposes a concept without PoC.

in specific tasks, such as question answering [Chang and et al. 2024]. While benchmarks are helpful to pre-evaluate LLMs, an additional evaluation in an open environment is required, like a proof-of-concept (PoC) [Chang and et al. 2024]. Human and sophisticated automatic evaluation can help to deliver such PoC. While human evaluation is based on human judgments or experiments [Askell et al. 2021; Chiang et al. 2024], sophisticated automatic evaluation leverages LLMs to judge the responses of other models, such as GPTScore or LLM-Eval [Fu et al. 2023; Lin and Chen 2023]. An LLM evaluator is prompted to provide feedback and scores to assess the output of another LLM [Chang and et al. 2024]. Automatic evaluation does not require intensive human participation, saving time and reducing the impact of human subjective factors [Chang and et al. 2024].

2.5 Related Work

Analyzing recent studies results in various approaches to implement LLM-assistants with knowledge bases that are mostly developed for general usage and evaluated on general question-answering benchmarks [Gao et al. 2024; Li et al. 2024; Zong et al. 2024]. These studies analyze general architectures, mostly evaluated on standard benchmarks. Therefore, this study only considers approaches that implement PoCs for specific use cases of LLM-based assistance systems. As stated in Table 1, multiple PoC studies of LLM-based assistance systems have been attempted in various domains, such as textile code generation and debugging [Liu et al. 2024] but also legal [Cui et al. 2023], or even cooking [Chan et al. 2023]. Although most studies do not explicitly name it, they all evaluate the honesty of their LLM-based assistance systems, except two studies that only test the impressiveness of the answers [Kernan Freire et al. 2023] and the user satisfaction [Chan et al. 2023]. Furthermore, all studies implement prompting (see Table 1), while some also implement the chain-of-thought (CoT) prompting method, which prompts the model to reason before responding to the main task [Wu et al. 2023]. Some studies further fine-tune their generator LLMs [Cui et al. 2023; Jiang et al. 2024; Kernan Freire et al. 2023], while other studies retain using the pre-trained LLMs due to their already high prompting capabilities [Chan et al. 2023; Liu et al. 2023; Shi et al. 2023]. While all recent PoC studies use existing datasets or manually generate their datasets, a common data generation method is utilizing LLMs as data generators [Yu et al. 2023]. Although an initial study exists on the development of LLM-based assistance systems to support the maintenance of the electricity distribution grid in Germany [zur Heiden and Kaltenpoth 2024], the authors do not apply a PoC. This is probably based on lacking existing LLM-based assistance systems studies on German LLMs. Thus, we conducted additional literature research on German LLMs, resulting in Table 2.

Table 2 lists all possible LLMs that were provided by different research groups and provided appropriate model descriptions on the HuggingFace platform [Wolf et al. 2020] or research papers, respectively. In fact, the first model in Table 2 is very small. However, the *German GPT2* [Bayr. Staatsbibliothek 2021] model is longly researched and serves as a baseline for this study. Although all models are German LLMs, except GPT2, all models are based on foundation models that were pre-trained on primary English language or multilingual data and are fine-tuned on German datasets to refactor their main language [Harries et al. 2023; Ostendorff and Rehm 2023; Plüster 2023].

Table 2. German LLM candidates for the PoC

Model*	Base	#Params	Source
GPT2	GPT2	1,5	[Bayr. Staatsbibliothek 2021]
IGEL	Bloom	6,4	[Ostendorff and Rehm 2023]
LeoLM	Mistral	7	[Plüster 2023]
	LLaMA2	13	
	LLaMA2	70	
DiscoLM	LeoLM	7	[Harries et al. 2023]
	LLaMA3	8	

* The model names are linked to the HuggingFace platform.

This is based on the high amount of computational resources needed to pre-train LLMs. While pre-training is only possible at full model size, prompting, SFT, and RLHF can be applied using recent parameter efficient fine-tuning and quantization techniques such as low-rank adaptation (LoRA) [Hu et al. 2022], and Q-LoRA [Dettmers et al. 2023].

3 METHODOLOGY

Figure 2 illustrates the pipeline of our PoC study and is explained in the following. Our PoC study is based on a RAG architecture that is based on prompting, as this is the most common strategy in

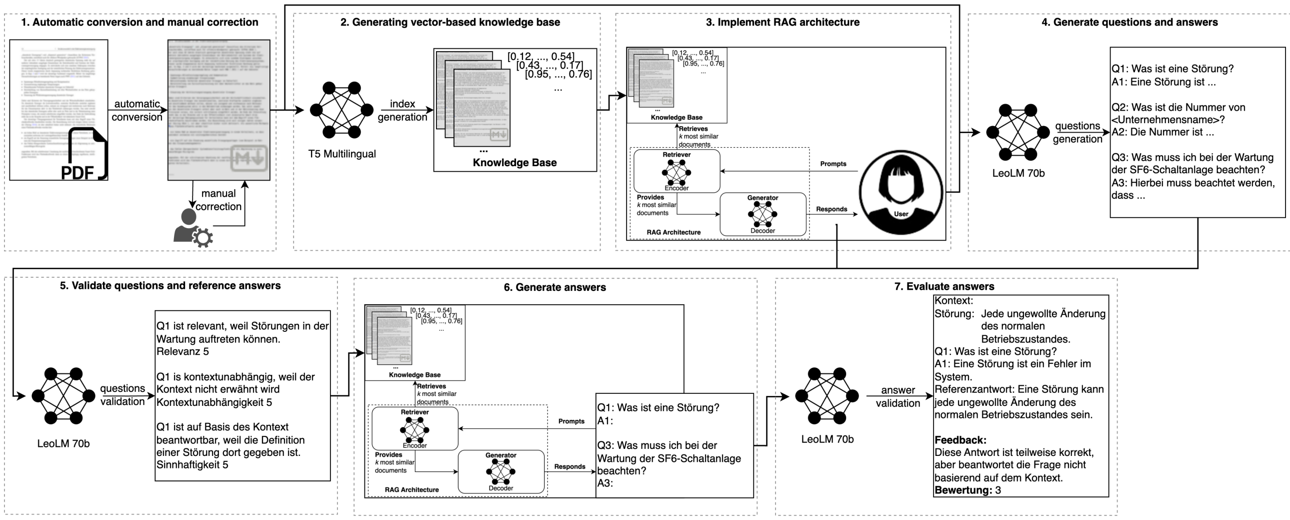


Fig. 2. PoC Pipeline[1]

the related work (see Table 1). All prompts used in this work are adapted from Roucher [2024], translated, and fitted to the task[1].

3.1 Data Acquisition and Preparation

First, we automatically convert the PDF files provided by the DSO into an LLM readable Markdown format using *PyMuPDF* [PyMuPDF 2024]. We then manually correct all the Markdown documents, removing unnecessary symbols and correcting header formatting in order to make them transferable into the RAG vector-based database.

As we aim to provide a PoC with an analysis of the German LLMs' alignment, we do not compare different retrievers and use the recent SOTA retriever model Multilingual-E5 [Wang et al. 2024]. We then leverage the retriever to generate a vector database using chunks of 512 tokens with 200 tokens overlap in order to provide an appropriate context without too much information loss to our LLM in the RAG architecture.

Afterward, we generate synthetic questions and answers in exchanging the use of GPT-4 with the German LeoLM-70b [Plüster 2023] model. The model exchange is necessary, as we cannot share the critical infrastructure data with external APIs. Therefore, we use the LeoLM-70b model to generate questions and answers based on the generated vector base chunks, which is more capable than the other recent German open-source models. We use an instruction-style prompt adapted to DSO-related terminology from Roucher [2024]. The instruction-style prompt first explains the task and provides a template that needs to be filled by the model to make the generated questions and answers easily processable. We generate 500 initial questions based on randomly drawn chunks from the vector database.

As the final step of the data generation, we automatically assess the generated questions using the LeoLM-70b. We assess the generated questions using groundedness, relevance, and context independence. Groundedness measures the extent to which the question is answerable by the provided context chunk. Relevance measures how relevant the question is for a technician working in maintenance. Context independence evaluates the extent to which the question contains direct or indirect questions to the text. An example is the question, "What is the appropriate action in case of a disturbance according to the text?". The reference "according to the text" is problematic, as no real person would reference a text instead of simply asking for the next step. We use single prompts for every question assessment dimension. They are in an instruction style that first instructs the review of the dimension, followed by a template containing a field for textual feedback and another field for a rating from 1-5. The preliminary feedback exploits the chain-of-thought (CoT) training of the models [Wu et al. 2023], improving the rating results.

After the scoring, we manually investigate a random subsample of the LLM assessments and set the groundedness and context independence minimum score to 4, while we decrease the minimum relevance score to 3, as the model tends to overshoot the relevance to the term "maintenance" during the assessment. We then filter the questions and answers resulting in a set of 97 test questions and answers.

3.2 The RAG architecture

As already mentioned, we use the Multilingual-E5 model [Wang et al. 2024] as a retriever providing the 5 most similar documents for the generator LLMs, except the German GPT2, which can only handle one document of 512 tokens [Bayr. Staatsbibliothek 2021]. Then, we employ a prompting template exemplarily illustrated in Figure 3 to answer the question based on the retrieved documents. The prompt is instruction-based [Zhou et al. 2023], instructing the general behavior and providing a context, followed by the question.

[1]We can only provide prompt templates and synthetic examples due to data protection requirements. Thus, an additional evaluation using open-source data, supporting the best performances of LlaMA3-DiscoLM-8b and LeoLM-Mistral-7b, can be found in our GitHub repository: https://github.com/skaltenp/dont_touch_the_powerline

Mit den Informationen aus dem Kontext geben Sie eine umfassende Antwort auf die Frage. Antworten Sie nur auf die gestellte Frage, Ihre Antwort sollte knapp und relevant zur Frage sein. Geben Sie die Nummer des Quelldokuments an, wenn dies relevant ist. Wenn die Antwort nicht aus dem Kontext abgeleitet werden kann, geben Sie keine Antwort.

Kontext:
{context}

Hier ist nun die Frage, die Sie beantworten müssen.

Frage: *{question}*

Fig. 3. RAG generator prompt

We compare five of the seven German LLMs on the RAG question-answering task to determine the most aligned and resource-efficient model. We exclude the IGEL model [Ostendorff and Rehm 2023], as the current implementation on the Huggingface platform has weight-loading issues. Additionally, we exclude the LeoLM-70b, as it is utilized to generate data and evaluate the other LLMs. As LLMs tend to prefer their own answers [Panickssery et al. 2024], a self-evaluation by the LLM would lead to evaluation bias.

We track the responses of all five considered models using the 97 questions generated in the data generation. Afterward, we start the evaluation using the following scheme.

3.3 Evaluation Scheme

We evaluate the answers of the five models using the automatic evaluation considering two factors: honesty and helpfulness. Thus, we specifically prompt the LeoLM-70b model. Figure 4 shows our prompt for the evaluation. Here, we again leverage a combination of instruction- and CoT-prompting to improve the LLM assessment.

Aufgabenbeschreibung:
Eine Anweisung (kann eine Eingabe enthalten), eine zu bewertende Antwort, eine Referenzantwort, die eine Punktzahl von 5 erhält, und ein Bewertungsschema, das ein Bewertungskriterium darstellt, werden gegeben.
Schreiben Sie ein detailliertes Feedback, das die Qualität der Antwort streng basierend auf dem gegebenen Bewertungsschema bewertet, ohne allgemein zu bewerten.
Nach dem Schreiben eines Feedbacks geben Sie eine Punktzahl an, die eine ganze Zahl zwischen 1 und 5 ist.
Sie sollten sich auf das Bewertungsschema beziehen.
Das Ausgabeformat sollte wie folgt aussehen: "Feedback: {{schreibe ein Feedback für die Kriterien}} [ERGEBNIS] {{ganze Zahl zwischen 1 und 5}}"
Bitte erzeugen Sie keine anderen Einleitungen, Abschlüsse und Erklärungen. Stellen Sie sicher, dass [ERGEBNIS] in Ihrer Ausgabe enthalten ist.

Die zu bewertende Anweisung:
{instruction}

Zu bewertende Antwort:
{response}

Referenzantwort (Punktzahl 5):
{reference_answer}

Bewertungskriterien:
[Ist die Antwort korrekt, genau und faktisch basierend auf der Referenzantwort?]
Punktzahl 1: Die Antwort ist völlig falsch, ungenau und/oder nicht faktisch.
Punktzahl 2: Die Antwort ist überwiegend falsch, ungenau und/oder nicht faktisch.
Punktzahl 3: Die Antwort ist teilweise korrekt, genau und/oder faktisch.
Punktzahl 4: Die Antwort ist größtenteils korrekt, genau und faktisch.
Punktzahl 5: Die Antwort ist vollständig korrekt, genau und faktisch.

Feedback:

Fig. 4. Evaluation Prompt

Our prompt provides the instruction, the LLM response, the (correct) reference answer, and the assessment criteria. We use mixed criteria of honesty and helpfulness, as their distinguishment is complex humans and LLMs. Therefore, a correct answer that explicitly answers the question based on the contexts is rated with a 5 (best value). Neglecting one dimension, like answering most correctly with minimal errors (i.e., losing honesty) based on the context but obtaining precision for the question (i.e., obtaining helpfulness), is rated with 4 points.

4 RESULTS

The results of our RAG architecture variants are summarized in Table 3. Table 3 shows that while the LLaMA3 version fine-tuned by Harries et al. [2023] performs best, most models perform nearly equally, except the significantly smaller German GPT2 model, which only reaches an average score below 3.

Table 3. Results of the automatic evaluation

Model	Normalized Score	Actual Score
German GPT2	0.466	2.866
LeoLM-Mistral-7b	0.629	3.515
LeoLM-LaMA2-13b	0.634	3.536
DiscoLM-Mistral-7b	0.660	3.639
LlaMA3-DiscoLM-8b	0.683	3.732

Figure 5 reveals more obvious differences between the automatic evaluation answer scores. While the German GPT2 model has a maximum score of 4 and a median score of 3, all other models have a median score of 4. However, most LLMs tend to be left-skewed, with most answer scores below or equal to the median. The LlaMA3-DiscoLM-8b model's score seems to be more centristic to its median.

Fig. 5. Distribution of LeoLM automatic evaluation scores

Considering the inference time stated in Table 4, the LlaMA3-DiscoLM-8b model outperforms all others.

Table 4. Inference time

Model	Inference time (s)	GPU util. (GB)
German GPT2	0.5	2
LeoLM-Mistral-7b	4	7.5
LeoLM-LlaMA2-13b	6	21.5
DiscoLM-Mistral-7b	18	7.5
LlaMA3-DiscoLM-8b	3.5	14.5

However, it utilizes times two GPU resources compared to the other mid-size LLMs between 7 and 13 billion parameters. Comparing the models considering inference time, GPU utilization, and alignment score, the LeoLM-Mistral-7b LLM performs second best due to its fast inference.

5 DISCUSSION AND OUTLOOK

As a possible solution to the demographic change and the resulting knowledge loss due to retirements in the Energy sector [Federal Ministry for Economic Affairs and Energy 2021; zur Heiden and Kaltenpoth 2024], this study aimed to develop the first PoC for LLM-based assistance systems to support the maintenance of the electricity distribution grid. Based on the existing PoC studies on LLM-based assistant systems from other domains (see Table 1), we developed a generic pipeline to implement and evaluate PoCs for LLM-based assistance systems in new domains. Our pipeline contains an LLM-based data generation strategy based on documents, an RAG architecture utilizing prompting techniques on existing German LLMs, and an LLM-based automatic evaluation strategy.

We then realized our pipeline using documents of a German DSO. We found that the LlaMA3-DiscoLM-8b and the LeoLM-Mistral-7b provide appropriate solutions for the LLM-based assistance system, while the LLaMA3-DiscoLM-8b model provides the best alignment according to our joined helpfulness and honesty metric. Although a high alignment is important when utilizing LLMs in real-world cases, GPU utilization and inference time are also important factors [Dettmers et al. 2023; Hu et al. 2022]. Considering the inference time, the LLaMA3-DiscoLM model outperforms the other models significantly, but considering the GPU resources, it needs twice as much GPU resources as the other models. Based on this, we recommend implementing the pipeline with one of the LlaMA3-DiscoLM-8b models when alignment of the model is most important (e.g., in the support of the distribution grid) while the LeoLM-Mistral-7b model is an appropriate alternative in environments with resource restrictions.

Naturally, this study is not without limitations. Firstly, this study only considered a prompting-based approach to RAG architectures for LLM-based assistance systems. While existing studies such as Jiang et al. [2024] and Liu et al. [2024] show that fine-tuning increases the performance of generator models in RAG architectures, human demonstrations are required to enable SFT [Ouyang et al. 2022], which simply were not available in the PoC phase. Another limitation is the automatic evaluation. While it significantly decreases evaluation time and human resources, the LLM-based evaluation can be biased [Chang and et al. 2024]. However, this study only provides a PoC, and an exhaustive human evaluation will follow during the piloting of the LLM-based assistance system.

While the mentioned limitations exist, to the best of our knowledge, this is the first general pipeline for new PoC. Furthermore, our PoC is the first for LLM-based assistance systems supporting the maintenance of the electricity distribution system in Germany. With the proposal of our pipeline, we hope to encourage research on LLM-based assistance systems in various domains, while our future research will further investigate LLM-based assistance systems in the energy sector, especially those for supporting maintenance in the electricity distribution grid. Based on the known limitations of this study, we will conduct additional evaluations and a piloting experiment to gather human demonstrations and preferences to supervised fine-tune the existing pipeline and incorporate human preferences. With further evaluations, we hope to gather insights on the alignment of LLMs in technical areas, such as the support of maintenance in the electricity distribution grid.

ACKNOWLEDGMENTS

This contribution was developed within the research and development project “AProSys – KI-gestützte Assistenz- und Prognosesysteme für den nachhaltigen Einsatz in der intelligenten Verteilnetztechnik”. The project is funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) under the funding code 03EI60090E and is supervised by Project Management Jülich.

REFERENCES

Amanda Askell, Yuntao Bai, Anna Chen, Dawn Drain, Deep Ganguli, Tom Henighan, Andy Jones, Nicholas Joseph, Ben Mann, Nova DasSarma, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Jackson Kernion, Kamal Ndousse, Catherine Olsson, Dario Amodei, Tom Brown, Jack Clark, Sam McCandlish, Chris Olah, and Jared Kaplan. 2021. A General Language Assistant as a Laboratory for Alignment. arXiv:2112.00861 [cs.CL]

Muhammad Abdul Basit, Saad Dilshad, Rabiah Badar, and Syed Muhammad Sami ur Rehman. 2020. Limitations, challenges, and solution approaches in grid-connected renewable energy systems. *International Journal of Energy Research* 44, 6 (May 2020), 4132–4162. https://doi.org/10.1002/er.5033

Bayr. Staatsbibliothek. 2021. https://huggingface.co/DiscoResearch

Rishi Bommasani and et al. 2022. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258 [cs.LG]

Tom Brown et al. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Szeyi Chan, Jiachen Li, Bingsheng Yao, Amama Mahmood, Chien-Ming Huang, Holly Jimison, Elizabeth D Mynatt, and Dakuo Wang. 2023. "Mango Mango, How to Let The Lettuce Dry Without A Spinner?": Exploring User Perceptions of Using An LLM-Based Conversational Assistant Toward Cooking Partner. arXiv:2310.05853 [cs.HC]

Yupeng Chang and et al. 2024. A Survey on Evaluation of Large Language Models. *ACM Trans. Intell. Syst. Technol.* (jan 2024). https://doi.org/10.1145/3641289 Just Accepted.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. arXiv:2403.04132 [cs.AI]

Jiaxi Cui, Zongjian Li, Yang Yan, Bohua Chen, and Li Yuan. 2023. ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases. arXiv:2306.16092 [cs.CL]

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient Finetuning of Quantized LLMs. In *Thirty-seventh Conference on Neural Information Processing Systems*. https://openreview.net/forum?id=OUIFPHEgJU

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

European Commission Joint Research Center. 2024. Do we have sufficient skills for the energy transition in the changing labour market? https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/do-we-have-sufficient-skills-energy-transition-changing-labour-market-2024-01-16_en

European Parliament and the Council. 2023. Directive (EU) 2023/2413 of the European Parliament and of the Council of 18 October 2023 amending Directive (EU) 2018/2001, Regulation (EU) 2018/1999 and Directive 98/70/EC as regards the promotion of energy from renewable sources, and repealing Council Directive (EU) 2015/652. *Official Journal of the European Union L series* (2023), 1–77. http://data.europa.eu/eli/dir/2023/2413/oj

Federal Ministry for Economic Affairs and Energy. 2021. The Energy of the Future: 8th Monitoring Report on the Energy Transition – Reporting Years 2018 and 2019. https://www.bmwk.de/Redaktion/EN/Publikationen/8th-monitoring-report-the-energy-of-the-future.html

Jinlan Fu, See-Kiong Ng, Zhengbao Jiang, and Pengfei Liu. 2023. GPTScore: Evaluate as You Desire. arXiv:2302.04166 [cs.CL]

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv:2312.10997 [cs.CL]

Dylan Hadfield-Menell and Gillian K. Hadfield. 2019. Incomplete Contracting and AI Alignment. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 417–422. https://doi.org/10.1145/3306618.3314250

Jan Philipp Harries, Björn Plüster, and Daniel Auras. 2023. https://huggingface.co/DiscoResearch

Martin W. Hoffmann, Stephan Wildermuth, Ralf Gitzel, Aydin Boyaci, Jörg Gebhardt, Holger Kaul, Ido Amihai, Bodo Forg, Michael Suriyah, Thomas Leibfried, Volker Stich, Jan Hicking, Martin Bremer, Lars Kaminski, Daniel Beverungen, Philipp zur Heiden, and Tanja Tornede. 2020. Integration of Novel Sensors and Machine Learning for Predictive Maintenance in Medium Voltage Switchgear to Enable the Energy and Mobility Revolutions. *Sensors* 20, 7 (April 2020), 2099. https://doi.org/10.3390/s20072099

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*. https://openreview.net/forum?id=nZeVKeeFYf9

Feihu Jiang, Chuan Qin, Kaichun Yao, Chuyu Fang, Fuzhen Zhuang, Hengshu Zhu, and Hui Xiong. 2024. Enhancing Question Answering for Enterprise Knowledge Bases using Large Language Models. arXiv:2404.08695 [cs.CL]

Zachary Kenton, Tom Everitt, Laura Weidinger, Iason Gabriel, Vladimir Mikulik, and Geoffrey Irving. 2021. Alignment of Language Agents. arXiv:2103.14659 [cs.AI]

Samuel Kernan Freire, Mina Foosherian, Chaofan Wang, and Evangelos Niforatos. 2023. Harnessing Large Language Models for Cognitive Assistants in Factories. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (<conf-loc>, <city>Eindhoven</city>, <country>Netherlands</country>, </conf-loc>) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 44, 6 pages. https://doi.org/10.1145/3571884.3604313

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 9459–9474. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

Zhenyu Li, Sunqi Fan, Yu Gu, Xiuxing Li, Zhichao Duan, Bowen Dong, Ning Liu, and Jianyong Wang. 2024. FlexKBQA: A Flexible LLM-Powered Framework for Few-Shot Knowledge Base Question Answering. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 17 (March 2024), 18608–18616. https://doi.org/10.1609/aaai.v38i17.29823

Yen-Ting Lin and Yun-Nung Chen. 2023. LLM-Eval: Unified Multi-Dimensional Automatic Evaluation for Open-Domain Conversations with Large Language Models. In *Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023)*, Yun-Nung Chen and Abhinav Rastogi (Eds.). Association for Computational Linguistics, Toronto, Canada, 47–58. https://doi.org/10.18653/v1/2023.nlp4convai-1.5

Mingjie Liu, Teodor-Dumitru Ene, Robert Kirby, Chris Cheng, Nathaniel Pinckney, Rongjian Liang, Jonah Alben, Himyanshu Anand, Sanmitra Banerjee, Ismet Bayraktaroglu, Bonita Bhaskaran, Bryan Catanzaro, Arjun Chaudhuri, Sharon Clay, Bill Dally, Laura Dang, Parikshit Deshpande, Siddhanth Dhodhi, Sameer Halepete, Eric Hill, Jiashang Hu, Sumit Jain, Ankit Jindal, Brucek Khailany, George Kokai, Kishor Kunal, Xiaowei Li, Charley Lind, Hao Liu, Stuart Oberman, Sujeet Omar, Sreedhar Pratty, Jonathan Raiman, Ambar Sarkar, Zhengjiang Shao, Hanfei Sun, Pratik P Suthar, Varun Tej, Walker Turner, Kaizhe Xu, and Haoxing Ren. 2024. ChipNeMo: Domain-Adapted LLMs for Chip Design. arXiv:2311.00176 [cs.CL]

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* 55, 9, Article 195 (jan 2023), 35 pages. https://doi.org/10.1145/3560815

OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

Malte Ostendorff and Georg Rehm. 2023. Efficient Language Model Training through Cross-Lingual and Progressive Transfer Learning. https://doi.org/10.48550/ARXIV.2301.09626

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 27730–27744. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf

Arjun Panickssery, Samuel R. Bowman, and Shi Feng. 2024. LLM Evaluators Recognize and Favor Their Own Generations. arXiv:2404.13076 [cs.CL]

Björn Plüster. 2023. LeoLM: Igniting German-Language LLM Research. https://laion.ai/blog/leo-lm/

PyMuPDF. 2024. https://pymupdf.readthedocs.io/en/latest/

Aymeric Roucher. 2024. https://huggingface.co/learn/cookbook/rag_evaluation

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347 [cs.LG]

Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* (11 2023). https://doi.org/10.1038/s41586-023-06647-8

Yucheng Shi, Shaochen Xu, Zhengliang Liu, Tianming Liu, Xiang Li, and Ninghao Liu. 2023. MedEdit: Model Editing for Medical Question Answering with External Knowledge Bases. arXiv:2309.16035 [cs.CL]

Oliver Smith, Oliver Cattell, Etienne Farcot, Reuben D. O'Dea, and Keith I. Hopcraft. 2022. The effect of renewable energy incorporation on power grid stability and resilience. *Science Advances* 8, 9 (March 2022). https://doi.org/10.1126/sciadv.abj6734

Timm Teubner, Christoph M. Flath, Christof Weinhardt, Wil van der Aalst, and Oliver Hinz. 2023. Welcome to the Era of ChatGPT et al.: The Prospects of Large Language Models. *Business amp; Information Systems Engineering* 65, 2 (March 2023), 95–101. https://doi.org/10.1007/s12599-023-00795-x

Hugo Touvron et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Multilingual E5 Text Embeddings: A Technical Report. arXiv:2402.05672 [cs.CL]

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Qun Liu and David Schlangen (Eds.). Association for Computational Linguistics, Online, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

World Economic Forum. 2024. The Global Risks Report 2024: Insight Report. https://www.weforum.org/publications/global-risks-report-2024/

Dingjun Wu, Jing Zhang, and Xinmei Huang. 2023. Chain of Thought Prompting Elicits Knowledge Augmentation. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 6519–6534. https://doi.org/10.18653/v1/2023.findings-acl.408

Yue Yu, Yuchen Zhuang, Jieyu Zhang, Yu Meng, Alexander J Ratner, Ranjay Krishna, Jiaming Shen, and Chao Zhang. 2023. Large Language Model as Attributed Training Data Generator: A Tale of Diversity and Bias. In *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., 55734–55784. https://proceedings.neurips.cc/paper_files/paper/2023/file/ae9500c4f5607caf2eff033c67daa9d7-Paper-Datasets_and_Benchmarks.pdf

Wenxuan Zhou, Sheng Zhang, Hoifung Poon, and Muhao Chen. 2023. Context-faithful Prompting for Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 14544–14556. https://doi.org/10.18653/v1/2023.findings-emnlp.968

Chang Zong, Yuchen Yan, Weiming Lu, Jian Shao, Eliot Huang, Heng Chang, and Yueting Zhuang. 2024. Triad: A Framework Leveraging a Multi-Role LLM-based Agent to Solve Knowledge Base Question Answering. arXiv:2402.14320 [cs.CL]

Philipp zur Heiden and Sascha Kaltenpoth. 2024. Knowledge Management for Service and Maintenance on the Distribution Grid—Conceptualizing an Assistance System based on a Large Language Model. *HMD Praxis der Wirtschaftsinformatik* (May 2024). https://doi.org/10.1365/s40702-024-01074-3

Philipp zur Heiden, Jennifer Priefer, and Daniel Beverungen. 2024. Predictive Maintenance on the Energy Distribution Grid–Design and Evaluation of a Digital Industrial Platform in the Context of a Smart Service System. *IEEE Transactions on Engineering Management* 71 (2024), 3641–3655. https://doi.org/10.1109/TEM.2024.3352819

Practical Application of Energy-related Use Case Methodology and SGAM Framework: A Study of User Perspectives and Challenges

RENÉ KUCHENBUCH, PRANIT NALE, LAURA NIEMANN, and OLIVER WERTH, OFFIS e.V., Germany
JÜRGEN SAUER, University of Oldenburg, Germany

The IEC 62559 Use Case Methodology and the Smart Grid Architecture Model Framework (SGAM) play a crucial role in energy-related national and European research and innovation projects, facilitating a shared understanding among stakeholders in ICT-based Smart Grid solutions. However, their adoption presents challenges due to the complexity and diverse expert backgrounds involved, which can lead to ambiguous IEC 62559 Use Cases and SGAM Models, and consequently, potentially wrong decisions in the implementation by inaccurate descriptions or models. This research identifies the challenges faced by IEC 62559 Use Cases and SGAM Model modelers, as well as their working processes utilizing the toolchain. To achieve this, a survey was conducted among experts experienced in modeling IEC 62559 Use Cases and SGAM Models, followed by qualitative content analysis of the collected data. The findings enhance the usage of the standard/framework and toolchain by improving the understanding of creators' challenges. The results aim to enhance both energy-related and methodological research projects by leveraging modelers' insights into project conditions and processes. This includes ensuring that required information are available on creating, establishing collaborative structures for SGAM modeling, and advancing the maturity of the Use Case Management Repository. The insights support the development of future qualitative SGAM Models and IEC 62559 Use Case descriptions, thereby raising awareness of potential issues.

CCS Concepts: • **Hardware → Power and energy**; • **Software and its engineering → Requirements analysis**.

Additional Key Words and Phrases: Power and energy, Requirement Engineering, Interoperability, Smart Grid Architecture Model, IEC 62559 Use Case Methodology

Availability of Data and Material:
The material is available on https://osf.io/x4cwg/?view_only=4f3ebe8af26f4c2682d1abea5e519fe3

1 INTRODUCTION

The utilization of the IEC 62559 Use Case methodology managed by the IEC TC8 [IEC TC 8 2015] in conjunction with the Smart Grid Architecture Model (SGAM) Framework developed under the EU mandate M/490 [CEN-CENELEC-ETSI 2012] during the Requirements Engineering phase of Information and Communication Technology (ICT) energy projects represents a pivotal stride in ensuring a shared comprehension of the System-of-Interest. Principal outcomes resulting from the application of both methodologies encompass IEC 62559 Use Case descriptions and SGAM Models, which not only encapsulate a depiction of the System-of-Interest but also elucidate its intricacies in relation to other systems (in context of System-of-Systems) [Uslar et al. 2013]. Enabling tools, such as the Use Case Management Repository (UCMR) and the SGAM Toolbox, serve to facilitate authors in the development and deployment of these artifacts [Gottschalk et al. 2017; Neureiter et al. 2014].

Authors' addresses: René Kuchenbuch, rene.kuchenbuch@offis.de; Pranit Nale, pranit.jagannath.nale@offis.de; Laura Niemann, laura.niemann@offis.de; Oliver Werth, oliver.werth@offis.de, OFFIS e.V., Escherweg 2, Oldenburg, Lower Saxony, Germany, 26121; Jürgen Sauer, University of Oldenburg, Ammerländer Heerstraße 114-118, Oldenburg, Lower Saxony, Germany, 26129, juergen.sauer@uol.de.

Developing IEC 62559 Use Case descriptions and SGAM Models is a non-trivial task [Gottschalk et al. 2017; Paustian et al. 2022], resulting in inherent challenges. Furthermore the quality of IEC 62559 Use Case descriptions and SGAM Models profoundly influences the development of interoperable ICT-based energy systems, underscoring the imperative to optimize their quality [Kuchenbuch et al. 2023]. Previous investigations have delved into the quality attributes of IEC 62559 Use Case descriptions and SGAM Models, as well as strategies for their amelioration [Kuchenbuch et al. 2023]. The challenges associated with their creation can adversely affect their quality. By identifying these challenges, an awareness of the pertinent issues can be cultivated, serving as a foundation for subsequent research endeavors. Consequently, this research contribution seeks to discern the challenges encountered during the formulation of IEC 62559 Use Case descriptions and the utilization of the associated toolchain. This problem statement begets the following research inquiries:

RQ 1 What challenges can be identified in the process of creating IEC 62559 Use Case descriptions and SGAM Models?

RQ 2 Which tools are used in the creation of IEC 62559 Use Case descriptions and SGAM Models, and what challenges are encountered when utilizing the toolchain for these artifacts?

2 THEORETICAL BACKGROUND

To provide a better understanding, we provide some background on the research context in the following. The shift to Smart Grids presents challenges in digitization due to increased diversity in energy sector actors, technologies, and business models [Uslar et al. 2013]. Effectively managing this complexity requires identifying new standards and employing robust Requirement Engineering practices. This chapter introduces the IEC 62559 Use Case Methodology and the SGAM Framework roughly, established approaches for energy-related ICT projects [Uslar et al. 2013]. Emphasizing static and dynamic aspects, these methodologies elucidate the behavior of the System-of-Concern within the System-of-Systems [Gottschalk et al. 2017], with SGAM Models relying on IEC 62559 Use Case descriptions for essential information.

2.1 IEC 62559 Use Case Methodology

Use cases play a pivotal role in the conventional requirements engineering process for the identification of stakeholders' desires [Cockburn 2001]. A use case is defined as the "specification of a set of actions performed by a system, which yields an observable result that is, typically, of value for one or more actors or other stakeholders of the system" [IEC TC 8 2015]. The IEC 62559 Use Case Methodology is an established standard from the energy sector, which defines the overall process, (tool) environment and also the structure of the template (in particular IEC 62559-2 Use Case Template) [IEC TC 8 2015]. The standard consists of the following sub-standards:

- IEC 62559-1: Concept and Processes in Standardization
- IEC 62559-2: Definition of the templates for use cases, actor list and requirements list
- IEC 62559-3: Definition of use case template artefacts into an XML serialized format
- IEC 62559-4: Best Practices in Use Case Development for IEC processes and company projects (which also outlines the definition of the Use Case Management Repository (UCMR) for the management of use cases)

The template is defined by the IEC 62559-2 standard and provides the structure for IEC 62559 Use Case descriptions and is structured as follows [IEC TC 8 2015]:

(1) Description of the Use Case (metadata)
(2) Diagrams of Use Case
(3) Technical Details (actors and references)
(4) Step-by-Step Analysis of Use Case (scenario description)
(5) Information Exchanged
(6) Requirements (Non-Functional Requirements)
(7) Common Terms and Definitions
(8) Custom Information

2.2 Smart Grid Architecture Model (SGAM)

The SGAM serves as a reference model for technology-neutral analysis, visualization, and comparison of Smart Grid-related use cases within a framework, facilitating collaboration with diverse stakeholders [CEN-CENELEC-ETSI 2012]. It depicts implementation scenarios within SGAM Models to illustrate the scope of energy-related projects and their dependencies within the System-of-Systems [CEN-CENELEC-ETSI 2012]. Utilizing a three-dimensional representation (see Figure 1), the SGAM consists of [CEN-CENELEC-ETSI 2012]:

Interoperability Layers : Derived from the GWAC Interoperability Stack, five interoperability layers are defined. These include the Component Layer, representing involved components; the Communication Layer, specifying interfaces for information exchange; the Information Layer, depicting semantic information exchange; the Function Layer, illustrating functionalities of use cases; and the Business Layer, representing business goals and regulations.

SGAM Domains Components are situated within context-specific domains based on the energy value chain. These domains encompass Generation, Transmission, Distribution, Decentralized Energy Resources, and Customer Premises.

SGAM Zones Components are placed within SGAM zones, including Process, Field, Station, Operation, Enterprise, and Market.

3 METHODOLOGY

An empirical investigation into both the working and challenges associated with the development of IEC 2559 use case descriptions and SGAM Models can provide further insights from the creator's perspective. To gain a comprehensive understanding of the subject matter, descriptive data from users with varying levels of experience

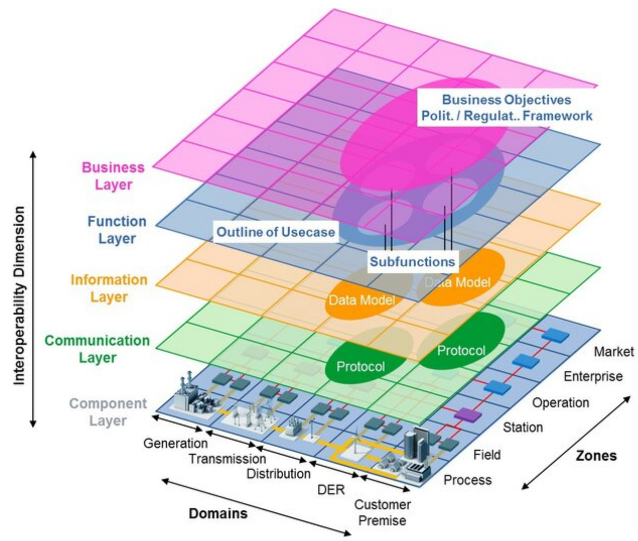


Fig. 1. Smart Grid Architecture Model [CEN-CENELEC-ETSI 2012]

are gathered. To address the research inquiries, an initial survey are conducted to unearth the challenges faced by IEC 62559 Use Case description creators and SGAM Model developers, as well as their challenges with the accompanying toolchain. We collected data through an online survey, containing open-and closed questions, which requires the application of a mix of quantitative analysis methods and qualitative content analysis (following the Kuckartz method [Kuckartz and McWhertor 2014]) for the interpretation of descriptive data.

3.1 Survey Design

The survey undertaken in this study was designed as an online questionnaire, implemented via the Limesurvey tool, with one instance provided by *OFFIS e.V.*. The online survey was selected to utilize this method due to its easy accessibility to potential respondents, the ability to use various question formats, and the specific provision of question items [Jamsen and Corley 2007]. It is possible to obtain an objective opinion of a mixed group of participants. To ensure reliability, participants must already have experience with the creation of IEC 62559 Use Case descriptions and SGAM Models, what is taken into account within the questionnaire. The goal of this study is to capture the experiences of a diverse cohort of IEC 62559 Use Case description and SGAM Model creators, spanning various levels of expertise. Participants were encouraged to articulate the challenges they confronted both in the work of the given standards and the application of the toolchain.

Conducted throughout the month of June/July 2023, the survey deliberately eschewed any extrinsic incentives, such as voucher-based inducements. The survey was deployed via mail distribution lists within the confines of the *OFFIS e.V.*. and current ongoing energy-context projects *int:net* and *Redispatch 3.0* and therefore provided both in German and English. As the survey also deals comprehensively with the recording of requirements for the development of an AI system, only a subset of the survey is used for this study.

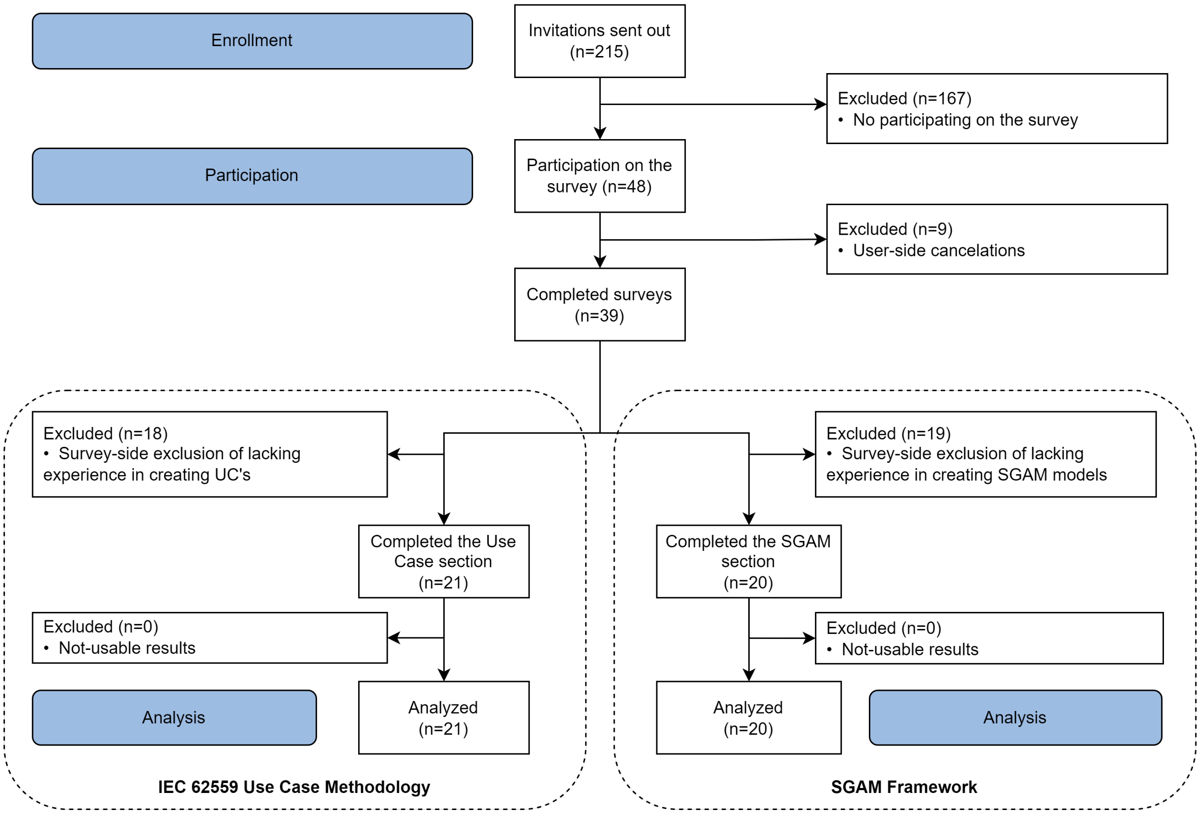


Fig. 2. Flow diagram of survey participants

Section A + B (General profile information UC / SGAM), Section C + E (Creation of IEC 62559 UC / SGAM Models) and Section G (Finalization of the questionnaire / Metadata) represent the relevant data for the study, whereby Section C, E and G can only be entered if it is stated in A + B that IEC 62559 Use Case descriptions and/or SGAM Models have already been created. The estimated approximate duration for the completion of the entire survey, based on the responses of two pilot participants, is 28 minutes. The collected data has been anonymized by Limesurvey, which involved deactivating IP tracking and not storing timestamps. As a result, no personally identifiable information is available.

The number of survey participations and the filtering is shown in Figure 2 as a flow diagram. n_{invite}=215 individuals were invited, determined from estimates based on the e-mail lists of *OFFIs e.V.* (74), *int:net* project (59), and *Redispatch 3.0* project (82). Of these, $n_{part} = 48$ participated in the survey. Given the survey's focus on experience in creating IEC 62559 Use Case descriptions and/or SGAM Models, only $n_{UC} = 21$ qualified for the IEC 62559 UC-part, and n_{SGAM}=20 for the SGAM part. Further details on survey respondent metadata are available in section 4.

3.2 Data Analysis

Qualitative content analysis is a systematic method for describing collected material and can be used for data analysis. This method is particularly useful for interpreting self-collected data [Schreier 2012]. The data analysis methodology in this study closely adhered to the principles of ’"Qualitative Text Analysis" by Udo Kuckartz [Kuckartz and McWhertor 2014]. This approach places particular importance on developing categories from the material [Schreier 2014], which was considered particularly important in this context.

The initial phase of our analysis involved a thorough familiarization with the dataset obtained from our survey. This dataset, rich in qualitative data, was methodically reviewed to gain a foundational understanding and to inform the subsequent analytical processes. The dataset included structured data (predefined answer options, such as metadata or demographic-related data), and unstructured data (open responses). Deductive coding was applied to structured data. This step facilitated the organization of the dataset into clear, quantifiable categories. However, the core of our qualitative analysis lies in the inductive coding of open-ended responses, particularly those detailing challenges and tool challenges users face in creating IEC 62559 Use Case descriptions and SGAM Models.

Utilizing the qualitative content analysis acc. Kuckartz [Kuckartz and McWhertor 2014], a coding process is executed. Each textual element (responses) is segmented into frames through sentences, comma-separated lists, and bullet points. These segments are coded according to existing codes when applicable; alternatively, codes are inductively added to the coding scheme based on semantics. The coded frames are then aggregated across the entirety of the

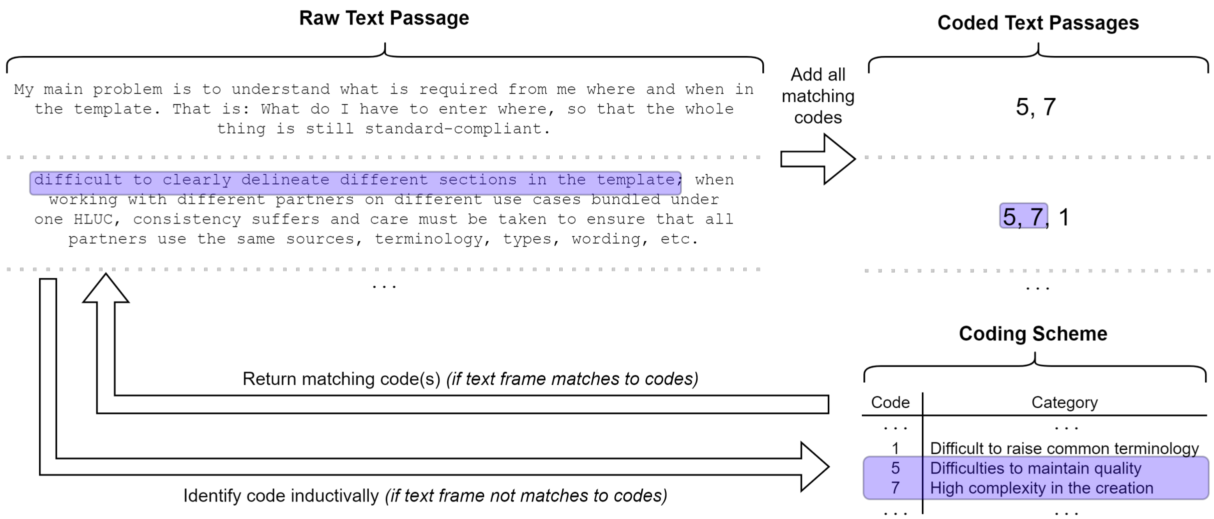


Fig. 3. Exemplary Coding for "IEC 62559 Use Case Creation Challanges"

response, thereby facilitating quantification. A depiction of this process is provided in Figure 3.

Following the coding phase, we engaged in thematic analysis to discern patterns, themes, and relationships within the data, aligning with our research questions. Each code was carefully categorized into broader thematic categories that best represented the collective insights. This thematic categorization was essential for structuring our qualitative data into analyzable units. Moreover, to ensure the intersubjectivity, minimizing biases, comprehensive data exploration, and reliability of our findings, individual coding was initially performed by different authors (n=2) of this study. The themes and categories for the final coding scheme were established through reflective discussions and a consensus-building process among the research team. Further, the dataset was manually coded according to this final scheme in "Microsoft Excel".

Due the envision usage of tools acc. [Kuckartz and McWhertor 2014], we utilized Python including Pandas, Numpy libraries for frequency analysis and Matplotlib and Seaborn for graph plotting as as established Python data analytics libraries [Stančin and Jović 2019].

Subsequent to the frequency analysis, we conducted a detailed code analysis to interpret the frequency of each code in relation to our research questions.

4 SURVEY RESPONDENTS' METADATA

Out of the 39 surveys evaluated, 21 were eligible for the IEC 62559 Use Case section, while 20 were eligible for the SGAM section, as depicted in Figure 2. Table 1 indicates that only surveys from respondents who had already created IEC 62559 Use Cases were evaluated in the use case section. For the SGAM section, individuals who have created SGAM Models or work with them evaluated the surveys. This guarantees that the survey participants possess adequate experience and expertise to conduct the survey effectively. Table 1 includes metadata on the participants' experience with these methods and their specific areas of focus. It should therefore be noted that even those who identify as "Beginners" have either created a IEC 62559 Use Case and/or worked with SGAM Models, indicating some level of experience in this area.

As the last section of the survey contains the demographic section, the collected demographic data is only that of qualified participants who have also completed the survey in full. The demographic data relates on the age groups, but also level of education, the program of study and the preferred language on the survey. These are listed in Table 2.

Table 1. Overview about the unfiltered methodology usage metadata

Characteristic	UC: n (%)	SGAM: n (%)
Experience		
Yes, also created IEC 62559 Use Cases / SGAM Models	21 (53.85 %)	20 (54.05 %)
Yes, but not created IEC 62559 Use Cases / SGAM Models yet	3 (7.69 %)	9 (24.32 %)
No, but generally with use cases	10 (25.64 %)	-
No	5 (12.82 %)	8 (21.62 %)
Experience Level		
Expert	2 (8.33 %)	3 (10.34 %)
Advanced	16 (66.67 %)	9 (31.03 %)
Beginner	6 (25 %)	17 (58.62 %)
Working Intensity		
High	3 (12.50 %)	2 (6.90 %)
Moderate	14 (58.33 %)	17 (58.62 %)
Little	7 (29.17 %)	10 (34.48 %)
Roles		
Method-oriented	21 (67.74 %)	18 (51.43 %)
Technical-oriented	7 (22.58 %)	14 (40 %)
Part of standardization bodies	3 (9.68 %)	3 (8.57 %)
Created types of IEC 62559 Use Cases		
High Level Use Case	13 (38.24 %)	-
Low Level Use Case	5 (14.71 %)	-
Business Level Use Case	6 (17.65 %)	-
System Use Case	10 (29.41 %)	-

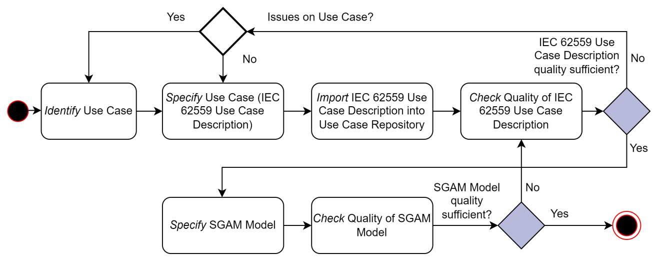


Fig. 4. Process of creating IEC 62559 Use Case descriptions and SGAM Models

5 SURVEY RESULTS

This section delves into the empirical findings from our survey, focusing on the IEC 62559 Use Case Methodology and the Smart Grid Architecture Model (SGAM) development from the perspectives of their creators. Through a comprehensive analysis of survey responses, this section sheds light on the prevalent issues in use case and SGAM creation and the effectiveness and limitations of the tool chain, offering a comprehensive overview of the current landscape in requirement engineering of ICT-related energy systems. The process (depicted in a UML flowchart) in Figure 4 is assumed when creating IEC 62559 Use Case descriptions and SGAM Models. This study deals with the challenges within the creation but also the quality assurance of IEC 62559 Use Case descriptions and SGAM Models phases. The challenges in the "Identify Use Case" phase are also considered, as this phase has a direct impact on the IEC 62559 Use Case descriptions and SGAM Models quality. This study also analyses tool usages and their challenges, which could also have an impact on quality in the phases.

5.1 IEC 62559 Use Case Methodology

Our analysis within this subsection explores the challenges faced during the creation of IEC 62559 Use Cases and the tools employed to facilitate this process. This subsection reveals insights into the complexities and toolchain efficacy, highlighting areas for improvement in practice and methodology.

5.1.1 Challenges. In the context of Requirement Engineering of ICT-related energy systems, several challenges surface during the creation of IEC 62559 Use Cases could be identified. Figure 5 shows

Table 2. Overview about the demographic information

Characteristic	Participants	Characteristic	Participants
Age		**Level of Education**	
21 - 29 y.	8 (42.11 %)	PhD	4 (20 %)
30 - 39 y.	6 (31.58 %)	Master	14 (70 %)
40 - 49 y.	3 (15.79 %)	Diploma	1 (5 %)
50 - 59 y.	2 (10.53 %)	Bachelor	1 (5 %)
Pref. Language		**Program of Study**	
German	25 (64 %)	Electrical engineering	6 (31.59 %)
English	14 (36 %)	Computer science	6 (31.59 %)
		Business science	3 (15.79 %)
		Information Systems Research	2 (10.53 %)
		Physics	1 (5.26 %)
		Chemistry	1 (5.26 %)

that "**High complexity in the creation**" holds the highest frequency at 29.17 % in our data set. Notably, as reported by participants in our survey, the linear progression inherent in use case development often necessitates revisiting and reworking initial ideas. This way of working is envisioned and necessary for precise use case development [Cockburn 2001]. Moreover, as one of the advanced (in accordance with the metadata) use case developers stated, the lack of clarity in technical knowledge across all facets of the use case remains a persistent issue, makes it difficult for one of the survey participants to choose the right level of abstraction. Furthermore, that advance user also expressed concern about the difficulty of creating diagrams and pointed out a void in enhanced tool support. Additionally, the user has also described the viewpoint about the fundamental issue of standard, that segregating short and long descriptions within use cases may be unnecessary, advocating for consolidating these descriptions into a single comprehensive narrative.

Simultaneously, "**Difficulties to maintain quality**" with a frequency of 20.83 %, is directly concerned on the maintenance of quality which could lead to misunderstanding of the System-of-Concern and therefore could result in consequential errors and higher costs, even at late stages of projects [Haskins et al. 2004]. Moreover, as one of the method expert noted, "Difficult to clearly delineate different sections in the template" may be one of the important factors that consistency suffers. Furthermore, the user also indicated that the involvement of multiple partners adds complexity as the variety of resources used for creation increases.

Moreover, the other challenge of "**Difficult to raise common terminology**" appeared with a frequency of 16.67 %. This challenge revolves around using a consistent vocabulary and language among all experts. Achieving the consensus is complex due to the diversity of backgrounds and perspectives among experts, making it challenging to effectively align terminologies [Paustian et al. 2022]. Moreover, one of the survey participants who had worked regularly on use cases also remarked that "When working with different partners on different use cases bundled under one High-Level Use Case, consistency suffers and care must be taken to ensure that all partners use the same sources, terminology, types, wording, etc".

The challenge of "**Information unavailability**" appears with a frequency of 12.5 % in our data set. This challenge pivots around essential information that may not always be accessible or available during the initial stages of creating use case descriptions, as one of the participants states in the survey. This initial uncertainty calls for adapting and remaining flexible in the use case description creation process e.g. via high-level describing, search for sufficient information from other use cases [Cockburn 2001]. Use case creators must be prepared to revise and refine use cases as additional information becomes available, ensuring their continued relevance and alignment with evolving project requirements [Cockburn 2001]. This challenge highlights the relevance of effective stakeholder coordination to ensure that essential information is shared when needed. Among other challenges, "**Heterogeneity on (prior) knowledge**" occurs 8.33 % of the time, stemming from the varying levels of expertise among IEC 62559 Use Case description creators engaged in this process. Furthermore, "**Time-consuming creation**" occurs 8.33 % of the time. The method oriented advanced user states that it

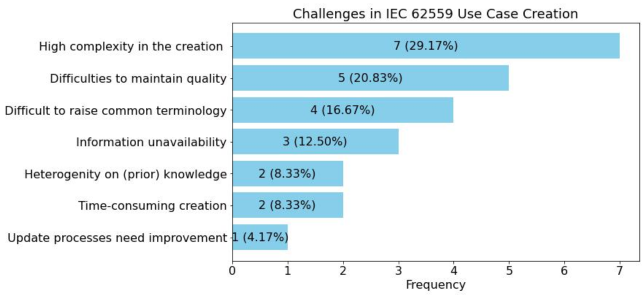


Fig. 5. Frequency analysis of challenges in creating IEC 62559 Use Case descriptions (n=24)

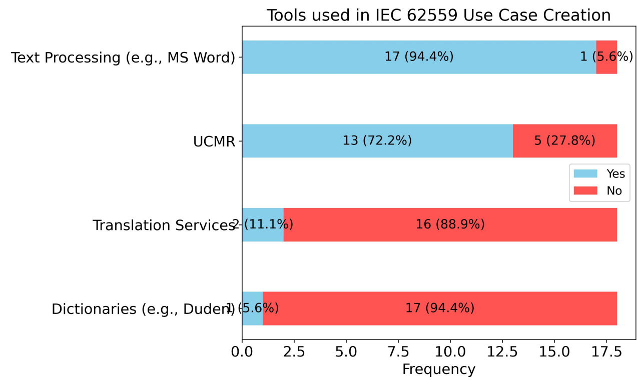


Fig. 6. Frequency analysis of used tools in creating IEC 62559 Use Case descriptions (n=32)

is challenging to synchronize use cases effectively due to high interpretational scope and a lack of predefined processes for achieving uniform wording and understanding. Lastly, "**Update processes need improvement**", reported 4.17 % of the time, and the user reports, "Continuous updating does not happen due to the nature of projects".

5.1.2 Toolchain. In this section, we delve into the comprehensive analysis of the tools employed in the creation of IEC 62559 Use Cases. As visualized in Figure 6, our investigation into the data about the tools utilized for developing and facilitating IEC 62559 Use Case descriptions revealed that a majority, 94.4 %, of respondents utilized text processing tool [IEC TC 8 2015], as the standard envisions the usage of the IEC 62559-2 Use Case Word-Template. Furthermore, the Use Case Management Repository (UCMR) was another popular choice, used by 72.2 % of the respondents. In contrast, translation services were notably less utilized, with only 11.1 % of respondents incorporating them into their workflow. Dictionaries, such as Duden, were the least used resource, with only 5.6 % of respondents reporting their usage.

In addition to the tools listed, other tools and resources were named, such as the analysis of other use cases or the open source tool Modelio. Reference was also made to various specialist literature (in particular [Gottschalk and Sauer 2015; Uslar et al. 2013]) and the practice-focused application aid from [Faller et al. 2020].

Use Case Management Repository. This section comprehensively analyzes the Use Case Management Repository (UCMR), a pivotal tool envisioned in managing, providing and storing IEC 62559 Use Case [IEC TC 8 2019]. With a high frequency of approximately 70 % dedicated to the storage of use cases. One utility of UCMR observed at a 10 % frequency, pertains to the "Generation of SGAM Models", which was envisioned that the IEC 62559 Use Case descriptions are helpful for creating the corresponding SGAM Model manually. Another aspect, also appearing at a 10 % frequency, involves "Follow-up control". This facet of UCMR usage denotes its role in facilitating the ongoing updating and maintenance of IEC 62559-2 Use Cases. Furthermore, UCMR serves a vital purpose in the "Setup" phase, accounting for 10 % of frequency.

The IEC 62559-1 standard envisions the use of the UCMR as a tool for managing and deploying IEC 62559 Use Case descriptions, rather than for creating them [IEC TC 8 2019]. This might lead to the question why the UCMR is not used in the direct creation of IEC 62559 Use Case descriptions. In our survey analysis, we found out that only 30.77 % of participants used UCMR for direct IEC 62559 Use Case descriptions creation. Given that most creators of use case descriptions prefer using the Word template over the structured digital support provided by UCMR, it is valuable to investigate the reasons behind this preference.

With a frequency of 50 %, it was stated that the **Use Case Management Repository was often considered to be "still immature"**. Users cite several factors, including the lack of a operational collaboration tool, the lack of version control capabilities, and a general lack of maturity that hinders easy data entry and modification. Another challenge, reported at a frequency of 30 %, revolves around the "**No access to the UCMR**". Since UCMR instances are usually deployed at the project or community level [Gottschalk et al. 2017], this is not a given at all times, as decision-making processes are also involved. Additionally, "**Word template is more accessible**" reported at a frequency of 20 %, offers an alternative perspective. One user mentions, in some instances, using a Word template may be more accessible and user-friendly for editing with partners. Another user commented that the Word Template provides increased speed and offline availability when compared to UCMR systems.

Challenges in the toolchain. In addition to the challenges of working with the IEC 62559 Use Case Methodology, challenges could also be identified when using the toolchain in general.

General challenges arose when working with the UCMR, which are represented as frequencies in Figure 7. Among these challenges is the "**Technical maturity of the UCMR**", identified with a frequency of 17.39 %, which is characterized by issues such as versioning discrepancies and technical bugs, which sometimes render the UCMR system unstable. This challenge can also be linked to issues addressed in detail in Section 5.1.2, emphasizing its observation in the context of UCMR challenges. Moreover, "**Complexity issues of the UCMR**", also occurring at 8.70 %, entail challenges related to the initial familiarization and setup of the UCMR system. Additionally, "**Missing automation in the UCMR**", identified with a frequency of 4.35 %, raises concerns about the manual effort required to create use cases involving multiple tools like PowerPoint, Word, and PlantUML. A survey participant who regularly works on use cases and SGAM pointed out that integrating "PlantUML", a tool for swift and straightforward diagram creation, into UCMR can be a potential solution to generate diagrams.

Concerning the challenges related to IEC 62559-2 Use Case Microsoft Word templates, "**Missing UX/UI of Word Templates**", with a frequency of 13.04 %, emphasizes the template's user experience and interface shortcomings. A survey participant who works almost daily with use cases says these word templates are relatively easy to understand and use. However, they lack essential guidance elements such as examples and can present challenges in comprehending specific fields. Further, the "**High effort with Word document exchange**" challenge, at a frequency of 8.70 %, revolves around the intricate process of exchanging files when utilizing Word documents for use case creation. A survey participant who works regularly on use cases says this challenge arises due to the fragmented nature of Word files, often comprising numerous individual documents that necessitate meticulous checking for interconnections. With a similar frequency of 8.70 %, the challenge, "**Lack of restriction by the word template**", appeared in our analysis. A user who worked daily with use cases stated that while the template's flexibility allows for customization, it also presents a potential for inconsistencies and redundancies, potentially leading to deviations from established standards.

Regarding challenges related to the general tool chain, a challenge occurring at a frequency of 13.04 % is labeled "**Insufficient enthusiasm of project partners**". One of the Expert users comments, "Specialized tools lack the acceptance with partners". Other advanced users also register "A lot of effort that partners avoid". Moreover, "**Incompliance with standard**" also reported at a frequency of 13.04 %, addresses the challenges of creating a standards-compliant IEC 62559 Use Case description by using only the IEC 62559-2 Use Case Template, which consequently compromises completeness, consistency, and uniqueness, as reported by a survey participant. Furthermore, it is also intrinsically linked with the challenge mentioned earlier in this subsection of "**Lack of restriction by the word template**", which might be a possible reason for the effects on the quality properties. Similarly, "**Comprehension problems with the standard**" present another challenge at a frequency of 13.04 %. Users report difficulty understanding certain fields and a lack of guidance in inputting information. An advanced user also remarks varying interpretations of terms in each use case, posing a challenge to standardizing use cases from different partners.

5.2 Smart Grid Architecture Model

This subsection examines the challenges of developing SGAM Models and the variety of tools utilized to aid these processes. Feedback from participants unveils key challenges in model creation and tool application, providing a detailed perspective on the practical aspects of SGAM deployment.

5.2.1 Challenges. In the process of creating SGAM Models, several challenges have been identified presented in Figure 8. The foremost challenge, occurring at a frequency of 30.77 %, revolves around the **issues in precise classification of elements into SGAM Domains and Zones**. A user highlights the ambiguity in the classifying zone as they intricately linked to processes. Being inherently dynamic, processes don't always neatly fit into predefined categories, making precise categorization complicated. This challenge is further worsened by the occasional absence of specific information or necessary technical terminology for distinct categorization. Simultaneously, another challenge is the "**Missing reference information**" with a frequency of 23.08 %. A survey participant with regular SGAM working intensity mentions the challenge in constructing a compliant SGAM Model when referring to external use cases due to gaps in necessary information.

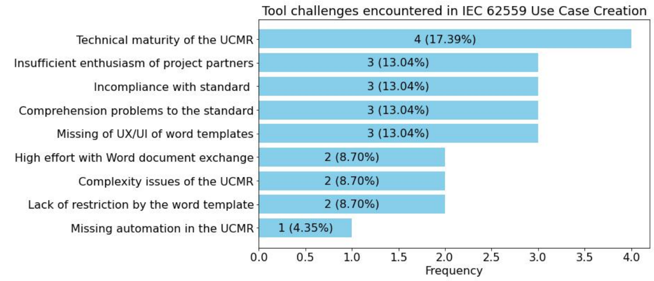


Fig. 7. Frequency analysis of tool challenges encountered in creating IEC 62559 Use Case descriptions (n=23)

Moreover, the challenge of "**Difficult to raise common terminology**" occurred 15.38 % of the time. An expert user emphasizes the challenge of establishing and utilizing uniform technical terms, actors, and a shared language within the SGAM framework, which is crucial for representing use cases in a common SGAM, especially in collaborative projects with multiple partners. Users also report that there is no equal understanding when working with diverse partners, coupled with the absence of a common language or shared understanding regarding information exchange, leading to delays and the need for rework.

Additionally, the issue of "**Challenges on proper display**" was identified 7.69 % of the time. An advanced method expert in SGAM development pointed out that the display fields are quite small, requiring manual size adjustments. Similarly, the problem of "**Missing guidance in creation**" was observed in 7.69 % of instances. An advanced user with high SGAM working intensity noted the importance of receiving assistance for more effectively justifying decisions during creation. Furthermore, "**Missing comprehensibility of the SGAM framework**" appeared with a frequency of 7.69 %. A domain expert in SGAM creation expressed, "Comprehensibility of the SGAM framework first had to be established". Lastly, the challenge of "**Lack of knowledge**" also stands at 7.69 %. This challenge is not rooted in specific SGAM problems but rather in the deficiency of knowledge and understanding. Moreover, one of the user also commented "A fool with a tool is still a fool" underscoring that many issues may stem from a lack of profound expertise rather than inherent problems within the SGAM framework and the tools around.

5.2.2 Toolchain. Figure 9 illustrates that a substantial 89.5 % of participants utilized the SGAM Microsoft PowerPoint PPTX templates.

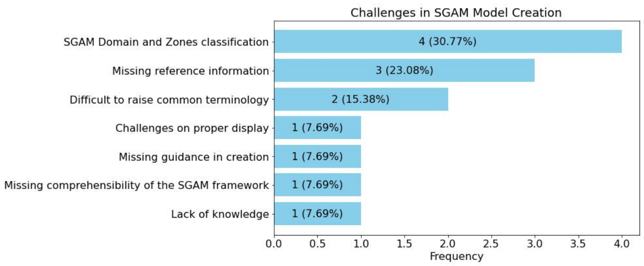


Fig. 8. Frequency analysis of challenges in creating SGAM Models (n=13)

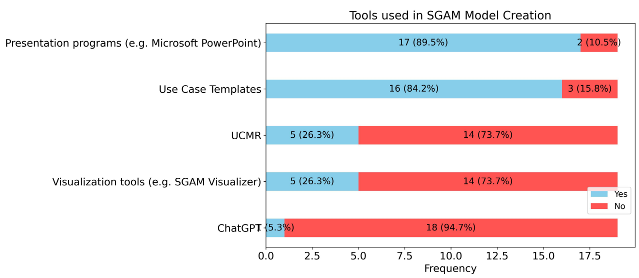


Fig. 9. Frequency analysis of used tools in creating SGAM Models (n=46)

As documented by users, most rely on platforms like Microsoft PowerPoint to develop and visualize SGAM Models, effectively conveying the intricate architecture of smart grid systems. Additionally, 84.2 % of respondents used completed IEC 62559 Use Case descriptions as the basis for creating SGAM Models, which describe the static view of the use case at the system-of-systems level [Faller et al. 2020]. The UCMR, which respondents described as an aid for an organized overview of the IEC 62559 Use Case descriptions, was used by 26.3 % to support the creation of SGAM Models. In addition, visualization tools such as SGAM Visualizer was also used by 26,3 % of the participants.

The use of generative AI tools (e.g., ChatGPT) was less common; with only 5.53 % of respondents incorporating them into their workflow (study was conducted in June 2023). Lastly, supplementary tools such as SGAM Toolbox [Neureiter et al. 2014], Conceptboard, other SGAM Models, and Literature are employed for reference purposes and collaborative coordination.

Challenges in the toolchain. In addition to the challenges of creating SGAM Models, challenges can also be identified when working with the associated toolchain as depicted in Figure 10. One of the predominant challenges, occurring with a frequency of 28.57 %, revolves around "**Challenges on proper display**". Users emphasize that this issue arises when multiple intricate and interconnected components need to be depicted within a confined field. For instance, in handling very complex systems, several layers must be overlaid within a single field, a situation often encountered with use cases involving many actors, such as prosumers. These complexities translate into a Time-consuming creation process, further exacerbating the challenges posed by proper display [CEN-CENELEC-ETSI 2012]. Another challenge, reported at a frequency of 21.43 %, involves the

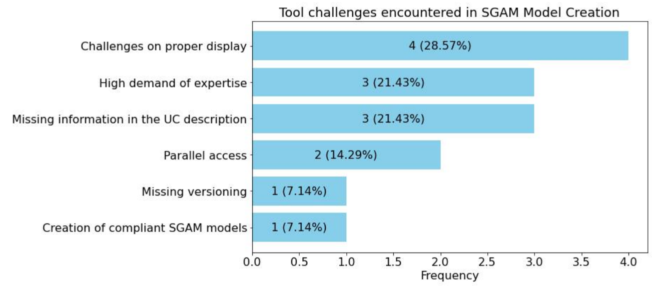


Fig. 10. Frequency analysis of tool challenges encountered in creating SGAM Models (n=14)

"**High demand of expertise**" needed for creating correct SGAM Models and to realize a proper mapping between the objects as well as display content correctly on the designated layers, domains, and zones, as per user reports. Moreover, the advanced user also cites, "Plausibility checks must continue to be carried out by expert interview". "**Missing information in the UC description**", also occurring at a frequency of 21.43 %, has been already pointed out at the general challenges part in Section 5.2.1 as it has strong overlaps. "**Parallel access**" presents another challenge, with a frequency of 14.29 %. Collaboration becomes complex, impeding workflow efficiency and giving rise to various interconnected issues. One of the users states that parallel access of all partners is essential for effective collaboration, yet unavailability and access difficulties can hinder this crucial aspect. "**Missing versioning**" has been also identified as challenge reported at 7.14 %, which can lead to inefficiencies and misunderstandings in the creation of SGAM Models. Furthermore, "**Creation of compliant SGAM**" also at 7.14 %, highlights challenges associated with tools like PowerPoint, which may not inherently assist in model creation. One of the advanced users with high SGAM working intensity mentioned that freedom offered by PPTX SGAM templates may lead to the construction of non-syntactically adequate SGAM Models and raises concerns about placing entities correctly from a stylistic perspective, emphasizing the need for more robust and supportive tools.

6 DISCUSSION

From the survey and analysis, challenges in creating IEC 62559 Use Case descriptions and SGAM Models, as well as tools and their associated challenges, could be identified. The findings from the data suggest that despite the perceived ease of understanding the standards and frameworks, which was confirmed by the participant knowledge survey, there is an observable complexity in the creation of IEC 62559 Use Case descriptions and SGAM Models. Challenges in ensuring quality and the necessity for a high level of expertise become apparent, which has been already stated in [Paustian et al. 2022]. Additionally, the data reveals that IEC 62559 Use Case descriptions do not always provide all the required information for the immediate creation of SGAM Models. The conditions (in particular the time planning, lack of time, type of tasks) within the projects can lead to the fact that a high-quality implementation of IEC 62559 Use Case descriptions or SGAM Models, among other things, can be hindered because revision and ongoing processes are necessary acc. [Cockburn 2001]. Furthermore, a negative quality of IEC 62559 Use Case description can have an impact on the quality of SGAM Models, since these are usually used for creation acc. [Faller et al. 2020; Gottschalk et al. 2017]. A cascade effect thus follows. While consistency of IEC 62559 Use Case descriptions has been subject of various studies ([Gottschalk and Sauer 2015; Kuchenbuch et al. 2023; Neureiter et al. 2014]), it remains a persistent challenge, as reflected in the encountered difficulties. Furthermore, the data indicates that the appropriate utilization of the UCMR tools presents challenges regarding their maturity. As the UCMR matures, consistency (even across multiple templates) could be improved, as the UCMR should be able to reference elements via centralized component management [Gottschalk and Sauer 2015] (e.g. actors, glossary). In this regard, it is essential to enhance the maturity of the respective deployed UCMR instances.

However a limitation of this study to be noted is the low rate of qualified participants (see section 3.1) who were eligible to respond to the survey's content-specific sections. This restriction was essential to ensure a qualified representation of actual creators. Further analyses (e.g. correlation analyses by considering cohorts) are difficult to carry out due to the small number of qualified survey participants. This may be due to the fact that the IEC 62559 Use Case methodology and the SGAM Framework, while recognized in requirements management for energy-related projects, are specific and utilized (in creating concrete IEC 62559 Use Case descriptions and SGAM Models) by a smaller fraction of employees in these projects and the typical effects of participating in surveys. Nonetheless, important insights have emerged from the qualitative data and an image of the practical application could be captured. Future research should explore the limitations, emphasizing the intentional gathering of quantitative data as a basis for in-depth analyses, particularly correlation studies. Subsequent studies are encouraged to focus on addressing the identified challenges, thereby contributing to the enhancement of practices in this domain

7 CONCLUSION

The goals of this research endeavour encompassed the identification of challenges related to the development of IEC 62559 Use Case descriptions and SGAM Models, as well as the identification of the tools utilized in this context and their associated difficulties. In this context, various challenges have been identified, including high complexity, the high heterogeneity of contributors, and the high requisite domain expertise, all of which may potentially have adverse effects on quality. In addition, a lower quality of the IEC 62559 Use Case descriptions has a direct impact on the quality of the SGAM Models, creating a cascade effect in which deficiencies in the former are reflected in the latter. The results of the research also show that the toolchain but also processes offers further potential to enhance acceptance and, notably, reduce complexity, especially the maturity within the Use Case Management Repository.

This paper identifies several implications for future initiatives, encompassing both energy-related project applications and the advancement of methodologies. It suggests a reassessment of future project planning to ensure the thorough creation and continual updating of SGAM Models and IEC 62559 Use Case descriptions. Introducing processes that facilitate collaborative and concurrent work on compliant SGAM Models could be advantageous. Moreover, engaging all relevant stakeholders in requirements management and establishing a common terminology (potentially through a glossary) could enhance quality. Additionally, this paper highlights the toolchain's potential to improve the User Experience (e.g. by employing Artificial Intelligence) to assist modelers in developing precise IEC 62559 Use Case descriptions and SGAM Models.

Further research can address the challenges in both the application of methodologies and tools to further improve requirements engineering within energy-related projects.

ACKNOWLEDGMENTS

This research project was conducted as part of the int:net project (Funding Code No. 101070086) for supporting the development of interoperable solutions by enforcing energy-related knowledge base and standards [Reif et al. 2023]

REFERENCES

CEN-CENELEC-ETSI. 2012. Smart Grid Reference Architecture.

Alistair Cockburn. 2001. *Writing effective use cases*. Addison-Wesley, Boston.

Sebastian Faller, Johann Schütz, Alexander Bogensperger, and Mathias Uslar. 2020. Anwendungshilfe SGAM: Smart Grid Use Cases modellieren. https://doi.org/10.13140/RG.2.2.24347.95525

Marion Gottschalk, Christina Delfs, and Mathias Uslar. 2017. *The Use Case and Smart Grid Architecture Model Approach: The IEC 62559-2 Use Case Template and the SGAM applied in various domains*. Springer. https://doi.org/10.1007/978-3-319-49229-2

Marion Gottschalk and Jurgen Sauer. 2015. Towards identifying an approach for consistency checks to smart grid descriptions. In *2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST)*. IEEE, Vienna, Austria, 380–385. https://doi.org/10.1109/SEDST.2015.7315238

Bill Haskins, Jonette Stecklein, Brandon Dick, Gregory Moroney, Randy Lovell, and James Dabney. 2004. 8.4.2 Error Cost Escalation Through the Project Life Cycle. *INCOSE International Symposium* 14, 1 (June 2004), 1723–1737. https://doi.org/10.1002/j.2334-5837.2004.tb00608.x

IEC TC 8. 2015. *Use case methodology - Part 2: Definition of the templates for use cases, actor list and requirements list*. Standard IEC 62559-2:2015. International Organization for Standardization.

IEC TC 8. 2019. *Use case methodology - Part 1: Concept and processes in standardization*. Standard IEC 62559-1:2019. International Organization for Standardization.

J. Jamsen and K. Corley. 2007. *E-Survey Methodology*. IGI Global, 1–8. https://doi.org/10.4018/978-1-59140-792-8.ch001

René Kuchenbuch, Johann Schütz, and Jürgen Sauer. 2023. Quality properties of IEC 62559 use cases and SGAM models. *Energy Informatics* 6, S1 (Oct. 2023), 38. https://doi.org/10.1186/s42162-023-00280-5

Udo Kuckartz and Anne McWhertor. 2014. *Qualitative text analysis: a guide to methods, practice & using software*. SAGE, Los Angeles. OCLC: ocn875376801.

Christian Neureiter, Dominik Engel, Jörn Trefke, Rafael Santodomingo, Sebastian Rohjans, and Mathias Uslar. 2014. Towards consistent smart grid architecture tool support: From use cases to visualization. In *IEEE PES Innovative Smart Grid Technologies, Europe*. 1–6. https://doi.org/10.1109/ISGTEurope.2014.7028834 ISSN: 2165-4824.

Sabrina Paustian, Julia Köhlke, Jannika Mattes, and Sebastian Lehnhoff. 2022. The (still unexplored) social side of smart grid development: towards a social layer for the smart grid architecture model (SGAM). In *Proceedings of the Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*. Association for Computing Machinery, New York, NY, USA, 521–531. https://doi.org/10.1145/3538637.3539585

Valerie Reif, Thomas I. Strasser, Joseba Jimeno, Marjolaine Farre, Oliver Genest, Amélie Gyrard, Mark McGranaghan, Gianluca Lipari, Johann Schütz, Mathias Uslar, Sebastian Vogel, Arsim Bytyqi, Rita Dornmair, Andreas Corusa, Gaurav Roy, Ferdinanda Ponci, Alberto Dognini, and Antonello Monti. 2023. Towards an interoperability roadmap for the energy transition. *e & i Elektrotechnik und Informationstechnik* 140, 5 (Aug. 2023), 478–487. https://doi.org/10.1007/s00502-023-01144-2

Margrit Schreier. 2012. *Qualitative Content Analysis in Practice*. SAGE Publications Ltd.

Margrit Schreier. 2014. Ways of Doing Qualitative Content Analysis: Disentangling Terms and Terminologies. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 15, 1 (Jan. 2014). https://doi.org/10.17169/fqs-15.1.2043

I. Stančin and A. Jović. 2019. An overview and comparison of free Python libraries for data mining and big data analysis. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2019-05). 977–982. https://doi.org/10.23919/MIPRO.2019.8757088 ISSN: 2623-8764.

Mathias Uslar, Michael Specht, Christian Danekas, Jorn Trefke, Sebastian Rohjans, Jose M. Gonzalez Vazquez, Christine Rosinger, and Robert Bleiker. 2013. *Standardization in smart grids: introduction to IT-related methodologies, architectures and standards*. Springer, Berlin ; New York. OCLC: ocn842074536.

Choosing the right Ontology to describe Research Data in the Energy Domain

ALEXANDRO STEINERT, OFFIS - Institute for Information Technology, Germany and Department of Computer Science, Carl von Ossietzky Universität of Oldenburg, Germany
STEPHAN FERENZ, OFFIS - Institute for Information Technology, Germany and Department of Computer Science, Carl von Ossietzky Universität of Oldenburg, Germany
ASTRID NIESSE, OFFIS - Institute for Information Technology, Germany and Department of Computer Science, Carl von Ossietzky Universität of Oldenburg, Germany

As in all disciplines, increasing the FAIRness (findability, accessibility, interoperability, and reusability) of research data and software is a goal in energy research. In order to achieve this, it is important to identify the most appropriate ontology for the description of research data and software. However, despite the importance of this task, it still presents a significant challenge. While there are some comparisons of ontologies, a gap exists in assessing their usefulness according to ontology metadata. This paper fills this gap by defining 21 criteria sorted into four categories to help researchers choose ontologies in the energy domain. The criteria are used to compare eight ontologies for energy research to showcase their use and analyze the ontologies. The analysis reveals the Open Energy Ontology (OEO) as the top-ranked ontology. This underscores the importance of metadata comparison in ontology selection and highlights the benefits of incorporating metadata criteria into ontology terminology services to support researchers.

CCS Concepts: • **Information systems → Information integration**; • **Applied computing → Digital libraries and archives**.

Additional Key Words and Phrases: FAIR principles, Ontology, Energy Research, Semantic Web, Energy Research Data Management

1 INTRODUCTION

Data and software are an integral part of energy research [15]. As the amount of data and software involved in research keeps rising, there is an even greater urgency to produce reusable research data and software, facilitating collaborative research [41]. The umbrella term "digital research object" is used to signify any digital entity, including data, software, or other digital research artifacts.

The FAIR principles, developed by Wilkinson et al. [56], present guidelines to aid researchers in handling digital research objects responsibly. The principles aim to enable better research [8, 56]. FAIR is an acronym standing for the four principles: findable, accessible, interoperable, and reusable.

To enable FAIR digital research objects, it is important to define well-used terms and the relations between them. This facilitates better findability and easier machine interpretability of digital research object. A general accepted concept from the semantic web which is usable for this purpose is ontologies. An ontology is, in simple terms, a representation of knowledge in the form of a structured, finite collection of terms and the relationships between them. Ontologies can be represented as graphs, with the vertices being the terms and the edges their relationships [2]. They are categorised as either Upper Ontologies, which describe general concepts, or Domain Ontologies, which describe concepts specific to a particular subject area [2, 3]. Moreover, ontologies enable seamless merging of data from diverse sources [2].

The energy research domain offers a wide variety of ontologies, each with unique priorities and coverage [55]. These ontologies frequently originate from different sources which complicates the task for researchers to identify the most suitable choice for their needs, e.g., when they want to use ontology terms to describe their dataset fields.

To address this challenge, we compare multiple existing ontologies in the energy domain within this paper to answer our research question: How can an ontology be chosen to be best usable in energy research to describe digital research objects?

While there are existing comparisons for ontologies in energy research, they only focus on the scope of the ontology. The comparisons come in two forms: papers, in which the comparison is not a focus, and papers, like the paper from Pritoni et al. [45], comparing ontologies regarding their use in a well defined subdomain of energy research, in this case for building energy applications [45]. This comparison fills a gap in existing works by focusing on comparing ontology metadata.

Our contributions are as following:

- In section 2, we summarize the related work on general comparison of ontologies.
- Afterwards, we introduce multiple ontologies in the energy domain in section 3.
- Based on the methods described in section 4, we compare multiple ontologies in section 5.
- Finally, we draw some conclusions in section 6.

2 RELATED WORKS

This section highlights several research efforts devoted to addressing the complex issue of how ontologies can be compared. One common approach towards this is through the comparison of ontologies within specific domains, such as Grubic and Fan's study on supply chain ontologies [20]. These reviews typically focus on examining and comparing the content and scope, as well as other factors relevant to the content and scope of an ontology. Another

Authors' addresses: Alexandro Steinert, alexandro.steinert@offis.de, OFFIS - Institute for Information Technology, Escherweg 2, Oldenburg, Lower Saxony, Germany, 26121 and Department of Computer Science, Carl von Ossietzky Universität of Oldenburg, Ammerländer Heerstraße 114-118, Oldenburg, Lower Saxony, Germany, 26111; Stephan Ferenz, stephan.ferenz@offis.de, OFFIS - Institute for Information Technology, Escherweg 2, Oldenburg, Lower Saxony, Germany, 26121 and Department of Computer Science, Carl von Ossietzky Universität of Oldenburg, Ammerländer Heerstraße 114-118, Oldenburg, Lower Saxony, Germany, 26111; Astrid Nieße, astrid.niesse@offis.de, OFFIS - Institute for Information Technology, Escherweg 2, Oldenburg, Lower Saxony, Germany, 26121 and Department of Computer Science, Carl von Ossietzky Universität of Oldenburg, Ammerländer Heerstraße 114-118, Oldenburg, Lower Saxony, Germany, 26111.

paper, by Lourdusamy and John, primarily delves into metrics for assessing ontologies, with a clear focus on their contents [35]. In their work the number of classes and attributes, as well as other such content focused criteria, are compared.

In addition to papers which compare the content of ontologies in a subdomain, there are instances where new ontologies are introduced. In the course of the introduction, similar ontologies are compared based on feature set and content. Often, these comparisons are not the main focus of these papers, but rather highlight what essential information is lacking from other ontologies according to the authors. This can be seen in papers like Haghgoo et al.'s [21] presentation of different ontologies in the same domain, which outlines the missing content.

Within the energy domain, ontologies are compared similarly to how they are analyzed in other domains. For instance, there are papers that review a broader spectrum of ontologies within a specific domain, such as Lygerakis et al. [36] work on energy efficiency in buildings. Other papers, including the paper by Pritoni et al. [45], focus on reviewing ontologies for building energy applications.

A broader comparison of energy ontologies represents the paper from Wierling et al. [55] in which the authors review the landscape of metadata in energy research, listing different energy research ontologies and also comparing the practices used for their creation. Furthermore, other papers typically introduce a new energy ontology and compare similar ones in terms of content, with a focus on identifying missing information from other ontologies. This is exemplified in the works of Haghgoo et al. [21] and Cuenca et al. [11], who present and compare different ontologies in the same target domain and outline the missing content.

For this specific comparison, however, the focus is not on ontology content, but it encompasses the metadata associated with the ontologies. The criteria for evaluating these aspects of ontologies can be found in section 4.

3 ONTOLOGIES IN ENERGY RESEARCH

This section focuses on the ontologies compared in this paper. First, the selection process is explained and afterward short introductions for the selected ontologies are given.

3.1 Selection

It is crucial to select a suitable subset of ontologies for comparison. The intention of this section is to explain the selection process used. This paper adopts a systematic approach. The method for identifying ontologies is adapted from a narrative review and shown in a modified Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) process [37]. Figure 1 illustrates this process, indicating the number of ontologies/papers identified at each step.

First, an initial selection of ontologies relevant to the energy domain is made. This process involves utilizing three main sources: ① an article database search using Google Scholar[1], Elsevier[2], IEEE[3], and Scopus[4] with the keywords "energy" and "ontology", ② an ontology database search through Linked Open Vocabularies (LOV[5]) with the same keywords, and ③ using curated lists found in relevant papers, which are Wierling et al. [55], Cuenca et al. [11], Fernández-Izquierdo et al. [16] and Haghgoo et al. [21]. After the removal of duplicates, a total of 98 ontologies and papers were identified.

Afterward, a preliminary screening is applied, where only ontologies in the English language with an energy domain focus, or a significant part related to the energy domain, are considered. This step also involves filtering and excluding papers focused on working with existing ontologies, that were already identified, or ontologies unrelated to the energy domain. After this first screening, 80 ontology papers remain, while 18 papers were removed.

Next, a second screening based on relevancy is conducted. An ontology is considered relevant based on three criteria,① is the ontology considered a standard in the industry, ② the age of the ontology and ③ the citations of an ontology. A newer ontology is considered more relevant than an old one, as well as an ontology with more citations than one with less. These citations are taken through Google scholars statistics. Using the defined criteria, 35 ontologies were selected after this stage, which can be found in Table 7. Out of the 35 ontologies, two are standards and 21 received government funding. The selected ontologies cover various aspects of the energy domain, such as smart homes, the general energy domain, the smart grid, energy management, wind energy, solar energy, and the energy market.

After this step, the final selection is made through an eligibility screening. In this, eight ontologies are chosen for comparison. To get an overview over the existing ontologies and answer the research question formulated in the beginning, only ontologies were chosen that either had the energy domain as their scope or had a energy subdomain as their focus that was not yet the focus of another ontology chosen. The full list of chosen ontologies can be seen in Table 1.

3.2 Selected Ontologies

In this section, the selected ontologies are presented to give an overview over their scope and also give further reasoning on their selection.

Brick: Brick is an ontology created for smart homes and was first presented by Balaji et al. [4] in 2016. According to Google Scholar, it has been referenced 373 times. It is designed to represent sensors and subsystems within a building as well as their relations, this widely cited ontology is a relevant choice for the smart home domain. A multitude of other ontologies exist in this field, but Brick stands out as a good representation of it with industry support.

Common Information Model (CIM). CIM, an ontology created to serve as an industry-standard model for facilitating information exchange among entities, was presented in 2003 by the International Electrotechnical Commission (IEC)[6] [25]. According to Google Scholar, it has been cited two times. This is most likely an oversight in this method and does not reflect the multiple research

[1] See: https://scholar.google.de/ Last checked 29.04.2024.
[2] See: https://www.sciencedirect.com/ Last checked 29.04.2024.
[3] See: https://ieeexplore.ieee.org/Xplore/home.jsp Last checked 29.04.2024.
[4] See: https://www.scopus.com Last checked 29.04.2024.
[5] See: https://lov.linkeddata.es/dataset/lov Last checked 29.04.2024.
[6] See: https://www.iec.ch/who-we-are Last visited: 29.04.2024

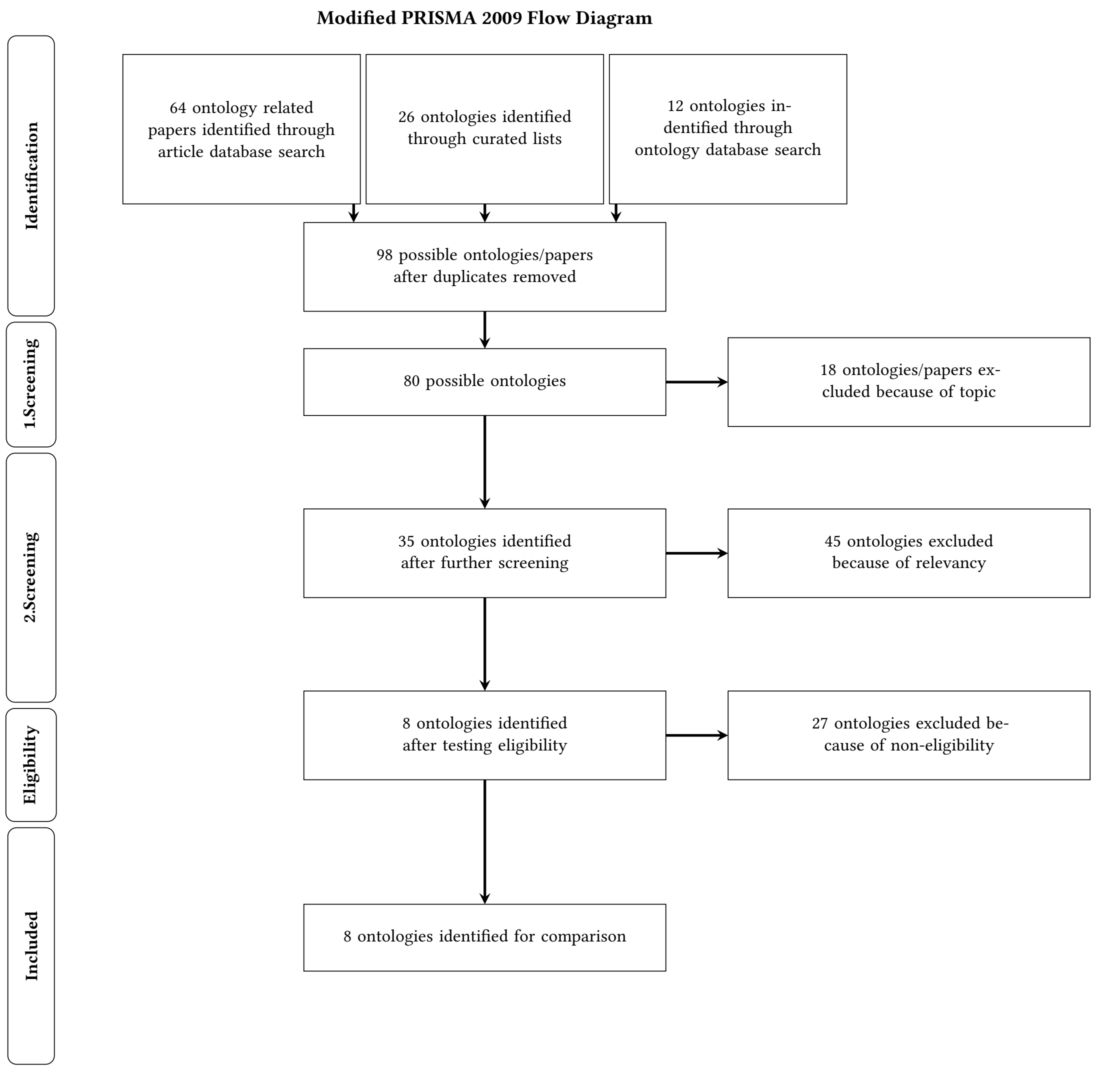


Fig. 1. Modified PRISMA approach used for ontology procurement [37].

papers using CIM, as it can be considered widely used in the industry. The core CIM standard allows users to represent every object required in a typical electricity grid. It is selected as a result of its widespread use as a standard in Europe. This ontology is adaptable and has various extensions, official and otherwise.

Domain Analysis-Based Global Energy Ontology (DABGEO). DABGEO, an ontology network created as a general energy domain ontology, was presented in 2020 by Cuenca et al. [11]. According to Google Scholar, it has been cited 26 times. It was developed as a successor to the OEMA ontology, which was developed by the same authors [10]. The network consists of numerous smaller domains, each featuring their own ontology. The top-level domains covered in DABGEO include energy equipment, infrastructure, energy performance data, energy external factors, and smart grid stakeholders.

DABGEO is chosen for being designed as an ontology network, offering a unique structure. It is examined instead of the OEMA ontology because it provides a better reusability and usability balance, according to the authors. Additionally, DABGEO is a more recent ontology than OEMA.

Electricity Markets Ontology (EMO). EMO, an ontology developed for modeling energy markets, was presented in 2016 by Santos et al. [49]. According to Google Scholar, it has been cited 23 times. It is designed to support agent-based simulation of electricity markets. This ontology is used in six other ontologies. These include three electricity markets ontologies: a call for proposal ontology, a results ontology, and an interoperability ontology.

EMO is chosen because it represents a domain with limited ontologies. Additionally, its approach to examining the energy domain from a market perspective makes it a valuable resource in this field

Table 1: Ontologies identified for comparison. Citations taken from Google Scholar.
Citations last checked: 29.04.2024. Sorted alphanumerical by ontology name.

Ontology	Domain	Year	Reference	Standard	Citations	Gov. Funding
Brick	Smart Home	2016	[4]	✗	373	✓
CIM	Energy	2003	[25]	✓	2	✓
DABGEO	Energy	2020	[11]	✗	26	✓
EMO	Energy Markets	2016	[49]	✗	23	✓
EM-KPI	Smart Grid	2019	[33]	✗	45	✓
OEO	Energy	2021	[6]	✗	66	✓
Sargon	Energy	2020	[21]	✗	34	✓
SEAS	Energy	2017	[32]	✗	85	✓

and its small size make it an outlier in this selection, which could be promising for using it to create new ontologies.

Energy Management-Key Performance Indicator Ontology (EM-KPI). EM-KPI, an ontology developed for exchanging performance data between buildings to enhance energy management, was presented in 2019 by Li et al. [33]. According to Google Scholar, it has been cited 45 times. Constructed by reusing existing ontologies and introducing new elements, this ontology is composed of seven modules: the Key Performance Index module, the Observation Model, the Location module, the Infrastructure module, the Occupancy module, the Weather module, and the Energy Parameter module.

It is chosen as it represents an ontology for the smart grid domain.

Open Energy Ontology (OEO). OEO, an energy domain ontology, was presented in 2021 by Booshehri et al. [6]. According to Google Scholar, it has been cited 66 times. The researchers developed this ontology based on the Basic Formal Ontology (BFO), which serves as an upper ontology. Designed around the Open Energy Platform (OEP), OEO emphasizes open access and community collaboration, as it is freely available on GitHub for contributors to work on. Its primary focus lies in modeling energy systems and analyzing them. Comprising four modules: oeo-model (entities for models, data, and more.), oeo-physical (physically existing entities), oeo-social (social relations and people), and oeo-shared (concepts needed in all other modules) [6]. The OEO is chosen due to its recent development and emphasis on community collaboration.

SmArt eneRGy dOmain oNtology (Sargon). Sargon, an energy domain ontology, was presented in 2020 by Haghgoo et al. [21]. According to Google Scholar, it has been cited 34 times. Comprising five modules and built as an extension of the Smart Application Reference Ontology (SAREF), Sargon is designed to model the smart energy domain and serves as a non-official extension of SAREF, an Internet of Things (IoT) ontology, which is an official European Telecommunications Standards Institute (ETSI) standard. Notably, SAREF is an IoT ontology, so Sargon inherits some IoT focus.

Smart Energy Aware Systems Ontology (SEAS). SEAS, an energy domain ontology, was presented in 2017 by Lefrançois [32]. According to Google Scholar, it has been cited 85 times. Comprising four core modules and several vertical extensions that focus on specific domains within the energy domain, SEAS also inherits some IoT focus as it is built upon SAREF. It is chosen due to its basis in SAREF and for potential comparisons with the Sargon ontology.

4 COMPARISON METHOD

This section introduces the method to compare the ontologies. The criteria and the scoring system for simpler comparison are explained.

4.1 Criteria Categories

The criteria have been categorized into four groups based on the overall subject they aid in analyzing. This is done to make comparison easier and the information more manageable. These categories are not fixed and should not be seen as an objective division, as this division exists to showcase the possibility of grouping the criteria. The division can be seen in Table 2.

The criterion devised in this section are partly derived from literature describing ontology creation such as Smith and Spear [3] or Antoniou and Van Harmelen [2] and partly by looking at available ontology metadata and choosing useful criterion for comparison.

4.2 Scoring for Ease of Comparison

As comparing something in-depth takes a lot of time, this paper showcases a simple scoring function to use for easier comparison. This score is meant to help in comparison and show the possibility of comparing ontologies in an easier way. While this score can be used unmodified, it might be best to modify it to the subjective use case of the person comparing different ontologies. For this paper, the score is based on the open science principles, rewarding more open ontologies. Not all criteria are scored, as they are unsuitable for scoring according to the basis of this score.

The score is calculated by giving certain criterion an integer value between zero and three, with zero being the worst possible and three the best possible value, then the mean of each scored criterion rounded to the first decimal place is calculated.

4.3 Automated Script

A Python script helps with compiling data for some of the criteria. The script uses the Owlready2 package[7] to load ontologies and choose five classes for the description quality and description size criteria, using the random module. If the script can not load an ontology with Owlready2, the classes are manually picked. The same package is also used to compile the number of classes and

[7] See: https://github.com/pwin/owlready2, last checked 29.04.2024.

Table 2: The 21 ontology criteria divided into the 4 categories.

Category	Description	Criteria
Best Practice	Criteria that are associated with best practices to be followed in ontology creation.	Used Upper Ontologies, Scope, Creation Type, Modularity, Extensibility, Validation
Practical Implementation	Criteria that concern metadata that arises during the implementation of the ontology.	Available Languages, Available Ontology Languages, Description Size, Description Quality, Number of Terms, Used Ontologies
Maintenance and Accessibility	Criteria that concern the maintenance of the released ontology.	Sourcecode, License, Accesibility, Maintenance, Latest Release, Automatic Analysis
Governance	Criteria that concern the governance of the ontology.	Funding, Governing Instances, Citations

properties for the criterion number of terms. In the case that the ontology can not be loaded, the ontology editor Protége[8] is used[39].

The rdflib library[9] is used to compile the used ontologies for the used ontologies criterion. If not possible, Protége is used.

Additionally, the request package and rdflib library are used to automatically generate OOPS[10] , FOOPS![11] and FairChecker[12] reports for the automatic analysis criterion. The script also calculates a mean FAIR score from the FairChecker report [17, 18, 44].

It is also used for the DABGEO ontology network to sum up the classes and used ontologies, as well as calculate the average of the reports for the automatic analysis. This is done, because it is an ontology network made up of many ontologies and a single report cannot be created.

The Python script can be found on GitHub[13]. During the data collection, a few problems occurred, which are mentioned in the following.

In OEO class descriptions are not found as *rdfs:comment* tags but as *obo:IAO_0000115*, as it is based on the Basic Formal Ontology [6]. Additionally, the automated script was unable to load Brick, SEAS, and 40 of the 97 ontologies included in DABGEO. As a result, the classes used for determining description size and quality were selected manually. For Brick and SEAS the number of terms was identified through Protége. The numbers for DABGEO were taken from Cuenca et al.'s [11] introductory paper. For analyzing DABGEO automatically, each ontology in the network was examined individually, and a mean score was calculated for the FOOPS! and FairChecker scores. Only 23 of the DABGEO ontologies did produce an OOPS! score, because the REST service of Poveda-Villalón et al. [44] was not able to generate a report for the rest of them[14] are improperly configured, such as the link for the air pollutants ontology[15]. The EMO website is not accessible via https, and EMO cannot be automatically downloaded through the provided link. To access the source code, it either needs to be copied into a file or downloaded using wget[16] or other similar methods.

4.4 Comparison Criteria

This section will go over the criteria and also the given score, if applicable, for a criterion. The criteria will be split up into the different categories, a register of which can be seen in Table 2.

4.4.1 Best Practice.

Used Upper Ontology. This criterion lists the upper ontology a given ontology is based on. If only a part of an upper ontology is used, it will not be listed here.

Not Found: No Upper Ontology used. (Score: 0)
Upper Ontology Used: Upper Ontology used. (Score: 3)

Scope. This criterion lists the planned scope of the ontology according to the authors.

Creation Type. This criterion lists the method used to create the ontology. The following creation methods are identified:

Automatic: Ontology created automatically by algorithm or other automated measure. Without or with little human interference. (Score: 0)
Manual: Ontology created manually by a human. (Score: 3)
Hybrid: Ontology created semi-automatically by using algorithms or automated measures and humans. (Score: 3)

If the creation method used is a well known method, it is listed.

Modularity. This criterion quantifies the modularity of the ontology. As ontologies are modular by design, this criterion shows how modular a given ontology is. It is divided into two cases:

Not (as) Modular: Ontology does not consist of easily differentiable modules. (Score: 0)
Modular: Ontology consists of different modules that can easily be used independently. (Score: 3)

[8]See: https://protege.stanford.edu/ Last checked 29.04.2024
[9]See: https://github.com/RDFLib/rdflib, last checked 29.04.2024.
[10]See: https://oops.linkeddata.es/, last checked 29.04.2024.
[11]See: https://foops.linkeddata.es/FAIR_validator.html, last checked 29.04.2024.
[12]See: https://fair-checker.france-bioinformatique.fr/check, last checked 29.04.2024.
[13]See: https://github.com/a-steinert/OntMetaScript under the MIT license. Last checked 29.04.2024.
[14]See: https://oops.linkeddata.es/webservice.html. This is also the case for the Brick, SEAS and CIM Ontology.
Furthermore, some links on the DABGEO websitehttps://innoweb.mondragon.edu/ontologies/dabgeo/ Last checked 29.04.2024.
[15]See: https://innoweb.mondragon.edu/ontologies/dabgeo/domain-task/application_type/city_energy_performance_assessment/airpollutants/1.0/ontology/ Last checked 29.04.2024.
[16]See: https://www.gnu.org/software/wget/ Last checked 29.04.2024

Extensibility. This criterion quantifies how extensible the ontology is. It is in partly based on the Modularity criterion and also based on how well the ontology is explained, e.g. explanations of the structure in additional documents. This criterion is divided into the following cases:

Not (as) Extensible: Ontology is not that easily extensible. (Score: 0)
Explanations given: Ontology has explanations that help newcomers to extend it by explaining the ontology or the method used to create it further. (Score: 2)
Modular: Ontology is modular and through this more extensible. (Score: 2)
Extensible: Ontology is modular and has explanations that help newcomers to extend it by explaining it/the method used to create it further. (Score: 3)

Validation. This criterion explains the steps the ontology creators used to validate the ontology. It is divided into the following:

Not Validated No Validation method given. (Score: 0)
Validated through other means: Some other means were used in the validation of the ontology. (Score: 1)
Validated through automated Tools: Automated tools, e.g. OOPS! [44], were used in validation (Score: 2)
Validated through Experts: Ontology was validated by experts in the field either using case studies, coverage studies or some other way. (Score: 2)
Validated: Automated tools and experts were consulted in the validation (Score: 3)

4.4.2 Practical Implementation.

Available Languages. This criterion lists the different natural languages the ontology is available in.

Available Ontology Languages. This criterion lists the different ontology languages the ontology is available in. It is divided into three different categories:

Other: Ontology is available in other languages. (Score: 0)
RDF/XML: Ontology is available in RDF/XML. (Score: 2)
OWL: Ontology is not available in RDF/XML, but in other OWL ontology languages. (Score: 2)
RDF/XML+: Ontology is available in RDF/XML and other ontology languages. (Score: 3)

Description Size. This criterion looks at a random subset of 5 terms to determine the size of the descriptions. The terms are compiled using a Python script as outlined in subsection 4.3. If there are classes with no descriptions given in the selection, the rest of the terms are analyzed, but one point less is given for the score with a lower limit of zero. It is divided into three possible categories based on what the majority of the five terms descriptions are:

Large Over five sentences of description. (Score: 0)
Medium: Up to five sentences of description. (Score: 2)
Small: Up to two sentences of description. (Score: 3)

Description Quality. This criterion looks at the same random subset of terms used to determine the description size to determine the quality of the descriptions. The terms are compiled using a Python script as outlined in subsection 4.3. If there are classes with no descriptions given in the selection, the rest of the terms are analyzed, but one point less is given for the score with a lower limit of zero. It is divided into three possible categories based on what the majority of the five terms descriptions are, and is more subjective:

Hard: Only understandable by domain experts. (Score: 0)
Neutral: Understandable by people with a slight understanding of the energy domain. (Score: 2)
Easy: Easily understandable even by people that are not experts in the energy domain. (Score: 3)

Number of Terms. This criterion lists the dimension of the ontology. In this paper the dimension means the number of classes the ontology envelops as well as the number of properties. The data is compiled using a Python script as outlined in subsection 4.3.

Used Ontologies. This criterion lists all ontologies used in the ontology being compared. This criterion differs from the Used Upper Ontologies criterion laid out above as it lists all ontologies used and not only the upper ontology the compared ontology is based on. A used ontology is identified as given an ontology that is used with a namespace in the compared ontology. The data is compiled using a Python script as outlined in **??**.

4.4.3 Maintenance and Accessibility.

Sourcecode. This criterion lists how many sources the ontology has. The ontology is looked for in different ontology database searches (Linked Open Vocabulary[17], TIB Terminology Service[18], Ontology Lookup Service[19]). Furthermore, it is looked for on GitHub[20] as well as sources listed in the paper presenting the ontology, if it is available. This criterion is divided into the following categories:

0: Ontology has no available sources. (Score: 0)
1S (One Self-hosted Source): Ontology has one available source, the source is self-hosted. (Score: 2)
1D (One Distributed Source): Ontology has one available source, the source is not hosted by the authors. (Score: 2)
2+: Ontology has more than one available source. (Score: 3)

License. This criterion defines the license of the ontology, it is divided into two categories:

Not Found: Ontology license not found. (Score: 0)
License Found: Ontology license found. (Score: 3)

Accessibility. This criterion defines the accessibility of a given ontology. It is divided into four categories:

Closed-Access: Ontology is not openly accessible. (Score: 0)
Open-Access: Ontology is openly accessible. (Score: 1)
Open-Source: Ontology source code is openly available and openly accessible. (Score: 2)
Open-Source (OSI): Ontology is open-source with a license following OSI[21] and openly accessible. (Score: 3)

[17] See: https://lov.linkeddata.es/dataset/lov/ Last checked 29.04.2024.
[18] See: https://service.tib.eu/ts4tib/index Last checked 29.04.2024.
[19] See: https://www.ebi.ac.uk/ols/index Last checked 29.04.2024.
[20] See: https://github.com/ Last checked 29.04.2024.
[21] See: https://opensource.org/licenses/ Last checked 29.04.2024.

Table 3: A comparison of the adherence to best practices for energy research ontologies. *Not found:* ✗.

Metric / Ontology	Best Practice						
	Used Upper Ontology	Scope	Creation Type	Modularity	Extensibility	Validation	Score
Brick	✗	Buildings	Manual Own Method	Modular	Extensible	Validated through Experts	2.2
CIM	✗	General	Manual Own Method	Modular	Modular	Validated through Experts	2.0
DABGEO	✗	General	Manual OntoCape Method [38]	Modular	Extensible	Validated through Experts	2.2
EMO	✗	General	Manual Own Method	Not as Modular	Explanations given	Not Validated	1.0
EM-KPI	✗	Electricity Market	Manual NeOn Method [53]	Not as Modular	Extensible	Validated through Experts	1.6
OEO	BFO	General	Manual OBO Foundry Method [51]	Modular	Extensible	Validated through Experts	2.8
Sargon	✗	General	Manual Linked open terms Method [43]	Modular	Modular	Validated	2.2
SEAS	✗	General	Manual Own Method	Modular	Extensible	Not Validated	1.8

Maintenance. This criterion lists the persons or institution maintaining the ontology. Furthermore, it lists if the ontology is actively maintained. An actively maintained ontology is identified as either discourse taking place on GitHub in the issues tab or on the ontology website, or the ontology was updated/revised within 2023, up to April 2024. The criterion is divided into two categories:

No Maintenance found: No Maintenance found. (Score: 0)
Active Maintenance: Maintenance found. (Score: 3)

Latest Release. This criterion lists the date of the latest release as well as the version. It is useful to determine how current the ontology is.

Automatic Analysis. This criterion uses the tools OOPS![22], FOOPS[23] and Fair-Checker[24] to automatically analyze the ontology[17, 18, 44]. A mean of the Fair-Checker score is computed for each of the four elements (F,A,I,R). Furthermore, the HermiT reasoner is run to check for inconsistencies. The data is compiled using a Python script as outlined in subsection 4.3. If the ontology can not be analyzed as one, all parts of the ontology are analyzed and a mean is calculated. The criterion is divided into three categories:

Problems: Problems detected. (Score: 0)
Minor Problems: Less than two minor problems detected, e.g. minor pitfall according to OOPS!. Overall FOOPS score over 50 percent and less than two scores of zero in the mean Fair-Checker score. (Score: 1)
No Problems: No problems detected in the analysis. (Score: 3)

4.4.4 *Governance.*

Funding. This Criterion outlines where the funding that was used in the development of the ontology came from. It is differentiated between the number of source of funding.

No Funding found: No funding found. (Score: 0)
Single Source of Funding Found: Single source of funding found. (Score: 1)
Multiple Sources of Funding Found: Multiple sources of funding found. (Score: 3)

[22] See: https://oops.linkeddata.es/ Last checked 29.04.2024.
[23] See: https://foops.linkeddata.es/FAIR_validator.html Last checked 29.04.2024.
[24] See: https://fair-checker.france-bioinformatique.fr/check Last checked 29.04.2024.

Table 4: A Comparison of the practical implementation of energy research ontologies.

Metric / Ontology	Practical Implementation						
	Available Languages	Available Ontology Languages	Descr. Size	Descr. Quality	Number of Terms	Used Ontologies	Score
Brick	en	OWL Formats: ttl	Small*	Easy*	1623	20	2.0
CIM	en, fr†	RDF/XML+ Formats: owl,ttl	Small	Hard	10922	6	2.0
DABGEO	en	RDF/XML+ Formats: owl,ttl,n3	Small*	Easy*	2523	68	2.3
EMO	en	RDF/XML Formats: owl	Small	Easy	52	5	2.7
EM-KPI	en	RDF/XML+ Formats: owl,ttl,n3	Small	Easy	281	6	3.0
OEO	en	RDF/XML+ Formats: owl,omn	Small	Easy	1518	12	3.0
Sargon	en	RDF/XML+ Formats: owl,ttl	Small*	Easy*	268	9	2.3
SEAS	en	RDF/XML+ Formats: owl,ttl,n3,vowl	Small	Easy	901	13	3.0

* The picked classes include some with no description given
† Only an english version of CIM is available as sourcecode.

Governing Instances. This criterion lists the governing Instances for the ontology. A governing instance being a person or institution with the final say on the content of the ontology.

No Governing Instances found: No Governing Instances found. (Score: 0)
Governing Instances: Governing Instances found. (Score: 3)

Citations. This criterion lists the citation of the paper introducing the ontology as determined via Google Scholar[25].

5 COMPARISON

In this section, we compare the selected ontologies using the categories listed in section 4. The results are summarized in four tables that display the criteria values and use color coding to indicate high scores (green), medium scores (yellow), and low scores (red) according to the scoring. The raw data can be found in an unordered format on Zenodo[26], while a structured comparison is available on the Open Research Knowledge Graph (ORKG)[27].

5.1 Best Practice

In this category, ontologies are compared based on their adherence to best practices in ontology design. As shown in Table 3, all ontologies except for the OEO do not use an upper ontology as their base.

While all the listed ontologies were manually created, only half of them followed a specific creation method established in research literature, while the other half did not use any previously established methods. Among those that used an established method, none employed the same one. Only Sargon uses automated validation tools additionally to a reasoner for validation. Furthermore, evaluation for the ontologies varies significantly. Brick and EMKPI use cases/examples to validate their ontologies [4, 33]. On the other hand, Booshehri et al. [6], who developed OEO, used a coverage study, competency questions, and an agreement study for validation. The creators of Sargon also employ use cases for validation and use automated validation tools through a tool called OnToology [1]. DABGEO is validated based on the reusability and usability of the ontology, as Cuenca et al. [11] set these as their main objectives. CIM is validated through the IEC.

[25] See: https://scholar.google.de/ Last checked 29.04.2024.
[26] See: https://doi.org/10.5281/zenodo.11141516 Last checked 29.04.2024.
[27] See: https://orkg.org/comparison/R690319 Last checked 29.04.2024.

Table 5: A Comparison of the maintenance and accessibility of energy research ontologies. Last checked: 15.02.2023. *Open-Source (OSI):* OSI, *Open-Source/Active Maintenance (Accessibility/Maintenance):* ✓, *Not found:* ✗.

Metric / Ontology	Maintenance and Accessibility								
	Sourcecode	License	Accessibility	Maintenance	Latest Release	Automatic Analysis			Score
						FOOPS	Fair-Checker	OOPS	
Brick	2+	BSD-3-Clause	OSI	✓	V1.4 (2024-04-15)	0.15	"F": 0.5, "A": 1, "I": 0.0, "R": 0.0	unable to generate	2.4
CIM	2+	Apache-2.0	OSI	✓	V7.1 (2022-02-22)ϟ	0.04	"F": 0.75, "A": 0, "I": 1, "R": 0.67	unable to generate	2.4
DABGEO	1S	CC-BY-4.0	✓	✗	V1.0 (2019-02-22)	0.23	"F": 0.05, "A": 0.06, "I": 0.06, "R": 0.08	†Minor:48 Important:36	1.4
EMO	1S	Not Found	✓	✗	V1.0 (✗)	0.04	"F": 1.25, "A": 1, "I": 1.0, "R": 0.67	Minor:2 Important:1	0.8
EM-KPI	2+	CC-BY-4.0	✓	✗	V1.1 (2017-07-07)	0.47	"F": 1.5, "A": 1, "I": 1.67, "R": 0.67	Minor:3 Important:1	1.6
OEO	2+	CC0-1.0	OSI	✓	V2.2.0 (2024-03-04)	0.10	"F": 0.0, "A": 1, "I": 0.0, "R": 0.0	Minor:4 Important:1	2.4
Sargon	2+	CC-BY-4.0	✓	✗	V1.0.1 (✗)	0.69	"F": 1.5, "A": 2, "I": 2.0, "R": 2.0	Minor:3 Important:3	1.6
SEAS	2+	Apache-2.0	OSI	✗	V1.1 (2017-08-29)	0.40	"F": 1.5, "A": 2, "I": 2.0, "R": 2.0	unable to generate	1.8

† Only 23 Oops reports were generated.

ϟ The sourcecode for CIM is not officially available as RDFXML through IEC. While the newest version is given here, the rest of the data is taken from https://github.com/cimug-org/IEC-CIM-Ontology/tree/main Last Checked: 08.05.2024.

5.2 Practical Implementation

This category examines the practical implementation of the ontologies. As shown in Table 4 all ontologies have source code with English language descriptions available.

Most ontologies follow similar implementation practices. Except for Brick and EMO, all the ontologies offer more than one ontology language format. Except for CIM, most ontologies have descriptions that are small and easy to read. It is worth noting that some ontologies do not provide descriptions for certain terms. The median number of terms used is 1210 terms with half of the ontologies (Brick, CIM, DABGEO, OEO) having more than one thousand terms.

Except for DABGEO and OEO, all ontologies use between five and twelve other ontologies. The median number of ontologies used is eleven.

5.3 Maintenance and Accessibility

In this category, the ontologies are compared according to how well they are maintained and how accessible they are. As shown in Table 5, the source code for the ontologies is available from more than one source, except for DABGEO and EMO. For DABGEO and EMO, the source code can be found on their respective websites.

Except for EMO, all of the ontologies have a license, but half of them are not licensed through an OSI-approved license. Every ontology has openly available source code.

Of the eight ontologies, only three (Brick, CIM, and OEO) have seen active maintenance in 2023/2024. The other ontologies do not seem to be maintained anymore. All of the compared ontologies have issues with automatic analysis. Although Sargon was tested with OOPS! when being created, not all mistakes were fixed. Furthermore,

Table 6: A comparison of governance criteria between ontologies. *Not found:* ✗.

Ontology \ Metric	Governance			
	Funding	Governing Instances	Citations	Score
Brick	USA Saudi Arabia Denmark EU Intel Corporation	Found Brick Consortium	373	3.0
CIM	EU USA	Found IEC	2	3.0
DABGEO	Basque Government	✗	26	0.3
EMO	EU Portugal	✗	23	1.5
EM-KPI	China EU	✗	45	1.5
OEO	Germany	Found OEO Steering Comitee	66	2.0
Sargon	Germany	✗	34	0.3
SEAS	Austria Belgium Canada Finland France Germany Hungary Israel Netherlands Norway Portugal Spain Sweden Türkiye ENGIE R&D	Found https://www.maxime-lefrancois.info/	85	3.0

none of the other ontology creators used automated tools to check their ontologies for errors.

5.4 Governance

In this category, we compare how well the ontologies are governed. As shown in Table 6, all of the ontologies in this comparison received funding for their creation.

Most of the ontologies receive funding from governments. Brick and SEAS also received funding from Intel and ENGIE R&D respectively. Only three ontologies receive funding from a single source, while the others receive funding from multiple sources. Only half of the ontologies have a governing instance identified. While OEO and Brick are governed by a group of people, SEAS is governed by a single person and CIM is governed by IEC, a standards organization.

5.5 Evaluation

The comparison displayed in this chapter shows how the central research question of how an ontology can be chosen to be best usable in energy research to describe digital research objects can be answered. It shows the strength and weaknesses of the different ontologies based on their metadata and, while the right ontology depends on the use case it is applied to, the comparison can help to ease the selection of ontologies. This comparison should be seen as a helpful tool and indication.

A score was calculated for each criterion category to assist researchers in quickly choosing the right ontology. This score helps to identify the ontology that is best suited while keeping open science principles in mind. The results of this can be seen in the spiderweb

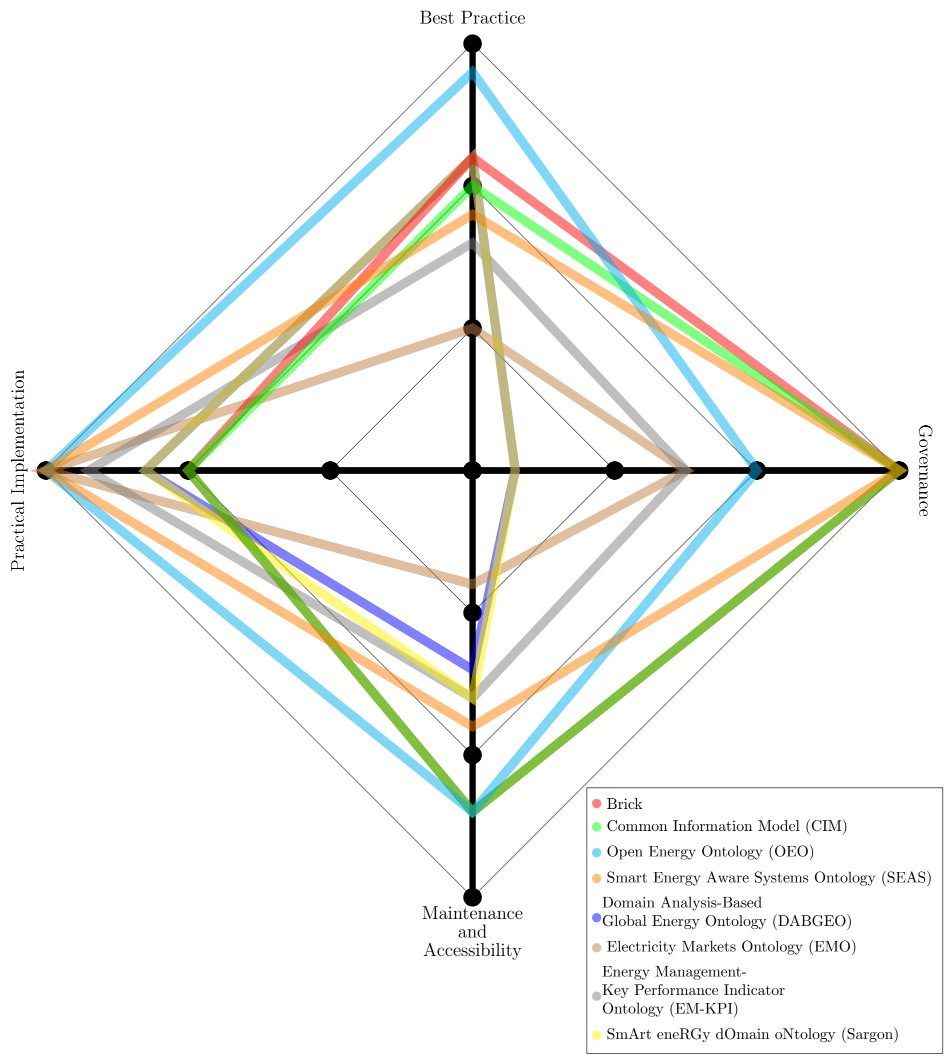


Fig. 2. Spiderweb diagram of the ontology scores.

diagram in Figure 2. The scores for each ontology can be seen in detail in Table 8.

From the score, it can be seen that the top-scoring ontologies are Brick, CIM, OEO, and SEAS. The ontology with the highest accumulated score is the OEO. It also has the highest score in every category except governance. No ontology got full points and there are some flaws in the higher-scoring ones.

The OEO does not use automated tools for validation, and it is not as big as other ontologies like CIM, which means some concepts might be missing. Problems with the SEAS ontology include that it is no longer actively maintained, and it does not use automated analysis tools either. The Brick ontology is only available in one ontology language format for download, and it is mainly an ontology for smart buildings. Because of this focus, updates related to energy topics might be slower developed than in other ontologies. The descriptions in the CIM ontology are harder to understand than others, and it is maintained by the IEC, which makes it less transparent what new features will be added. Furthermore, the ontology is not directly provided in a regular ontology format and needs to be converted.

The ontology recommended in this paper is the OEO. Even though it has flaws, it seems to be the best choice according to the score. The OEO is actively maintained and openly developed on GitHub. If the OEO does not fit the needs, Brick or CIM are viable alternatives. The SEAS ontology is not recommended because it is no longer being updated.

The most important thing when choosing an ontology is the use case, so there is a need to review these recommendations carefully and see if they fit the specific situation.

5.6 Limitations

This work answers how an ontology can be chosen to be best usable in energy research to describe digital research objects. While no definitive answer can be given to the question of which ontology is best to use, as such is the nature of ontologies, recommendations are made based on the comparison. There are however, some limitations to this paper, which are discussed in the following.

This comparison mostly covers metadata of ontologies and does not consider the scope in detail. This is the case, because to look at the scope of an ontology in detail and compare it, a rigid definition of different subdomains in the energy domain are needed and a list of concepts belonging to the subdomains has to be defined, which is outside the scope of this work.

Furthermore, a small amount of ontologies in the energy domain are compared in this paper and this list can be extended in future work. There might also be relevant ontologies missing in the ontology selection step, as is the case with any literature search.

As already discussed in section subsection 4.2, the score that is computed might not be agreeable to all use cases, this is the case for any comparison that shows a focus and the score might need to be adjusted or disregarded.

6 CONCLUSION

In this paper, we explored how ontologies, which can help manage digital research objects in the field of energy research, can be chosen. A comparison of eight energy research ontologies was conducted to find the most suitable one for researchers. To do so, four categories of criteria: best practice, practical implementation, maintenance and accessibility, and governance were developed. These categories include 21 criteria that allow the comparison of ontologies and the identification of their strengths and weaknesses. A script was created to collect data for the comparison and the ontologies were scored based on their openness with a simple scoring function.

As can be seen in Figure 2, the OEO was found to be the most suitable ontology for an energy research project based on the comparison and scoring, thus answering the question postulated in the introduction. Other ontologies that can be recommended are CIM, SEAS, and Brick. However, as the use case is key in choosing an ontology, it is important to make sure to review these recommendations carefully and see if they fit the specific use case.

For future work the list of criteria should be refined, with criteria added or removed based on expert consensus and the comparison should be adjusted based on this refinement. The selection of an ontology could be aided, by integrating more criteria into ontology aggregation/lookup portals like the TIB Terminology Service[28]. Furthermore, the comparison of ontologies designed in the future can be aided, by automating the comparison process where possible. Some of the criteria listed in this paper are collected automatically and more can hopefully be collected automatically in future work.

ACKNOWLEDGMENTS

The authors would like to thank Luca Manzek for proofreading.

The authors would like to thank the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy consortium. The work was partially funded by the German Research Foundation (DFG) 501865131 within the German National Research Data Infrastructure (NFDI, www.nfdi.de).

This research was partially funded by the Lower Saxony Ministry of Science and Culture under grant number 11-76251-13-3/19–ZN3488 (ZLE) within the Lower Saxony "Vorab" of the Volkswagen Foundation. It was supported by the Center for Digital Innovations (ZDIN).

REFERENCES

[1] Ahmad Alobaid, Daniel Garijo, María Poveda-Villalón, Idafen Santana-Perez, Alba Fernández-Izquierdo, and Oscar Corcho. 2019. Automating ontology engineering support activities with OnToology. *Journal of Web Semantics* 57 (2019), 100472. https://doi.org/10.1016/j.websem.2018.09.003

[2] G. Antoniou and Frank Van Harmelen. 2008. *A Semantic Web Primer* (2nd ed ed.). MIT Press.

[3] Robert Arp, Barry Smith, and Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology.* Massachusetts Institute of Technology.

[4] Bharathan Balaji, Arka Bhattacharya, Gabriel Fierro, Jingkun Gao, Joshua Gluck, Dezhi Hong, Aslak Johansen, Jason Koh, Joern Ploennigs, Yuvraj Agarwal, Mario Berges, David Culler, Rajesh Gupta, Mikkel Baun Kjærgaard, Mani Srivastava, and Kamin Whitehouse. 2016. Brick: Towards a Unified Metadata Schema For Buildings. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments* (Palo Alto CA USA, 2016-11-16). ACM, 41–50. https://doi.org/10.1145/2993422.2993577

[5] A L Beck and V Rai. 2019. Solar Soft Cost Ontology: A Review of Solar Soft Costs. 2, 1 (27 Dec 2019), 012001. https://doi.org/10.1088/2516-1083/ab59be

[6] Meisam Booshehri, Lukas Emele, Simon Flügel, Hannah Förster, Johannes Frey, Ulrich Frey, Martin Glauer, Janna Hastings, Christian Hofmann, Carsten Hoyer-Klick, Ludwig Hülk, Anna Kleinau, Kevin Knosala, Leander Kotzur, Patrick Kuckertz, Till Mossakowski, Christoph Muschner, Fabian Neuhaus, Michaja Pehl, Martin Robinius, Vera Sehn, and Mirjam Stappel. 2021. Introducing the Open Energy Ontology: Enhancing Data Interpretation and Interfacing in Energy Systems Analysis. 5 (Sep 2021), 100074. https://doi.org/10.1016/j.egyai.2021.100074

[7] Paolo Brizzi, Dario Bonino, Alberto Musetti, Alexandr Krylovskiy, Edoardo Patti, and Mathias Axling. 2016. Towards an Ontology Driven Approach for Systems Interoperability and Energy Management in the Smart City. In *2016 International Multidisciplinary Conference on Computer and Energy Science (SpliTech)* (Split, Croatia, 2016-07). IEEE, 1–7. https://doi.org/10.1109/SpliTech.2016.7555948

[8] Neil P. Chue Hong, Daniel S. Katz, Michelle Barker, Anna-Lena Lamprecht, Carlos Martinez, Fotis E. Psomopoulos, Jen Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez, and Tom Honeyman. 2021. FAIR Principles for Research Software (FAIR4RS Principles). (2021). https://doi.org/10.15497/RDA00065

[9] Edward Corry, Pieter Pauwels, Shushan Hu, Marcus Keane, and James O'Donnell. 2015. A Performance Assessment Ontology for the Environmental and Energy Management of Buildings. 57 (Sep 2015), 249–259. https://doi.org/10.1016/j.autcon.2015.05.002

[10] Javier Cuenca, Felix Larrinaga, and Edward Curry. 2017. A Unified Semantic Ontology for Energy Management Applications.. In *WSP/WOMoCoE@ ISWC*. 86–97.

[11] Javier Cuenca, Felix Larrinaga, and Edward Curry. 2020. DABGEO: A Reusable and Usable Global Energy Ontology for the Energy Domain. 61–62 (Mar 2020), 100550. https://doi.org/10.1016/j.websem.2020.100550

[12] Laura Daniele, Monika Solanki, Frank den Hartog, and Jasper Roes. 2016. Interoperability for Smart Appliances in the IoT World. In *The Semantic Web – ISWC 2016* (Cham) *(Lecture Notes in Computer Science)*, Paul Groth, Elena Simperl, Alasdair Gray, Marta Sabou, Markus Krötzsch, Freddy Lecue, Fabian Flöck, and Yolanda

[28] See: https://service.tib.eu/ts4tib/index Last checked 29.04.2024.

Gil (Eds.). Springer International Publishing, 21–29. https://doi.org/10.1007/978-3-319-46547-0_3

[13] Houssem Eddine Degha, Fatima Zohra Laallam, Bachir Said, and Djamel Saba. 2018. Onto-SB: Human Profile Ontology for Energy Efficiency in Smart Building. In *2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)* (2018-10). 1–8. https://doi.org/10.1109/PAIS.2018.8598509

[14] Aravind Devanand, Gourab Karmakar, Nenad Krdzavac, Rémy Rigo-Mariani, Y. S. Foo Eddy, Iftekhar A. Karimi, and Markus Kraft. 2020. OntoPowSys: A Power System Ontology for Cross Domain Interactions in an Eco Industrial Park. 1 (01 Aug 2020), 100008. https://doi.org/10.1016/j.egyai.2020.100008

[15] Stephan Ferenz, Annika Ofenloch, Fernando Penaherrera Vaca, Henrik Wagner, Oliver Werth, Michael H. Breitner, Bernd Engel, Sebastian Lehnhoff, and Astrid Nieße. 2022. An Open Digital Platform to Support Interdisciplinary Energy Research and Practice—Conceptualization. 15, 17 (02 Sep 2022), 6417. https://doi.org/10.3390/en15176417

[16] Alba Fernández-Izquierdo, Andrea Cimmino, Christos Patsonakis, Apostolos C. Tsolakis, RaúÍ García-Castro, Dimosthenis Ioannidis, and Dimitrios Tzovaras. 2020. OpenADR Ontology: Semantic Enrichment of Demand Response Strategies in Smart Grids. In *2020 International Conference on Smart Energy Systems and Technologies (SEST)* (2020-09). 1–6. https://doi.org/10.1109/SEST48500.2020.9203093

[17] Alban Gaignard, Thomas Rosnet, Frédéric De Lamotte, Vincent Lefort, and Marie-Dominique Devignes. 2023. FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards. *Journal of Biomedical Semantics* 14, 1 (01 Jul 2023), 7. https://doi.org/10.1186/s13326-023-00289-5

[18] Daniel Garijo, Oscar Corcho, and Marıa Poveda-Villalón. 2021. FOOPS!: An Ontology Pitfall Scanner for the FAIR Principles. 2980 (2021). http://ceur-ws.org/Vol-2980/paper321.pdf

[19] Syed Gillani, Frederique Laforest, Gauthier Picard, et al. 2014. A Generic Ontology for Prosumer-Oriented Smart Grid.. In *EDBT/ICDT Workshops*, Vol. 1133. 134–139.

[20] Tonci Grubic and Ip-Shing Fan. 2010. Supply Chain Ontology: Review, Analysis and Synthesis. 61, 8 (01 Oct 2010), 776–786. https://doi.org/10.1016/j.compind.2010.05.006

[21] Maliheh Haghgoo, Ilya Sychev, Antonello Monti, and Frank H.P. Fitzek. 2020. SARGON – Smart Energy Domain Ontology. 2, 4 (Dec 2020), 191–198. https://doi.org/10.1049/iet-smc.2020.0049

[22] J. L. Hippolyte, Y. Rezgui, H. Li, B. Jayan, and S. Howell. 2018. Ontology-Driven Development of Web Services to Support District Energy Applications. 86 (01 Feb 2018), 210–225. https://doi.org/10.1016/j.autcon.2017.10.004

[23] Tianzhen Hong, Simona D'Oca, Sarah C. Taylor-Lange, William J. N. Turner, Yixing Chen, and Stefano P. Corgnati. 2015. An Ontology to Represent Energy-Related Occupant Behavior in Buildings. Part II: Implementation of the DNAS Framework Using an XML Schema. 94 (01 Dec 2015), 196–205. https://doi.org/10.1016/j.buildenv.2015.08.006

[24] Tianzhen Hong, Simona D'Oca, William J. N. Turner, and Sarah C. Taylor-Lange. 2015. An Ontology to Represent Energy-Related Occupant Behavior in Buildings. Part I: Introduction to the DNAs Framework. 92 (01 Oct 2015), 764–777. https://doi.org/10.1016/j.buildenv.2015.02.019

[25] IEC IEC. 2003. *61970-301: Energy management system application program interface (EMS-API)-Part 301: Common Information Model (CIM) Base*. Technical Report. Technical report, IEC-International Electrotechnical Commission.

[26] Fadime Kaya and Jaap Gordijn. 2021. DECENT: An Ontology for Decentralized Governance in the Renewable Energy Sector. In *2021 IEEE 23rd Conference on Business Informatics (CBI)* (2021-09), Vol. 01. 11–20. https://doi.org/10.1109/CBI52690.2021.00012

[27] Mario J. Kofler, Christian Reinisch, and Wolfgang Kastner. 2012. A Semantic Representation of Energy-Related Information in Future Smart Homes. 47 (Apr 2012), 169–179. https://doi.org/10.1016/j.enbuild.2011.11.044

[28] Joanna Kott and Marek Kott. 2019. Generic Ontology of Energy Consumption Households. 12, 19 (Jan 2019), 3712. Issue 19. https://doi.org/10.3390/en12193712

[29] Serge P. Kovalyov and Olga V. Lukinova. 2023. Integrated Heat and Electric Energy Ontology for Digital Twins of Active Distribution Grids. 2552, 1 (05 Jan 2023), 080005. https://doi.org/10.1063/5.0111541

[30] Dilek Küçük and Doğan Küçük. 2018. OntoWind: An Improved and Extended Wind Energy Ontology. arXiv:1803.02808 [cs] http://arxiv.org/abs/1803.02808

[31] Dilek Küçük, Özgül Salor, Tolga İnan, Işık Çadırcı, and Muammer Ermiş. 2010. PQONT: A Domain Ontology for Electrical Power Quality. 24, 1 (Jan 2010), 84–95. https://doi.org/10.1016/j.aei.2009.06.009

[32] Maxime Lefrançois. 2017. Planned ETSI SAREF Extensions Based on the W3C&OGC SOSA/SSN-compatible SEAS Ontology Paaerns. (2017), 9.

[33] Yehong Li, Raúl García-Castro, Nandana Mihindukulasooriya, James O'Donnell, and Sergio Vega-Sánchez. 2019. Enhancing Energy Management at District and Building Levels via an EM-KPI Ontology. 99 (Mar 2019), 152–167. https://doi.org/10.1016/j.autcon.2018.12.010

[34] Clement Lork, Vishal Choudhary, Naveed Ul Hassan, Wayes Tushar, Chau Yuen, Benny Kai Kiat Ng, Xinyu Wang, and Xiang Liu. 2019. An Ontology-Based Framework for Building Energy Management with IoT. 8, 5 (30 Apr 2019), 485. https://doi.org/10.3390/electronics8050485

[35] Ravi Lourdusamy and Antony John. 2018. A Review on Metrics for Ontology Evaluation. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)* (Coimbatore, 2018-01). IEEE, 1415–1421. https://doi.org/10.1109/ICISC.2018.8399041

[36] Filippos Lygerakis, Nikos Kampelis, and Dionysia Kolokotsa. 2022. Knowledge Graphs' Ontologies and Applications for Energy Efficiency in Buildings: A Review. 15, 20 (Jan 2022), 7520. Issue 20. https://doi.org/10.3390/en15207520

[37] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and The PRISMA Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. 6, 7 (21 Jul 2009), e1000097. https://doi.org/10.1371/journal.pmed.1000097

[38] Jan Morbach, Andreas Wiesner, and Wolfgang Marquardt. 2009. OntoCAPE—A (Re)Usable Ontology for Computer-Aided Process Engineering. 33, 10 (14 Oct 2009), 1546–1556. https://doi.org/10.1016/j.compchemeng.2009.01.019

[39] Mark A. Musen. 2015. The protégé project: a look back and a look forward. *AI Matters* 1, 4 (2015), 4–12. https://doi.org/10.1145/2757001.2757003

[40] Fedor S. Nepsha, Alexei A. Nebera, Alexander A. Andrievsky, and Mikhail I. Krasilnikov. 2022. Development of an Ontology for Smart Distributed Energy Systems *. 55, 9 (01 Jan 2022), 454–459. https://doi.org/10.1016/j.ifacol.2022.07.079

[41] Astrid Nieße, Stephan Ferenz, Sören Auer, Stefan Dähling, Stefan Decker, Johannes Dorfner, Reinhard German, Moritz Gütlein, Veit Hagenmeyer, Sarah Henni, Sebastian Lehnhoff, Johan Lilliestam, Linna Lu, Franziska Mey, Antonello Monti, Christoph Muschner, Jan Reinkensmeier, Mascha Richter, Mirko Schäfer, Philipp Staudt, Wolfgang Süß, Berthold Vogel, Christof Weinhardt, Anke Weidlich, and Julia Zilles. 2022. Nfdi4energy – National Research Data Infrastructure for the Interdisciplinary Energy System Research. (06 Jul 2022). https://doi.org/10.5281/zenodo.6772013

[42] Leif Oppermann, Simon Hirzel, Alexander Güldner, Karoline Heiwolt, Joachim Krassowski, Ulrich Schade, Christoph Lange, and Wolfgang Prinz. 2021. Finding and Analysing Energy Research Funding Data: The EnArgus System. 5 (Sep 2021), 100070. https://doi.org/10.1016/j.egyai.2021.100070

[43] María Poveda-Villalón. 2012. A Reuse-Based Lightweight Method for Developing Linked Data Ontologies and Vocabularies. In *The Semantic Web: Research and Applications* (Berlin, Heidelberg) *(Lecture Notes in Computer Science)*, Elena Simperl, Philipp Cimiano, Axel Polleres, Oscar Corcho, and Valentina Presutti (Eds.). Springer, 833–837. https://doi.org/10.1007/978-3-642-30284-8_66

[44] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. 2014. OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. 10, 2 (2014), 7–34. https://doi.org/10.4018/ijswis.2014040102

[45] Marco Pritoni, Drew Paine, Gabriel Fierro, Cory Mosiman, Michael Poplawski, Avijit Saha, Joel Bender, and Jessica Granderson. 2021. Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis. 14, 7 (Jan 2021), 2024. Issue 7. https://doi.org/10.3390/en14072024

[46] Christian Reinisch, MarioJ Kofler, Félix Iglesias, and Wolfgang Kastner. 2011. ThinkHome Energy Efficiency in Future Smart Homes. 2011, 1 (2011), 104617. https://doi.org/10.1155/2011/104617

[47] Djamel Saba, Fatima Zohra Laallam, Hocine Belmili, Fonbeyin Henry Abanda, and Ahmed Bouraiou. 2017. Development of an ontology-based generic optimisation tool for the design of hybrid energy systems. *International Journal of Computer Applications in Technology* 55, 3 (2017), 232–243.

[48] Djamel Saba, Youcef Sahli, and Abdelkader Hadidi. 2021. An Ontology Based Energy Management for Smart Home. 31 (Sep 2021), 100591. https://doi.org/10.1016/j.suscom.2021.100591

[49] Gabriel Santos, Tiago Pinto, Zita Vale, Isabel Praça, and Hugo Morais. 2016. Electricity Markets Ontology to Support MASCEM's Simulations. In *Highlights of Practical Applications of Scalable Multi-Agent Systems. The PAAMS Collection*, Javier Bajo, María José Escalona, Sylvain Giroux, Patrycja Hoffa-Dąbrowska, Vicente Julián, Paulo Novais, Nayat Sánchez-Pi, Rainer Unland, and Ricardo Azambuja-Silveira (Eds.). Communications in Computer and Information Science, Vol. 616. Springer International Publishing, 393–404. https://doi.org/10.1007/978-3-319-39387-2_33

[50] Nazaraf Shah, Kuo-Ming Chao, Tomasz Zlamaniec, and Adriana Matei. 2011. Ontology for Home Energy Management Domain. In *Digital Information and Communication Technology and Its Applications*, Hocine Cherifi, Jasni Mohamad Zain, and Eyas El-Qawasmeh (Eds.). Communications in Computer and Information Science, Vol. 167. Springer Berlin Heidelberg, 337–347. https://doi.org/10.1007/978-3-642-22027-2_28

[51] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J. Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H. Scheuermann, Nigam Shah, Patricia L. Whetzel, Suzanna Lewis, and The OBI Consortium. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25, 11 (01 Nov 2007), 1251–1255. https://doi.org/10.1038/nbt1346

[52] Thanos G. Stavropoulos, Dimitris Vrakas, Danai Vlachava, and Nick Bassiliades. 2012. BOnSAI: A Smart Building Ontology for Ambient Intelligence. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12* (Craiova, Romania). ACM Press, 1. https://doi.org/10.1145/2254129.2254166

[53] Mari Carmen Suárez-Figueroa. 2010. NeOn Methodology for Building Ontology Networks:Specification, Scheduling and Reuse. https://doi.org/10.20868/UPM.thesis.3879

[54] Nikola M. Tomašević, Marko Č. Batić, Luis M. Blanes, Marcus M. Keane, and Sanja Vraneš. 2015. Ontology-Based Facility Data Model for Energy Management. 29, 4 (01 Oct 2015), 971–984. https://doi.org/10.1016/j.aei.2015.09.003

[55] August Wierling, Valeria Jana Schwanitz, Sebnem Altinci, Maria Bałazińska, Michael J. Barber, Mehmet Efe Biresselioglu, Christopher Burger-Scheidlin, Massimo Celino, Muhittin Hakan Demir, Richard Dennis, Nicolas Dintzner, Adel el Gammal, Carlos M. Fernández-Peruchena, Winston Gilcrease, Paweł Gładysz, Carsten Hoyer-Klick, Kevin Joshi, Mariusz Kruczek, David Lacroix, Małgorzata Markowska, Rafael Mayo-García, Robbie Morrison, Manfred Paier, Giuseppe Peronato, Mahendranath Ramakrishnan, Janeita Reid, Alessandro Sciullo, Berfu Solak, Demet Suna, Wolfgang Süß, Astrid Unger, Maria Luisa Fernandez Vanoni, and Nikola Vasiljevic. 2021. FAIR Metadata Standards for Low Carbon Energy Research—A Review of Practices and How to Advance. 14, 20 (15 Oct 2021), 6692. https://doi.org/10.3390/en14206692

[56] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. 3, 1 (Dec 2016), 160018. https://doi.org/10.1038/sdata.2016.18

[57] Sheng-Yuan Yang. 2013. Developing an Energy-Saving and Case-Based Reasoning Information Agent with Web Service and Ontology Techniques. 40, 9 (Jul 2013), 3351–3369. https://doi.org/10.1016/j.eswa.2012.12.044

A ONTOLOGIES IDENTIFIED AFTER SCREENING

Table 7: Ontologies identified after second screening.

Ontology	Topic	Year	Ref.	Standard
DNAs ontology	Smart Home	2019	[24] [23]	✗
Brick	Smart Home	2016	[4]	✗
ThinkHome	Smart Home	2011	[46]	✗
A comprehensive smart home ontology	Smart home	2012	[27]	✗
Bonsai	Smart Home	2012	[52]	✗
A performance assessment ontology	Smart home	2015	[9]	✗
SEAS Ontology	Energy Domain	2017	[32]	✗
OEO	Energy Domain	2021	[6]	✗
SAREF4ENER	Smart Home	2016	[12]	✓
Facility ontology	Energy Management	2015	[54]	✗
DEHEMS	Energy Management	2011	[50]	✗
Energy-saving ontology	Energy Domain	2013	[57]	✗
OntoPowSys	Smart Grid	2020	[14]	✗
OEMA	Energy Domain	2017	[10]	✗
EM-KPI	Smart Grid	2019	[33]	✗
ProSGV3	Smart Grid	2014	[19]	✗
Building Energy Ontology	Smart Home	2019	[34]	✗
EE-district ontology	Smart Grid	2016	[22]	✗
Sargon	Energy Domain	2020	[21]	✗
DIMMER ontology	Smart Grid	2016	[7]	✗
PQONT	Energy Domain	2010	[31]	✗
Hybrid energy systems ontology	Smart Grid	2017	[47]	✗
DABGEO	Energy Domain	2020	[11]	✗
NewOSEIM	Smart Home	2021	[48]	✗
OpenADR Ontology	Smart Grid	2020	[16]	✗
Electricity Markets Ontology	Energy Markets	2016	[49]	✗
Generic Ontology of Energy Consumption Households	Smart Home	2019	[28]	✗
Solar soft cost ontology	Solar Energy	2019	[5]	✗
OntoWind	Wind Energy	2018	[30]	✗
Onto-SB	Smart Home	2018	[13]	✗
DECENT	Energy Domain	2021	[26]	✗
EnArgus ontology	Energy Domain	2021	[42]	✗
Integrated heat and electric energy ontology	Smart Grid	2023	[29]	✗
CIM	Energy Domain	2003	[25]	✓
∀ Platform ontology	Energy Domain	2022	[40]	✗

B ACCUMULATED SCORE

Table 8: Accumulated scores for each ontology.

Accumulated Score	Best Practice	Practical Implementation	Maintenance and Accessibility	Governance	Ontology
10.2	2.8	3.0	2.4	2.0	OEO
9.6	2.2	2.0	2.4	3.0	Brick
9.6	1.8	3.0	1.8	3.0	SEAS
9.4	2.0	2.0	2.4	3.0	CIM
7.7	1.6	3.0	1.6	1.5	EM-KPI
6.4	2.2	2.3	1.6	0.3	Sargon
6.3	2.2	2.3	1.4	0.3	DABGEO
6.0	1.0	2.7	0.8	1.5	EMO

Received 10 May 2024; accepted 14 June 2024

Utilizing autonomous airport parking facilities as virtual power plants to provide frequency containment reserve

DOMINIK VERENO, RENÉ SCHÜTZ, and KATHARINA POLANEC, Josef Ressel Centre for Dependable System-of-Systems Engineering, Salzburg University of Applied Sciences, Austria
FRANCESCO MALDONATO, CARIAD SE, Germany and Technical University of Berlin, Germany
FLORIAN MIELKE-SULZ, Vienna University of Technology, Austria
DAMIEN KÖGLER, ROLF NICODEMUS, and JOHANNA WEBER, Robert Bosch GmbH, Germany
JOUNES-ALEXANDER GROSS and CHRISTIAN NEUREITER, Josef Ressel Centre for Dependable System-of-Systems Engineering, Salzburg University of Applied Sciences, Austria
AMIN KHODAEI, Electrical and Computer Engineering Department, University of Denver, USA
SEBASTIAN LEHNHOFF, R&D Division Energy, OFFIS, Germany

With the rise of inverter-based renewables, grid stability is increasingly strained. Electric vehicles (EVs) offer unique potential in enhancing this stability through intelligent (dis-)charging control; by aggregation, they can provide significant balancing services to network operators. To exploit this opportunity, airport parking is particularly well suited due to the large number of vehicles with predictable departure times. By using the novel Automated Valet Charging (AVC) technology, parking facilities can optimally schedule EVs—autonomously maneuvering cars to and from charge points and connecting them robotically. We propose using AVC-capable parking facilities as virtual power plants (VPPs), providing fast-acting frequency containment reserve (FCR). This study's key contributions are: 1) presenting the AVC VPP concept and formalizing its two core optimization tasks, 2) developing a modular co-simulation architecture for comprehensive evaluation, 3) performing preliminary experiments, to assess feasibility and to identify simulation challenges, and 4) outlining strategic recommendations for practical AVC VPP implementation. The results suggest the concept's potential in utilizing airport EV parking for FCR, promising additional revenue for parking operators without sacrificing customer satisfaction. However, for conclusive assessment, extended simulation models are necessary. Furthermore, prequalification emerges as a significant obstacle to profitable real-world applications. Ultimately, AVC VPPs may help with transforming EVs from being a burden on the grid to a valuable asset.

CCS Concepts: • **Hardware → Smart grid**; **Smart grid**; • **Applied computing → Engineering**; • **Computing methodologies** → *Discrete-event simulation*; *Simulation environments.*

Additional Key Words and Phrases: cyber-physical energy system, Automated Valet Parking, Automated Valet Charging, co-simulation, power-systems simulation, smart charging, electric vehicle charging, optimization, balancing service

Authors' Contact Information: Dominik Vereno, dominik.vereno@fh-salzburg.ac.at; René Schütz; Katharina Polanec, Josef Ressel Centre for Dependable System-of-Systems Engineering, Salzburg University of Applied Sciences, Puch/Salzburg, Austria; Francesco Maldonato, CARIAD SE, Wolfsburg, Germany and Technical University of Berlin, Berlin, Germany; Florian Mielke-Sulz, Vienna University of Technology, Vienna, Austria; Damien Kögler; Rolf Nicodemus; Johanna Weber, Robert Bosch GmbH, Gerlingen-Schillerhöhe, Germany; Jounes-Alexander Gross; Christian Neureiter, Josef Ressel Centre for Dependable System-of-Systems Engineering, Salzburg University of Applied Sciences, Puch/Salzburg, Austria; Amin Khodaei, Electrical and Computer Engineering Department, University of Denver, Denver, Colorado, USA; Sebastian Lehnhoff, R&D Division Energy, OFFIS, Oldenburg, Germany.

1 Introduction

Power grids are transforming to address the energy challenges of the 21st century and to exploit its opportunities. By interlinking electricity infrastructure with information and communications technology (ICT) for monitoring, control, and planning, traditional grids are transitioning to smart grids [11]. The push for renewables and electric vehicles (EVs), driven by global CO_2 reduction goals, promises sustainable and affordable energy supply, and low-emission transport. On the flipside, these trends are straining grid stability due to their hard-to-predict nature and lack of inertia. However, EVs present unique opportunities in dealing with these challenges: Recent innovations in smart charging promise intelligent control of the charging process in response to grid conditions [7], acting as a form of demand-side management [16]. Beyond that, advancements in vehicle-to-grid (V2G) technologies promise to further amplify this potential, by enabling vehicles to inject power back to the grid at times of particularly high demand [23]. If properly exploited, this flexibility can play a critical role in increasing sustainability without sacrificing dependability.

To leverage smart charging and V2G, airport parking facilities emerge as excellent venues, given their high volume of vehicles with long availability. A particularly compelling aspect for this study is the predictable departure schedule [29], which allows for ensuring cars are sufficiently charged when picked up while offering flexibility in managing the state of charge in the meantime. By aggregating this large number of EVs into a virtual power plant (VPP), an airport parking garage can offer significant balancing services to network operators via energy markets. Such an arrangement can be a win-win situation for all involved parties [29]. Specifically, the quick-acting nature of EV batteries makes them ideal for providing frequency containment reserve (FCR); it is the operating reserve required to continuously manage frequency deviations [10]. The VPP can reduce the charging power of its vehicles or even request their temporary discharge to react to sudden frequency drops. Even though research has already dealt with using EVs for FCR, reliability and availability have been strong concerns, as not being able to fulfill balancing contracts may lead to severe financial penalties [40]. To address these concerns, we focus on the novel

Automated Valet Charging (AVC) [4] technology. Building on Automated Valet Parking (AVP) [43] for autonomous, garage-driven vehicle guidance, AVC summons vehicle to charge points to then be plugged in robotically—enabling fully autonomous management of EVs and their charging, without the need for human intervention. Using AVC, a large number of EVs can be distributed among a small number of high-powered chargers, opening up the possibilities of intelligently scheduling vehicles to always ensure sufficient FCR capabilities without sacrificing the garage's ability to charge all vehicles in time for their departure. This paper introduces the AVC VPP concept, whereby an AVC-enabled parking facility is used to provide ancillary grid services.

An AVC VPP offers significant promise for enhancing grid stability and generating revenue for parking operators. Successful deployment relies on thorough testing. In power-systems engineering, simulation is a critical tool for validation and analysis [38]; it is vital for early and low-risk tests. However, complex cyber-physical systems with a high degree of heterogeneity and independence of subsystems, are especially challenging to simulate. Co-simulation is particularly well suited to these challenges [56]. With co-simulation, independent subsystem simulators are jointly executed [47]. The approach is likely effective in establishing a robust and modular simulation model that can serve as a testbed for various studies on the AVC VPP, including determining financial viability, optimal number of chargers, and its effects on grid operation.

This work establishes a solid foundation for in-depth exploration of the AVC VPP concept by addressing the following objectives:

(1) Detailed presentation of the AVC VPP concept, including the formulation of the two key optimization tasks: vehicle scheduling and charging control (Chapter 3).
(2) Development of a co-simulation architecture tailored to assess the concept in a way that is flexible and extensible to support diverse analyses (Chapter 4).
(3) Evaluation of the concept's feasibility through initial proof-of-concept experiments, including a detailed discussion of relevant data and assumptions (Chapter 5).
(4) Provision of strategic recommendations for future research to advance the concept towards real-world application (Chapter 6).

This study lays the groundwork for analyzing and establishing AVC-capable parking facilities as VPPs to provide fast-acting balancing services. We thus take a step in ensuring improved grid stability while increasing revenue and tending to customer needs. This effort represents a significant part in unlocking the potential in the ever-increasing number of EVs, to transform them from a burden on the grid to a helpful asset.

2 Research foundations

Studying the usage of an AVC garage as a VPP requires background across various disciplines. This chapter insights into AVP and AVC, existing work on using multiple EVs in VPPs, basics on frequency control, and finally the state of research in smart grid co-simulation.

2.1 Automated Valet Parking and Automated Valet Charging

Valet parking makes parking more convenient to the customer, allowing them to drop their vehicle off at a designated location. With AVP, no human operator is needed; instead the parking facility guides the vehicle to a free parking spot fully autonomously. Using a wide array of sensors, the garage plans a safe path for each vehicle and feeds it step-by-step movement instructions, avoiding obstacles along the way. The first fully-operational implementation of AVP to has received SAE Level 4 certification was developed by Bosch in cooperation with Mercedes-Benz at a multistorey car park at Stuttgart Airport [43], shown in Figure 1.

Fig. 1. Automated Valet Parking at Stuttgart Airport

For EV drivers, charging their car during long-term airport parking is important. Therefore, EV charging must be integrated with AVP, ideally without re-introducing the need for human operators. Consequently, Bosch, together with Volkswagen subsidiary CARIAD, presented AVC—set to begin testing in early 2024 [4]. AVC utilizes AVP technology to guide cars from their long-term spot to an empty charge point. The vehicles are then automatically connected to the charging infrastructure via robots, as demonstrated in Figure 2. The concept will likely use a small number of high-powered DC chargers. One alternative would be outfitting many parking spots with low-powered chargers. However, either the cost of procuring an equal number of robots for plugging in would be prohibitive, or a human operator would be needed to plug the cars in. Consequently, DC fast charging emerges as the appropriate option.

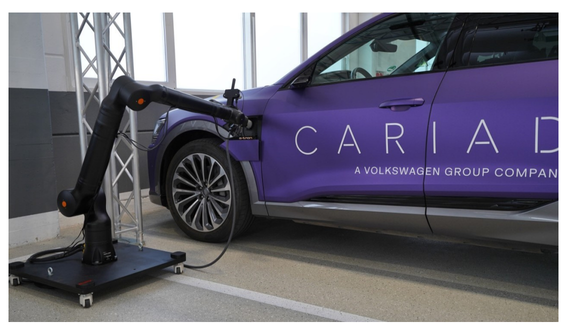


Fig. 2. Automated Valet Charging demonstration

2.2 Electric vehicles in virtual power plants

The increase in distributed energy resources (DERs) requires actively integrating them in the energy system [41]. Such DERs are typically small-scale (3–10,000 kW [50]) and include resources like solar panels, wind turbines, and energy storage systems. A VPP represents a set of DER units to allow them to actively participate in electricity markets while facilitating efficient network operation [41]. VPPs present a wide variety of benefits to operators of DER units, network operators, policymakers, and aggregators [46]. With VPPs, the constituent resources are usually coordinated through advanced software. By aggregating many smaller sources, VPPs mimic larger conventional power plants but with greater flexibility and responsiveness to market demands. The market interaction of VPPs was extensively reviewed by Naval *et al.* [32]: ranging from participation in the day-ahead market to the provision of ancillary servics.

One rapidly advancing type of DER is the EV. As noted by Pudjianto *et al.* [41], DERs are not just generators but also controllable loads. Via smart charging, the charging time and power of EVs can be scheduled in a coordinated way [7] as a form of demand-side management [16]. This can help lessen the negative impact of hard-to-predict high-powered EV charging on the grid. Beyond smart charging, advances in V2G technology potentially allow vehicles to discharge, providing power to the grid during times of high demand and charging during times of lesser demand [23], possibly profiting from energy price arbitrage. Tan *et al.* [54] review V2G technologies for EV–grid integration and various optimization algorithms. Finally, battery-swapping stations may also be operated in grid-aligned way that offers additional revenue to operators without direct customer impact [28]. The difficult challenges and great potential of EV–grid integration has prompted much research into bundling vehicles capable of smart charging and V2G charging in VPPs as reviewed by Yang *et al.* [61] and İnci *et al.* [63].

2.3 Basics of frequency containment reserve

In power grids, even small frequency disturbances can lead to a cascading network failure. The stable and reliable operation therefore requires extensive control measures on various time scales, ranging from a seconds to minutes. Primary control, also known as FCR, is the fastest type of control measure; it rapidly counteracts sudden frequency changes. The grid's frequency is supposed to stay around a nominal frequency f_{nom}, 50 Hz in Europe. If the deviation

$$\Delta f = f - f_{\text{nom}}$$

stays within a previously defined band, no control action is taken. In Europe, this band is typically $[f_{\text{nom}} - 10\text{ mHz}, f_{\text{nom}} + 10\text{ mHz}]$; if it is exceeded, resources tasked with FCR intervene within seconds to contain these fluctuations. Further control measures then act to restore the frequency to its nominal value.

With FCR, this monitoring and control happens completely decentralized, requiring no centralized prompt by the network operator. Instead, a plant tasked with FCR monitors the grid's frequency and injects positive or negative power, depending on the deviation, usually following a linear relationship. Usually, FCR is provided symmetrically, meaning a unit responds to overfrequency by decreasing net injection and to underfrequency by increasing net injection. The power output P of a plant participating in FCR is defined as

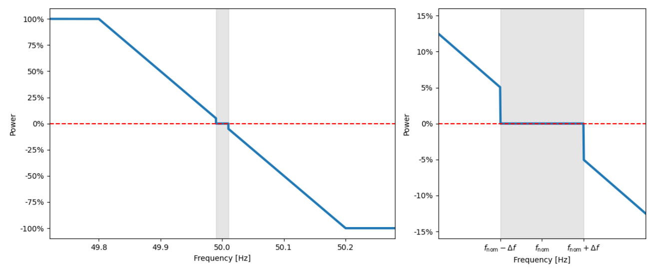


Fig. 3. Illustration of symmetrical frequency containment reserve

$$P = P_{\text{nom}} - K \cdot (f - f_{\text{nom}}) \,, \tag{1}$$

where P_{nom} is the nominal power output, and K is the *droop* coefficient. Figure 3 illustrates symmetrical FCR. For power electronics–based resources, deliberately choosing an appropriate droop coefficient is pivotal; a higher K indicates increased sensitivity to deviations, leading to quicker and more significant power adjustments. Conversely, a lower K results in less sensitivity and smaller power alterations. The choice of K must strike a balance between effective frequency control and avoiding grid instability or overcompensation. In summary, FCR is critical to ensuring stable grid operation, especially in times of increasing inverter-based renewables.

2.4 Smart grid co-simulation

A co-simulation jointly executes multiple models with varying representations and runtime environments [47]. Therefore, simulation models can be developed for individual subsystems 'without having the coupled problem in mind' [47, p. 516]. The paradigm is thus particularly well suited to simulating systems of systems, like smart grids [56]. One can either connect simulators directly, or have them managed by a central orchestrating framework [51]. Such a co-simulation platform is primarily tasked with synchronizing the individual simulations and enabling their data exchange [37]. This study focuses on orchestrated co-simulation, since it allows for a simpler architecture when faced with many simulators [34] and helps to maintain consistency among involved simulation tools [37].

Co-simulation has seen increasing research attention, highlighting its effectiveness in handling complex multidisciplinary energy systems [37]. Beyond power systems, Hafner and Popper [21] review the state of the art of co-simulation in general, and Gomes *et al.* [19] focus on technical aspects. For the energy domain, which seems to be a strong research focus [18], Schweiger *et al.* [48] provide an empirical analysis, whereas a extensive literature review can be found in [30]. Furthermore, Palensky *et al.* [38] provide an extensive primer on co-simulation in cyber-physical energy systems.

3 The Automated Valet Charging virtual power plant

We propose using an AVC-capable airport garage as a VPP to provide ancillary grid services. Via AVC, a limited number of high-powered chargers can deal with a substantial number of EVs. These charge points can be adjusted to rapidly change power, either curtailing charging speeds or discharging vehicles temporarily. This capability presents significant potential for supplying quick-response FCR power to the market, thereby generating a viable revenue stream without compromising the vehicles' charging needs.

Long-term airport parking, with its large number of vehicles and predictable departure times, presents unique advantages for EV charging management [29]. AVC leverages these benefits by scheduling and redistributing vehicles across a limited number of high-powered chargers, ensuring sufficient charging while also being able to provide grid services. To participate in the market for balancing services, such as FCR, an AVC VPP must first undergo prequalification to prove its capabilities to the transmission system operator (TSO). The plant can then submit bids offering FCR services. Usually, FCR is provided symmetrically [27], meaning the plant can supply both positive and negative power in case of under- or overfrequency, respectively. For an AVC VPP, asymmetrical bids are preferable (see Figure 5), as it relies on maximizing the use of a small number of chargers. Negative FCR would require a significant reduction of standard charging operations. This assumption is substatiated by Pavic *et al.* [40], suggesting the asymmetrical case of V2G is the most attractive option financially. After an asymmetrical bid is accepted by the TSO, the VPP is obligated to constantly monintor grid frequency to detect drops below a threshold. Upon detection, counteracting FCR power must be provided. Importantly, the FCR power is the purposeful deviation from standard charging operation, meaning it can be provided by curtailing charging power or by even discharging vehicles temporarily. The plant operator receives financial compensation for their readiness to provide FCR, known as the capacity price, regardless of the actual power delivered [27]. A failure to meet the contractual obligation may lead to signficiant penalties [40]. If properly executed, AVC VPP operation could yield substantial revenue for operators. To ensure profitable operation, it is crucial to implement effective measures ensuring the garage is able to deliver power changes when necessary. This problem can be split into two main tasks (illustrated in Figure 4):

(1) **Vehicle scheduling:** Selecting which vehicles to summon and connect to chargers. This combinatorial decision needs to be periodically reevaluated. It is instrumental for making sure that all cars are sufficiently charged upon departure, and that we can supply enough FCR power at all times.
(2) **Charging control:** Adjusting the charging power of each connected vehicle in response to detected frequency drops. This continuous optimization problem must be repeated frequently to quickly react to grid frequency drops.

3.1 Vehicle scheduling

The first core task is vehicle scheduling, deciding when which car is going to be plugged in. This is necessary for every AVC garage, even without VPP capabilities. However, considering balancing services increases the complexity of the optimization problem. The task requires an overview of all (electric) cars in the garage, including permanent characteristics (e.g. capacity), and temporary properties such as the state of charge. The challenge lies in optimally selecting vehicles in a way that ensures potent and reliable FCR services without compromising the garage's ability to charge vehicles in time

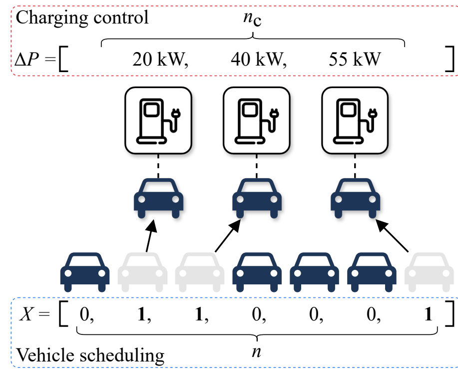


Fig. 4. Vehicle scheduling and charging control

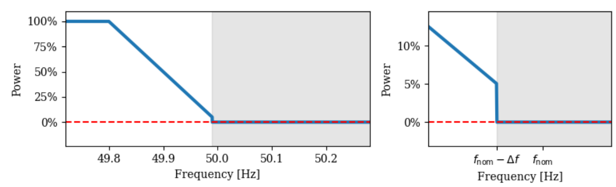


Fig. 5. Illustration of asymmetrical frequency containment reserve

for their pick-up. The task boils down to a binary combinatorial optimization problem. It can be likened to unit commitment, one of the most important power-systems problems whereby a schedule of generating units is decided [36]. Similarly to unit commitment, there are various approaches for timing schedule changes; for this study, we have decided on a simple approach whereby the schedule is decided in advance for a certain time slice of length Δt, e.g. 5 min.

For each of the n EVs in the garage, we only consider a subset of available information for scheduling. For every vehicle $i \in \{1, \ldots, n\}$ we look at the following attributes:

(1) $x_i \in \{0, 1\}$: binary selection variable, represents whether the vehicle is selected for charging.
(2) $x_{\text{prev},i} \in \{0, 1\}$: vehicle was previously selected (i.e. it is currently plugged in).
(3) $c_i \in [0.0, 1.0]$: state of charge.
(4) $P_{\max,i} \in \mathbb{R}^+$: maximum charging power in kW.
(5) $t_{\text{dep},i} \in \mathbb{R}^+$: time to departure in s.

Our goal is to find an optimal configuration of

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \{0, 1\}^n .$$

For this task, we define the objective function $f(X)$ that we aim to minimize. It consists of different sub-objectives f_k:

O1: Vehicles with upcoming departure are selected preferentially.

$$\min_X \left(f_1(X) = \sum_{i=1}^{n} x_i \cdot t_{\text{dep},i} \right)$$

O2: Emphasize plugging in vehicles with low state of charge.

$$\min_X \left(f_1(X) = \sum_{i=1}^{n} x_i \cdot c_i \right)$$

O3: Maximize power availability (disincentivize plugging in only or mostly low-powered cars).

$$\max_X \sum_{i=1}^{n} x_i \cdot P_{\max,i} \quad \rightarrow^{\text{inv}} \quad \min_X \left(f_2(X) = \sum_{i=1}^{n} x_i \cdot (U - P_{\max,i}) \right)$$

$$\text{with } U \geq \max_{i=1}^{n} P_{\max,i}$$

O4: Minimize switching to reduce garage traffic and avoid unnecessary idling of charge points during switch (cars that are already plugged in are prioritized).

$$\max_X \sum_{i=1}^{n} x_{\text{prev},i} \cdot x_i \quad \rightarrow^{\text{inv}} \quad \min_X \left(f_3(X) = \sum_{i=1}^{n} x_{\text{prev},i} \cdot (1 - x_i) \right)$$

O5: Ensure high utilization of charge points (maximize number of selected vehicles).

$$\max_X \sum_{i=1}^{n} x_i \quad \rightarrow^{\text{inv}} \quad \min_X \left(f_4(X) = \sum_{i=1}^{n} (1 - x_i) \right)$$

Using weights $w_1, \ldots, w_5$ we can construct the total cost function $f(X)$ for the scheduling problem:

$$\begin{aligned} \min_X f(X) \quad \text{with } f(X) &= \sum_{k=1}^{5} w_k f_k(X) \\ &= w_1 \cdot \sum_{i=1}^{n} x_i \cdot t_{\text{dep},i} \\ &+ w_2 \cdot \sum_{i=1}^{n} x_i \cdot c_i \\ &+ w_3 \cdot \sum_{i=1}^{n} x_i \cdot (U - P_{\max,i}) \\ &+ w_4 \cdot \sum_{i=1}^{n} x_{\text{prev},i} \cdot (1 - x_i) \\ &+ w_5 \cdot \sum_{i=1}^{n} (1 - x_i) \end{aligned} \tag{2}$$

adhering to the following constraint:

C1: The number of selected vehicles must not exceed the number of available charge points n_c.

$$\sum_{i=1}^{n} x_i \leq n_c$$

3.2 Charging control

The second key task required to realize the AVC VPP concept, is controlling the power of the individual plugged-in vehicles. Whereas vehicle scheduling is necessary even for an AVC garage without balancing capabilities, charging control is required specifically for advanced energy management—either to simply optimize consumption or to provide ancillary services. The task is similar to the economic dispatch problem in real-time operation of power grids to find the output power of all commited generating units [5]. This further strengthens the comparison: whereby vehicle scheduling can be likened to unit commitment, and charging control to economic dispatch. Similarly, the charging control works on a much smaller time scale, and heavily depends on the vehicle selection done by the scheduler. The time scale of charging management must be sufficiently small to act in quick frequency control, making use of the responsiveness of EV batteries.

The control mechanism's job is to supply the required power for FCR. Based on the measured grid frequency f, the required balancing power P_{FCR} is computed. It is proportional to the frequency deviation, as defined in Equation 1. It is the difference to the baseline generation/consumption, so for this study

$$P_{\text{FCR}} = -K \cdot (f - f_{\text{nom}}) \, .$$

Since we focus on asymmetrical operation for financial viability, our baseline operation (leading to nominal power f_{nom}) is to charge all connected vehicles as quickly as possible. When FCR services are requested charging speeds are temporarily reduced, or—if necessary—vehicles are discharged.

The goal of the charging control is to determine a change in charging power Δp_i for each of the plugged-in vehicles $i \in \{1, \ldots, n_c\}$, ideally matching (in sum) the required balancing power P_{FCR} while keeping other objectives in mind and not violating operational constraints. For this task, we must estimate the currently possible charging and discharging power for each vehicle, which, in practice, can be a difficult task, hinging on various aspects like the current state of charge, battery chemistry, temperature, degradation etc. For this formulation, the exact approach for estimating them is not relevant however.

For charging control, the following variables are considered:

(1) $\Delta p_i \in \mathbb{R}$: optimization variable, represents change in power output in kW.
(2) $c_i \in [0.0, 1.0]$: state of charge.
(3) $p_{\text{c,max},i} \in \mathbb{R}^-$: estimated current maximum charging power in kW, since charging represents energy consumption, it is defined as negative.
(4) $p_{\text{d,max},} \in \mathbb{R}_0^+$: estimated current maximum discharging power in kW, we define as positive since net injection, generation (0 if the car does not have V2G capabilities).
(5) $p_{\text{nom},i} \in \mathbb{R}$: nominal power in kW, represents the baseline operation, in our case where we want to charge cars as quickly as possible and not intend to provide negative FCR power, therefore always $p_{\text{nom},i} = p_{\text{c,max},i}$.

Our goal is to find an optimal configuration of

$$\Delta P = \begin{bmatrix} \Delta p_1 \\ \Delta p_2 \\ \vdots \\ \Delta p_{n_c} \end{bmatrix} \in \mathbb{R}^{n_c} \, . \tag{3}$$

For this task, we define the objective function $f(\Delta P)$ which we aim to minimize. Its sub-objectives f_k are:

O1: Our primary objective is achieving the required balancing power P_{FCR} (the sum of changes should be as close as possible to the required value).

$$\min_{\Delta P}\left(f_1(\Delta P) = (P_{\text{FCR}} - \sum_{i=1}^{n_c} \Delta p_i)^2\right)$$

O2: Vehicles with low state of charge (high depth of discharge) should receive less reduction in charging power.

$$\min_{\Delta P}\left(f_2(\Delta P) = \sum_{i=1}^{n_c}(1 - c_i) \cdot \Delta p_i\right)$$

O3: The required balancing power P_{FCR} should be spread fairly across the plugged-in vehicles, in proportion to a vehicle's current maximum available charging power (the ratio between Δp_i and $p_{\text{c,max},i}$ should be similar across all vehicles).

$$\min_{\Delta P}\left(f_3(\Delta P) = \sum_{i=1}^{n_c}\sum_{j=i+1}^{n_c}\left(\frac{\Delta p_i}{p_{\text{c,max},i}+\epsilon} - \frac{\Delta p_j}{p_{\text{c,max},j}+\epsilon}\right)^2\right)$$

$$\text{with } 0 < \epsilon \ll 1$$

So using weights $w_1, \ldots, w_3$ we construct the cost function we want to minimize:

$$\begin{aligned}\min_{\Delta P} f(\Delta P) \quad \text{with } f(\Delta P) &= \sum_{k=1}^{3} w_k f_k(\Delta P) \\ &= w_1 \cdot (P_{\text{FCR}} - \sum_{i=1}^{n_c}\Delta p_i)^2 \\ &+ w_2 \cdot \sum_{i=1}^{n_c}(1 - c_i) \cdot \Delta p_i \\ &+ w_3 \cdot \sum_{i=1}^{n_c}\sum_{j=i+1}^{n_c}\left(\frac{\Delta p_i}{p_{\text{c,max},i}+\epsilon} - \frac{\Delta p_j}{p_{\text{c,max},j}+\epsilon}\right)^2\end{aligned} \tag{4}$$

Adhering to the following constraints:

C1: The output power must not exceed the maximum charging power, so do not go below.

$$p_{\text{nom},i} + \Delta p_i \geq p_{\text{c,max},i}, \quad \forall i \in \{1, \ldots, n_c\}$$

C2: The output power must not exceed the maximum discharging power.

$$p_{\text{nom},i} + \Delta p_i \leq p_{\text{d,max},i}, \quad \forall i \in \{1, \ldots, n_c\}$$

4 Co-simulation architecture

The complex AVC VPP use case requires a capable tool for analysis and validation. We propose co-simulation, whereby independent subsystem simulators are coupled. This chapter presents a co-simulation approach for the AVC VPP, and outlines our architectural decision making as well as tool selection.

4.1 Architectural drivers

Based on the properties of the presented concept and the planned simulation studies (both for this paper and future projects), we have established the following drivers for the co-simulation architecture:

(1) **Modularity**: The co-simulation must facilitate a high degree of interchangeability of components (see [26] for an in-depth discussion on simulator interchangeability).
(2) **Diverse time paradigms**: The significant interplay between ICT and electrical components in the simulation requires support for different time paradigms: fixed time step, variable time step, and event-based [38].
(3) **Information flow transparency**: The co-simulation must enable a straight-forward and transparent way of assessing the various information flows between subsystem simulators.
(4) **Extensibility**: The architecture of the co-simulation must facilitate the addition of further simulators, to be prepared for future extension with augmented capability.

4.2 Tool selection

The complex tasks involved in managing and coordinating multiple simulators usually require using a framework with a dedicated orchestrator algorithm to initialize simulators, synchronize them, and exchange data between them [38]. Vogt *et al.* [59] present an extensive overview of the variety of frameworks specifically for smart grid co-simulation. We draw from this selection of frameworks. In addition to being suitable for power-systems analysis, our selection must fulfill the following requirements:

- The framework must support large-scale scenarios, including numerous simulated entities, since we plan on scaling the co-simulation to incorporate other aspects such as a comprehensive power grid simulation.
- The framework must have an easy-to-use programming interface to connect simulators with the orchestrator.
- The framework must allow for versatile programmatic specification of co-simulation scenarios.
- The framework must be free and open-source, to facilitate accessibility and reproducibility of our work.

These criteria are addressed by the framework Mosaik. It was originally introduced by Schütte *et al.* [49] and Rohjans *et al.* [45]. We now use the current major version 3.0, presented in [35]. Mosaik provides a high-level API for selected languages; it further provides a low-level API designed around sending JSON-encoded packages via TCP. The flexible framework allows for integrating a variety of heterogeneous simulators; Python-based simulators can even be run within the same process. Mosaik offers various time paradgims, namely time-based (either fixed or variable step), event-based, or hybrid. With hybrid simulators, both time-based and event-based stepping is supported. This paradigm has proven highly useful for our simulation needs.

4.3 Multi-controller architecture

The multi-controller is an architectural pattern which is particularly important for our simulation scenario. The need for a multi-controller arises from the following situation: A controller controls an arbitrary number of simulated systems. For example, a charging station management system controlls the available charging power for multiple charge points [56]. The controller receives data from all of them and sends individual data to each. To realize this architecture in Mosaik, we introduce the *multi-controller*, which consists of the controller itself and channels for each individual system it controls (see Figure 6). The number of instantiated channels is based on the number of controlled systems and can be scaled arbitrarily. In Mosaik, the concept is realized by having the channels be so-called *children* entities of the controller.

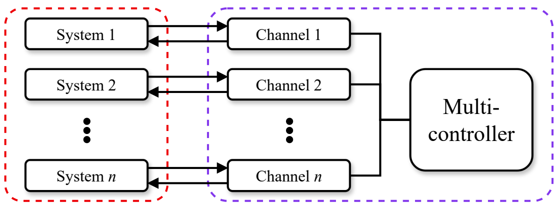


Fig. 6. Multi-controller architecture pattern

4.4 Architecture overview and components

Here, we present the co-simulation architecture, depicted in Figure 7. The architecture overview highlights which simulated entities are considered part of the garage. It further shows that for the charge point simulator, multiple entities exist; each entity being connected to the two multi-controllers; the vehicle connector and the garage energy management system (EMS).

4.4.1 External components. Some simulators fulfill the role of replicating outside influences on the AVC VPP garage. They are subsystems outside the direct control of the garage.

- **Passenger flow** represents the passenger volume handled by the airport. It is considered proportional to vehicle arrivals. The time scale could be variable, data on a day-to-day basis or smaller time scale.
  Time paradigm: time-based (constant), providing passenger numbers in regular, pre-determined intervals (e.g. 24 hours).
- **Vehicle dispatcher** is responsible for generating the (electric) vehicles arriving at the garage. It scales its output depending on the airport's total passenger volume. The dispatcher generates vehicles with various attributes—some of them permanent characteristics (e.g. capacity), others are temporary properties (e.g. state of charge).
  Time paradigm: time-based (variable), irregularly generating new vehicles with random time intervals in between.
- **Power grid** is a simplified representation of the power grid. For this study, it is solely responsible for generating an appropriate frequency profile for testing.
  Time paradigm: time-based (constant), simulating the ever-changing grid frequency in small time steps (e.g. 1 second).

4.4.2 Internal components. The other simulators represent subsystems of the garage itself. They simulate the garage's capability to move cars from their parking space to a charger, the garage's ability to charge cars, and its ability to respond to frequency deviations in the grid.

- **Vehicle connector** represents the garage's capability of moving vehicles throughout the garage and maintains inventory of the EVs within. Functioning as a multi-controller, the connector receives dispatched vehicles, guides selected vehicles to their charge point, and manages their charging. It provides a list of EVs and receives a subset of selected vehicles.
  Time paradigm: event-based, is externally triggered by incoming vehicle selections, arriving vehicles, and reacts to changing charging power of the charge points.
- **Charge point** simulates the behavior of the individual chargers. It relays the requested power to the plugged-in EV, and—conversly—sends the actual charging power from the EV.
  Time paradigm: event-based, acting solely as a relay for external data streams.
- **Garage EMS** manages the parking garage's various energy-related aspects. It is the heart of the garages AVC VPP capabilities, since it includes the two key tasks: vehicle scheduling (Section 3.1) and charging control (Section 3.2). The garage EMS receives the list of vehicles and then calls the internal scheduler to make a selection, which is then sent back. Moreover, the EMS uses the grid's frequency to compute the required balancing power P_{FCR}, which—combined with the information about the currently plugged-in vehicles—is used to determine the appropriate charging power for each vehicle. The garage EMS further outputs the total power to the grid.
  Time paradigm: hybrid, being prompted externally by incoming vehicle data, but also in small time steps (e.g. 1 second) measuring frequency and determining charging power.

5 Proof-of-concept experiments

The co-simulation architecture presented in the previous chapter can now serve as a testbed for experiments. At this stage, the focus is on establishing a baseline assessment of the concept's feasibility and to lay the groundwork for more extensive experimentation.

5.1 Assumptions and data

The simulation setup requires a broad range of data for parameterization; additionally, various assumptions have to be made. To facilitate transparency and reproducibility, we lay out the data used and assumptions made. The four main areas discussed are:

- Plausible distribution of vehicle types, encompassing various EV characteristics (Section 5.1.1).
- Arrival data:
  - Arrival time distribution, to create a random and appropriate distribution of arrivals (Section 5.1.2).
  - Parking duration and departure time (Section 5.1.3).
  - State of charge at arrival (Section 5.1.4).
- Charging and discharging characteristics, the power depends heavily on current battery state, modeled here (Section 5.1.5).

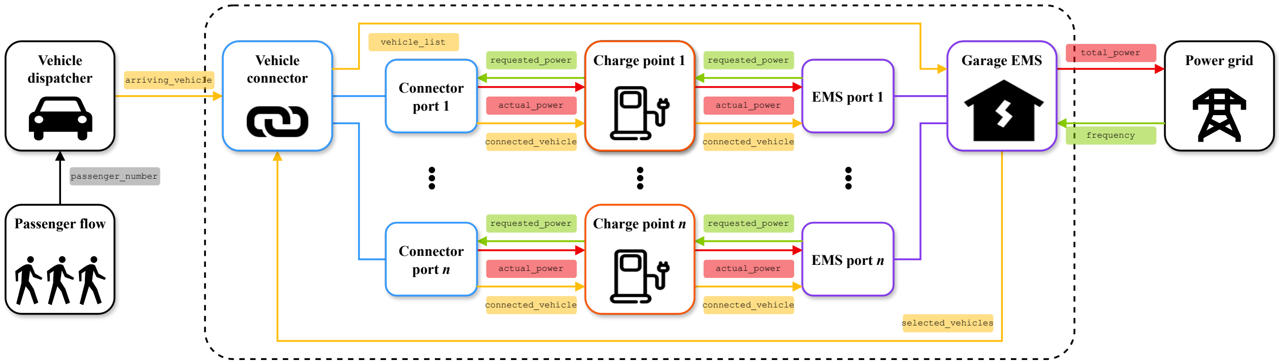


Fig. 7. Co-simulation architecture overview

- Time series of realistic grid frequency (Section 5.1.6).

5.1.1 Vehicle types. We aim to replicate a realistic distribution of vehicle types for our study. Since the AVC pilot is planned in Germany, we focus on official German registration statistics from the year 2022 from the Kraftfahr-Bundesamt [25], the German motor transport authority. From this data, we extract the 20 most popular models and exclude any models that lack DC fast-charging capabilities. We use the identifiers in that list (the TSN, a vehicle type number) to get further vehicle details from a vehicle database maintained by the ADAC [2], Germany's leading automobile club. We specifically extract the following info for each type:

- net battery capacity in kWh,
- maximum charging power in kW, and
- charging duration from empty to 80 % in min.

Furthermore, the registration numbers of the vehicle types are used to determine their probability (normalized for the types used in this study). The vehicle types, their characteristics and probabilities are compiled in Table 1.

Name	Prob.	Capacity [kWh]	Power [kW]	Duration [min]
Fiat 500e	0.12	37.30	85	35
Tesla Model Y Perf.	0.10	79	250	22
Tesla Model 3	0.07	62	170	21
Opel Corsa	0.06	46	100	30
Dacia Spring	0.06	26.80	30	56
VW ID.4 PRO	0.05	77	125	38
VW e-up!	0.05	32.30	40	60
Opel Mokka Electric	0.05	46	100	30
Mini Cooper SE	0.05	28.90	50	36
Tesla Model 3 LR	0.04	79	250	22
Renault Zoe	0.04	52	50	70
BMW i3	0.04	27.20	50	39
Peugeot e-208	0.04	46	100	30
Tesla Model Y LR	0.04	79	250	22
VW ID.3 PRO	0.04	58	100	35
Hyundai Kona Elektro (39,2 kWh)	0.04	39.20	100	47
Renault Megane E-Tech	0.03	52	130	42
Hyundai Kona Elektro (64 kWh)	0.03	64	100	47
Cupra Born	0.03	58	120	35

Table 1. List of considered EV types

5.1.2 Arrival distribution. A critical factor in our simulation is the distribution of EV arrivals. The number of arrivals is subject to significant fluctuations, the major ones being 1) intra-day, and 2) seasonal. For the intra-day variability, we use a highly simplified sinusoidal approximation, as in [3]. This will be our basis for our daily density function $d_{\text{arr}}(t)$, with $t \in [0, 24)$ being the time in h. The periodicity is 24 h, therefore $\omega = 2\pi/24 = \pi/12$. The peak is at 13:00; hence, $T_{\text{peak}} = 13$, therefore $\tau = T_{\text{peak}} - 6 = 7$. We further shift the function, so that $\forall t \in [0, 24) | d_{\text{arr}}(t) \geq 0$, and normalize it so that $\int_0^{24} d_{\text{arr}}(t)\mathrm{d}t = 1$. We further make sure the density function is always $> 0 \; \forall \; t \in [0, 24)$. Therefore, we introduce the additional parameter α, ratio between the sinusoidal and the constant part. For example, for $\alpha = 3$ the sinusoidal part is be responsible for $1/3$ of the arrivals, while the constant part makes up $2/3$. For our experiments, we assume $\alpha = 2$. Let us denote $\min d_{\text{arr}}$ the minimum of our density function in the interval $[0, 24)$, then

$$\frac{1}{\alpha} = \frac{\int_0^{24} d_{\text{arr}}(t)\mathrm{d}t - \min d_{\text{arr}}}{\int_0^{24} d_{\text{arr}}(t)\mathrm{d}t} .$$

We further scale the density function by n_{arr}, the expected number of daily arrivals. We assume it to be proportional to the airports total number of passengers for a certain day n_{pass}, which we get from the annual report of Stuttgart Aiport [13]. The selection of $r = n_{\text{pass}}/n_{\text{arr}}$ heavily depends on the scope of the experiments; for example, we focus on a single level of a multi-storey parking facility and assume $r = 1/3 \cdot 10^4$.

Our full density function (depicted in Figure 8) is

$$d_{\text{arr}}(t, n_{\text{pass}}) = \frac{1}{r} n_{\text{pass}} \cdot \frac{\sin(\omega \cdot (t - \tau)) + \alpha}{24\alpha} . \tag{5}$$

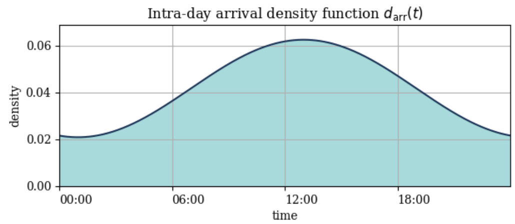


Fig. 8. Intra-day vehicle arrival distribution for $\alpha = 2$

Finally, we devise a mechanism to randomly produce vehicles in accordance with this distribution. Ater each arriving vehicle, we determine the wait time to the next, following an exponential distribution. The time between arrivals is denoted as $t_{\text{wait}} \sim \text{Exp}(\lambda)$, with λ being the expected wait time. Consequently, $1/\lambda$ is the expected arrival rate in vehicles per hour, which depends on the time of day according to $d_{\text{arr}}(t, n_{\text{pass}})$. Consequently, the wait time is

$$t_{\text{wait}} \sim \text{Exp}\left(\frac{1}{d_{\text{arr}}(t, n_{\text{pass}})}\right) . \tag{6}$$

5.1.3 Parking duration and departure time. We model the distribution of flight trip durations according to the estimations in [44]. The natural-language categories are interpreted as day ranges (Table 2).

Duration	Interpreted day range	Probability
shorter than 2 days	$[1, 2)$	2 %
2–3 days	$[2, 4)$	45 %
4–6 days	$[4, 7)$	34 %
1–2 weeks	$[7, 15)$	14 %
longer than 2 weeks	$[15, 29)$	5 %

Table 2. Trip durations with associated probabilities

We apply a Gaussian kernel density estimation with a smoothing bandwidth of $h = 0.3$ to the distribution. The density estimator $\hat{f}(d)$, with d being the trip duration in days, is truncated to the interval $d \in [1, 29)$ and then renormalized to preserve its property as a probability density function. The original data and the estimated probability density function are compared in Figure 9.

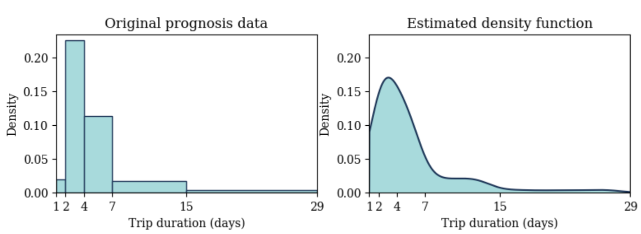


Fig. 9. Trip duration density estimation

The resulting distribution is similar to the one used in [17], where short trips (1–3 days) are very frequent, up-to-6-day trips occur quite often, and the probability of longer trips is significantly lower. In contrast, our distribution goes up to 4 weeks, whereas the one used in [17] is limited to 2 weeks.

5.1.4 State of charge at arrival. When arriving at the parking facilities, the cars exhibit various states of charge. In reality, this value strongly depends on a multitude of factors, e.g. from where the car is traveling. For this study, however, we assume a simplified independent distribution of the state of charge that is based on a report by the U.S. Department of Energy [14] using the example of away-from-home charging of Nissan Leaf cars; this data was also used for analyses in [17]. Similarly to that study, we assume a Gaussian distribution and estimate $\hat{\mu} \approx 0.493$, $\hat{\sigma} \approx 0.177$. Additionally, we limit the distribution to the interval $[0.0, 1.0]$, meaning c can only fall within that range. For this study, the state of charge c therefore follows a truncated normal distribution:

$$c \sim \text{TN}(\hat{\mu}, \hat{\sigma}, 0.0, 1.0) . \tag{7}$$

Figure 10 shows the distribution from the report and the approximated truncated normal distribution.

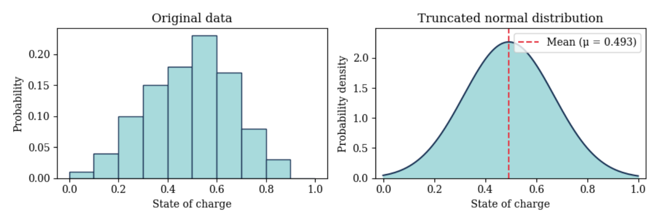


Fig. 10. State of charge at arrival

5.1.5 Charging and discharging characteristics. The maximum charging and discharging rates of EV batteries vary depending on several factors including the state of charge and environmental conditions such as temperature. Therefore, (dis-)charging rate variability must be considered when simulating EV charging. For our experiments, we focus solely on the impact of the state of charge and create a simple linearized model, forgoing more detailed analytical models like the kinetic battery model [22]. The main reasons for the simplified approach are the difficulty and expenses of gathering detailed battery data, implementation complexity, and computational cost. Our model is based on three widely available attributes: net capacity, maximum charging power, and charging duration to 80 %. These are the data gathered in Section 5.1.1. For both the charging rate and the discharging rate, we create a piecewise linear model.

For the charging rate, we define two linear functions, one up to 80 %, one above that. It is a reasonable cut-off since the charging rate is often considered to drop significantly above 80 %; therefore, the charging duration is often given up to that point. We define $P_c(c)$ as the maximum possible charging power dependent on the state of charge c, using the maximum charging power $P_{\max}$ in kW, the expected charging time to 80 % T in h, net capacity C in kWh. We further introduce P_{avg} as the average charging power from 0 % to 80 %, and $\kappa \coloneqq P_c(c)(1.0)/P_c(0.8)$ as a measure for the final drop-off in charging power with $0 < \kappa \ll 1$.

$$P_c(c) = \begin{cases} P_{c,1}(c) = m_{c,1}c + b_{c,1} & \text{if } 0.0 \leq c \leq 0.8 \\ P_{c,2}(c) = m_{c,2}c + b_{c,2} & \text{if } 0.8 < c \leq 1.0 \end{cases} \tag{8}$$

with

$$m_{c,1} = \frac{P_{avg} - P_{max}}{40}, \quad b_{c,1} = P_{max}, \text{ and}$$

$$m_{c,2} = \frac{\kappa P_{c,1}(0.8) - P_{c,1}(0.8)}{0.2}, \quad b_{c,2} = P_{c,1}(0.8) - 0.8 m_{c,2} \ .$$

Moreover, we have to define our discharging power function $P_d(c)$. Here, we face major additional issues:

(1) Only few EVs today have V2G capabilities.
(2) Many V2G-capable vehicles discharge only at a slow rate, too insignificant for balancing services.
(3) Reliable and detailed data about discharging characteristics is difficult to acquire.

Therefore, we have to make significant assumptions for this experiment. We assume sizeable V2G capabilities that are proportional to the vehicle's charging rates. Specifically, we assume a mirroring of charging behavior: low discharge rates for low charge that increases with rising c. We define $P_d(c)$ as the discharging rate in kW. For $P_d(c)$, we set the cut-off of our piecewise linear definition at 20 %, and furthermore assume scaled-down charging behavior by a factor of $\delta = 0.65$. Similar to $P_c(c)$, we get

$$P_d(c) = \begin{cases} P_{d,1}(c) = m_{d,1}c + b_{d,1} & \text{if } 0.0 \leq c \leq 0.2 \\ P_{d,2}(c) = m_{d,2}c + b_{d,2} & \text{if } 0.2 < c \leq 1.0 \end{cases},$$

with

$$m_{d,2} = \frac{\delta \cdot (P_{max} - P_{avg})}{0.4}, \quad b_{d,2} = \delta \cdot P_{avg} - 0.6 m_{d,2} \ , \text{ and}$$

$$m_{d,1} = \frac{P_{d,2}(0.2) \cdot (1 - \kappa)}{0.2}, \quad b_{d,1} = P_{d,2}(0.2) - 0.2 m_{d,1} \ .$$

Figure 11 shows the charging and discharging curves for an Opel Corsa Electric, with net capacity $C = 46$ kWh, maximum power of $P_{max} = 100$ kW, and a charge time of $T = 30$ min.

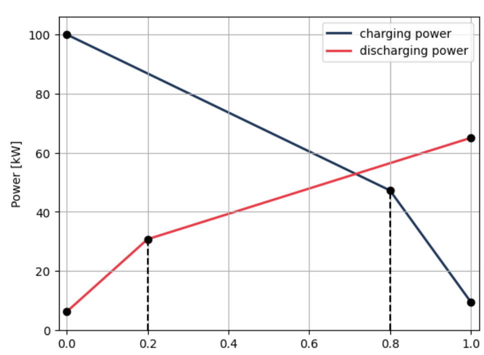


Fig. 11. Piecewise linear (dis-)charging curves for an Opel Corsa Electric

5.1.6 Grid frequency. To analyze the FCR capabilities of the AVC VPP garage, we need the grid frequency as it determines the required output power P_{FCR} (see Section 2.3). For our experiments, we take a one-hour time series (Figure 12) an replicate it for all simulated times. Specifically, we use the measurements of 01.07.2022, 10:30–11:30 from a low-voltage grid in Germany provided by the Fraunhofer Institute for Solar Energy Systems [15]. They have compiled a multi-year, high-frequency (1 measurement per second) database.

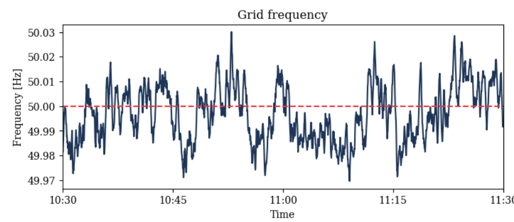


Fig. 12. Examplary grid frequency timeseries used in experiments

5.2 Optimization implementation and droop configuration

For the core optimization tasks, the vehicle scheduling (Section 3.1) and charging control (Section 3.2), we have selected the CPLEX modeling toolset, a commonly used optimization software for academic and industrial applications. For the scheduling task, we use binary optimization, and for the charging control, we use continuous optimization. The optimization objectives and constraints described in sections 3.1 and 3.2 are implemented. Based on preliminary experiments, we have chosen the following weights:

- vehicle scheduling (Equation 2): $w_1 = 10^4$, $w_2 = 10^{-3}$, $w_3 = 5$, $w_4 = 7$, $w_5 = 5 \cdot 10^7$, and
- charging control (Equation 4): $w_1 = 1$, $w_2 = 3 \cdot 10^{-2}$, $w_3 = 1$.

Parameters for droop-based FCR control:

- droop coefficient $K = 5$ kW/mHz,
- maximum power $P_{FCR,max} = 100$ kW, and
- allowable frequency deviation $\Delta f = 10$ mHz.

5.3 Simulation experiments and results

First, we observe the number of vehicles in the parking garage. Specifically, we only focus on a part of the garage comperable in size to a single level of a multi-storey parking facility. We simulate an exemplary 2-month period (July and August 2022), depicted in Figure 13. The time series shows initial increase, followed by stable (yet fluctuating) operation. The initial increase is due to the simulated garage needing to fill up with vehicles before reaching normal operation, this initial period is dictated by the maximum stay duration of up to 4 weeks, as defined in Section 5.1.2. Figure 13 further compares the vehicle number to the passenger number, highlighting the impact of intra-week passenger fluctuations on the number of vehicles.

As for the number of chargers, we run our experiments for $n_c = 5$, a plausible number for a single parking level. It should be noted, that this number is significantly lower than the number of EVs. Each charger sees multiple different cars during operation; for this

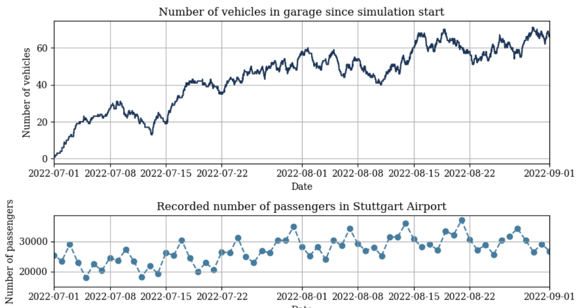


Fig. 13. Total number of EVs in garage

study, we configured the scheduling so that the connected vehicle may change every 5 minutes. Depending on the state of the vehicles, a single vehicle may remain connected for several consecutive 5-minute periods. Figure 14 shows an exemplary schedule. Each colored bar represented the connection period of a particular vehicle. The figure further shows that even though the connection periods are often quite a lot longer than the 5-minute resolution, multiple switches may still occur simulatenously.

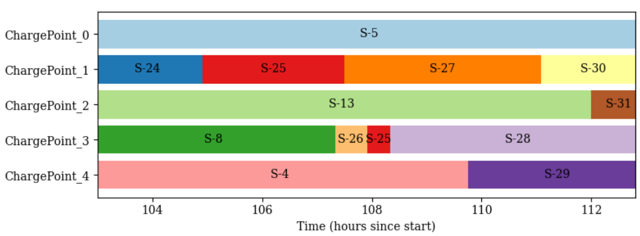


Fig. 14. Exemplary EV schedule

Furthermore, we illustrate the relationship between the grid's frequency and the requested FCR power P_{FCR} in Figure 15. For our FCR configuration, which is asymmetrical, only underfrequency leads to requested FCR power, specifically positive P_{FCR}, to counteract the frequency drop. The figure also highlights that the output power is capped to the previously defined maximum of 100 kW. It further shows that the minimum power output other than 0 is $K \cdot \Delta f = 5 \text{ kW/mHz} \cdot 10\text{mHz} = 50 \text{ kW}$, since FCR is only activated once the threshold is surpassed.

The total requested FCR power is distributed among the $n_c = 5$ charge points, and consequently the respective connected vehicles. If properly calculated by the Garage EMS, the power requested of each vehicle should be the same as the actual power it provides; it could, however, deviate.

The key criteria for these preliminary experiments is if the garage is able to provide the necessary FCR power while still properly charging the vehicles in the garage. Regarding the charging, we evaluate the distribution of final state of charge of the vehicles. Our experiments yielded the distribution depicted in Figure 16. The histogram shows that no vehicles depart with a low state of charge, while most are above 90 %. When compared to the distribution of

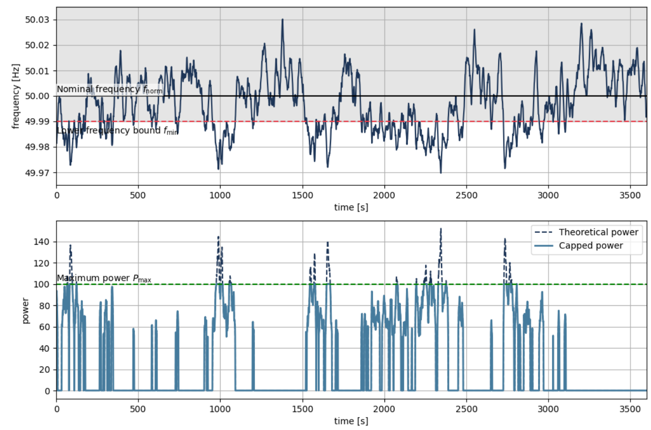


Fig. 15. Grid frequency and FCR power comparison

initial state of charge (see Figure 10), this can be seen as a promising result, if not fully satisfactory. Regarding the garage's FCR-providing capabilities, we compare the total power output of the garage with the requested FCR power in Figure 17. It shows that the changes in power output closely match the spikes in FCR power, meaning the garage is largely able to change its power output according to grid demand. One can further observe, that FCR power request can lead to not just reduction in charging power, but actually a net discharge of vehicles for short durations.

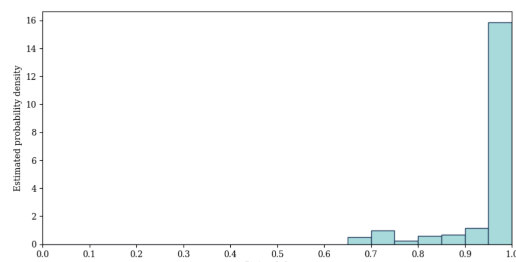


Fig. 16. Distribution of state of charge at departure

5.4 Discussion

The preliminary proof-of-concept experiments have provided valuable insights into the AVC VPP concept, offering an initial feasibility evaluation and revealing various challenges that have to be overcome. Our key findings are:

- **Feasibility**: Our study suggests that the AVC VPP concept is feasible. The experiments show that the parking facilitiy is able to provide the necessary FCR power without sacrificing its ability to charge the vehicles sufficiently. Furthermore, the experiments demonstrate the functionality and efficacy of the key tasks: vehicle scheduling and charging control. The study suggests that the unique advantages of airport parking combined with AVC capabilities are ideal for usage as VPP. However, even though early results are promising, they are not yet conclusive.

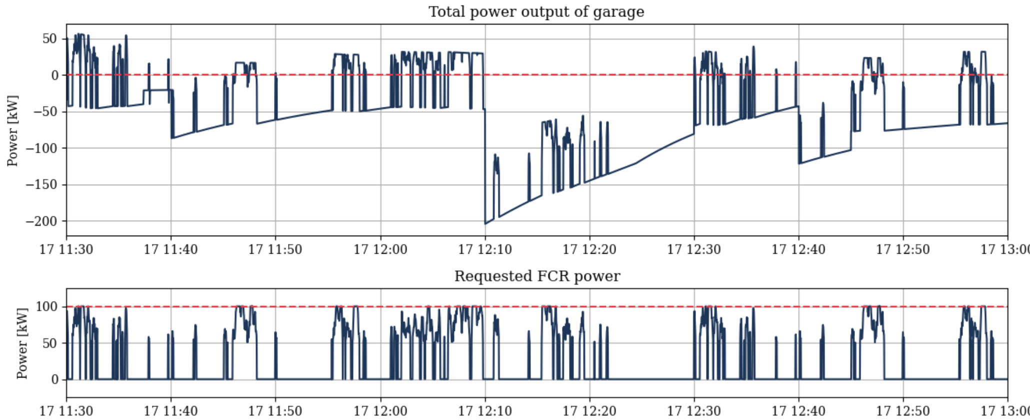


Fig. 17. Comparison between requested FCR power and actual power output

- **Simulatenous re-scheduling**: Our current approach reschedule vehicles every 5 min. At some chargers, the car will stay plugged in, at others, there will be a switch. This simultaneous switching of multiple cars leads to short but significant dips in FCR-providing capabilities, potentially causing temporary loosing the ability to fulfill contractual balancing obligation. To properly understand this effect, it will be necessary to augment the vehicle connector model to incorporate realistic switching times (duration of replacing a car at a charger). As opposed to unit commitment in powergrid operations, overlap of consecutive units is not possible. We therefore suggest substituting the current approach with a more flexible scheduling procedure (e.g. event-based) that prevents simulatenous switching of multiple cars.
- **Computing time**: Running the simulation for many simulated months takes considerable computing time. In fact, on a consumer-grade computer, it can take up to an entire day, proving to be an obstacle for fine-tuning the optimization. When applying complex optimization methods that need extensive hyperparameter tuning, this could be prohibitive. There is likely great potential to accelerate parts of the simulation, requiring profiling to determine performance bottlenecks and issues. Furthermore, the step time of some simulators could be altered to further optimize performance. Co-simulation suffers even more from performance issues due to additional overhead; however, it also offers the potential for parallelized or distributed execution [53], as highlighted by [31] and [42] who emphasizes the role of High-Level Architecture (HLA) [9]. However, to reap the performance rewards of distributed co-simulation, further research is needed [52].
- **Long initialization**: Upon starting the simulation, it takes a few simulated weeks for the parking facility to fill up and reach a reasonable occupancy (see Figure 13). In combination with the issue of computational performance, this hinders effectively exploring the scenario, due to the long wait time. Regardless of optimizing the entire simulation's performance, it would be highly beneficial to devise a quick way of initializing the simulated garage with a plausible vehicle occupancy (including plausible states of charge), making simulation analyses quicker and more efficient.

6 Recommendations and future work

We have explored the concept of using an AVC-enabled garage to provide balancing services as a VPP by detailing the core optimization problems, suggesting a workable co-simulation architecture, and conducting preliminary proof-of-concept experiments. Advancing the concept from theory to practical application demands thorough investigation across multiple research domains. Here, we pinpoint key areas requiring further exploration to not only establish feasibility but also ensure financial and operational viability.

6.1 Simulation refinement and augmentation

So far, we have established a solid simulation basis for various analyses of the AVC VPP concept. However, the simulation model must be improved and expanded to allow for more in-depth studies that go beyond feasibility-oriented experiments. There are four main areas for improvement: refining simulation data, using more advanced sub-models, and integrating complex simulators for the electricity grid and market.

6.1.1 Data refinement. The various data and assumptions have a significant impact on the simulation results. We therefore suggest conducting sensitivity analyses and studying the influence of the parameters; this would result in a set of particularly influential variables, warranting prioritized refinement. The main candidates for improvement are all the data described in Section 5.1. However, we assume that the details regarding vehicle arrival are especially impactful, ranging from vehicle types, arrivals, and stay durations; real-world data by a parking garage operator is valuable here. Furthermore, proper data about discharging capabilities would make studies involving V2G functionality vastly more accurate. For the near-term future, working with forecasts may be necessary, since the current EVs on the road are significantly limited in their V2G support.

6.1.2 *Advanced analytical models.*

Our current simulations use simplified models, which offer a good starting point but have limitations. For example, for EV charging behavior, we have devised a linearized model in Section 5.1.5. There are various more complex alternatives, including the kinetic battery model [22]. Furthermore, the complexities of the AVP-enabled garage could be considered to get a more accurate assessment of the impact of such an autonomous parking system. For such extensions, co-simulation is an ideal paradigm. It allows for easily coupling independent subsystem simulations, with only minimal changes to the rest of the simulation.

6.1.3 *Power grid integration.*

The AVC VPP interacts with the power grid by measuring its frequency and reacting to deviations with a change in net injection: negative during charging, potentially positive when discharging. To properly capture the complex interactions between grid and garage, we need to integrate a grid simulator. Steady-state simulators, such as MATPOWER [62], are not appropriate since they do not encompass transient disturbances. Here, software such as PowerFactory is appropriate. Farrokhseresht *et al.* [12] have demonstrated integrating transient analysis in a Mosaik co-simulation.

6.1.4 *Electricity market integration.*

The operator of the VPP would interact in the market for balancing services, placing bids to provide FCR to a TSO. A realistic simulation of this market is critical to analyze the financial aspects of the concept. It is also necessary to devise a profitable yet competitive pricing strategy. Among others, Cutler *et al.* [8] demonstrate the integration of transactive market mechanism with co-simulation. Furthermore, the HELICS co-simulation framework [39] offers advanced support for integrating market simulators, alongside transmission, distribution, and communication simulators.

6.2 Prequalification assessment

For the AVC VPP to provide FCR services, it must take part in the market for balancing services. To qualify for said markets, the plant operator must prove effective and reliable balancing capabilities to the network operator. In Germany, the plant must undergo prequalification. The German TSOs lay out the prequalification requirements in a publicly available document [1]. To offer FCR services, a VPP generally must

(1) activate the reserve fully within 30 s [6],
(2) dispatch power automatically for frequency deviations from ±20 mHz [27], and
(3) be available for at least 15 min [27].

For the AVC VPP, it is necessary to deal with the following two challenges: First, the VPP must be able to reliably fulfill the prequalification requirements. Second, the operator must devise a trustworthy and efficient means of assessing fulfillment. Due to the quick-acting nature of EV batteries, low reaction times are likely easy to achieve in practice. However, *proving* the reliable provision of sufficient power is difficult. The characteristics of the plant vary significantly with time, as with other aggregate VPPs [41]. It is thus hard to demonstrate reliable FCR capabilities in a variety of conditions. A model-based engineering approach can help in making sure the AVC VPP complies with prequalification requirements from early in system development [58]. In summary, prequalification is very challenging for non-conventional and time-varying plants, such as the one proposed in this study. The endeavor therefore requires extensive research and possibly regulatory changes.

6.3 Financial feasibility study

To explore the financial viability of an AVC-enabled parking garage operating as a VPP, it is essential to assess both the potential revenue streams and associated costs. Revenue primarily comes from capacity price, which is paid by the network operator compensating the readiness to supply FCR power. The market operates on a competitive bidding system where various plant operators participate. For more information on the German electricity market specifically, see [27]. It is, therefore, necessary to have an accurate model of such markets to optimize a bidding strategy, and to assess revenue realistically. Regarding cost, we assume a parking facility operator already has or is planning to implement AVP and AVP anyways, e.g. for customer convenience and consequently increased rates or competitive advantage. The main sources of cost are:

(1) potentially very costly process of conducting prequalification,
(2) maintaining and updating software and systems for automated operation of the plant,
(3) reliable technology for controlling the charging process, high-cost software and systems development, and
(4) opportunity cost incurred by artificially charging cars less, either in the event of underfrequency, but also potentially having to use a lower baseline to be able to provide negative FCR to be able to react to overfrequency.

To analyze the financial viability fully, it is necessary to consider the customer's perspective. For a similar scenario as ours, Maigha *et al.* [29] say it is a win-win situation for both drivers and parking operators, since the parking fees may be reduced in return for using the vehicle to provide profitable grid services. Furthermore, Wang *et al.* [60] present a framework tailored to ensuring mutually beneficial operation of a VPP using EVs. However, drivers may be worried about battery degradation and object to their car being used not just for flexible charging, but actually discharging (V2G).

6.4 Multi-domain architecture modeling

An AVC VPP brings together power-systems engineering, automotive technology, and building automation. The latter is a key part of enabling AVP, since the building is responsible for guiding vehicles through the garage. The study highlights that functional interoperability between these engineering fields is critical. Managing the inherent complexity of these domains on their own is already a challenge, for which model-based systems engineering is a valuable approach [33]. However, managing the complexities across different domains—each with their own ontologies, frameworks, and tools—is difficult, but necessary to realize such multi-domain concepts as an AVC VPP. In response to the missing work on consolidating domain-specific engineering approaches, [20] introduced an approach aimed at reconciling granularity differences and facilitating cross-domain interoperability. Moving forward, we suggest modeling the AVC

VPP with its respective domain-specific frameworks and tools, and using these models to probe multi-domain interoperability.

6.5 Application of advanced algorithms

At the heart of the AVC VPP are two optimization tasks: scheduling EVs for charging (Section 3.1), and managing the charging power of plugged-in vehicles (Section 3.2). Beyond our simple baseline implementation, there are various promising alternatives. Artificial intelligence has shown great promise in power-systems engineering, from long-term grid planning to safety-critical grid control. For the AVC VPP concept, Reinforcement Learning (RL) seems particularly interesting; with RL, an agent learns optimal behavior by interacting with its environment and being incentivized for benefitial outcomes. In [56], the integration of RL in a smart grid co-simulation was shown. Furthermore, quantum computing has recently been gaining traction in the power-systems domain, promising computational speed-up for various fundamental problems [55]. Such quantum computing–based approaches may be integrated into an existing co-simulation like ours via quantum–classical co-simulation [57]. For example, Koretsky *et al.* [24] use a quantum algorithm for unit commitment, which could be applied to large-scale vehicle scheduling problems.

7 Conclusion

The energy sector's shift towards sustainability and renewable integration introduces challenges and opportunities for the power grid. This transition has sparked interest in smart charging and V2G technologies, positioning EVs as active participants in enhancing grid operations. Our research explores the potential of AVC-enabled airport facilities to act as VPPs, offering FCR services to maintain grid stability. This study focuses on introducing and exploring the concept, developing simulation tools to assess its feasibility, and paving the way for its practical implementation. Our investigations reveal several key insights across four areas:

(1) The AVC VPP concept requires sophisticated optimization tasks akin to unit commitment and economic dispatch: First, scheduling the EVs for charging, and second, controling the power of each plugged-in vehicle within a short amount of time to allow quick response to frequency deviations. We devise formulations for both.
(2) The adaptable and modular nature of co-simulation has proven indispensable, especially as our model's complexity grew. This approach is well-suited for future enhancements and the incorporation of additional subsystems due to its flexibility in handling diverse time paradgims and modeling formalisms.
(3) Preliminary experiments have confirmed the feasibility of the AVC VPP concept: providing significant FCR power while not sacrificing the primary responsibility of ensuring sufficiently charged cars. While the current, simplified model lays a solid foundation for more detailed and realistic studies, it highlights the dependence on suitable data and grounded assumptions.
(4) Bringing the concept to profitable real-world operation demands extensive additional research-and-development effort. Initially, assessing financial viability requires comprehensive simulations and analyses. Then, prequalification for market participation emerges as a pivotal yet challenging milestone.

This work highlights EVs as valuable assets for a sustainable and resilient energy ecosystem, rather than liabilities for the grid. By using AVC-equipped airport parking as VPP, the study demonstrates how we can improve grid stability, facilitate the integration of renewables, and open new revenue opportunities, while ensuring EVs are fully charged upon their owners' return. We take a step towards leveraging the potential of electric mobility for a mutually benefitial energy lanscape where inverter-based renewables do not compromise secure and reliable grid operation.

Acknowledgments

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development and the Christian Doppler Research Association as well as the Federal State of Salzburg is gratefully acknowledged.

Availability of Materials

The code and simulation results associated with this study are available from the corresponding author upon reasonable request. Interested researchers and practitioners are encouraged to contact the corresponding author to gain access to the detailed computational resources and data necessary for replicating or extending the analyses presented in this paper.

References

[1] 50hertz, Amprion, TenneT, and TransnetBW. 2022. Prequalification Process for Balancing Service Providers (FCR, aFRR, mFRR) in Germany. https://www.regelleistung.net/xspproxy/api/StaticFiles/Regelleistung/Infos_f%C3%BCr_Anbieter/Wie_werde_ich_Regelenergieanbieter_Pr%C3%A4qualifikation/Pr%C3%A4qualifikationsbedingungen_FCR_aFRR_mFRR/PQ-Bedingungen-03.06.2022(englisch).pdf Accessed: [24.01.2024].
[2] Allgemeine Deutsche Automobil-Club e. V. (ADAC). 2024. Autokatalog: Automarken & Modelle. https://www.adac.de/rund-ums-fahrzeug/autokatalog/marken-modelle Accessed: [20.01.2024].
[3] Samiur Arif, Stephan Olariu, Jin Wang, Gongjun Yan, Weiming Yang, and Ismail Khalil. 2012. Datacenter at the Airport: Reasoning about Time-Dependent Parking Lot Occupancy. *IEEE Transactions on Parallel and Distributed Systems* 23, 11 (2012), 2067–2080. https://doi.org/10.1109/TPDS.2012.47
[4] CARIAD SE. 2024. Driverless navigation to charge spots – thanks to Bosch and CARIAD. cariad.technology/de/en/news/stories/driverless-navigation-to-charge-spots-thanks-to-bosch-and-cariad.html. [Accessed Jan 27, 2024].
[5] Antonio J. Conejo and Luis Baringo. 2018. *Unit Commitment and Economic Dispatch.* Springer International Publishing, Cham, Chapter 7, 197–232. https://doi.org/10.1007/978-3-319-69407-8_7
[6] Consentec. 2014. Description of load-frequency control concept and market for control reserves. Commissioned by German TSOs.
[7] Constance Crozier, Thomas Morstyn, and Malcolm McCulloch. 2020. The opportunity for smart charging to mitigate the impact of electric vehicles on transmission and distribution systems. *Applied Energy* 268 (2020), 114973. https://doi.org/10.1016/j.apenergy.2020.114973
[8] Dylan Cutler, Ted Kwasnik, Sivasathya Balamurugan, Tarek Elgindy, Siddharth Swaminathan, Jeff Maguire, and Dane Christensen. 2021. Co-simulation of transactive energy markets: A framework for market testing and evaluation. *International Journal of Electrical Power & Energy Systems* 128 (2021), 106664. https://doi.org/10.1016/j.ijepes.2020.106664

[9] Judith S. Dahmann. 1997. High Level Architecture for simulation. In *Proceedings of the 1st International Workshop on Distributed Interactive Simulation and Real Time Applications*. IEEE, Eliat, 9–14. https://doi.org/10.1109/IDSRTA.1997.568652
[10] Emissions-EUETS. 2023. Frequency Containment Reserve (FCR). https://emissions-euets.com/internal-electricity-market-glossary/793-frequency-containment-reserve. Accessed: [10.02.2024].
[11] H. Farhangi. 2010. The path of the smart grid. *IEEE Power and Energy Magazine* 8, 1 (2010), 18–28. https://doi.org/10.1109/MPE.2009.934876
[12] Nakisa Farrokhseresht, Arjen A. van der Meer, José Rueda Torres, and Mart A. M. M. van der Meijden. 2021. MOSAIK and FMI-Based Co-Simulation Applied to Transient Stability Analysis of Grid-Forming Converter Modulated Wind Power Plants. *Applied Sciences* 11, 5 (2021), 2410. https://doi.org/10.3390/app11052410
[13] Flughafen Stuttgart GmbH. 2023. Statistischer Jahresbericht 2022. https://www.flughafen-stuttgart.de/media/309310/230503_statistischer-jahresbericht-2022-flughafen-stuttgart.pdf Accessed: [20.01.2024].
[14] James Francfort. 2014. *2014 DOE Vehicle Technologies Office Review - EV Project Data & Analytic Results*. Technical Report. Idaho National Laboratory. https://www.energy.gov/eere/vehicles/articles/vehicle-technologies-office-merit-review-2014-ev-project-data-analytic Accessed: [26.01.2024].
[15] Fraunhofer Institute for Solar Energy Systems ISE. 2023. Messung der Netzspannung und -frequenz am 01.07.2022 und potenzielle Bereitstellung von Reserveleistung durch NBWR. https://energy-charts.info/charts/frequency/chart.htm Accessed: [04.02.2024].
[16] J. García-Villalobos, I. Zamora, J.I. San Martín, F.J. Asensio, and V. Aperribay. 2014. Plug-in electric vehicles in electric distribution networks: A review of smart charging approaches. *Renewable and Sustainable Energy Reviews* 38 (2014), 717–731. https://doi.org/10.1016/j.rser.2014.07.040
[17] Rishabh Ghotge, Ad van Wijk, and Zofia Lukszo. 2021. Off-grid solar charging of electric vehicles at long-term parking locations. *Energy* 227 (2021), 120356. https://doi.org/10.1016/j.energy.2021.120356
[18] Cláudio Gomes, Casper Thule, David Broman, Peter Gorm Larsen, and Hans Vangheluwe. 2017. Co-simulation: State of the art. https://arxiv.org/abs/1702.00686
[19] Cláudio Gomes, Casper Thule, David Broman, Peter Gorm Larsen, and Hans Vangheluwe. 2018. Co-Simulation: A Survey. *Comput. Surveys* 51, 3, Article 49 (2018), 33 pages. https://doi.org/10.1145/3179993
[20] Jounes-Alexander Gross, Katharina Polanec, Dominik Vereno, Christoph Binder, and Christian Neureiter. 2024. Addressing Cross-Domain Interoperability between Automotive and Smart Grid Architecture Models. In *INCOSE International Symposium*, Vol. 34. INCOSE, Dublin. to appear.
[21] Irene Hafner and Niki Popper. 2021. An Overview of the State of the Art in Co-Simulation and Related Methods. *SNE Simulation Notes Europe* 31, 4 (2021), 185–200.
[22] M.R. Jongerden and B.R. Haverkort. 2009. Which battery model to use? *IET Software* 3 (December 2009), 445–457(12). Issue 6. https://digital-library.theiet.org/content/journals/10.1049/iet-sen.2009.0001
[23] Willett Kempton and Jasna Tomić. 2005. Vehicle-to-grid power fundamentals: Calculating capacity and net revenue. *Journal of Power Sources* 144, 1 (2005), 268–279. https://doi.org/10.1016/j.jpowsour.2004.12.025
[24] Samantha Koretsky, Pranav Gokhale, Jonathan M. Baker, Joshua Viszlai, Honghao Zheng, Niroj Gurung, Ryan Burg, Esa Aleksi Paaso, Amin Khodaei, Rozhin Eskandarpour, and Frederic T. Chong. 2021. Adapting Quantum Approximation Optimization Algorithm (QAOA) for Unit Commitment. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, Broomfield, CO, 181–187. https://doi.org/10.1109/QCE52317.2021.00035
[25] Kraftfahr-Bundesamt. 2023. Statistik: Neuzulassungen von Kraftfahrzeugen und Kraftfahrzeuganhängern nach Herstellern und Handelsnamen, Jahr 2022 (FZ 4). https://www.kba.de/DE/Statistik/Nachrichten/2023/Statistik/fz_4_2022.html Accessed: [16.01.2024].
[26] Sebastian Lehnhoff, Okko Nannen, Sebastian Rohjans, Florian Schlogl, Stefan Dalhues, Lena Robitzky, Ulf Hager, and Christian Rehtanz. 2015. Exchangeability of power flow simulators in smart grid co-simulations with mosaik. In *2015 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*. IEEE, Seattle, WA, 1–6. https://doi.org/10.1109/MSCPES.2015.7115410
[27] Théobald LeLouarn. 2017. Optimization Of A Virtual Power Plant In The German Electricity Market. Master's Thesis.
[28] Mohsen Mahoor, Zohreh S. Hosseini, and Amin Khodaei. 2019. Least-cost operation of a battery swapping station with random customer requests. *Energy* 172 (2019), 913–921. https://doi.org/10.1016/j.energy.2019.02.018
[29] Maigha Maigha and M. L. Crow. 2018. A Transactive Operating Model for Smart Airport Parking Lots. *IEEE Power and Energy Technology Systems Journal* 5, 4 (2018), 157–166. https://doi.org/10.1109/JPETS.2018.2876453
[30] Peter Mihal, Martin Schvarcbacher, Bruno Rossi, and Tomáš Pitner. 2022. Smart grids co-simulations: Survey & research directions. *Sustainable Computing: Informatics and Systems* 35 (2022), 100726. https://doi.org/10.1016/j.suscom.2022.100726
[31] Markus Mirz, Steffen Vogel, Bettina Schäfer, and Antonello Monti. 2018. Distributed real-time co-simulation as a service. In *2018 IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES)*. IEEE, Hamilton, 534–539. https://doi.org/10.1109/IESES.2018.8349934
[32] Natalia Naval and Jose M. Yusta. 2021. Virtual power plant models and electricity markets - A review. *Renewable and Sustainable Energy Reviews* 149 (2021), 111393. https://doi.org/10.1016/j.rser.2021.111393
[33] Christian Neureiter, Christoph Binder, and Goran Lastro. 2020. Review on Domain Specific Systems Engineering. In *2020 IEEE International Symposium on Systems Engineering (ISSE)*. IEEE, Vienna, 1–8. https://doi.org/10.1109/ISSE49799.2020.9272214
[34] Van Hoa Nguyen, Yvon Besanger, Quoc Tuan Tran, and Tung Lam Nguyen. 2017. On Conceptual Structuration and Coupling Methods of Co-Simulation Frameworks in Cyber-Physical Energy System Validation. *Energies* 10, 12 (2017), 1977. https://doi.org/10.3390/en10121977
[35] Annika Ofenloch, Jan Soren Schwarz, Deborah Tolk, Tobias Brandt, Reef Eilers, Rebeca Ramirez, Thomas Raub, and Sebastian Lehnhoff. 2022. MOSAIK 3.0: Combining Time-Stepped and Discrete Event Simulation. In *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*. IEEE, Aachen, 1–5. https://doi.org/10.1109/osmses54027.2022.9769116
[36] N.P. Padhy. 2004. Unit commitment-a bibliographical survey. *IEEE Transactions on Power Systems* 19, 2 (2004), 1196–1205. https://doi.org/10.1109/TPWRS.2003.821611
[37] Peter Palensky, Pierluigi Mancarella, Trevor Hardy, and Milos Cvetkovic. 2024. Cosimulating Integrated Energy Systems With Heterogeneous Digital Twins: Matching a Connected World. *IEEE Power and Energy Magazine* 22, 1 (2024), 52–60. https://doi.org/10.1109/MPE.2023.3324886
[38] Peter Palensky, Arjen A. van der Meer, Claudio David Lopez, Arun Joseph, and Kaikai Pan. 2017. Cosimulation of Intelligent Power Systems: Fundamentals, Software Architecture, Numerics, and Coupling. *IEEE Industrial Electronics Magazine* 11, 1 (2017), 34–50. https://doi.org/10.1109/MIE.2016.2639825
[39] Bryan Palmintier, Dheepak Krishnamurthy, Philip Top, Steve Smith, Jeff Daily, and Jason Fuller. 2017. Design of the HELICS high-performance transmission-distribution-communication-market co-simulation framework. In *2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*. IEEE, Pittsburgh, 1–6. https://doi.org/10.1109/MSCPES.2017.8064542
[40] Ivan Pavić, Hrvoje Pandžić, and Tomislav Capuder. 2020. Electric Vehicles as Frequency Containment Reserve Providers. In *2020 6th IEEE International Energy Conference (ENERGYCon)*. IEEE, Gammarth, 911–917. https://doi.org/10.1109/ENERGYCon48941.2020.9236585
[41] Danny Pudjianto, Charlotte Ramsay, and Goran Strbac. 2007. Virtual power plant and system integration of distributed energy resources. *IET Renewable Power Generation* 1 (2007), 10–16.
[42] Kasim Rehman, Orthodoxos Kipouridis, Stamatis Karnouskos, Oliver Frendo, Helge Dickel, Jonas Lipps, and Nemrude Verzano. 2019. A Cloud-based Development Environment using HLA and Kubernetes for the Co-simulation of a Corporate Electric Vehicle Fleet. In *2019 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, Paris, 47–54. https://doi.org/10.1109/SII.2019.8700423
[43] Robert Bosch GmbH. 2023. Automated valet parking. www.bosch-mobility.com/en/solutions/parking/automated-valet-parking/. [Accessed May 25, 2023].
[44] Lindsey Roeschke. 2022. Morning Consult: The State of Travel & Hospitality Report: Q1 2022. https://pro.morningconsult.com/analyst-reports/state-of-travel-hospitality-q1-2022 Accessed: [21.01.2024].
[45] S. Rohjans, S. Lehnhoff, S. Schütte, S. Scherfke, and S. Hussain. 2013. mosaik - A modular platform for the evaluation of agent-based Smart Grid control. In *IEEE PES ISGT Europe 2013*. IEEE, Lyngby, 1–5. https://doi.org/10.1109/ISGTEurope.2013.6695486
[46] H. Saboori, M. Mohammadi, and R. Taghe. 2011. Virtual Power Plant (VPP), Definition, Concept, Components and Types. In *2011 Asia-Pacific Power and Energy Engineering Conference*. IEEE, Wuhan, 1–4. https://doi.org/10.1109/APPEEC.2011.5749026
[47] Florian Schloegl, Sebastian Rohjans, Sebastian Lehnhoff, Jorge Velasquez, Cornelius Steinbrink, and Peter Palensky. 2015. Towards a classification scheme for co-simulation approaches in energy systems. In *International Symposium on Smart Electric Distribution Systems and Technologies*. IEEE, Vienna, 516–521. https://doi.org/10.1109/SEDST.2015.7315262
[48] G. Schweiger, C. Gomes, G. Engel, I. Hafner, J. Schoeggl, A. Posch, and T. Nouidui. 2019. An empirical survey on co-simulation: Promising standards, challenges and research needs. *Simulation Modelling Practice and Theory* 95 (2019), 148–163. https://doi.org/10.1016/j.simpat.2019.05.001
[49] Steffen Schütte, Stefan Scherfke, and Martin Tröschel. 2011. Mosaik: A framework for modular simulation of active components in Smart Grids. In *2011 IEEE First International Workshop on Smart Grid Modeling and Simulation (SGMS)*. IEEE, Brussels, 55–60. https://doi.org/10.1109/SGMS.2011.6089027

[50] Smart Grid Coordination Group. 2012. *Smart Grid Reference Architecture.* Technical Report. CEN/CENELEC/ETSI, Brussels, Belgium.

[51] Cornelius Steinbrink, Marita Blank-Babazadeh, André El-Ama, Stefanie Holly, Bengt Lüers, Marvin Nebel-Wenner, Rebeca Ramírez Acosta, Thomas Raub, Jan Schwarz, Sanja Stark, Astrid Nieße, and Sebastian Lehnhoff. 2019. CPES Testing with MOSAIK: Co-Simulation Planning, Execution and Analysis. *Applied Sciences* 9, 5 (2019), 923. https://doi.org/10.3390/app9050923

[52] C. Steinbrink, S. Lehnhoff, S. Rohjans, T. I. Strasser, E. Widl, C. Moyo, G. Lauss, F. Lehfuss, M. Faschang, P. Palensky, A. A. van der Meer, K. Heussen, O. Gehrke, E. Guillo-Sansano, M. H. Syed, A. Emhemed, R. Brandl, V. H. Nguyen, A. Khavari, Q. T. Tran, P. Kotsampopoulos, N. Hatziargyriou, N. Akroud, E. Rikos, and M. Z. Degefa. 2017. Simulation-Based Validation of Smart Grids – Status Quo and Future Research Trends. In *Industrial Applications of Holonic and Multi-Agent Systems*, Vladimír Mařík, Wolfgang Wahlster, Thomas Strasser, and Petr Kadera (Eds.). Springer International Publishing, Cham, 171–185.

[53] Cornelius Steinbrink, Florian Schlögl, Davood Babazadeh, Sebastian Lehnhoff, Sebastian Rohjans, and Anand Narayan. 2018. Future perspectives of co-simulation in the smart grid domain. In *IEEE International Energy Conference (ENERGYCON)*. IEEE, Cyprus, 1–6. https://doi.org/10.1109/ENERGYCON.2018.8398830

[54] Kang Miao Tan, Vigna K. Ramachandaramurthy, and Jia Ying Yong. 2016. Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques. *Renewable and Sustainable Energy Reviews* 53 (2016), 720–732. https://doi.org/10.1016/j.rser.2015.09.012

[55] Md Habib Ullah, Rozhin Eskandarpour, Honghao Zheng, and Amin Khodaei. 2022. Quantum computing for smart grid applications. *IET Generation, Transmission & Distribution* 16, 21 (2022), 4239–4257. https://doi.org/10.1049/gtd2.12602

[56] Dominik Vereno, Jonas Harb, and Christian Neureiter. 2023. *Paving the Way for Reinforcement Learning in Smart Grid Co-simulations.* Springer International Publishing, Cham, 242–257. https://doi.org/10.1007/978-3-031-26236-4_21

[57] Dominik Vereno, Amin Khodaei, Christian Neureiter, and Sebastian Lehnhoff. 2023. Quantum–classical co-simulation for smart grids: a proof-of-concept study on feasibility and obstacles. *Energy Informatics* 6, S1 (Oct. 2023). https://doi.org/10.1186/s42162-023-00292-1

[58] Dominik Vereno, Katharina Polanec, and Christian Neureiter. 2024. Compliance by Design for Cyber-Physical Energy Systems: The Role of Model-Based Systems Engineering in Complying with the EU AI Act. In *Proceedings of the 12th International Conference on Model-Based Software and Systems Engineering - MBSE-AI Integration.* Science and Technology Publications, Rome, 365–370. https://doi.org/10.5220/0012623000003645

[59] Mike Vogt, Frank Marten, and Martin Braun. 2018. A survey and statistical analysis of smart grid co-simulations. *Applied Energy* 222 (2018), 67–78. https://doi.org/10.1016/j.apenergy.2018.03.123

[60] Han Wang, Youwei Jia, Mengge Shi, Chun Sing Lai, and Kang Li. 2023. A Mutually Beneficial Operation Framework for Virtual Power Plants and Electric Vehicle Charging Stations. *IEEE Transactions on Smart Grid* 14, 6 (2023), 4634–4648. https://doi.org/10.1109/TSG.2023.3273856

[61] Xiyun Yang and Yanfeng Zhang. 2021. A comprehensive review on electric vehicles integrated in virtual power plants. *Sustainable Energy Technologies and Assessments* 48 (2021), 101678. https://doi.org/10.1016/j.seta.2021.101678

[62] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. 2011. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Transactions on Power Systems* 26, 1 (2011), 12–19.

[63] Mustafa İnci, Murat Mustafa Savrun, and Özgür Çelik. 2022. Integrating electric vehicles as virtual power plants: A comprehensive review on vehicle-to-grid (V2G) concepts, interface topologies, marketing and future prospects. *Journal of Energy Storage* 55 (2022), 105579. https://doi.org/10.1016/j.est.2022.105579

How Satisfactory Can Deep Reinforcement Learning Methods Simulate Electricity Market Dynamics? Benchmarking via Bi-level Optimization

NICK HARDER*, University of Freiburg, Germany
LESIA MITRIDATI, Technical University of Denmark, Denmark
FARZANEH POURAHMADI, Technical University of Denmark, Denmark
ANKE WEIDLICH, University of Freiburg, Germany
JALAL KAZEMPOUR, Technical University of Denmark, Denmark

Various factors make electricity markets increasingly complex, making their analysis challenging. This complexity demands advanced analytical tools to manage and understand market dynamics. This paper explores the application of deep reinforcement learning (DRL) and bi-level optimization models to analyze and simulate electricity markets. We introduce a bi-level optimization framework incorporating realistic market constraints, such as non-convex operational characteristics and binary decision variables, to establish an upper-bound benchmark for evaluating the performance of DRL algorithms. The results confirm that DRL methods do not reach the theoretical upper bounds set by the bi-level models, thereby confirming the effectiveness of the proposed model in providing a clear performance target for DRL. This benchmarking approach demonstrates DRL's current capabilities and limitations in complex market environments but also aids in developing more effective DRL strategies by providing clear, quantifiable targets for improvement. The proposed method can also identify the information gap cost since DRL methods operate under more realistic conditions than optimization techniques, given that they don't need to assume complete knowledge about the system. This study thus provides a foundation for future research to enhance market understanding and possibly its efficiency in the face of increasing complexity in the electricity market. Our methodology's effectiveness is further validated through a large-scale case study involving 150 power plants, demonstrating its scalability and applicability to real-world scenarios.

CCS Concepts: • **Computing methodologies** → **Model verification and validation**; *Simulation evaluation*; • **Applied computing** → *Decision analysis.*

Additional Key Words and Phrases: agent-based modeling, electricity markets modeling, bi-level optimization, reinforcement learning

1 INTRODUCTION

The electricity market landscape is undergoing a profound transformation, driven by integrating new energy supply and consumption technologies, the evolution of regulatory frameworks, and geopolitical risks [29, 32]. This shift is part of the broader evolution of the energy system, characterized by the adoption of new technologies, increased variability, the emergence of new consumption patterns, and the expansion of cross-country interconnections and multi-energy sector coupling initiatives. These developments add layers of complexity to the dynamics of electricity markets [38]. Additionally, integrating emerging technologies and new market participants necessitates regulatory adjustments [8]. As a result, electricity markets must continuously adapt, becoming more short-term focused, increasingly integrated, and inclusive of innovative products that enhance flexibility and adaptability in bidding strategies. This ongoing evolution underscores the importance of developing tools and simulation models to explore and comprehend the intricate interactions within markets and to anticipate potential adverse outcomes, such as market power abuse.

In response to these challenges, agent-based models and DRL have emerged as powerful tools for understanding and navigating the complexities of electricity markets [17]. Agent-based models can simulate individual market actors with distinct strategies [42], while DRL provides a framework for developing individual strategies through the interactions among multiple market participants [18]. These approaches can enhance our understanding of market dynamics and offer insights into potential market power being exercised by strategic players, which help inform the development of more effective regulatory and operational strategies. However, effectively applying DRL in the analysis of electricity markets is not straightforward and requires designing models capable of making meaningful contributions and producing good results. Among others, one reason is that DRL algorithms are susceptible to the settings of their hyperparameters and the choices made in their architectural design, which can significantly influence their performance and behavior. This sensitivity means that small changes in hyperparameters or architecture can lead to vastly different outcomes, potentially affecting the stability, convergence speed, and overall effectiveness of the models [24]. Given the complexity of real-world electricity markets, where numerous factors make it challenging to discern the cause-and-effect relationships between bidding strategies and market outcomes, there is a need for robust benchmarking methods capable of effectively evaluating the performance of DRL agents within these complex environments.

Current research often focuses on comparing different DRL algorithms with each other, overlooking valuable insights from established optimization models. Equilibrium and bi-level models have traditionally been cornerstones for analyzing market dynamics, offering insights grounded in economic theory into achievable strategic equilibrium states [12]. These models clarify the effects of different market structures and behaviors, providing a perspective often lacking in purely algorithmic approaches.

To address these benchmarking challenges, this study introduces a framework using bi-level optimization models to establish an upper bound for DRL methods' performance. These models provide a theoretically grounded performance ceiling based on economic

Authors' Contact Information: Nick Harder, nick.harder@inatech.uni-freiburg.de, University of Freiburg, Freiburg im Breisgau, Germany; Lesia Mitridati, lemitri@dtu.dk, Technical University of Denmark, Kgs. Lyngby, Denmark; Farzaneh Pourahmadi, farpour@dtu.dk, Technical University of Denmark, Kgs. Lyngby, Denmark; Anke Weidlich, anke.weidlich@inatech.uni-freiburg.de, University of Freiburg, Freiburg im Breisgau, Germany; Jalal Kazempour, jalal@dtu.dk, Technical University of Denmark, Kgs. Lyngby, Denmark.

theory and strategic equilibrium states. Incorporating non-convex operational constraints and binary decision variables offers a realistic benchmark for DRL methods. The bi-level optimization framework is an upper bound because it assumes perfect information and optimal decision-making by market participants—conditions under which the highest possible performance can be achieved. In contrast, DRL methods operate under more realistic conditions with limited information and adaptive learning. This framework allows for effective evaluation of DRL methods against established upper bounds, highlighting areas for improvement and guiding the development of more sophisticated strategies.

Our primary contributions in this paper are as follows:

(1) **Enhanced Bi-Level Optimization Model**: We extend existing bi-level optimization approaches by incorporating non-convex constraints and binary variables, creating a realistic upper bound for DRL performance in complex electricity markets with non-convex operational characteristics.
(2) **Benchmarking Framework for DRL**: We demonstrate the application of our model as a benchmarking tool for evaluating and improving DRL strategies in complex market environments, providing clear, quantifiable targets for performance improvement.
(3) **Practical Implementation and Analysis**: We present a comparative analysis between our approach and DRL methods, along with a complete mathematical formulation and implementation code (available at *GitHub repository*[1]), enabling researchers and practitioners to apply, validate, and extend our approach in their work.

To demonstrate the scalability of our approach, we extend our analysis to a larger case study involving 150 power plants with a demand profile representative of the European electricity market. This expansion aims to validate our findings in a more complex, realistic scenario and illustrate the applicability of our methodology to larger-scale market simulations.

The remainder of this paper is organized as follows: Section 2 reviews the state-of-the-art DRL applications for electricity market modeling and existing benchmarking approaches. Section 3 outlines our bi-level optimization model and the rationale for selecting the specific DRL algorithm. Section 4 discusses the results of our experimental comparisons. Finally, Section 5 concludes the paper with a summary of key findings, their implications for theory and practice, and suggestions for future research directions.

2 RELATED WORK

Bi-level optimization problems feature a hierarchical decision-making structure where decisions at the upper level (such as individual profit maximization) directly influence outcomes at the lower level (such as social welfare maximization). These frameworks are essential for modeling the strategic behaviors of market participants who act in their self-interest in electricity markets [21, 31]. However, bi-level optimization models often face challenges with non-convexities and the incorporation of binary variables, which complicate their solution method [3]. The assumption of perfect information about operational parameters and competitors' strategies limits their practical use [43]. A significant limitation exists in many well-established bi-level optimization frameworks, such as those proposed by [20] and [22]. These models often omit binary decision variables in the lower-level problem. As a result, these models generally account for variable costs, maximum output limits, and ramp rates but overlook non-convex cost components and constraints tied to binary commitment decisions. These can include start-up and shut-down costs and minimum stable generation limits. Such complex operating characteristics greatly influence market-clearing outcomes and the strategic decisions of market participants. This omission can lead to sub-optimal bidding strategies and inaccurate market equilibria, particularly in markets with market clearing mechanisms that require solving unit commitment (UC) problems, as seen in the USA and Europe [26].

On the other hand, single-agent DRL focuses on optimizing a single objective over time, adapting autonomously to changing environments through methods like dynamic programming. This approach makes the modeling process more flexible than bi-level optimization, making DRL a popular choice for developing bidding strategies in electricity markets [7, 23]. Recent developments have introduced state-of-the-art DRL techniques to refine bidding strategies across various market setups. Notable examples include the use of proximal policy optimization for sequential market bidding [1], twin delayed deep deterministic policy gradients (TD3) for continuous intra-day market operations [41], and deep deterministic policy gradient algorithms for pool-based markets [46]. These studies benchmark proposed DRL approaches against existing DRL methods or against rule-based systems that usually serve as a lower performance bound. Bi-level optimization is sometimes used to establish an upper-performance benchmark [46], which gives a much better understanding of the performance of the developed DRL method. This benchmarking is essential as it allows researchers to compare the performance of DRL methods against theoretically optimal outcomes, highlighting areas for improvement and confirming the algorithm's robustness. However, its effectiveness diminishes when binary variables are included in the model because these variables introduce non-linearities and discontinuities that complicate the optimization process, often requiring computationally intensive methods and approximations to be able to solve the problem. The need to use approximation techniques to solve bi-level optimization problems with binary variables in the lower level problem often leads to suboptimal solutions, making these models less competitive compared to DRL methods [46]. Consequently, using more effective bi-level optimization approaches such as the one proposed by [44] that can reliably handle binary variables is crucial for establishing the upper-performance bounds in market simulations.

Building on the discussion of bi-level optimization problems, equilibrium-based models also represent the market as a bi-level problem [28]. The main difference between bi-level and equilibrium models is their handling of interdependencies and decision-making. While bi-level models explicitly structure the decision-making hierarchy and interaction between levels, equilibrium models inherently assume that all market participants are simultaneously optimizing their strategies; namely, each market participant takes into account the strategic behavior of its competitors, thus leading to a market equilibrium [37]. This distinction makes equilibrium models

[1] GitHub repository: `https://github.com/INATECH-CIG/bilevel_opt`

essential for analyzing strategic behavior and market dynamics and evaluating the impact of policy interventions on market structures [28, 34]. Despite their comprehensive analytical capabilities, these models' high mathematical and computational demands present significant challenges, particularly in large environments [2].

Similar to applications of bi-level optimization techniques and single-agent DRL algorithms, the challenges of applying equilibrium models to market simulations have sparked a new direction in utilizing multi-agent DRL to approximate solutions of these analytical models. The multi-agent DRL approach aims to leverage the capabilities of DRL algorithms in capturing the strategic interactions and dynamics among multiple agents in a market setting, offering a promising tool set to study and understand the equilibrium outcomes in electricity markets. For instance, [45] employed a multi-agent DRL approach using a deep policy gradient method to model the electricity market with ten strategic agents over 24 simulation time steps. Further extending this, [9] utilized a multi-agent deep deterministic policy gradient algorithm to approximate the equilibrium in a setup involving nine agents, three of whom employed DRL strategies. Their research highlights the capability of DRL agents to exploit market conditions such as grid congestion to maximize profits. Reference [36] utilized a semi-distributed DRL algorithm combined with a graph convolutional neural network to model electricity market participants aware of their location in the network and the overall network structure.

In multi-agent setups, similar to single-agent studies, performance benchmarks typically compare the results to other DRL algorithms, as illustrated in the studies mentioned above [9, 36, 45]. Notably, some studies employed a diagonalization technique after convergence to assess the robustness of the strategies. For example, [45] conducted a bi-level optimization for each agent while keeping the strategies of all others fixed to determine if an agent's strategy deviates from the reached convergence point. If no deviation occurs, the strategies are deemed at Nash equilibrium. Similarly, [9] applied a diagonalization technique starting from scratch, comparing the converged solution to the DRL-derived solution to validate its effectiveness. These approaches underscore the reliance on diagonalization algorithms where bi-level optimization formulation plays a central role. Yet, these benchmarks are confined to market environments and bidding strategies that lack non-convex constraints, including binary variables. When implementing more sophisticated methods, performing similar benchmarking procedures for complex market environments becomes essential. Additionally, it is important to acknowledge that the existence of a unique Nash equilibrium in competitive multi-agent games is not guaranteed [25], and there is a possibility that multi-agent DRL algorithms might converge to different Nash equilibria. This potential variability highlights the need to consider and validate the equilibrium outcomes when employing these DRL techniques in market simulations.

Building on the insights above, this paper contributes to the field by proposing a novel benchmarking framework for DRL methods using a bi-level optimization model. This approach addresses the computational complexities associated with non-convexities and binary variables and enhances the robustness and accuracy of evaluating DRL strategies in complex market environments.

3 METHODOLOGY

This study introduces a bi-level optimization problem formulation designed to estimate an upper-performance boundary for DRL algorithms in market environments with non-convex operating characteristics. The discussion begins with a detailed mathematical description of the problem, presented in Section 3.1. This is followed by a concise overview of the single-agent and multi-agent DRL algorithms employed, detailed in Section 3.2. Finally, Section 3.3 outlines the modeling framework and provides additional details on the implementation.

3.1 Bi-level optimization problem formulation

We model a pool-based, energy-only market with a non-convex clearing mechanism in the present study. In this market, each supplier submits price-quantity bids and detailed technical constraints, including must-run power, maximal capacity, ramping constraints, and start-up and shut-down costs. The market-clearing algorithm follows the framework proposed by [33]. A UC problem is initially addressed using a mixed-integer linear programming formulation to determine the optimal commitment (on/off) and dispatch. Subsequently, a continuous relaxed version of the problem is solved to establish the clearing prices. Although the market-clearing mechanism employed does not mimic those used in actual markets, it is a generalized representation of prevalent algorithms and can be tailored to specific scenarios. For this study, no side payments are considered.

We build upon the work by [44], which introduces a modeling approach incorporating non-convexities into bi-level optimization. Specifically, their methodology is based on the relaxation and primal-dual reformulation of the original, non-convex lower-level problem, coupled with the penalization of the associated duality gap. Their case studies illustrate the model's capability to approximate the actual market-clearing solutions of the UC algorithm, enabling strategic producers to make more profitable bidding decisions than traditional bi-level optimization methods, ignoring UC constraints. However, a notable limitation of [44] is that the bi-level optimization, when reformulated as a single-level problem, only approximates the market-clearing price anticipated by the strategic producer. This approximation arises due to the inherent limitations in minimizing the duality gap.

As our primary goal was to use the bi-level optimization method to establish an upper-performance benchmark for DRL agents, the limitation mentioned above became apparent when, under certain conditions, the DRL agents outperformed the bi-level optimization formulation, as demonstrated in Section 4. This observation challenges the initial purpose of our study. It is important to understand the source of this behavior to address this. In the original formulation, the upper-level problem is defined as:

$$\max_{k_{i,t}} \sum_t \left((\lambda_t - \Lambda_i^G) g_{i,t} - F_i u_{i,t} - c_{i,t}^U - c_{i,t}^D \right), \tag{1}$$

where $k_{i,t}$ represents the strategic bidding price variable for producer i at time t, λ_t is the market-clearing price (€/MWh), Λ_i^G denotes the marginal cost (€/MWh), $g_{i,t}$ is the power output (MW), $u_{i,t}$ is the binary unit commitment (UC) status of producer i, F_i is the

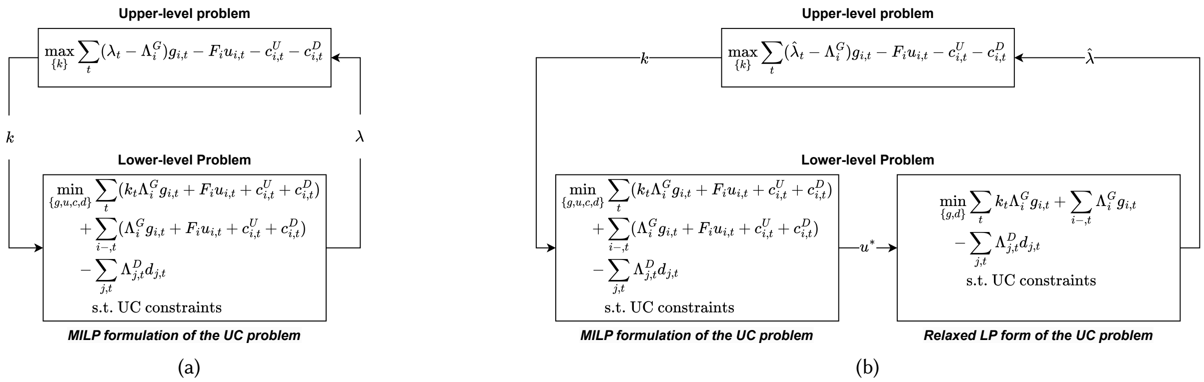


Fig. 1. Illustration of the proposed bi-level problem formulation as described by [44] (a) and proposed in the current work (b). The UC constraints are defined according to [44]. Here, Λ^D and $d_{j,t}$ represents marginal benefit (€/MWh) and power input (MW) of demand j at t, respectively. For the detailed original problem formulation, please refer to Appendix A.

no-load cost (€/h), and $c_{i,t}^U$ and $c_{i,t}^D$ are the start-up and shut-down costs (€), respectively.

The strategic agent determines the bid price as $B_{i,t} = k_{i,t}\Lambda_G$ using the variable $k \in [1, 2]$ to achieve a market-clearing price λ_t that maximizes its profits. This market-clearing price is derived from the lower-level problem by minimizing the duality gap between the primal and dual formulations of the original UC problem. While this approximation is generally accurate, it can still produce a market-clearing price that differs from the "actual" market-clearing, obtained when the lower-level market-clearing problem is solved independently. This discrepancy can lead the strategic agent to make sub-optimal decisions, resulting in lower profits than the DRL agent's.

To overcome this issue, we introduce the pricing mechanism, which uses the continuous version of the UC problem with status binaries u fixed to their optimal values u^*. To implement these adjustments, several steps are required. First, the Karush-Kuhn-Tucker (KKT) conditions of the original UC-constrained market-clearing are derived while treating the binary status variables u as constants. In this formulation, the variable u is shared with the original formulation, while a new market-clearing variable $\hat{\lambda}$ is introduced. This new market-clearing price represents the accurate market-clearing price, where no UC constraints are enforced. Second, the market-clearing price variable in the upper-level objective function is exchanged for the newly derived value $\hat{\lambda}$ to enable the strategic bidder to consider the actual market outcome when optimizing the bids.

To address the discrepancy between the approximated and actual market-clearing prices, we introduce a refined pricing mechanism that employs a continuous version of the UC problem, with the binary status variables u fixed at their optimal values, denoted as u^*. Fig. 1 illustrates the differences between the original and the proposed problem formulations. Fig. 1b presents our proposed approach, extending the lower-level problem with a relaxed LP model with constant binary status variables at u^*. In this approach, the lower-level problem maximizes social welfare and determines the unit commitment variables u. As a byproduct of this optimization, the market clearing price λ is derived. However, instead of directly using λ in the upper-level optimization, we solve a second relaxed problem using the optimal values of u to find $\hat{\lambda}$, which is then employed in the upper-level optimization. By decoupling this process, $\hat{\lambda}$ influences the strategic agent's bidding decision variables k and can be seen as an extension of the original UC problem without directly affecting the unit commitment variables u. The implementation of this refined mechanism involves the following steps:

(1) **Derivation of KKT Conditions:** We begin by deriving the KKT conditions for the original UC market-clearing problem, treating the binary status variables u as fixed constants. In this adapted formulation, the variable u remains aligned with the original model, but we introduce a new market-clearing price variable, $\hat{\lambda}$. This $\hat{\lambda}$ represents the accurate market-clearing price, assuming that the UC constraints do not influence the pricing.
(2) **Derivation of KKT Conditions:** We begin by deriving the KKT conditions for the original UC market-clearing problem (Appendix B), treating the binary status variables u as fixed constants. In this adapted formulation, the variable u remains aligned with the original model, but we introduce a new market-clearing price variable, $\hat{\lambda}$. This $\hat{\lambda}$ represents the actual market-clearing price, assuming that the UC constraints do not influence the pricing.
(3) **Update of Upper-Level Objective Function:** In the upper-level problem as specified in Eq. 1, we substitute the market-clearing price variable λ with the newly derived $\hat{\lambda}$. This modification allows the strategic bidder to base their bidding decisions on the actual market outcomes, optimizing their bids with a more accurate representation of market dynamics.
(4) **Update of Upper-Level Objective Function:** In the upper-level problem as specified in Eq. 1, we substitute the market-clearing price variable λ with the newly derived $\hat{\lambda}$. This modification allows the strategic bidder to base their bidding decisions on the actual market outcomes, optimizing their bids with a more accurate representation of market

dynamics. For the complete mathematical formulation of the final optimization problem, please refer to Appendix C.

In our proposed approach, as illustrated in Fig. 1b, we solve three optimization problems sequentially, but this remains a bi-level optimization problem rather than a three-level optimization. The structure consists of an upper-level optimization where the strategic agent determines the optimal bidding prices, followed by a lower-level optimization that models the market-clearing process as a UC problem with non-convex constraints and binary decision variables. The second optimization problem within the lower level represents the same market clearing as the first one but in its relaxed form, ensuring more accurate market-clearing prices by fixing the binary variables and solving a continuous LP problem. This relaxed LP problem is embedded into the optimization using its KKT conditions. It effectively extends the first market-clearing optimization problem where additional constraints define the refined market-clearing price, $\hat{\lambda}$. Thus, this refinement step does not introduce a new hierarchical level but enhances the accuracy of the lower-level solution within the existing bi-level framework.

This revised approach enhances the model's realism by better aligning the strategic bidding decisions with real-world market conditions. This adjustment ensures that strategic bidders respond to a market-clearing price that more accurately reflects the actual economic environment, mitigating the distortions introduced by the approximations in the original bi-level model. Ultimately, this leads to a more robust comparison between the bi-level optimization models and DRL agents, potentially contributing to more effective and realistic simulations of electricity markets.

3.2 DRL algorithm description

Continuing from the previous section and moving into the description of the DRL algorithm used in our comparative analysis, we focus on the TD3 algorithm, introduced by [14]. It is an advanced form of the deep deterministic policy gradient algorithm [27], specifically designed to address issues related to the overestimation bias and the variance in the Q-value estimation, which are prevalent in deep deterministic policy gradient. In the context of electricity markets, TD3's ability to handle continuous and high-dimensional action spaces allows it to effectively model market participants' strategic bidding behavior. Agents can learn to optimize their bids based on various inputs, such as market prices and demand forecasts. Another reason for choosing the TD3 algorithm is the ability to extend it to a multi-agent setup similar to multi-agent deep deterministic policy gradient [30], where a centralized critic is introduced to evaluate the actions of all agents collectively, enhancing the learning process in environments with multiple interacting agents. This setup allows each agent to maintain its policy network (actor) while sharing information through the centralized critic, fostering cooperation or competition as dictated by the market scenario. The reasoning behind the choice of the algorithm, the complete algorithm flow description, the utilized hyper-parameters, and the methodology validation are presented in [17].

The environment for our DRL setup is modeled analogously to the bi-level optimization problem. Market clearing is initially conducted using the UC problem, after which the binary status variables are fixed at their optimal values, and a linear programming problem is solved to determine the market-clearing price. This separation between the environment and the agent in the DRL framework simplifies the integration of market-clearing algorithms.

In terms of the observation space, at time-step t each agent receives a vector $[L_t^{\mathrm{h}}, L_t^{\mathrm{f}}, M_t^{\mathrm{h}}, M_t^{\mathrm{f}}, \Lambda_i]$. Here, L_t^{h} and L_t^{f} represent historical and forecasted residual loads, respectively, while M_t^{h} and M_t^{f} denote historical and forecasted market prices. Additionally, the agent receives data on its current marginal cost, Λ_i. All inputs are normalized to enhance the efficiency of the training process.

In this study, the forecasted market prices are assumed to have perfect foresight, meaning the agent can access the true prices for the bidding horizon. This ensures that the DRL agent has access to the same information as in the bi-level optimization setting. The information about the residual load is also part of the bi-level optimization framework, where the strategic agent has a full overview of all states.

A significant differentiation between the DRL agent and the bi-level optimization problem is that, in the bi-level setting, the strategic agent is assumed to have full knowledge about the actions of other agents as well as a complete understanding of the market clearing mechanism. In contrast, the DRL agent does not possess this level of information, making the DRL agent's representation more realistic. This distinction highlights the practical advantage of the DRL approach, as it operates under more realistic conditions than the idealized assumptions of bi-level optimization. Moreover, while this study uses perfect foresight for forecasted prices and loads, future research could switch to more realistic forecasts to further enhance the closeness to real-world conditions. However, incorporating realistic forecasts is beyond the scope of this theoretical study.

The actions taken by the agents involve submitting their UC constraints alongside their price bids for each hour of the trading period. Specifically, the agent determines a vector of 24 price values, one for each hour. These price bids are calculated using the same equation as in the case of the bi-level optimization problem as $B_{i,t} = k_{i,t}\Lambda_G$, where $k_t \in [1, 2]$ is a multiplier applied to the generator's marginal cost Λ_i. This action space is designed to align closely with the strategic bidding decisions modeled in the bi-level optimization framework. The choice of this particular action space facilitates benchmarking and ensures that the strategic behavior modeled in the bi-level problem is comparable to that of the DRL agent. The action space can be tailored to best suit specific market conditions and strategic objectives in practical applications. It is worth noting that k values of less than one could also be applicable in some cases, such as for conventional power plants during low residual load to prevent shutdown, but this scenario is not investigated in this study.

The reward function is straightforward, focusing on the pure profit of the generator, defined identically to the objective function in Eq. 1. This simplification avoids complicating the reward structure and ensures a direct and fair comparison between DRL and bi-level optimization. We deliberately avoid enhancing the reward function as detailed in [17] to maintain focus on comparing methodologies rather than optimizing DRL performance. The reward of agent i at time step t is defined as:

$$R_{i,t} = \beta \left((\lambda_t - \Lambda_i^G) g_{i,t} - F_i u_{i,t} - c_{i,t}^U - c_{i,t}^D \right) \tag{2}$$

Here, β is a scaling factor used to normalize the reward. This factor is determined as the product of the maximum power and bid price, ensuring that the rewards are scaled appropriately for the learning process. Due to significant differences in their marginal costs, different scales of rewards among agents can arise. In the context of DRL and our utilized TD3 algorithm, each agent independently maximizes its reward without considering the rewards of other agents. This approach ensures that the variation in reward magnitudes does not lead to any numerical issues or impact individual agents' learning process. This applies to both single-agent and multi-agent setups.

These steps ensure that we model the market environment consistently with the setup described previously, facilitating a clear and direct comparison. The primary goal of this paper is not solely to demonstrate the superiority of DRL but to provide a detailed comparative analysis that aids in selecting and enhancing DRL algorithms. By comparing DRL outcomes with the precise upper bounds from bi-level optimization, researchers can refine their models to better approach these optimal solutions, thereby advancing DRL's application in simulating strategic behaviors in electricity markets.

3.3 Implementation details

For implementing the bi-level optimization model, we utilized the Pyomo framework [6, 19]. As solver, we used Gurobi [15] under academic license. For simulations involving DRL agents, we utilized the ASSUME framework [16]. ASSUME is an open-source tool designed for simulating and analyzing markets using DRL techniques implemented in Python3. The learning algorithm employed is a multi-agent variation of the TD3, which utilizes a centralized critic. The development of learning agents and the DRL algorithm was inspired by and built upon the single-agent DRL algorithm implementations from the Stable-Baselines3 project [35]. The DRL agents are trained until no more changes in their actions are observed or the total reward stays within some margin for an extended number of episodes.

The computational experiments were conducted on a conventional laptop, demonstrating the feasibility and accessibility of these methods for prototyping and agile model development. This practicality is particularly important in scenarios where rapid prototyping and agile model development are needed. The complete mathematical derivation of the optimization problem is given in the Appendix A-D. Additionally, the Python code and input data are available under the following *GitHub repository*[2]. This ensures transparency and reproducibility, facilitating further research and validation by the academic community.

4 RESULTS

This section presents the results of our study in two parts. First, we examine a small-scale case study to illustrate our proposed methodology's fundamental principles and effectiveness. Following this, we extend our analysis to a larger, more realistic scenario to demonstrate our approach's scalability and practical applicability in real-world settings.

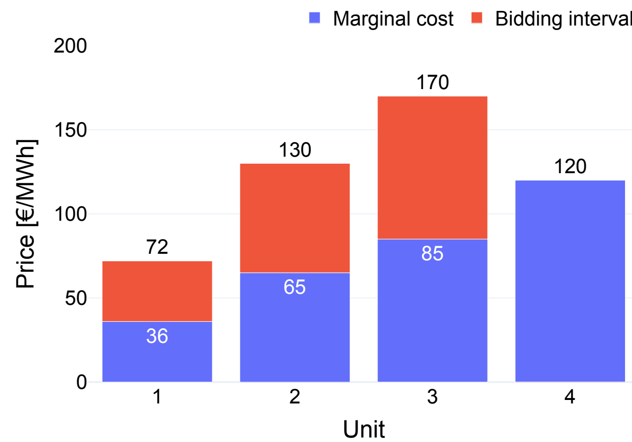


Fig. 2. Input data: Marginal costs and bidding range of power generation units.

4.1 Small scale case study

This case study employs a dataset analogous to that used by [44] but is simplified to include only four generation units. Three of these units act as strategic agents within the study, while the fourth serves as a non-strategic power plant. This non-strategic unit is restricted to bidding only its marginal cost, a constraint implemented to prevent the strategic agents from adopting a trivial solution of consistently bidding at maximum price and instead to compel the agents to develop a more complex bidding strategy.

The marginal costs of the units, depicted in Fig. 2, are deliberately set so that their bidding ranges overlap. This overlap ensures that the units must carefully consider the potential repercussions of overly aggressive bids, such as the risk of being outbid by competitors and selling a reduced volume of energy.

The demand profile is strategically chosen to reflect varying levels of operational complexity. Unit 1, the lowest-cost base-load unit, consistently operates without shutdowns, serving as a straightforward baseline where all methodologies should converge in their outcomes. Unit 2 introduces moderate complexity with occasional possible shutdowns, while Unit 3, facing frequent possible start-ups and shutdowns, represents the highest complexity. This gradation in complexity illustrates the capacity of the discussed methodologies to handle increasingly intricate operational scenarios, thereby providing insights into the practical implications and performance of the methods under study.

We employ three distinct methodologies to analyze the financial outcomes for individual generation units acting as strategic agents. Fig. 3 illustrates the profits accrued by each unit using Method 1 (by [44]), Method 2 (current bi-level formulation), and Method 3 (DRL). For Unit 1, profits remain the same across all methods at k€494, demonstrating that all methods can capture the optimal bidding strategy for this straightforward operational scenario.

[2] GitHub repository: `https://github.com/INATECH-CIG/bilevel_opt`

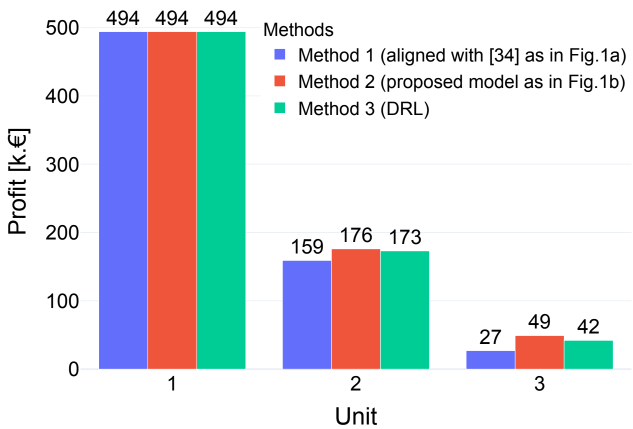


Fig. 3. Profits of individual units in the small-scale case study using the method by [44] (Method 1), the proposed bi-level method as in Fig. 1b (Method 2), and DRL (Method 3). Each unit assumes the role of a strategic agent sequentially, while others bid non-strategically.

The analysis for Unit 2 and Unit 3, however, is particularly insightful for evaluating the performance of DRL against established methods. For Unit 2, DRL achieves a profit of k€173, situating it between Method 1 (k€159) and Method 2 (k€176). This placement of DRL shows its capability to surpass the performance of Method 1 and its near-optimal performance, trailing Method 2 by a mere 2%. The proximity of DRL's performance to Method 2 validates Method 2's role in providing a reliable benchmark that helps fine-tune and assess DRL agents' capabilities within a competitive market framework.

Unit 3, which presents a more significant operational challenge, further underscores the point above. Here, DRL secures a profit of k€42, significantly outperforming Method 1 (k€27) by approximately 55%, yet it falls short of the k€49 profit by Method 2 by about 15%. The substantial gap between the performances of Method 1 and Method 2 and the intermediate yet closer-to-optimal positioning of DRL profits highlights the effectiveness of Method 2 as a benchmark. It provides a clear performance target for DRL, indicating areas where DRL can still improve to reach the upper bounds of market strategy optimization.

It is essential to point out that the performance of the bi-level optimization algorithms discussed herein is influenced by several hyperparameters, primarily the weight assigned to the optimality gap minimization formulation and the size of the discretization set used for binary expansion. Consequently, researchers and practitioners intending to utilize this method should carefully consider these factors, as they are case-specific and require individual selection.

4.2 Large-scale case study with realistic market conditions

We conducted a large-scale case study to evaluate the scalability and real-world applicability of our proposed approach. This study utilized a comprehensive dataset comprising 120 power plants, a realistic demand profile, and variable renewable generation data. Based on the German electricity market in 2021, the dataset was previously employed in [18]. Table 1 provides a detailed overview of the data sources used in this case study.

Table 1. Data sources used for input data

Type	Data description	Reference
Prices	Natural gas	[11]
	CO_2 EU-ETS certificates	[10]
	Other energy carriers	[39]
Plants	Power plant characteristics	[13]
		[40]
		[4]
Net load	VRE generation	[5]
	Inelastic demand	[5]
	Net load forecast	[5]

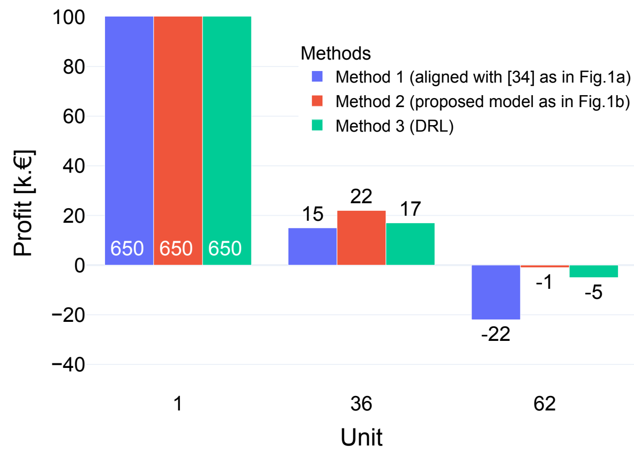


Fig. 4. Profits of representative units in the large-scale case study using the method by [44] (Method 1), the proposed bi-level method as in Fig. 1b (Method 2), and DRL (Method 3). Each unit assumes the role of a strategic agent sequentially, while others bid non-strategically.

We selected three representative power plants from the 150-unit dataset for analysis to provide a detailed comparison similar to our small-scale study. These units were chosen to represent different positions in the typical merit order: a nuclear power plant (generally at the beginning of the merit order), a hard coal power plant (typically in the middle), and a gas-fired unit (usually at the end of the merit order, acting as a peak power plant). This selection allows us to examine our methodology's performance across various generation technologies and operational characteristics. Fig. 4 illustrates the profits accrued by each of these units using Method 1 (by [44]), Method 2 (our proposed bi-level formulation), and Method 3 (DRL). The results mirror our findings from the small-scale study, demonstrating the consistency of our methodology across different scales and technology types.

When comparing the overall performance of DRL agents to the bi-level optimization results in this realistic setting, we observed varying outcomes across the three representative units. All methods for the nuclear unit (Unit 1) achieved the same profit of 650 k€, indicating optimal performance across the board for this base-load plant. For the coal unit (Unit 36), the DRL method (Method 3) achieved a profit of 17 k€, surpassing Method 1 (15 k€) but falling short of Method 2 (22 k€). The results were more complex for the gas-fired peaking unit (Unit 62). Method 2 performed best with a profit of -1 k€, while Method 1 and Method 3 (DRL) resulted in a lower profit of -22 k€ and -5 k€, respectively. The negative profit for Unit 62 is attributable to its role as a peaking unit, which often requires frequent starts and stops. These operational characteristics incur significant start-up costs, which can outweigh the revenue generated during short periods of high demand, resulting in an overall negative profit. In this case, while DRL outperformed Method 1, it did not approach the performance of Method 2, which managed to minimize losses more effectively. The computational time for this large-scale optimization remained manageable, taking approximately 20 minutes on a standard workstation for each unit.

In addition to finding the performance upper bound for DRL agents, the proposed algorithm also offers a method to verify if an equilibrium has been reached in multi-agent DRL setups. By fixing the strategies of all but one agent and optimizing that agent's strategy using our bi-level model, we can compare the optimized performance to that achieved through learning. If the performances (or performance gaps) are similar across all agents, the multi-agent system has converged to an equilibrium.

However, it's worth noting that while this method can be used for equilibrium verification, applying a diagonalization algorithm to a multi-agent setup using the proposed method to compute the equilibrium would be computationally challenging due to the large number of agents and the resulting high number of required iterations. This limitation underscores the value of DRL approaches in handling complex, multi-agent scenarios where traditional optimization methods may become intractable.

This large-scale case study validates our approach's scalability and demonstrates its practical value in developing, refining, and verifying DRL strategies for complex electricity market simulations. The consistency of results between our small-scale and large-scale studies further reinforces the robustness of our methodology across different market sizes and complexities.

5 DISCUSSION & CONCLUSION

The results presented above demonstrate the utility of DRL in navigating non-convex environments and the role of our proposed methodology as a benchmark tool for such environments. The provided method sets a performance ceiling that DRL aspires to reach and offers a comparative framework that enhances our understanding of DRL's relative performance. Having appropriate benchmarks is crucial for advancing DRL applications in real-world market scenarios, showcasing that these agents can achieve very high strategic performance and ensure potential users of these methods of their performance compared to optimal solutions.

By establishing a robust benchmark, our methodology is a valuable reference for researchers and practitioners aiming to develop and refine DRL agents for application in electricity markets. This benchmark can guide the development of more advanced bidding strategies by highlighting the areas where DRL currently underperforms, thereby identifying the potential for enhancements. It can also be used to validate solutions of multi-agent systems. For researchers, this provides guidance for investigating new algorithms or modifications that could close the performance gap identified in our study. For practitioners, the benchmark offers a critical tool for evaluating current DRL implementations' effectiveness and making informed decisions about adopting such technologies for market operations.

When examining the results, one might question the necessity of DRL when the optimization methods that are capable of achieving the ideal solution already exist. The rationale for employing DRL, however, is two-fold.

Firstly, while bi-level optimization problems operate under the assumption of having complete knowledge of all market participants' bids, DRL functions under more realistic conditions where such comprehensive data is not available or entirely accurate. This makes DRL a more realistic tool for simulating real-life market dynamics. Additionally, the performance disparity observed between the bi-level optimization model and DRL can be seen as the cost of incomplete information, termed the "information gap cost." This cost quantifies DRL's disadvantages due to its limited access to comprehensive market data, unlike the idealized scenarios presumed by bi-level optimization. Understanding this gap is crucial as it assists researchers and practitioners in developing mechanisms to manage better and mitigate the challenges posed by incomplete data.

Secondly, the mathematical complexity of the bi-level optimization method is significant, making it impractical for large-scale simulations. It is designed to be utilized during the prototyping phase to refine algorithms and understand the criticality of various input data. This complexity also limits its applicability in multi-agent setups, making diagonalization methods rather challenging, thus highlighting the importance of developing and employing more scalable methods like DRL for real-world applications.

As environments become increasingly complex, DRL methods may potentially incorporate more sophisticated models and adapt to scenarios that our current bi-level optimization framework cannot capture. This flexibility is indeed one of the key strengths of DRL, and researchers should be aware that in highly complex or rapidly changing environments, the upper bound established by bi-level optimization may not always hold its validity. The comparative framework we've presented enhances our understanding of DRL's relative performance in our studied scenarios. However, it's important to note that this framework should be seen as a starting point rather than a definitive measure. As market complexities increase, there may be scenarios where DRL outperforms the bi-level optimization model, particularly where the assumptions underlying the bi-level model no longer hold. This observation underscores the complementary nature of these approaches. While bi-level optimization provides valuable insights and benchmarks under certain conditions, DRL offers adaptability and the potential to handle unforeseen complexities. The choice between these methods – or their combined use – should be guided by the specific characteristics and requirements of the problem. One area of future research is creating hybrid models that merge predictive analytics with real-time adaptive learning to enhance the effectiveness of their methodologies, whether using DRL or other approaches. These hybrid models could leverage the strengths of both methods, using bi-level models to set strategic benchmarks and DRL to provide adaptive strategies that respond to real-time market changes.

In conclusion, this study underscores the importance of setting performance baselines and establishing upper-performance bounds using methods like those proposed here, particularly when applying DRL in complex market environments. This approach provides a more comprehensive understanding and assurance of DRL capabilities, fostering confidence in its broader adoption and trust in real-world applications. Nevertheless, viewing it as part of an evolving toolkit for understanding and optimizing electricity market strategies is crucial. The dynamic nature of these markets demands ongoing innovation in both optimization and learning approaches, with each method informing and complementing the other.

ACKNOWLEDGMENTS

This work was conducted in the context of the project "ASSUME: Agent-Based Electricity Markets Simulation Toolbox" funded by the German Federal Ministry for Economic Affairs and Energy under grant number BMWK 03EI1052A. It was also supported by the Scientific Society in Freiburg im Breisgau and the German Academic Exchange Service, which partially funded the research stay at the Technical University of Denmark.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

While preparing this work, the authors used tools based on Large Language Models to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

REFERENCES

[1] Muhammad Anwar, Changlong Wang, Frits De Nijs, and Hao Wang. 2022. Proximal policy optimization based reinforcement learning for joint bidding in energy and frequency regulation markets. In *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 1–5.

[2] Christoph Böhringer and Thomas F Rutherford. 2006. Combining top-down and bottom-up in energy policy analysis: a decomposition approach. *ZEW-Centre for European Economic Research Discussion Paper* 06-007 (2006).

[3] Stephen P Boyd and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[4] Bundesnetzagentur. 2021. BNetzA List of Power Plants.

[5] Bundesnetzagentur. 2021. SMARD - Download Market Data. https://www.smard.de/en/downloadcenter/download-market-data#!?downloadAttributes=%7B%22selectedCategory%22:false,%22selectedSubCategory%22:false,%22selectedRegion%22:false,%22from%22:1651356000000,%22to%22:1652306399999,%22selectedFileType%22:false%7Dhttps://www.smard.de/en/downloadcenter/download-market-data

[6] Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. 2021. *Pyomo–optimization modeling in python* (third ed.). Vol. 67. Springer Science & Business Media.

[7] Di Cao, Weihao Hu, Junbo Zhao, Guozhou Zhang, Bin Zhang, Zhou Liu, Zhe Chen, and Frede Blaabjerg. 2020. Reinforcement learning and its applications in modern power and energy systems: a review. *Journal of Modern Power Systems and Clean Energy* 8, 6 (2020), 1029–1042.

[8] Peter Cramton. 2017. Electricity market design. *Oxford Review of Economic Policy* 33, 4 (2017), 589–612.

[9] Yan Du, Fangxing Li, Helia Zandi, and Yaosuo Xue. 2021. Approximating Nash equilibrium in day-ahead electricity market bidding with multi-agent deep reinforcement learning. *Journal of Modern Power Systems and Clean Energy* 9, 3 (2021), 534–544.

[10] EEX. 2019. Market Data - Environmental Markets - Auction Market. https://www.eex.com/en/market-data/environmental-markets

[11] EEX - Market Data. 2019. EEX Group DataSource - Power - Natural Gas. https://www.eex.com/en/market-data

[12] Andreas Ehrenmann. 2004. *Equilibrium problems with equilibrium constraints and their application to electricity markets.* Ph. D. Dissertation. University of Cambridge.

[13] S&P Global Market Intelligence (Firm). 2017. World Electric Power Plants Database, March 2017. https://doi.org/10.7910/DVN/OKEZ8A

[14] Scott Fujimoto, Herke van Hoof, and David Meger. 2018. Addressing function approximation error in actor-critic methods. *35th International Conference on Machine Learning, ICML 2018* 4 (2018), 2587–2601. arXiv:1802.09477

[15] Gurobi Optimization, LLC. 2023. Gurobi Optimizer Reference Manual. https://www.gurobi.com

[16] Nick Harder, Florian Maurer, Kim K. Miskiw, Johanna Adams, and Manish Khanra. 2024. *assume-framework/assume: v0.3.7*. https://doi.org/10.5281/zenodo.10848203

[17] Nick Harder, Ramiz Qussous, and Anke Weidlich. 2023. Fit for purpose: modeling wholesale electricity markets realistically with multi-agent deep reinforcement learning. *Energy and AI* 14 (2023), 100295.

[18] Nick Harder, Anke Weidlich, and Philipp Staudt. 2023. Finding individual strategies for storage units in electricity market models using deep reinforcement learning. *Energy Informatics* 6, Suppl 1 (2023), 41.

[19] William E Hart, Jean-Paul Watson, and David L Woodruff. 2011. Pyomo: modeling and solving mathematical programs in Python. *Mathematical Programming Computation* 3, 3 (2011), 219–260.

[20] Benjamin F Hobbs, Carolyn B Metzler, and J-S Pang. 2000. Strategic gaming analysis for electric power systems: An MPEC approach. *IEEE Transactions on Power Systems* 15, 2 (2000), 638–645.

[21] S Jalal Kazempour, Antonio J Conejo, and Carlos Ruiz. 2014. Strategic bidding for a large consumer. *IEEE Transactions on Power Systems* 30, 2 (2014), 848–856.

[22] S Jalal Kazempour and Hamidreza Zareipour. 2013. Equilibria in an oligopolistic market with wind power production. *IEEE Transactions on Power Systems* 29, 2 (2013), 686–697.

[23] Alexander J. M. Kell, Stephen McGough, and Matthew Forshaw. 2022. Machine learning applications for electricity market agent-based models: a systematic literature review.

[24] Mariam Kiran and Melis Ozyildirim. 2022. Hyperparameter tuning for deep reinforcement learning applications. *arXiv preprint arXiv:2201.11182* (2022).

[25] Thilo Krause, Elena Vdovina Beck, Rachid Cherkaoui, Alain Germond, Goran Andersson, and Damien Ernst. 2006. A comparison of Nash equilibria analysis and agent-based modelling for power markets. *International Journal of Electrical Power & Energy Systems* 28, 9 (2006), 599–607.

[26] George Liberopoulos and Panagiotis Andrianesis. 2016. Critical review of pricing schemes in markets with non-convex costs. *Operations Research* 64, 1 (2016), 17–31.

[27] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2019. Continuous control with deep reinforcement learning. arXiv:1509.02971 [cs.LG]

[28] Xuan Liu and Antonio J Conejo. 2021. Single-level electricity market equilibrium with offers and bids in energy and price. *IEEE Transactions on Power Systems* 36, 5 (2021), 4185–4193.

[29] Andreas Löschel, Veronika Grimm, Felix Matthes, and Anke Weidlich. 2023. *Stellungnahme zum Strommarktdesign und dessen Weiterentwicklungsmöglichkeiten.* Technical Report. https://www.wirtschaftstheorie.rw.fau.de/files/2023/03/Stellungnahme-zum-Strommarktdesign-und-dessen-Weiterentwicklungsmoeglichkeiten.pdf

[30] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems* 30 (2017).

[31] Ekaterina Moiseeva and Mohammad Reza Hesamzadeh. 2017. Strategic bidding of a hydropower producer under uncertainty: modified Benders approach. *IEEE Transactions on Power Systems* 33, 1 (2017), 861–873.

[32] Council of the EU and the European Council. [n. d.]. Electricity market reform. https://www.consilium.europa.eu/en/policies/electricity-market-reform/. [Accessed 12-05-2024].

[33] Richard P O'Neill, Paul M Sotkiewicz, Benjamin F Hobbs, Michael H Rothkopf, and William R Stewart Jr. 2005. Efficient market-clearing prices in markets with nonconvexities. *European Journal of Operational Research* 164, 1 (2005), 269–285.

[34] Alberto Orgaz, Antonio Bello, and Javier Reneses. 62017. Multi-area electricity market equilibrium model and its application to the European case. In *2017 14th International Conference on the European Energy Market (EEM)*. IEEE, 1–6.

[35] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. 2021. Stable-baselines3: reliable reinforcement learning implementations. *Journal of Machine Learning Research* 22 (2021), 1–8.

[36] Pegah Rokhforoz, Mina Montazeri, and Olga Fink. 2023. Multi-agent reinforcement learning with graph convolutional neural networks for optimal bidding strategies of generation units in electricity markets. *Expert Systems with Applications* 225, February (2023), 120010. arXiv:2208.06242

[37] Carlos Ruiz, Antonio J Conejo, J David Fuller, Steven A Gabriel, and Benjamin F Hobbs. 2014. A tutorial review of complementarity models for decision-making in energy markets. *EURO Journal on Decision Processes* 2 (2014), 91–120.

[38] Ruggero Schleicher-Tappeser. 2012. How renewables will change electricity markets in the next five years. *Energy Policy* 48 (2012), 64–75.

[39] Statistische Bundesamt. 2021. Data on energy price trends - Long-time series from January 2005 to June 2021. https://www.destatis.de/EN/Themes/Economy/Prices/Publications/Downloads-Energy-Price-Trends/energy-price-trends-pdf-5619002.pdf

[40] Umwelt Bundesamt. 2020. Datenbank: Kraftwerke in Deutschland. https://www.umweltbundesamt.de/dokument/datenbank-kraftwerke-in-deutschland

[41] Flin Verdaasdonk, Sumeyra Demir, and Nikolaos G Paterakis. 2022. Intra-day electricity market bidding for storage devices using deep reinforcement learning. In *2022 International Conference on Smart Energy Systems and Technologies (SEST)*. IEEE, 1–6.

[42] Anke Weidlich and Daniel Veit. 2008. A critical survey of agent-based wholesale electricity market models. *Energy Economics* 30, 4 (2008), 1728–1759.

[43] Gaofeng Xiong, Tomonori Hashiyama, and Shigeru Okuma. 2002. An electricity supplier bidding strategy through Q-Learning. In *IEEE Power Engineering Society Summer Meeting,*, Vol. 3. IEEE, 1516–1521.

[44] Yujian Ye, Dimitrios Papadaskalopoulos, Jalal Kazempour, and Goran Strbac. 2020. Incorporating non-convex operating characteristics into bi-Level optimization electricity market models. *IEEE Transactions on Power Systems* 35, 1 (2020), 163–176.

[45] Yujian Ye, Dawei Qiu, Jing Li, and Goran Strbac. 2019. Multi-period and multi-spatial equilibrium analysis in imperfect electricity markets: A novel multi-agent deep reinforcement learning approach. *IEEE Access* 7 (2019), 130515–130529.

[46] Yujian Ye, Dawei Qiu, Mingyang Sun, Dimitrios Papadaskalopoulos, and Goran Strbac. 2020. Deep reinforcement learning for strategic bidding in electricity markets. *IEEE Transactions on Smart Grid* 11, 2 (2020), 1343–1355.

APPENDIX

Indices and Sets

$t \in T$:	Index and set of time periods
$i \in I$:	Index and set of producers
i^- :	Index of producers other than i
$j \in J$:	Index and set of demands
V^P :	Set of decision variables of the original lower level problem
V^D :	Set of decision variables of the dual of the relaxed lower level problem
V :	Set of decision variables of the final single-level optimization problem

Parameters

N_T :	Length of market horizon
Λ_i^G :	Marginal cost of producer i (EUR/MWh)
K_i^U, K_i^D :	Start-up/shut-down cost of producer i (EUR/h)
$G_i^{\min}, G_i^{\max}$:	Minimum/maximum stable generation limit of producer i (MW)
R_i^U, R_i^D :	Ramp-up/down limit of producer i (MW/h)
$G_{i,0}$:	Initial output of producer i (MW)
$U_{i,0}$:	Initial status of producer i
$K_i^{\max}$:	Upper limit of strategic bidding variable of producer i
$\Lambda_{j,t}^D$:	Marginal benefit of demand j at period t (EUR/MWh)
$D_{j,t}^{\max}$:	Maximum demand limit of demand j at period t (MW)

Variables

$g_{i,t}$:	Power output of producer i at period t (MW)
$u_{i,t}$:	Binary UC status of producer i at period t ($u_{i,t} = 1$ if it is on, $u_{i,t} = 0$ if it is off)
$c_{i,t}^U, c_{i,t}^D$:	Start-up/shut-down cost incurred by producer i at period t (£/h)
k_i :	Strategic bidding variable of producer i
$d_{j,t}$:	Power input of demand j at period t (MW)
λ_t :	Market clearing price at period t (EUR/MWh)
$x_{i,t,n}$:	Auxiliary variable to discrete the $g_{i,t}$
$z_{i,t,n}^{\lambda}, z_{i,t,n}^{k}$:	Dummy variables to represent quadratic terms after linearization

A ORIGINAL PROBLEM FORMULATION

The original problem formulation is based on the methodology proposed by [44]. The upper level represents the strategic bidding decisions of market participants, while the lower level models the market-clearing mechanism, including unit commitment constraints.

$$\text{(Upper level)} \quad \max_{\{k_i\}} \sum_t \left(\lambda_t g_{i,t} - \Lambda_i^G g_{i,t} - c_{i,t}^U - c_{i,t}^D\right) \tag{3a}$$

subject to:

$$1 \le k_i \le k_i^{\max} \tag{3b}$$

$$\text{(Lower level)} \quad \min_{V^P} \left[\sum_{i,t} \left(k_{i,t}\Lambda_i^G g_{i,t} + c_{i,t}^U + c_{i,t}^D\right) - \sum_{j,t} \Lambda_{j,t}^D d_{j,t}\right] \tag{3c}$$

where:

$$V^P = \{g_{i,t}, u_{i,t}, c_{i,t}^U, c_{i,t}^D, d_{j,t}\} \tag{3d}$$

subject to:

$$\sum_j d_{j,t} - \sum_i g_{i,t} = 0 : \lambda_t, \quad \forall t \tag{3e}$$

$$u_{i,t}G_i^{\min} \le g_{i,t} \le u_{i,t}G_i^{\max} : \mu_{i,t}^{\min}, \mu_{i,t}^{\max}, \quad \forall i, \forall t \tag{3f}$$

$$0 \le d_{j,t} \le D_{j,t}^{\max} : v_{j,t}^{\min}, v_{j,t}^{\max}, \quad \forall j, \forall t \tag{3g}$$

$$g_{i,t} - G_{i,0} \le R_i^U : \pi_{i,t}^U, \quad \forall i, t = 1 \tag{3h}$$

$$g_{i,t} - g_{i,(t-1)} \le R_i^U : \pi_{i,t}^U, \quad \forall i, \forall t > 1 \tag{3i}$$

$$G_{i,0} - g_{i,t} \le R_i^D : \pi_{i,t}^D, \quad \forall i, t = 1 \tag{3j}$$

$$g_{i,(t-1)} - g_{i,t} \le R_i^D : \pi_{i,t}^D, \quad \forall i, \forall t > 1 \tag{3k}$$

$$c_{i,t}^U \geq 0 : p_{i,t}^U, \quad \forall i, \forall t \tag{3l}$$

$$c_{i,t}^D \geq 0 : p_{i,t}^D, \quad \forall i, \forall t \tag{3m}$$

$$c_{i,t}^U \geq (u_{i,t} - U_{i,0})K_i^U : \sigma_{i,t}^U, \quad \forall i, t = 1 \tag{3n}$$

$$c_{i,t}^U \geq (u_{i,t} - u_{i,(t-1)})K_i^U : \sigma_{i,t}^U, \quad \forall i, \forall t > 1 \tag{3o}$$

$$c_{i,t}^D \geq (U_{i,0} - u_{i,t})K_i^D : \sigma_{i,t}^D, \quad \forall i, t = 1 \tag{3p}$$

$$c_{i,t}^D \geq (u_{i,(t-1)} - u_{i,t})K_i^D : \sigma_{i,t}^D, \quad \forall i, \forall t > 1 \tag{3q}$$

$$u_{i,t} = \{0, 1\}, \quad \forall i, \forall t \tag{3r}$$

B OPTIMIZATION PROBLEM FORMULATION OF THE UC MARKET CLEARING USING KKT CONDITIONS

To derive the KKT conditions for the given optimization problem, we first construct the Lagrangian by introducing Lagrange multipliers for each of the constraints. The KKT conditions provide the necessary conditions for optimality in non-linear programming problems, ensuring that the solution satisfies both the primal and dual feasibility as well as the complementary slackness conditions.

B.1 Lagrangian function of lower level problem with fixed binary dispatch variable and corresponding KKT conditions.

The Lagrangian function for the lower-level problem is constructed by incorporating the primal constraints and their associated Lagrange multipliers. This function encapsulates the objective function and the constraints of the lower-level problem, allowing us to derive the necessary conditions for optimality.

$$L(V^L) = \left[\sum_{i,t} k_{i,t}\lambda_i^G g_{i,t} - \sum_{j,t} \lambda_{j,t}^D d_{j,t}\right] - \sum_t \hat{\lambda}_t \left(\sum_j d_{j,t} - \sum_i g_{i,t}\right) + \sum_{i,t} \hat{\mu}_{i,t}^{\min} \left(u_{i,t}G_i^{\min} - g_{i,t}\right) + \sum_{i,t} \hat{\mu}_{i,t}^{\max} \left(g_{i,t} - u_{i,t}G_i^{\max}\right)$$
$$+ \sum_{j,t} \hat{\upsilon}_{j,t}^{\min} \left(d_{j,t}\right) + \sum_{j,t} \hat{\upsilon}_{j,t}^{\max} \left(d_{j,t} - D_{j,t}^{\max}\right) + \sum_{i,t} \hat{\pi}_{i,t}^U \left(g_{i,t} - g_{i,(t-1)} - R_i^U\right) + \sum_{i,t} \hat{\pi}_{i,t}^D \left(g_{i,(t-1)} - g_{i,t} - R_i^D\right)$$

where:

$$V^L = \{g, d, \hat{\lambda}, \hat{\mu}^{\min}, \hat{\mu}^{\max}, \hat{\upsilon}^{\min}, \hat{\upsilon}^{\max}, \hat{\pi}^U, \hat{\pi}^D\}$$

The first condition is the **stationary condition**, which requires that the derivative of the Lagrangian with respect to each primary variable is zero. This ensures that the solution is at a stationary point of the Lagrangian.

$$\frac{\partial L}{\partial g} = k_{i,t}\Lambda_i^G - \hat{\lambda}_t - \hat{\mu}_{i,t}^{\min} + \hat{\mu}_{i,t}^{\max} + \hat{\pi}_{i,t}^U - \hat{\pi}_{i,t+1}^U - \hat{\pi}_{i,t}^D + \hat{\pi}_{i,t+1}^D = 0, \quad \forall i, \forall t > 1 \tag{4a}$$

$$\frac{\partial L}{\partial g} = k_{i,t}\Lambda_i^G - \hat{\lambda}_t - \hat{\mu}_{i,t}^{\min} + \hat{\mu}_{i,t}^{\max} - \hat{\pi}_{i,t}^U + \hat{\pi}_{i,t}^D = 0, \quad \forall i, \forall t = 1 \tag{4b}$$

$$\frac{\partial L}{\partial d} = -\Lambda_{j,t}^D + \hat{\lambda}_t - \hat{\upsilon}_{j,t}^{\min} + \hat{\upsilon}_{j,t}^{\max} = 0, \quad \forall j, \forall t \tag{4c}$$

The second condition is the **primal and dual feasibility** condition. Primal feasibility ensures that the original constraints of the optimization problem are satisfied, while dual feasibility requires that the Lagrange multipliers are non-negative.

$$\hat{\mu}^{\min}, \hat{\mu}^{\max}, \hat{\upsilon}^{\min}, \hat{\upsilon}^{\max}, \hat{\pi}^U, \hat{\pi}^D \geq 0 \tag{5}$$

The third condition is the **complementary slackness** condition, which ensures that for each constraint, either the constraint is active or the corresponding Lagrange multiplier is zero. This condition is critical for ensuring that the solution is optimal.

$$0 \leq \hat{\mu}_{i,t}^{\min} \perp (u_{i,t}G_i^{\min} - g_{i,t}) \geq 0, \quad \forall i, \forall t \tag{6a}$$

$$0 \leq \hat{\mu}_{i,t}^{\max} \perp (g_{i,t} - u_{i,t}G_i^{\max}) \geq 0, \quad \forall i, \forall t \tag{6b}$$

$$0 \leq \hat{\upsilon}_{j,t}^{\min} \perp d_{j,t} \geq 0, \quad \forall j, \forall t \tag{6c}$$

$$0 \leq \hat{\upsilon}_{j,t}^{\max} \perp (d_{j,t} - D_j^{\max}) \geq 0, \quad \forall j, \forall t \tag{6d}$$

$$0 \leq \hat{\pi}_{i,t}^U \perp (g_{i,t} - g_{i,(t-1)} - R_i^U) \geq 0, \quad \forall i, \forall t \tag{6e}$$

$$0 \leq \hat{\pi}_{i,t}^D \perp (g_{i,(t-1)} - g_{i,t} - R_i^D) \geq 0, \quad \forall i, \forall t \tag{6f}$$

These complementary slackness constraints are quadratic in nature and can be reformulated using the big M method, which introduces auxiliary variables and large constants to linearize the constraints.

C FINAL FORM OF THE PROPOSED OPTIMIZATION PROBLEM

The bi-level optimization problem is reformulated as a single-level optimization problem by incorporating the KKT conditions derived above. This approach follows the methodology outlined in [44] but extends it by including the additional KKT conditions to more accurately model the market-clearing process.

$$\max_{V} \sum_{t} \left(\lambda_t g_{i,t} - \Lambda_i^G g_{i,t} - c_{i,t}^U - c_{i,t}^D\right)$$

$$- W \left[\left(\sum_{i,t} \left(k_{i,t}\Lambda_i^G g_{i,t} + c_{i,t}^U + c_{i,t}^D\right) - \sum_{j,t} \Lambda_{j,t}^D d_{j,t}\right)\right.$$

$$\left. + \left(\sum_{j,t} v_{j,t}^{\max} D_{j,t}^{\max} + \sum_{i,t} \pi_{i,t}^U R_i^U + \sum_{i,t} \pi_{i,t}^U G_{i,0} - \sum_{i,t} \pi_{i,t}^D G_{i,0} + \sum_{i,t} \pi_{i,t}^D R_i^D + \sum_{i,t} \sigma_{i,t}^U K_i^U U_{i,0} - \sum_{i,t} \sigma_{i,t}^D K_i^D U_{i,0} + \sum_{i,t} \psi_{i,t}^{\max}\right)\right] \tag{7a}$$

where:

$$V = \left\{k_i, g_{i,t}, u_{i,t}, c_{i,t}^U, c_{i,t}^D, d_{j,t}, \lambda_t, \mu_{i,t}^{\max}, v_{j,t}^{\max}, \pi_{i,t}^U, \pi_{i,t}^D, \sigma_{i,t}^U, \sigma_{i,t}^D, \psi_{i,t}^{\max}\right\} \tag{7b}$$

subject to:

upper-level problem constraint

$$1 \le k_i \le k_i^{\max} \tag{7c}$$

lower-level problem constraints

$$\sum_j d_{j,t} - \sum_i g_{i,t} = 0 : \lambda_t, \quad \forall t \tag{7d}$$

$$u_{i,t} G_i^{\min} \le g_{i,t} \le u_{i,t} G_i^{\max} : \mu_{i,t}^{\min}, \mu_{i,t}^{\max}, \quad \forall i, \forall t \tag{7e}$$

$$0 \le d_{j,t} \le D_{j,t}^{\max} : v_{j,t}^{\min}, v_{j,t}^{\max}, \quad \forall j, \forall t \tag{7f}$$

$$g_{i,t} - G_{i,0} \le R_i^U : \pi_{i,t}^U, \quad \forall i, t = 1 \tag{7g}$$

$$g_{i,t} - g_{i,(t-1)} \le R_i^U : \pi_{i,t}^U, \quad \forall i, \forall t > 1 \tag{7h}$$

$$G_{i,0} - g_{i,t} \le R_i^D : \pi_{i,t}^D, \quad \forall i, t = 1 \tag{7i}$$

$$g_{i,(t-1)} - g_{i,t} \le R_i^D : \pi_{i,t}^D, \quad \forall i, \forall t > 1 \tag{7j}$$

$$c_{i,t}^U \ge 0 : p_{i,t}^U, \quad \forall i, \forall t \tag{7k}$$

$$c_{i,t}^D \ge 0 : p_{i,t}^D, \quad \forall i, \forall t \tag{7l}$$

$$c_{i,t}^U \ge (u_{i,t} - U_{i,0}) K_i^U : \sigma_{i,t}^U, \quad \forall i, t = 1 \tag{7m}$$

$$c_{i,t}^U \ge (u_{i,t} - u_{i,(t-1)}) K_i^U : \sigma_{i,t}^U, \quad \forall i, \forall t > 1 \tag{7n}$$

$$c_{i,t}^D \ge (U_{i,0} - u_{i,t}) K_i^D : \sigma_{i,t}^D, \quad \forall i, t = 1 \tag{7o}$$

$$c_{i,t}^D \ge (u_{i,(t-1)} - u_{i,t}) K_i^D : \sigma_{i,t}^D, \quad \forall i, \forall t > 1 \tag{7p}$$

$$u_{i,t} = \{0, 1\}, \quad \forall i, \forall t \tag{7q}$$

relaxed lower level problem constraints when $0 \le u \le 1$

$$k_{i,t}\Lambda_i^G - \lambda_t + \mu_{i,t}^{\max} - \mu_{i,t}^{\min} + \pi_{i,t}^U - \pi_{i,(t+1)}^U - \pi_{i,t}^D + \pi_{i,(t+1)}^D = 0, \quad \forall i, \forall t < N_T \tag{7r}$$

$$k_{i,t}\Lambda_i^G - \lambda_t + \mu_{i,t}^{\max} - \mu_{i,t}^{\min} + \pi_{i,t}^U - \pi_{i,t}^D = 0, \quad \forall i, \forall t = N_T \tag{7s}$$

$$-\mu_{i,t}^{\max} G_i^{\max} + \mu_{i,t}^{\min} G_i^{\min} + (\sigma_{i,t}^U - \sigma_{i,(t+1)}^U) K_i^U - (\sigma_{i,t}^D - \sigma_{i,(t+1)}^D) K_i^D + \psi_{i,t}^{\max} \ge 0, \quad \forall i, \forall t < N_T \tag{7t}$$

$$-\mu_{i,t}^{\max} g_i^{\max} + \mu_{i,t}^{\min} g_i^{\min} + \sigma_{i,t}^U K_i^U - \sigma_{i,t}^D K_i^D + \psi_{i,t}^{\max} \ge 0, \quad \forall i, \forall t = N_T \tag{7u}$$

$$1 - \sigma_{i,t}^U \ge 0, \forall i, \quad \forall t \tag{7v}$$

$$1-\sigma^D_{i,t} \geq 0, \forall i, \quad \forall t \tag{7w}$$

$$-\Lambda^D_{j,t} + \lambda_t + \nu^{\max}_{j,t} \geq 0, \quad \forall j, \forall t \tag{7x}$$

$$\nu^{\max}_{j,t} \geq 0, \quad \forall j, \forall t \tag{7y}$$

$$\mu^{\max}_{i,t}, \mu^{\min}_{i,t}, \pi^U_{i,t}, \pi^D_{i,t}, \sigma^U_{i,t}, \sigma^D_{i,t}, \psi^{\max}_{i,t} \geq 0, \quad \forall i, \forall t \tag{7z}$$

KKT conditions of the lower level porblem when $u = u*$

$$k_{i,t}\Lambda^G_i - \hat{\lambda}_t - \hat{\mu}^{\min}_{i,t} + \hat{\mu}^{\max}_{i,t} + \hat{\pi}^U_{i,t} - \hat{\pi}^U_{i,t+1} - \hat{\pi}^D_{i,t} + \hat{\pi}^D_{i,t+1} = 0, \quad \forall i, \forall t > 1 \tag{7aa}$$

$$k_{i,t}\Lambda^G_i - \hat{\lambda}_t - \hat{\mu}^{\min}_{i,t} + \hat{\mu}^{\max}_{i,t} - \hat{\pi}^U_{i,t} + \hat{\pi}^D_{i,t} = 0, \quad \forall i, \forall t = 1 \tag{7ab}$$

$$-\Lambda^D_{j,t} + \hat{\lambda}_t - \hat{\upsilon}^{\min}_{j,t} + \hat{\upsilon}^{\max}_{j,t} = 0, \quad \forall j, \forall t \tag{7ac}$$

$$\hat{\mu}^{\min}, \hat{\mu}^{\max}, \hat{\upsilon}^{\min}, \hat{\upsilon}^{\max}, \hat{\pi}^U, \hat{\pi}^D \geq 0 \tag{7ad}$$

$$0 \leq \hat{\mu}^{\min}_{i,t} \perp (u_{i,t}G^{\min}_i - g_{i,t}) \geq 0, \quad \forall i, \forall t \tag{7ae}$$

$$0 \leq \hat{\mu}^{\max}_{i,t} \perp (g_{i,t} - u_{i,t}G^{\max}_i) \geq 0, \quad \forall i, \forall t \tag{7af}$$

$$0 \leq \hat{\upsilon}^{\min}_{j,t} \perp d_{j,t} \geq 0, \quad \forall j, \forall t \tag{7ag}$$

$$0 \leq \hat{\upsilon}^{\max}_{j,t} \perp (d_{j,t} - D^{\max}_j) \geq 0, \quad \forall j, \forall t \tag{7ah}$$

$$0 \leq \hat{\pi}^U_{i,t} \perp (g_{i,t} - g_{i,(t-1)} - R^U_i) \geq 0, \quad \forall i, \forall t \tag{7ai}$$

$$0 \leq \hat{\pi}^D_{i,t} \perp (g_{i,(t-1)} - g_{i,t} - R^D_i) \geq 0, \quad \forall i, \forall t \tag{7aj}$$

The final form of the proposed optimization problem integrates the upper-level strategic bidding decisions with the lower-level market-clearing mechanism, ensuring that the solution satisfies all necessary conditions for optimality. This formulation provides a more accurate benchmark for evaluating the performance of DRL algorithms in simulating electricity market dynamics.

D BI-LINEAR TERMS LINEARIZATION

To avoid solving bi-linear problems with binary terms, we employ the binary expansion approach. Let $g_{i,t,n}, n = 1, 2, \ldots, L$ be a set of discrete values within the range $[0, G^{\max}_i]$ as prescribed by the physical bounds, where $L = 2^K$. In the current work, an expansion multiplier $K = 10$ was used. This allows us to express the variable $g_{i,t}$ as a sum of binary variables:

$$g_{i,t} = \Delta_i \sum_{n=0}^{K} 2^n x_{i,t,n}, \quad \forall i, \forall t \tag{8}$$

Here, $x_{i,t,n}$ is an auxiliary binary variable that facilitates the linearization process.

By multiplying both sides by λ_t, summing for every t, and defining a dummy variable $z^\lambda_{i,t,n}$, we obtain:

$$\sum_t \lambda_t g_{i,t} = \Delta_i \sum_{t,n} 2^n z^\lambda_{i,t,n} \tag{9a}$$

$$z^\lambda_{i,t,n} = \lambda_t x_{i,t,n}, \quad \forall i, \forall t, \forall n \tag{9b}$$

This transformation allows us to replace the bi-linear term $\sum_t \lambda_t g_{i,t}$ with a linear expression. The product of variables in the equation can be converted into equivalent mixed-integer linear constraints:

$$0 \leq \lambda_t - z^\lambda_{i,t,n} \leq M(1 - x_{i,t,n}), \quad \forall i, \forall t, \forall n \tag{10a}$$

$$0 \leq z^\lambda_{i,t,n} \leq M x_{i,t,n}, \quad \forall i, \forall t, \forall n \tag{10b}$$

Similarly, we can linearize the term $\Lambda^G_i \sum_t k_{i,t} g_{i,t}$. By multiplying both sides of the original equation by $k_{i,t}$, summing for every t, and defining $z^k_{i,t,n}$, we get:

$$\lambda^G_i \sum_t k_{i,t} g_{i,t} = \lambda^G_i \Delta_i \sum_{t,n} 2^n z^k_{i,t,n} \tag{11a}$$

$$z^k_{i,t,n} = k_{i,t} x_{i,t,n}, \quad \forall i, \forall t, \forall n \tag{11b}$$

This process enables us to replace the bi-linear term $\sum_t k_{i,t} g_{i,t}$ with a linear expression. The product of variables can then be transformed into equivalent mixed-integer linear constraints:

$$0 \leq k_{i,t} - z^k_{i,t,n} \leq M(1 - x_{i,t,n}), \quad \forall i, \forall t, \forall n \tag{12a}$$

$$0 \leq z^k_{i,t,n} \leq M x_{i,t,n}, \quad \forall i, \forall t, \forall n \tag{12b}$$

Power-dependent price profiles - defining grid- and market-oriented incentives for building energy management systems

TOBIAS RIEDEL, FZI Research Center for Information Technology, Germany
CARL HAUSCHKE, FZI Research Center for Information Technology, Germany
HARTMUT SCHMECK*, Karlsruhe Institute of Technology, Germany

In order to consider the grid- and market-oriented situation of the energy system and provide corresponding price signals for building energy management systems (BEMS), we propose time- and power-dependent price profiles. Those incentive signals are meant to allow the optimization of BEMS towards dynamic prices while providing a clear incentive to limit power during grid congestion. We evaluate those signals with a generic battery optimizer that is capable of optimizing towards them. Simulations show that those time- and power-dependent grid signals can reduce peak loads in distribution grids while at the same time allowing more market orientation than volumetric dynamic grid fees.

CCS Concepts: • **Information systems** → *Process control systems.*

Additional Key Words and Phrases: market-orientation, grid-orientation, dynamic grid fees, BEMS, incentive signals, optimization, EMS, energy management

1 INTRODUCTION

On March 10th, 2024, consumers in southwest Germany could observe an interesting phenomenon in the energy system: The weather has been windy and sunny in Germany and the load has been low on this Sunday. The result have been negative spot prices in Germany between 12:00 and 15:00. The "Stromgedacht" app[1] of the transmission system operator (TSO) of Baden-Württemberg TransnetBW indicated a "super green" period for nearly the whole day up to 16:00, indicating that electricity should be used preferably within this time. However, during 11:00 and 14:00, when wind and solar generation has been at its highest, the app switched to "orange", which means that consumption should be lowered if possible. The reason for this were bottlenecks in the transmission grid, likely caused by high electricity export to Switzerland and France during that time, which led to high redispatch costs [50Hertz Transmission GmbH, Amprion GmbH, TransnetBW GmbH, TenneT TSO GmbH 2024] [Bundesnetzagentur 2024]. Events like this show that the energy system is more complex than it can be indicated by colors or simple prices. It is therefore not trivial which information in which degree of complexity should be passed to households in order to activate flexibility.

What happens in the transmission grid will likely appear in the distribution grids as well: As soon as a large number of building energy management systems (BEMS) optimize the schedules of batteries, electric vehicles (EV), heat pumps (HP), and other flexible loads with respect to dynamic electricity prices, those price signals might cause bottlenecks in distribution grids. Since it makes sense to use demand flexibility and distributed storages for the energy system as a whole, it is crucial to define feasible incentive signals that take into account grid and market situations. The field of tension lies between the balancing of the residual load caused by the misfit of volatile renewable generation and the load on the one hand (market orientation) and the avoidance of grid congestion on the other hand (grid orientation).

1.1 Integrating BEMS into the energy system

BEMS can be key elements in a decentralized energy system that is based on renewable energies. In order to keep the grid stable and balance fluctuations of solar and wind power generation, millions of decentralized storages and flexible loads have to be controlled in a feasible way. Nowadays, battery storages in households are usually used to increase self-sufficiency with on-site photovoltaic (PV) generation. Self-consumption and self-sufficiency are often the only use cases for demand-side flexibility options like EV and HP as well. This is simply done by minimizing energy exchange at the building's grid connection point (GCP). However, the energy system as a whole could benefit more from flexibility options in buildings [Eicke et al. 2024]. For example, batteries that are only used for PV self-sufficiency remain discharged the whole night even if there is excess wind power nearby that has to be curtailed due to grid bottlenecks. Furthermore, when the residual load and thus market prices are very high, households could be financially rewarded for feeding in electricity from batteries, reducing the demand for expensive power generation. However, these market-oriented incentives should not lead to grid congestion due to synchronized behavior.

1.2 Contribution of this paper

In order to define feasible incentives for BEMS that reflect both grid and market situation properly, we propose power-dependent price profiles, shortly *PowerPriceProfiles* (PPP). Those PPP define electricity prices over time and are limited to a certain power for a certain time. Through this power- and time dependency and the possibility to "stack" multiple PPP to a *ProfileStack* (PS), it is possible to project more complex information from the energy system into price signals. At the same time, PPP and PS can be handled by BEMS' optimizers in a very similar way as on-site PV generation, making the application easy. Briefly, the contribution of this paper is the following:

- We define the model of *PowerPriceProfiles* and the possibility to stack them to a *ProfileStack*. These models can reflect the (complex) situation of the energy system better than simple dynamic tariffs and provide respective incentives to the BEMS.

[1] https://api.stromgedacht.de/index.html

Authors' addresses: Tobias Riedel, riedel@fzi.de, FZI Research Center for Information Technology, Haid-und-Neu Str. 10-14, Karlsruhe, Germany; Carl Hauschke, hauschke@fzi.de, FZI Research Center for Information Technology, Haid-und-Neu Str. 10-14, Karlsruhe, Germany; Hartmut Schmeck, hartmut.schmeck@kit.edu, Karlsruhe Institute of Technology, Kaiserstraße 12, Karlsruhe, Germany.

- We evaluate these signals in two examples and five scenarios with a generic battery optimizer that is capable to use those signals as input parameters in order to optimize the operation of batteries and EV. As performance indicators, we use economic results for households, peak load in the distribution grid, and the power during the highest residual load in the market zone.
- We publish the source code of the generic battery optimizer and the described data models open source[2].

This paper is structured as follows. In Section 2, we review related publications with respect to different tariff designs, coordination mechanisms, and protocols. In Section 3, we describe our approach to solve the given problem. The evaluation methodology is carried out in Section 4, followed by the results (Sec. 5) and a discussion (Sec. 6), where we also conclude with a short outlook.

2 RELATED WORK

This paper contributes to tariff designs for households with BEMS in order to provide information and incentives from the energy system (grid and market) and the capability of BEMS to react to those tariffs. Therefore, we include literature according different tariff designs and coordination mechanisms as well as protocols in the field. As outlined by [Hillemacher et al. 2013], there are two basic approaches to integrate households into the energy system: Dynamic price incentives and direct load control from a central instance, e.g. an aggregator. In this paper, we focus only on the first option without the need for a central control instance.

2.1 Dynamic energy prices

A very common way to provide both information and incentives from the energy system to BEMS is the use of dynamic price signals, visualized by different colors in Fig. 1. Throughout this paper, we visualize low prices with cold colors (blue) and high prices with warm colors (red). Technically, dynamic prices are defined by a time series of prices for a certain amount of energy to a certain period of time. Therefore, they can be described as *volumetric*. Following the definition made by Amin et al. [Amin et al. 2020], both time-of-use (ToU) tariffs, real-time-pricing (RTP) and critical-peak-pricing (CPP) can be seen as dynamic energy prices, since the basis of calculation is always a price per kilowatt-hour (kWh). These tariffs only differ in terms of the duration of certain prices and the announcement horizon. RTP tariffs are often based on day-ahead spot market prices with taxes, surcharges, and a margin (e. g. [aWATTar 2024]). In order to respect the grid situation, variable grid fees are a possible instrument that is often discussed. The idea of dynamic grid fees is to reflect the local grid load in the grid fees, providing financial incentives to shift load in a way that is favorable to the local (usually low-voltage) grid. Since the final consumer electricity price consists of sum of the market price, grid fees and levies, surcharges and taxes, it is possible to define both dynamic market prices and dynamic grid fees at the same time. No matter if dynamic prices are based on spot markets, variable grid fees, or of both, BEMS have to optimize towards a single time series of prices, since grid fees are added to market prices. Optimization algorithms to do that have already been investigated in the past with respect to different flexible loads and storage options (e. g. [Müller et al. 2018]). Several publications also investigate the effects of dynamic grid fees that are added to spot market prices with the goal of reflecting both grid and market situation within one price signal. Wanapinit et al. [Wanapinit et al. 2022] investigate day-ahead prices combined with a three-level grid fee and their effects on an industry gravel plant. They conclude that while variable tariffs are a good starting point, the given three-level grid fees do not reliably limit grid load. They also mention the "avalanche effect" or "herding effect" as called by Gottwalt [Gottwalt 2015], meaning that price signals can cause load synchronization and lead to grid congestion that may have not occurred without those signals. This finding is in line with von Bonin et al. [von Bonin et al. 2022] who compare EV charging strategies with respect to dynamic prices and/or PV self-consumption. They point out that identical price signals can lead to critical charging peaks. The authors of [Agora Energiewende and Forschungsstelle für Energiewirtschaft e. V. 2023] come to the same conclusion that spot market prices alone can increase grid congestion. They propose dynamic grid fees that should be added to the market signals. However, the authors state that this instrument could only reduce the additional costs caused by market signals, while grid load with combined grid- and market-oriented prices is still higher than with a flat electricity price. Möller et al. [Möller et al. 2021] compare variable grid fees with other incentive models, such as flexibility markets, and states that tariffs for grid- and market-oriented flexibility usage are subject to further research. Another approach of dynamic tariffs is investigated by Mauser et al. [Mauser and Schmeck 2014], who combine a long-term time-of-use tariff with optional short-term tariffs that is offered by DSO and that BEMS can accept if it is favorable for the household.

Based on this literature, it can be concluded that while dynamic prices have their advantages in grid- or market-oriented load shifting, load synchronization due to simultaneous behavior of BEMS is a risk that can cause new grid bottlenecks. BEMS optimization algorithms tend to exploit the lowest price period with maximum power. For example, optimizing an EV schedule with respect to dynamic prices usually means charging with maximum power during the lowest price within the charging period. Variable prices do not provide an incentive to smooth load peaks, as they are independent of the power drawn from the grid. Eicke et al. [Eicke et al. 2024] also addressed this issue and concluded that while dynamic grid fees were a good starting point, in the long term dynamic prices should be combined with additional instruments. According to them, one feasible instrument is the use of situation-based power prices, making peaks unattractive.

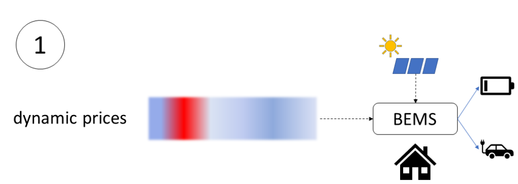


Fig. 1. Dynamic prices

[2]https://github.com/fzi-forschungszentrum-informatik/generic-battery-optimizer

2.2 Power limits

From grid operators' perspective, it is favorable to limit the power of buildings drawn from the grid during load-side grid congestion. Accordingly, during feed-in congestion, the power of buildings fed into the grid should be limited. Those limitations for power drawn from and fed into the grid over time form a hull curve for possible schedules of the building at the grid connection point (Fig. 2). BEMS can optimize within this corridor. The hull curve can be defined by the limitations of power (in kW). Alternatively, the hull curve is defined by a quota referring to the maximum power at the GCP, the maximum power of flexible devices, or the planned schedules of BEMS [Exner et al. 2020]. If prices are not varying over time, the optimization targets are usually maximizing self-consumption and avoiding curtailment of PV generation.

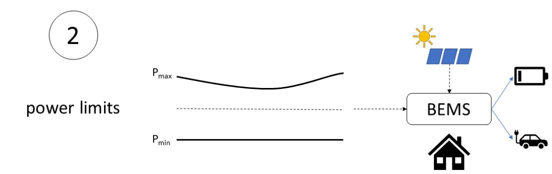


Fig. 2. Power limits forming a hull curve

2.3 Combination of variable prices and power limits

When combining those two approaches, the result is a price profile with limited power, as visualized in Fig 3. This is what a BEMS has to handle as optimization parameters. However, power limits usually come from the distribution system operator (DSO), while dynamic prices are sent by the electricity supplier.

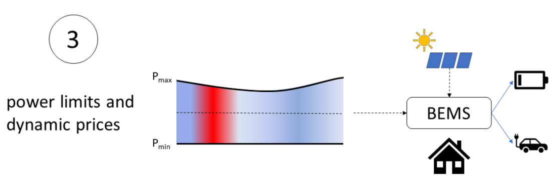


Fig. 3. Combination of power limits and variable prices

2.4 Load-variable tariffs

Load-variable tariffs can be based either on the highest peak power within a year or, as visualized in Fig. 4, change the energy price depending on the power of the building. Both options obviously provide a clear incentive to smooth the load and avoid peaks. Stute et al. [Stute and Klobasa 2024] investigated the combination of dynamic market prices and different grid charge designs, including capacity subscription charges. They conclude that those capacity charges significantly reduce grid reinforcement costs. The authors focus on economic choices of households and grid costs, but not the residual load. [Gansemer and Grossmann 2011] develop a demand response model for households in order to analyze the effects of time-variable or load-variable tariffs on the load curves. The consider the total energy demand as well as peak load reduction, but not the optimal use of volatile renewable energy sources.
Because of the lack of time dependence, load-variable tariffs provide no incentive to react dynamically to the availability of renewable energy. Furthermore, they prevent load increase during a supply-side grid bottleneck.

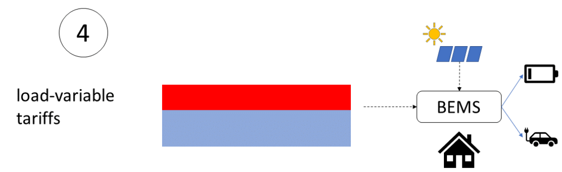


Fig. 4. Load-variable tariffs

2.5 Coordination mechanisms

In addition to the price and power limitation signals themselves, it is a relevant question how these signals are generated. Promising approaches are local flexibility markets that allow BEMS to change the originally assigned power limits via market mechanisms. For example, the "KOALA"-concept proposed in the research project unIT-e^2 envisions an auction of free network capacity [FfE 2024]. Power at the GCP is limited initially, but the limit can be raised in return for a payment. Another approach is to jointly update BEMS schedules in a high frequency based on a very dynamic price signal that depends on the schedules, which forms a closed loop [Gottwalt 2015]. Power-dependent price profiles as described in this paper can be the result of such coordination mechanisms, but do not depend on their existence. Instead, the DSO could also define power-dependent price profiles based on predictions. A reasonable horizon for defining those grid signals could be the day ahead, following the pattern of the spot markets. Differently from flexibility markets, power-dependent price profiles allow one to exceed a certain power limit (for example, when an EV has to be charged as soon as possible) without sending a request. Exceeding the limit only results in higher costs. Not requiring coordination mechanisms simplifies the necessary transactions and has advantages with respect to privacy protection.

2.6 Protocols

Protocols defining the communication interface between BEMS and the energy system have to build upon basic assumptions on how BEMS can interact with the energy system. In the following, the most relevant protocols for the interface between BEMS and the energy system are discussed. There are only protocols selected that address both grid- and market-oriented signals for BEMS. For example, the building communication protocol KNX could be used for power limits in Germany, as proposed by VDE FNN [FNN 2024a], but to the authors' knowledge, it is not meant to transfer dynamic electricity prices.

2.6.1 EEBus. EEBus is driven by the EEBus Initiative in Cologne, Germany. It defines communication interfaces to connect BEMS with grid and market operators. In their whitepaper they address solutions for power limitation due to curative grid measures, self-consumption optimization, as well as dynamic pricing and flexibility provision [Hollmann et al. 2024]. One part of the EEBus standard is the VDE application rule VDE-AR-E 2829-6-4, which standardizes the interface of the energy system to properties and customer's devices. Therefore, EEBus can become the most relevant communication protocol for the interface between BEMS and the energy system in Germany. The EEBus standard includes the use case "Incentive-Table based power consumption management" (IT-PCM). Those incentive tables are time- and power dependent price profiles. There are one or more tiers that define prices in a certain time period up to a certain power. Those tiers can be stacked, for example surplus PV power for 0 ct/kWh up to 1500 W, PV power for 12 ct/kWh up to 2500 W and (theoretically unlimited) grid power for 30 ct/kWh. In this example, if a power of 2000 W is needed, 1500 W are for free and 500 W cost 12 ct/kWh [EEBus 2022]. This means that PowerPriceProfiles can be implemented in EEBus. However, the described use cases of the EEBus incentive tables are focused on use cases "behind the meter", not on incentives coming from actors in the energy system sending incentives to the BEMS. To transfer the concept of this paper into practice, we propose to build upon the existing data models of the EEBus incentive tables to define grid- and market-oriented signals that are valid at the GCP.

2.6.2 OpenADR. Similarly to EEBus, the American standard OpenADR addresses dynamic prices as well as power limitations at the GCP [OpenADR 2023a]. It also allows direct load control (for example to operate virtual power plants) as well as critical peak pricing. Concepts like time- and power-dependent price profiles or incentive tables are not mentioned in the user guide [OpenADR 2023b].

2.7 Regulatory status quo in Germany

Since beginning of 2024, all new chargepoints, heat pumps, batteries, and air conditioners in households in Germany can be limited to 4.2 kW by the DSO according § 14a of the Energy Industry Act (EnWG) and the specification of the federal network agency (Bundesnetzagentur) ([Bundesnetzagentur 2023a]). This regulation differs to the power limits as described above in three aspects: Firstly, there is no "hull curve" over time but an immediate limitation. Secondly, only load is considered, not feed-in. And thirdly, not the power at the GCP is limited, but the effective power of flexible loads from the grid, ignoring inflexible demand. In the specification, variable grid fees are foreseen as well from April 2025. Furthermore, offering dynamic tariffs is obligatory for all electricity suppliers from the beginning of 2025 in Germany (§ 41a EnWG). This means that, from April 2025, every household with flexible loads can combine a variable grid fee with a dynamic tariff and power limits by the DSO. It follows the question which of these possible combinations is favorable to the energy system and how BEMS can optimize towards them. This paper aims at contributing to both questions.

3 SOLUTION APPROACH

In order to design grid and market oriented incentives for BEMS, we propose to build upon the combination of dynamic prices and power limits. Dynamic prices signal when *energy* is available for low costs due to a low residual load in the market zone. Power limits signal how much *power* the grid can transport at a certain time. The time-dependent combination of prices and power limits that is visualized in Fig. 3 is one PPP. The profile contains information on how much power is available at which energy price in a specific time period. It can be used to model local PV production, for example, which is available for zero marginal costs up to a certain power. Above this power, the power from the grid has to be used for another price. Following this idea, the power from the grid can also be limited to a certain power due to grid bottlenecks, which can be modeled by another PPP. The question arises what happens above this limit. We suggest that dynamic grid fees should be applied here: Exceeding a power limit should be priced with an unattractively high grid fee. This "penalty" grid fee, which is only valid above a certain limit, can be modeled by a PPP again. As a result, we get three stacked PPP, which we call *ProfileStack* (PS), as visualized in Fig. 5. The first PPP represents on-site PV power generation, the second one a dynamic energy price from the grid, and the third one an energy price plus a dynamic grid fee. The advantage of this approach compared to a fixed power limit is that households can exceed the limit if adhering to it would entail a major loss of comfort. An example could be an EV that has to be charged as soon as possible. Differently to the approach of a flexibility market, no additional request or transaction is required, since all information and potential costs have already been provided to the BEMS.

In this example, we assume that the technical maximum power of the household at the GCP is 30 kW. The PS contains all information on the energy sources available for the optimizer. Revenue for feeding electricity into the grid can be represented in the same manner.

PPP and PS could also be used to enable peer-to-peer trading, for example in energy communities. Each PPP could represent excess PV power generation, for example, and be traded locally. Another BEMS can optimize towards available PPP in the neighborhood or the combination of an own balcony PV plant and common building supply ("gemeinschaftliche Gebäudeversorgung") in flats. These use cases are not investigated further in detail in this paper, since we focus on the trade-off between grid- and market-orientation. Key elements for all described use cases are optimization algorithms in BEMS that can take PS as an input.

Additionally to prices for purchasing electricity, PS can also be used to describe prices for feed-in. This can be advantageous when PV or battery feed-in power should be limited due to grid congestion. By using PS, the feed in remuneration for power exceeding a limit can be reduced, which provides an incentive for smoothing the feed-in power.

3.1 Generic Battery Optimizer

In the following, we describe the functionality of the optimization model that is capable of optimizing with respect to multiple PPP stacked to a PS. It uses PPP for all data inputs that have a time-based reference. Fig. 6 shows the basic structure of the optimizer in the BEMS, with a PS consisting of three PPP for electricity from the grid (e.g. one for neighborhood solar energy, one from the supplier and one signaling grid congestion) and one PPP for local PV. Furthermore, a PS consisting of two feed-in PPP is available. Taking into

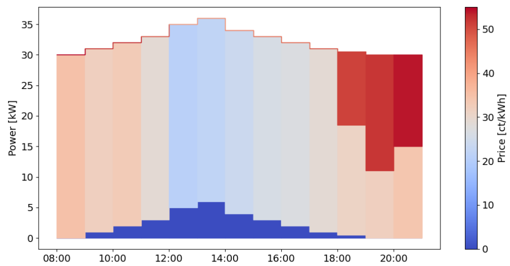


Fig. 5. Example ProfileStack consisting of 3 PowerPriceProfiles

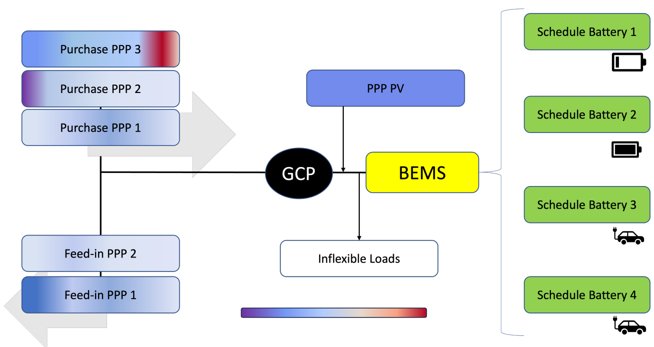


Fig. 6. Overview of PPP and BEMS

account inflexible loads as well, the optimizer generates optimal charging and discharging schedules for all batteries and EVs (in this case four).

On a high level the optimizer aims to calculate power schedules for all attached devices of an energy system given a set of energy sources, energy sell targets, consumption devices and batteries. These can be flexibly added to an optimization problem. The optimizer distinguishes between purchase profiles that contain PPP that allow a household to get energy from, sell profiles containing all PPP the household can use to sell energy, and inflexible consumption that cannot be controlled and must be met at all times.

3.1.1 Sources and Sinks. The different profiles the optimizer receives are treated internally as source profiles or sink profiles. Sinks consume energy and source profiles provide energy. Batteries are also interpreted as energy profiles. Their power input or output capabilities are determined by the SoC and technical specifications. Batteries can be both source or sink profiles depending on the capabilities. As the optimizer bases its decisions solely on the lowest possible price, it does not differentiate between different types of sources like PV or batteries. The decision was made to separate purchase and sell profiles in the inputs of the optimizer, since there is no implicit indication of what a profile represents. Both purchase and sell profiles can have negative price values, indicating that obtaining energy from that profile would create revenue or that feeding in energy would cost money. Both scenarios are feasible in current electricity networks, indicating either a surplus of supply or an excessive load on the grid where any additional feed-in would further worsen the situation of the grid.

3.1.2 Objective. The objective of the optimization problem is to minimize the electricity costs (C) for the household. The cost comprises the sum of all source energies ($e_{i,t}$) and their prices ($c_{i,t}$) and all sink energies ($e_{j,t}$) with their revenue ($c_{j,t}$). Sinks that are devices in a household that consume energy are treated as sinks that do not generate income. For clarity reasons, these devices are omitted from the cost calculation in the actual implementation. To account for the potential value of energy stored in a battery at the end of the optimization we multiply the highest revenue in the last time step (n) that is available from the sell profiles (sinks) ($max\{c_{j,n} : j \in J\}$) and multiply it with the energy stored in the batteries in the last time step ($E_{b,n}$).

$$C = \sum_{i \in I} \sum_{t \in T} e_{i,t} c_{i,t} - \sum_{j \in J} \sum_{t \in T} e_{j,t} c_{j,t} - \sum_{b \in B} E_{b,n} max\{c_{j,n} : j \in J\}$$

3.1.3 Energy. Throughout the optimizer, energy is used and not power to account for differing lengths between two time periods. This requires preprocessing of all input values and postprocessing of all output values but makes the calculations easier within the optimization, as the time component is removed from the calculations. If a feed-in profile allowed a maximum feed-in of 30 kW with a period length of 15 minutes, the preprocessing of the optimizer would constrain the energy usage to 7.5 kWh in that period which translates to a constant usage of 30 kW over 15 minutes. During a time period the power usage is always assumed to be constant, which allows us to calculate with energy units. To allow for granular schedules of power usage, a time series of differing intervals can be used. The input time series will be transformed to the union of all input time series. In this way a granular time series is created that covers all time steps of all input profiles and batteries.

3.1.4 Source modeling. Energy sources $i \in I$ are all profiles that allow a household to obtain energy. This can be a grid tariff, a PV system, or a battery. These profiles are restricted by the maximum energy $E_{i,t}$ that may be drawn from it within each individual time step. The optimizer uses all available energy sources $i \in I$ to meet the needs of the household and sells energy whenever feasible to reduce the overall costs for the household. The calculation relies on the energy matrix introduced in subsubsection 3.1.7 where $m_{i,j,t}$ is an energy flow from source i to sink j in period t.

$$E_{i,t} \geq \sum_{j \in J} m_{i,j,t} \geq 0$$

3.1.5 Sink modeling. Energy sinks $j \in J$ represent all profiles that absorb energy. This can be a feed-in tariff, a device that needs energy in a household such as a heat pump or dishwasher, or a battery that can or must be charged. Similar to the sources the sinks are restricted by the maximum energy they can absorb in each time step $E_{j,t}$.

$$E_{j,t} \geq \sum_{i \in I} m_{i,j,t} \geq 0$$

Fixed consumption profiles get an additional constraint that enforces the fulfillment of the required energy.

$$E_{j,t} \geq \sum_{i \in I} m_{i,j,t} \geq E_{j,t}$$

3.1.6 Battery modeling. Batteries $b \in B$ are modeled as an energy reservoir with a maximum amount of energy they can hold. Both the charging $e_{b,t}^{charge}$ and discharging energy $e_{b,t}^{discharge}$ of a battery can be limited to constrain power by either the battery's capabilities or its chargepoint. According EV batteries, both unidirectional and bidirectional charging can be considered. To account for inefficiencies in charging and discharging a battery charge η_b^{charge} and discharge $\eta_b^{discharge}$ efficiency can be specified. To cater for EV charging needs and especially the possible shorter period of it being plugged in than the optimization time period may be, a start and end time of the battery can be specified when charging should start or when charging should be finished. A terminal state-of-charge (SoC) can be set as well, which is used in combination with the charge end time to specify what SoC the battery should have at the end of the optimization.

The charging and discharging of batteries is realized with source and sink profiles like the purchase and sell profiles described above. The profiles created for the battery are limited by the maximum charge and discharge capabilities of the battery or chargepoint. To reflect the connection between the charging profile, the discharging profile, and the SoC of the battery, an additional restriction for a battery is added that calculates the SoC in every time step and is limited by the maximum and minimum energy a battery can hold. The SoC of a battery is the only aspect of the optimization problem that is dependent on previous decisions to charge or discharge a battery.

$$E_{b,t+1} = E_{b,t} + e_{b,t}^{charge} * \eta_b^{charge} - e_{b,t}^{discharge} * \frac{1}{\eta_b^{discharge}}$$

At the end of a time step, the amount of energy charged, multiplied by the efficiency, subtracted by the energy discharged, multiplied by its efficiency, amounts to the change in SoC compared to the previous time step. Because charging and discharging efficiencies are not 100 %, it is not cost-effective for the optimizer to charge and discharge the battery at the same time. In a scenario where the battery would be both charged and discharged at the same time it would be more cost effective to use the energy used for charging directly to feed the device that is being fed from that battery discharge instead of charging the battery first and discharging it afterwards. If both charge and discharge efficiencies are 100 %, binary variables are automatically added to the battery, enforcing it to be either charged or discharged. If efficiencies are lower, omitting this constraint reduces the complexity of the model considerably as the binary constraint turns the linear optimization problem (LP) into a mixed integer linear problem (MILP), which significantly increases the solving time.

Since EV usually do not charge below a certain minimum power, the battery model allows to set a minimum charge and discharge power. If the EV is charging, it has to charge with at least the minimum power. When using this constraint, binary variables are needed for the decision to charge or discharge a battery, making it a MILP as well. To sum up, if there is no minimum power required and battery efficiencies are below 100 %, the problem is linear, otherwise it is a MILP.

Table 1. Example Energy matrix for a single time step (e. g. 5 min). All values are in Wh

	grid feed-in	home battery charge	EV battery charge	fixed consumption
grid power	0	0	600	0
PV	0	804	300	7
home battery discharge	0	0	0	0

3.1.7 Energy Matrix. In each time step the optimizer calculates a matrix of energy flows between all sources and sinks. This ensures that in every time step, the energy consumed by the sinks is equal to the energy provided by the sources.

Each source calculates the energy drawn from it by adding up all energies it provides to the available sinks and the total energy consumption of a sink is calculated as the sum of energies it gets from the energy sources.

Table 1 shows the energy matrix M for an exemplary household for a single time step. $m_{i,j,t} \in M$ represents a single entry in the matrix where i represents sources, j represents sinks, and t is the time step. In combination with the formulas from subsubsection 3.1.4 and subsubsection 3.1.5 this matrix ensures that all sinks and sources stay within their set limits and that the energy usage of a household is balanced.

3.1.8 Technical implementation. The optimizer is implemented in Python and relies heavily on the Python optimization modeling language Pyomo [Hart 2017] [Bynum et al. 2021]. The optimization itself is done by default using the GNU Linear Programming Kit solver GLPK[3]. GLPK primarily uses the primal-dual interior point method for non-integer problems, but other, possibly more performant commercial solvers such as Gurobi or CPLEX could be used as well, especially for the MILP. The source code for the optimizer is provided with this paper under an open-source license.

4 EVALUATION

The effects of the defined PPP and PS are evaluated in two examples, considering a single building in a winter week and 74 prosumer households in a whole year. All buildings are single-family households, so we do not differentiate between households and buildings in this paper. In defining the examples and the technical equipment of the buildings, our aim is not to create representative compositions of buildings, but to show how the incentive signals affect the purchase and feed-in power of buildings in two realistic examples. We define five different scenarios, which differ only in terms of grid and market-oriented incentive signals. The simulation is conducted in 15-minute steps assuming perfect foresight. Since input data for PV generation is only available in a resolution of 15 minutes, simulating with a higher resolution would not lead to more insights.

[3]https://www.gnu.org/software/glpk/

4.1 Example 1: Single building in a winter week

The first example refers to a single building in a December week (December 8-15, 2023). The time period was chosen because of high fluctuation of the residual load and low local PV power generation. The goal is to understand the effects of the respective grid- and market-oriented signals during high load. The building is equipped with a 10 kWp rooftop PV, an 10 kWh battery storage with 8 kW charge/discharge power and 98 % charge/discharge efficiency, and a chargepoint with 11 kW power. It is assumed that in this week, the EV is plugged in three times at 7 PM and plugged out the next morning at 7 AM, while 20 kWh of energy need to be charged each time. The other data sources and reasons for the chosen technical data are described in Example 2.

The selected week has high fluctuations of the residual load and therefore the market price, as visualized in Fig. 7. During that time, rooftop PV generation is limited, making external price signals relevant to BEMS. It is assumed that between Dec 9, 21:00 and Dec 10, 7:00, there is a local grid congestion caused by herding effects (blue area in Fig. 7).

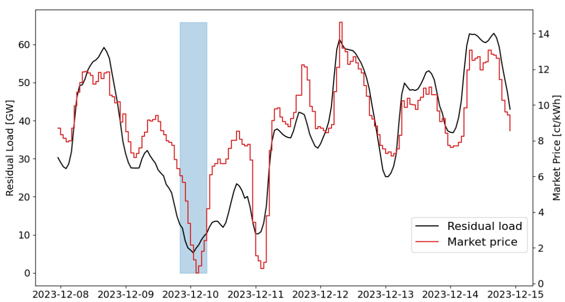


Fig. 7. Residual load and market price in investigated week of Example 1 [Bundesnetzagentur 2024]

4.2 Example 2: 74 households

In the second example, we model a district with 74 households for a period of the whole year 2023 of which 80 % have a PV plant and a battery. The simulation is conducted for every month assuming perfect foresight, which means that there is no rolling horizon with prediction errors.

4.2.1 Inflexible demand of households. For the inflexible demand of households, we use representative electricity demand profiles published by Tjaden et al. [Tjaden et al. 2015]. The average annual electricity demand of those households is 4683 kWh. We downsample the active power timeseries to a 15-minute resolution.

4.2.2 PV Generation. We use renewables.ninja [Pfenninger and Staffell 2016] to generate data for PV power output for the investigated time period. As a location for weather data, we choose the town Walldorf in southwest Germany. We assume that the maximum power of the PV plant is normally distributed around 6.9 kWp, based on the forecast of the responsible DSO of the considered location. Furthermore, the tilt is assumed to be normally distributed around 35 degrees, and the azimuth is normally distributed around 180 degrees (south). Those parameters are used as input parameters for the renewables.ninja API.

4.2.3 Batteries. We assume that every household with a PV plant is equipped with a battery. Its capacity is usually selected based on the size of the PV plant. As a sensible sizing for today's usual maximization of self-sufficiency is one kWh battery capacity for 1 kWp PV [Weniger and Quaschning 2013], we assume that the capacity of each battery is equal to $P_{max,PV} * 1h$. Both charge and discharge efficiencies are 98 %, resulting in a round-trip efficiency of around 96 % (comp. [BYD 2024]). Every battery has a C-rate of 0.8.

4.3 Scenarios

The scenarios only differ with respect to external signals from grid and market as described in the following. It should be noted that all of the grid signal parameters in Tab. 2 must be seen as example values for the given example and have to be adapted to the respective distribution grid.

4.3.1 A - Self consumption. In the scenario that is common today, there are no external power limits defined for purchasing and feeding in electricity. Certainly, the power limit is always the main fuse, which is assumed to be 30 kW per household. In this scenario, all prices are constant, thus providing no incentives other than self-consumption maximization. The battery is not charged from the grid since there is no financial incentive to do so. For feed-in, the remuneration of 8.2 ct/kWh is applied, based on the German Renewable Energies Act ("Erneuerbare-Energien-Gesetz", EEG). For purchasing electricity, a constant price that is composed of the monthly average of the EPEX Day Ahead price and taxes and surcharges of 17.7825 ct/kWh (valid for 2023 in Karlsruhe) is assumed.

4.3.2 B - Market orientation. Just like in Scenario A, there are no external power limits in the market oriented scenario. The price for feeding in electricity is the EEG remuneration of 8.2 ct/kWh. The price for purchasing electricity is the EPEX day ahead spot market price plus 17.7825 ct/kWh for taxes and surcharges, including grid fees.

4.3.3 C - Grid orientation. In order to define the grid oriented scenario, the same market prices as in Scenario A are used, which are constant for every month. The grid fee is reduced by 60 % as compensation for grid limitations, following the determination for § 14a EnWG [Bundesnetzagentur 2023b]. Therefore, taxes, grid fees and surcharges sum up to 12.2985 ct/kWh. Additionally, a power limit is generated based on the grid load (equal to the total power of all households at the GCP) in Scenario A. When the sum of the GCP power of all households $P_{exp,grid}$ exceeds the threshold $P_{thres,grid}$ in Scenario A, an additional grid fee is introduced for exceeding a certain power at the GCP during this time, which can be understood as a load quota w.r.t the GCP. The threshold $P_{thres,grid}$ can be seen as the load at the district's transformer. If the grid load is below this threshold, there are no grid signals. The quota at the grid connection point is $s_F \cdot \frac{P_{thres,grid}}{P_{exp,grid}}$ with a steepness factor s_F for purchase and feed in. The steepness factor defines how fast the quota declines from 1 to the minimum quota. An s_F of 1 means a slight decrease

Table 2. Grid signal parameters in Example 2

Parameter	Parameter description	Used value
$P_{thres,grid}$	Grid load threshold	50 kW
s_F	Steepness factor purchase	0.1
$P_{min,p}$	Minimum guaranteed power	2.4 kW
$q_{purchase,medium}$	Medium purchase quota	0.08
q_{feedin}	Feed-in quota (PV)	0.7
$c_{penalty}$	Penalty price (purchase)	10 ct/kWh
$f_{feed-in-red}$	Feed-in reduction factor	0.1
$c_{dyn,grid}$	Dynamic grid fee delta	5 ct/kWh

while an s_F of 0 leads directly to the minimum quota at the smallest exceedance of the threshold. This approach is based on the "operational simultaneity factor" as proposed by VDE FNN [FNN 2024b] in order to generate individual power limits based on the grid load. The quota refers to the maximum GCP power of 30 kW. At least $P_{min,p}$ purchase power is always guaranteed without additional grid fees. In the Example 1, the GCP power is limited to 5 kW during the assumed congestion period. For exceeding the purchase threshold, an additional grid fee (penalty price) $c_{penalty}$ is due for the power above the thresholds.
In order to reduce feed-in peaks, the remuneration for feed-in power exceeding 70 % of the installed PV power during an expected bottleneck is reduced by $f_{feed-in-red}$. This is following the abolished static "70 % rule", but allows higher feed-in and only has an effect during grid congestion.

4.3.4 D - Grid- and market orientation (power-dependent grid fees). Scenario D is the combination of Scenarios B and C. The market prices are equal to those in Scenario B while a limitation applies with the same rules as in Scenario C. Since we expect grid load peaks caused by market prices, the power limits are defined based on the load in Scenario B instead of Scenario A. Additionally, the pattern of the market price is analyzed: If the market price has a minimum with a prominence of more than the standard deviation, it is assumed that this price minimum can cause load peaks, even if they did not occur in Scenario B. During these times, a medium purchase quota $q_{purchase,medium}$ is sent to all households.

4.3.5 E - Grid- and market orientation (volumetric grid fees). In order to compare PS with the commonly discussed dynamic grid fees, this scenario has no power limits but a combination of dynamic prices (like Scenario B) and dynamic grid fees that are added to the market signal. Without expected bottlenecks, the grid fee is unchanged. During expected load-side bottlenecks, the grid fee rises by $c_{dyn,grid}$ and during feed-in bottlenecks, the grid fee is reduced by $c_{dyn,grid}$. The EEG feed-in remuneration remains unchanged, since there are usually no grid fees for generation units. Equally to Scenario D, the grid load of Scenario B is used to identify time periods of expected congestion. In Example 1, the grid fee delta is 20 ct/kWh.

4.4 Assumptions

We conduct the simulations assuming perfect foresight for both the households and the DSO and the supplier. We further assume that batteries are the only flexible devices while the other electricity consumption remains unaffected, making no change of consumer behavior necessary. Of course, the electricity demand of EV and HP could be optimized as well, but for the purpose of investigating the effects of external signals, the consideration of batteries is sufficient.

5 RESULTS

5.1 Example 1: Single building in a winter week

When looking at the congestion period in the different scenarios in Fig 8, it can be seen that the power at the GCP of the single building is different in every scenario. The power limit (valid in Scenarios C and D) during the congestion period is marked by the blue area and the market price is visualized by the black dotted line.
Without any external incentives **(Scenario A, blue)**, the lowest prices are not used by the building. Due to the flat tariff, the BEMS has neither information nor incentives to do so. Instead, the power arbitrarily rises during the congestion, since no grid information is provided to the BEMS.
Pure market prices **(Scenario B, orange)** lead to the exploitation of the lowest market prices and a peak during the grid congestion. Power-dependent prices **(Scenario C, green)** reliably limit the GCP power during the congestion. However, the lowest prices are not exploited by the BEMS.
Wen combining dynamic market prices with power-dependent grid prices **(Scenario D, red)**, it can be seen that the building uses energy during the cheapest prices but only up to the allowed power limit. Therefore, the combination of those signals leads to grid- and market orientation of BEMS.
Dynamic grid fees **(Scenario E, purple)** override the market prices. Another peak occurs outside the expected congestion period. The low energy market prices are not used at all and if all BEMS get the same dynamic grid fee, another congestion may be caused just before the expected congestion. Accordingly, adding volumetric dynamic grid fees to dynamic market prices without further coordination mechanisms is not necessarily targeted and can even be counterproductive.

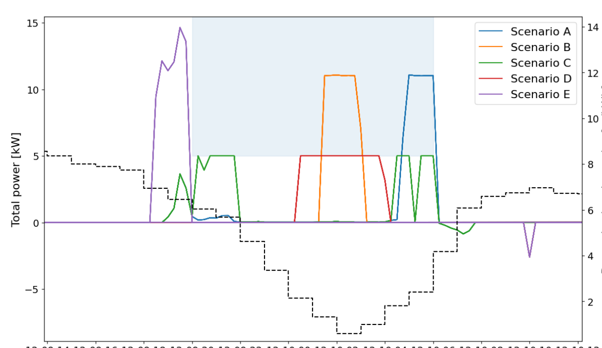


Fig. 8. GCP power in different scenarios (Example 1)

Table 3. Simulation results for Example 2

Scenario	A (baseline)	B (dyn. prices)	C (power limits)	D (PS)	E (dyn. grid fees)
Highest load [kW]	105.4	262.7	95.7	175.9	267.0
Highest feed-in [kW]	-197.4	-197.4	-185.3	-190.5	-201.7
Avg. costs per household [€]	806.4	799.1	647.3	802.7	795.1
Avg. revenue per household [€]	234.4	234.5	211.0	211.0	234.5
Avg. economic result per household [€]	-571.9	-564.7	-436.3	-591.7	-560.6
Total power during highest 5 % residual load [kW]	31.3	22.9	31.3	23.0	24.1
Total power during highest 1 % residual load [kW]	29.8	20.9	30.2	21.1	23.7
Total power during highest 0.1 % residual load [kW]	38.1	22.5	36.7	22.5	22.5

5.2 Example 2 - 74 prosumer households

Tab. 3 shows the highest load and feed in of the sum of all simulated households at their grid connection point over the whole year. This can be interpreted as the load on the district's transformer if no additional loads or generators are connected to it. It can be seen that in all scenarios with dynamic market prices (B, D, E), the load rises significantly, which is caused by the batteries charging from the grid during low prices.

The average total costs for purchasing electricity from and revenue for feeding in electricity to the grid under the respective price signals are calculated. The economic results are calculated by revenues minus costs. It has to be noted that those economic results highly depend on the used parameterization of grid signals in Tab. 2, the spot market prices in the investigated time period, and the technical equipment and energy demand of the considered buildings. Therefore, the numbers cannot be generalized.

The time periods when the residual load in the market zone (Germany/Luxembourg) [Bundesnetzagentur 2024] is highest are taken into account specifically. As these times define the demand from peak-load power plants, it is important that the grid consumption of households during these times is as low as possible or ideally negative. For the numbers in Tab. 3, we first summed up the power of all households at the grid connection point. In a second step, we calculated the average of this total power during the time steps of the highest residual load. It can be seen that dynamic price optimization in Scenarios B, D, and E lowers the total load when residual load is highest. In another publication [Riedel et al. 2023], we showed that dynamic prices for purchase and feed-in can also lead to negative power during those times.

6 DISCUSSION

The simulation results show that it is possible to provide both grid- and market oriented incentives for BEMS by using power-dependent price profiles. Dynamic prices incentivize the minimization of the residual load, while power limits prevent peak loads in distribution grids. In Scenario E it can be seen that volumetric dynamic grid fees do not necessarily avoid high feed-in or purchase peaks. They can shift the load away from the expected congestion, but this completely overrides the market signal, even if a load increase was favorable up to a certain extent due to a low residual load. Grid bottlenecks caused by herding effects are difficult to avoid with volumetric dynamic grid fees. The addition of dynamic grid fees to dynamic prices could even have counterproductive effects if a new load peak is caused outside the expected grid congestion period.

In Tab. 3, it can be seen that economic results for the households are more attractive in the baseline scenario (Scenario A) than in Scenario D. Since this result highly depends on grid fee reductions, height of remuneration, and the height of the additional (penalty) grid fee, it is not only a technical question but also a political decision about the extent to which prosumer households should shoulder a higher share of the grid costs or be rewarded for serving the energy system.

7 OUTLOOK

The trade-off between grid- and market orientation might be more effective by using time- and power-dependent grid fees than by using volumetric grid fees. However, the parameterization of those price signals is crucial. The results in Tab. 3 highly depend on the parameters in Tab. 2. Therefore, these parameters have to be defined in a feasible way for the corresponding grid and buildings. We used a very simple and pragmatic way to define quotas in Example 2. In order to find reasonable grid- and market-oriented signals, more sophisticated methods to parameterize PS with the use of predictions for the grid load is subject to further research.

ACKNOWLEDGMENTS

This paper is based on work in the project *SynergieQuartier (FKZ 03EI6031A)*, funded by the German Federal Ministry of Economic Affairs and Climate Action.

REFERENCES

50Hertz Transmission GmbH, Amprion GmbH, TransnetBW GmbH, TenneT TSO GmbH. 2024. Netztransparenz. https://www.netztransparenz.de/

Agora Energiewende and Forschungsstelle für Energiewirtschaft e. V. 2023. Haushaltsnahe Flexibilitäten nutzen. Wie Elektrofahrzeuge, Wärmepumpen und Co. die Stromkosten für alle senken können. https://www.agora-energiewende.de/fileadmin/Projekte/2023/2023-14_DE_Flex_heben/A-EW_315_Flex_heben_WEB.pdf

Adil Amin, Wajahat Ullah Khan Tareen, Muhammad Usman, Haider Ali, Inam Bari, Ben Horan, Saad Mekhilef, Muhammad Asif, Saeed Ahmed, and Anzar Mahmood. 2020. A Review of Optimal Charging Strategy for Electric Vehicles under Dynamic Pricing Schemes in the Distribution Charging Network. *Sustainability* 12, 23 (Dec. 2020), 10160. https://doi.org/10.3390/su122310160

aWATTar. 2024. aWATTar - Energy in Sync with Nature. https://www.awattar.at/

Bundesnetzagentur. 2023a. Beschluss BK6-22-300. https://www.bundesnetzagentur.de/DE/Beschlusskammern/1_GZ/BK6-GZ/2022/BK6-22-300/BK6-22-300_Beschluss.html?nn=877500

Bundesnetzagentur. 2023b. BK8-22/010-A. https://www.bundesnetzagentur.de/DE/Beschlusskammern/1_GZ/BK8-GZ/2022/2022_4-Steller/BK8-22-0010/BK8-22-0010-A_Festlegung_Download.pdf?__blob=publicationFile&v=5

Bundesnetzagentur. 2024. SMARD - Strommarktdaten, Stromhandel und Erzeugung in Deutschland. https://www.smard.de/home

BYD. 2024. BYD Battery Box Premium Datasheet HVS /HVM. https://bydbatterybox.com/uploads/downloads/BYD%20Battery-Box%20Premium_Datasheet_HV-AU%20V1.2%20EN-5eec6422498ad.pdf

Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. 2021. *Pyomo — Optimization Modeling in Python*. Springer Optimization and Its Applications, Vol. 67. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-68928-5

EEBus. 2022. EEBUS Use Case Specification - Incentive-Table based power consumption management. https://www.eebus.org/specifications/download-specifications/

Anselm Eicke, Lion Hirth, and Jonathan Mühlenpfordt. 2024. Mehrwert dezentraler Flexibilität. neon.energy/mehrwert-flex

Carmen Exner, Marc-Aurel Frankenbach, Alix Von Haken, Ariane Höck, and Martin Konermann. 2020. Practical implementation of the management of local flexible generation and consumption units using a quota-based grid traffic light approach. *CIRED - Open Access Proceedings Journal* 2020, 1 (Jan. 2020), 432–435. https://doi.org/10.1049/oap-cired.2021.0082

FfE. 2024. Der unIT-e² KOALA - Ein anreizbasierter Mechanismus zur Koordination netzorientierter Steuerungsvorgänge. https://www.ffe.de/veroeffentlichungen/ein-anreizbasierter-mechanismus-zur-koordination-netzorientierter-steuerungsvorgaenge-der-unit-e%c2%b2-koala/

VDE FNN. 2024a. Ausprägung der digitalen Schnittstelle an steuerbaren Einrichtungen oder an einem Energie-Management-System. https://www.vde.com/resource/blob/2292786/5d38accc5ab02ad04df7cab8a64f1f63/impuls--digitale-schnittstelle-data.pdf

VDE FNN. 2024b. Netzbetrieb mit Flexibilitäten: Umgang mit der kurativen Steuerung über iMSys und Ausblick auf mögliche vorrausschauende Steuerungsmaßnahmen. https://www.vde.com/de/fnn/aktuelles/netzorientierte-steuerung-richtig-umsetzen

S. Gansemer and U. Grossmann. 2011. Analysis on variable electricity pricing models and the influence on load curves of household customers. In *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, Vol. 2. IEEE, Prague, 532–535. https://doi.org/10.1109/IDAACS.2011.6072823

Sebastian Gottwalt. 2015. *Managing Flexible Loads in Residential Areas.* Ph. D. Dissertation. Karlsruher Institut für Technologie. http://digbib.ubka.uni-karlsruhe.de/volltexte/1000048803

William E. Hart. 2017. *Pyomo – optimization modeling in Python.* Springer Science+Business Media, New York, NY.

Lutz Hillemacher, Kai Hufendiek, Valentin Bertsch, Holger Wiechmann, Jan Gratenau, Patrick Jochem, and Wolf Fichtner. 2013. Ein Rollenmodell zur Einbindung der Endkunden in eine smarte Energiewelt. *Zeitschrift für Energiewirtschaft* 37, 3 (Sept. 2013), 195–210. https://doi.org/10.1007/s12398-013-0110-z

Christina Hollmann, Andreas Schwackenberg, and Annike Abromeit. 2024. EEBus Solutions Whitepaper. https://www.eebus.org/wp-content/uploads/2024/01/EEBUS-Whitepaper-Exploring-EEBUS-Solutions.pdf

Ingo Mauser and Hartmut Schmeck. 2014. Tarife zur Flexibilisierung des Stromverbrauchs in Haushalten mit Energiemanagementsystemen Tariffs for the Flexibilization of Energy Consumption in Households with Energy Management Systems. VDE Verlag, Frankfurt am Main.

Christian Möller, Kevin Kotthaus, Markus Zdrallek, and Fritz Schweiger. 2021. Comparison of Incentive Models for Grid-supporting Flexibility Usage of Private Charging Infrastructure. In *Das Gesamtsystem im Fokus der Energiewende 18. – 19. Mai 2021*. VDE Verlag GmbH, Wuppertal, 6.

Jan Müller, Mischa Ahrens, Ingo Mauser, and Hartmut Schmeck. 2018. Achieving Optimized Decisions on Battery Operating Strategies in Smart Buildings. In *Applications of Evolutionary Computation*, Kevin Sim and Paul Kaufmann (Eds.). Vol. 10784. Springer International Publishing, Cham, 205–221. https://doi.org/10.1007/978-3-319-77538-8_15 Series Title: Lecture Notes in Computer Science.

OpenADR. 2023a. OpenADR. openadr.org

OpenADR. 2023b. OpenADR 3.0 OpenADR 3.0.1 User Guide (non-normative). https://openadr.memberclicks.net/specification

Stefan Pfenninger and Iain Staffell. 2016. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy* 114 (Nov. 2016), 1251–1265. https://doi.org/10.1016/j.energy.2016.08.060

Tobias Riedel, Carl Hauschke, and Hartmut Schmeck. 2023. Dynamic feed-in from prosumer households with batteries considering spot market prices. In *ETG Congress 2023*. VDE, Kassel.

Judith Stute and Marian Klobasa. 2024. How do dynamic electricity tariffs and different grid charge designs interact? - Implications for residential consumers and grid reinforcement requirements. *Energy Policy* 189 (June 2024), 114062. https://doi.org/10.1016/j.enpol.2024.114062

Tjarko Tjaden, Joseph Bergner, Johannes Weniger, and Volker Quaschning. 2015. Repräsentative elektrische Lastprofile für Wohngebäude in Deutschland auf 1-sekündiger Datenbasis. https://solar.htw-berlin.de/elektrische-lastprofile-fuer-wohngebaeude/

Michael von Bonin, Elias Dörre, Hadi Al-Khzouz, Martin Braun, and Xian Zhou. 2022. Impact of Dynamic Electricity Tariff and Home PV System Incentives on Electric Vehicle Charging Behavior: Study on Potential Grid Implications and Economic Effects for Households. *Energies* 15, 3 (Feb. 2022), 1079. https://doi.org/10.3390/en15031079

Natapon Wanapinit, Jessica Thomsen, and Anke Weidlich. 2022. Find the balance: How do electricity tariffs incentivize different system services from demand response? *Sustainable Energy, Grids and Networks* 32 (Dec. 2022), 100948. https://doi.org/10.1016/j.segan.2022.100948

Johannes Weniger and Volker Quaschning. 2013. Begrenzung der Einspeiseleistung von netzgekoppelten Photovoltaiksystemen mit Batteriespeichern. In *28. Symposium Photovoltaische Solarenergie*. Bad Staffelstein. https://solar.htw-berlin.de/publikationen/begrenzung-einspeiseleistung-pv-systeme-batteriespeicher/

Received 12 May 2024; revised 16 August 2024; accepted 14 June 2024

Monitoring Germany's Core Energy System Dataset: A Data Quality Analysis of the Marktstammdatenregister

FLORIAN KOTTHOFF, OFFIS Institute for Information Technology, Germany and fortiss, Germany
CHRISTOPH MUSCHNER, Reiner Lemoine Institut gGmbH, Germany
DENIZ TEPE, fortiss, Research Institute of the Free State of Bavaria, Germany
ESTHER VOGT, Institute for Enterprise Systems, University of Mannheim, Germany
LUDWIG HÜLK, Reiner Lemoine Institut gGmbH, Germany

The energy system in Germany consists of a large number of distributed facilities, including millions of PV plants, wind turbines, and biomass plants. To understand and manage this system efficiently, accurate and reliable information about all facilities is essential. In Germany, the Marktstammdatenregister (MaStR) serves as a central registry for units of the energy system. The reliability of this data is critical for the registry's usefulness, but few validation studies have been published.

In this work we provide a review of existing literature that relies on data from the MaStR and thereby show the registry's importance. We then build a data and testing pipeline for relevant data of the registry, with a focus on the two aspects of facility's location and size. All test results are published online in a reproducible workflow. Hence, this work contributes to a reliable data foundation for the German energy system and starts an open validation process of the Marktstammdatenregister from an academic perspective.

CCS Concepts: • **Applied computing** → *Engineering*; • **General and reference** → **Evaluation**; **Validation**.

Additional Key Words and Phrases: Open Data, Data Validation, Data Quality Assessment, Renewable Energies, Photovoltaic systems, Wind power plants, Energy system analysis

Availability of Data and Material:

The examined dataset version can be found on zenodo [39]. Downloading and Preprocessing was done with the python package open-mastr [38]. The data pipeline and dashboard code can be found in our github repository [51].

1 INTRODUCTION

The current energy system is undergoing an important transformation, with a shift from a few large and mainly fossil power plants to an increasing number of small renewable power units. The wind and solar technologies, in particular, experience rapid growth rates, with 35.000 wind turbines and 3.9 million photovoltaic (PV) systems producing electricity in Germany as of early 2024 [10]. To understand and control this diverse and complex energy system, detailed and reliable data of all its components is essential. Such an accessible and up-to-date data source is a key enabler for the digitalization within the energy system [34].

For this reason, the German legislator has created a large and central power plant registry, the *Marktstammdatenregister*, or short, MaStR. It was released in 2019, is updated on a daily basis, and contains more than 22 million entries about electricity and gas producers, electricity and gas consumers, storages, grids, and actors from the energy market. The data is provided by the owners of the power plants themselves. In this sense, creating this dataset was very expensive and time-consuming, as millions of owners had to be contacted, needed to register on a website and enter their information. Validation and reliability assessment are of great importance, as the data is intended to play a central role in controlling the future German energy system and is based on the contribution and information of millions of individuals.

Considering the registry's increasing importance, the quality of the dataset should be monitored and discussed. In this paper, we start this work by building a data pipeline together with automated data tests for the MaStR. Since we do not have the rights to change the content of the dataset, we focus on finding and visualizing data errors that appear in the MaStR. By this, we want to answer the research question of *how reliable is the data of the Marktstammdatenregister*, with a special focus on the units' location and size.

The core contributions of this work are the following: We are the first work that reviews the usage of the MaStR dataset in research to identify its importance and we contribute with a first step to the urgently needed validation of this central German energy system dataset. Our code, together with all test results and visualization dashboards, are published openly and can be utilized by other MaStR users.

This paper is structured as follows: In section 2, we first describe the MaStR dataset. Then, in section 3, we review and categorize the usage of the MaStR dataset in existing literature. In section 4, we describe the data pipeline and the data tests. In section 5, we analyze the test results, discuss them in section 6 and draw our conclusion in section 7.

2 THE DATASET - MARKTSTAMMDATENREGISTER

The MaStR is operated by the German Federal Network Agency (ge: Bundesnetzagentur, short: BNetzA) since January 31, 2019 as a central online database of the German energy system. Owners of electricity or gas generating units are obliged to report master data on themselves and their units. Additionally, units consuming large amounts of electricity have to be registered if they are connected to at least a high-voltage electricity grid. Considering the type of information collected, the MaStR only includes master data like installed capacity or location information. It does not provide transaction data like actual production amounts or storage levels [11]. Following the introduction of the MaStR, all information had to be provided manually by plant owners using a registration website.

Authors' addresses: Florian Kotthoff, florian.kotthoff@offis.de, OFFIS Institute for Information Technology, Escherweg 2, Oldenburg, Lower Saxony, Germany, 26121 and fortiss, Guerickestraße 25, Munich, Bavaria, Germany, 80805; Christoph Muschner, Reiner Lemoine Institut gGmbH, Rudower Chaussee 12, Berlin, Berlin, Germany, 12489; Deniz Tepe, fortiss, Research Institute of the Free State of Bavaria, Guerickestraße 25, Munich, Bavaria, Germany, 80805; Esther Vogt, Institute for Enterprise Systems, University of Mannheim, Mannheim, Baden-Württemberg, Germany, 68131; Ludwig Hülk, Reiner Lemoine Institut gGmbH, Rudower Chaussee 12, Berlin, Berlin, Germany, 12489.

A transition period of two years until 01/31/2021 was granted for existing plants, but plants which were commissioned since then have to be registered within one month after commissioning [12].

Most information on units is openly accessible. The data is published under an open data license, the Data licence Germany – attribution – version 2.0 (DL-DE-BY-2.0) and can be downloaded, used and republished with no restriction if proper attribution to the Bundesnetzagentur is given. For units with a net capacity of up to 30 kW, some location information is restricted from publication. This applies to street name, house number, parcel designation and exact coordinates of units. The most granular location information accessible for all units is the postal code or the municipality. The MaStR dataset referenced in this work was downloaded at 2024-03-12 from the official website of the BNetzA using the *open-mastr* python package [38]. We have developed and published this package to provide a simplified and automated download, processing and storage in a local database. Additionally, we published the MaStR dataset from this date on zenodo [39].

Regarding the validation process of the data, information on units, plants and corresponding owners has to be verified by the Distribution System Operator (DSO). In [8], it is specified that the capacity and address, but not the coordinates of a power unit have to be validated by the DSO. Independent of the mandatory validation by the DSO, the BNetzA runs checks on the dataset as well. According to [12], they recently focused on dimensioning errors of capacities and location information of wind turbines. For units with a net capacity of more than 10 MW, the BNetzA checks all reported information. Obvious errors are fixed immediately and reported to the plant owners who may object. On average, the BNetzA fixes about 600 errors per month for the whole registry.

Besides this work, several other publications validated the MaStR dataset. Manske et al. developed a pipeline to extend and validate the entries for PV systems, wind turbines, biomass plants, and hydropower plants in the MaStR [67]. Mayer et al. applied a convolutional neural network (CNN) model to recognize PV systems from aerial images and provide an alternative PV dataset for one German state [70]. Expanding on this, they took 3D building data into consideration and focused on the orientation and tilt angle of rooftop PV systems. By comparing the MaStR to their own dataset, they pointed out the discrepancies between both datasets [69]. Schulz et al. mentioned the deficits in the MaStR and used a similar object recognition framework for PV, wind, biomass systems as well as solar thermal units. They aimed to supplement the official dataset and demonstrated the significance of unregistered units as case studies in two districts [83].

3 LITERATURE REVIEW

3.1 Relevance of the Marktstammdatenregister in Research

To understand the current relevance of the MaStR, we conducted a review of existing literature. Therefore, we used the search query "Marktstammdatenregister" and searched within the science databases of *Scopus*, *Web of Science*, *arxiv*, *IEEExplore*, *MDPI*, *SpringerLink*, *ScienceDirect*, and *GoogleScholar*. The result was a total of 532 documents by 1st April 2024. After removing duplicates and documents that were not peer reviewed or published within a scientific journal, 99 papers remained that contain the word "Markstammdatenregister". From these 99 works, a total of 83 works actually used the dataset and not only referred to it. These works, that build upon the MaStR dataset, are now described and categorized in further detail.

From the 83 papers that use data from the MaStR, 48 papers are Open-Access publications. Since it is a relatively new dataset, the first publications using the dataset appeared in 2019, followed by an increasing number of publications in the proceeding years. To get an idea which scientific fields use the MaStR, we assigned one research domain to each paper: *Sustainability Studies* if the paper focused on environmental impact, *Energy Politics* if the paper focused on energy politics and policy recommendations, *Energy Data* if the paper focused on creating or evaluating datasets, *Energy Economics* if the paper focused on economical aspects, and *Energy System Analysis* if the paper focused on modelling and analysing the energy system. In Fig. 1a we see that the dataset was most frequently used in the domains *Energy System Analysis* and *Energy Economics*. In Table 1 we list all identified research domains together with the research topics that utilize data from the MaStR dataset.

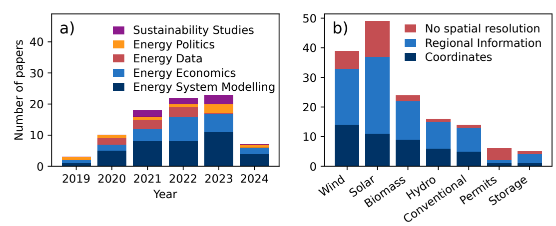


Fig. 1. Importance of MaStR in literature: In a), the number of published papers for the five identified research domains is plotted over the publication year (up to 1st April 2024). In b), the number of papers that use different tables from the MaStR is plotted. The papers are further subdivided according to the used location information, where *regional information* represents the use of aggregated data on zip-code, district, or state level.

Since one contribution of our work is a validation of power unit locations within the MaStR, we also highlight the usage of location information in the literature. Therefore, each paper is categorized in Fig. 1b according to one of the three categories: *No location information* if the location of power units was not used, *Regional information* if either the state, the NUTS-region, or the municipality of a power unit was used, and *Coordinates* if the actual longitude and latitude coordinates of a power unit were used within the publication. As a result, from the total of 83 papers we identified 24 papers which do not use location information, 33 papers which use regional information, and 26 papers which use coordinates.

As a next step, we want to evaluate which MaStR tables are most frequently studied by researchers when using the dataset. Therefore, we identify the tables used in each publication. In Fig. 1b one can see that the two technologies of solar and wind receive the largest attention, followed by biomass, hydropower plants, Combined Heat and Power (CHP) plants and conventional plants (coal, gas, oil, and nuclear) [1]. Both the tables on storages and permits are used

[1] In contrast to further categorizations, a paper can have multiple tables assigned.

Table 1. Description of the different domains and research questions where the MaStR dataset is used.

Research Domain	Research Topic
Sustainability Studies	- Land use of renewable energies [13, 18, 30, 55] - Climate gas emissions due to power plants [23] - Waste quantification of decommissioned renewables [95]
Energy Politics	- Local allocation of energy infrastructure [5, 16, 26, 65, 80, 84] - Effects of *Prosuming* [76, 97]
Energy Data	- Identifying renewable energy systems from aerial images [45, 67, 69, 70, 79, 83] - Building time series datasets for energy system modelling [6, 40, 77] - Developing ontologies for the energy domain [7] - Building datasets about energy cooperatives [100]
Energy Economics	- Evaluating auction designs for renewable energies [48, 53, 61, 62, 68, 105] - Evaluating costs of renewable energies [52] - Profitability and business models in the energy domain [21, 63, 73, 89, 101, 102] - Adoption of renewable energy technology [1, 25, 28, 29, 41, 64] - Local energy markets [47, 66]
Energy System Analysis	- Flexibility options [20, 50, 82, 88] - Curtailed energy potentials [27, 86] - Repowering [33, 87] - Analysis of models and methods [22, 37, 44, 93, 94] - Analysis of Electricity, Gas and Hydrogen Grids [4, 14, 36, 46, 58, 71, 72, 74, 91, 92, 98, 99, 104] - Forecasting [59, 60, 81] - Potential analysis of renewable energies, storages, or electrolysis [3, 42, 43, 49, 54, 56, 57, 90, 96]

relatively rare, whereas permits are mainly used in the scope of energy economics to evaluate the performance of auctions or the average construction time of power units. For the four renewable technologies wind, solar, biomass and hydro, most works also use location information of the power units, either on a higher level (regional information) or the exact coordinates.

3.2 Data Testing

To evaluate data quality, a set of evaluation criteria and metrics is needed. These criteria are independent of the specific application domain. Previous works have defined different dimensions of data quality, such as availability, usability, reliability, relevance, and presentation quality [15]. The field of linked data is especially suitable to define generic data tests and data quality metrics, as showcased by the Review of Zaveri et al [106]. Within the Energy Domain, some publications exist that examine data quality. Gotzens et al. [32] analyze conventional power plant data from 28 european coutries and verify attributes like capacities, coordinates, or fuel type by comparing different data sources. Hirth, Mühlenpfordt, and Bulkeley [35] evaluated the completeness and consistency of several data items from the ENTSO-E Transparency Platform. This evaluation was based on user surveys, interviews, a literature review, and a statistical analysis. In the field of Life Cycle Assessment (LCA) in the Energy Domain, Astudillo et al. [2] reviewed key challanges and opportunities arising from data quality issues. Data streams and time series data is also crucial for energy system research. In this domain, Chen et al. [17] analyzed the quality of electricity consumption time series data, where they find noise, outlier, and incompleteness as the three main sources of error.

4 METHOD

We now describe the whole procedure of obtaining, processing and testing the data. The details of the data pipeline are presented in section 4.1, whereas the data tests are presented in section 4.2.

4.1 Data Pipeline

The data pipeline consists of multiple steps. First, the MaStR dataset is downloaded and parsed to a PostGIS database using the open-mastr python package [38]. Then, data of geoboundaries from districts and municipalities [24] is downloaded and saved to the same database. The geoboundaries are later used for testing unit locations. After all raw data is successfully loaded, it is transformed to a common data model using the software dbt [19]. The transformation consists of deleting unused information and renaming columns. After all tables are transformed to the same data model, the data tests are performed. Those tests are described in detail in Section 4.2. All units that do not pass the data tests are saved to a seperate sqlite database. The sqlite database is then published online using the framework datasette [103]. This publication consists of two parts: First, the whole database of units that fail one or more tests can be searched online. Subsets of the data can also be downloaded as CSV for further use. Second, dashboards are published that visualize the main error metrics we have defined to monitor the MaStR. The whole pipeline is shown in Fig. 2.

4.2 Data Testing

Assuring high data quality is a very general problem, appearing both in science and industry. Pipino et al. [78] define 16 dimensions of data quality. In the scope of monitoring the data quality

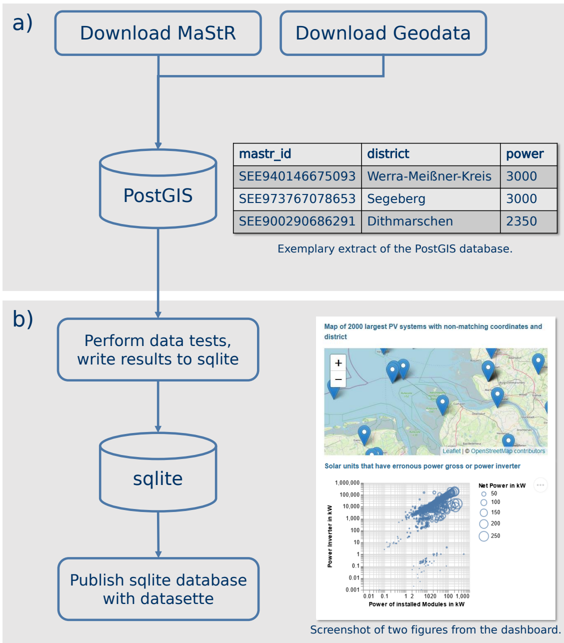


Fig. 2. Automated pipeline for downloading, processing, testing, and vizualizing the MaStR dataset. In the first step (a), the required raw data is downloaded and written to a PostGIS database. Afterwards in (b), the data transformation and testing is performed using the framework dbt. All units that fail at least one test are written to an sqlite database. The sqlite database together with monitoring dashboards is then published using the framework datasette.

of the MaStR dataset, two of those dimensions are of special interest: Completeness as "the extent to which data is not missing" and Free-of-Error as "the extent to which data is correct and reliable". Completeness can be defined as the fraction of entries that are not null and the total number of entries. The completeness of relevant parts of the MaStR is part of the the dataset description in the Appendix. Monitoring the "Free-of-Error" dimension however is complicated, as data correctness and reliability are closely coupled to the content of the data. To evaluate the "Free-of-Error" dimension, data unit testing comes into play. Data unit tests borrow the idea from software unit tests. They test small parts of the data to see if it behaves as expected. Several state-of-the-art tools implement data tests [19, 31]. The data pipelining framework dbt used in this work also integrates data testing.

Technically, each data test is represented by one SQL query. The query

SELECT * FROM wind WHERE mastr_id IS NULL;

represents the expectation that every MaStR unit has an ID. The test fails, if the query returns one or multiple rows.

In Tab. 2, all implemented tests are shown. The tests belong to one of four categories, where numbers in () denote the test IDs from Tab. 2:

- **Basic tests (ID 1-5):** For the MaStR dataset there are some baseline tests to ensure the data quality. We defined columns that are not allowed to have null values, checked that unit IDs are unique, that gross power and inverter power are larger then the net power, and that unit IDs, municipality IDs and zip codes match their respective regular expressions.
- **System size tests (ID 6-11):** The power of an electricity producing unit correlates with its system size. Hence we validated the rated power by comparing it with other quantities that represent system size. For storages and PV systems, the net power should be close to the power of the inverter. The power of PV systems should also correlate with the number of modules and, for ground-mounted PV systems, with the utilized area. For wind turbines, the rated power should correlate with the square of the rotor diameter. We also checked that rated powers are in acceptable range, ie. for biomass (0MW - 150MW), for combustion (0GW - 2GW), for hydro (0GW - 1.5GW), for solar (0MW - 500MW), for storages including pumped hydroelectric energy storages (0MW - 800MW), and for wind (0MW - 22MW) [75]. For wind turbines, upper boundaries for rated power can be established based on current technological specifications provided by manufacturers. In contrast, establishing upper boundaries for other technologies, such as PV farms and combustion plants, is more complex. The maximum size and capacity of these systems are highly dependent on site-specific planning and design considerations, which prevents the establishment of upper boundaries based solely on technological specifications. For those technologies we obtained upper boundaries from the largest currently installed plants in Germany. We manually checked that these plant capacities are correct and added a buffer. The upper boundaries of rated power values need to be adapted over time, as larger renewable power plants are built. Lastly, for PV systems we checked that balcony PV systems have a rated power $P < 1$kW.
- **Location tests (ID 12, 13):** For units with a power larger 30 kWp, exact coordinates as well as an address is given. We checked if the coordinates lie in the given districts and municipalities, where the geoboundaries were taken from [9]. Since geoboundaries do not have an infinite resolution, we added a buffer zone of 1.5km to each boundary, so that units that lie on the boundary (or a maximum of 1.5km outside the district / municipality) still pass the test.
- **Technology specific tests (ID 14, 15):** Some tests are specific to technologies and rely on domain knowledge. For each technology, we defined a range of years as accepted installation years. We chose 2030 as a maximum accepted year for planned installations. For wind, solar, and batteries we chose 1980 as earliest years. Since biomass (with solid biomass fuel), combustion, and hydro power plants are older technologies, we allowed earlier installation years. For wind turbines, we also checked that the hub height is larger than the rotor radius.

In total we have defined 90 data unit tests for the MaStR dataset.

Test ID	description	biomass	combustion	hydro	solar	storages	wind
1	Check for null values: unit ID, municipality ID, operating status, power	✓	✓	✓	✓	✓	✓
2	Check that unit IDs are unique	✓	✓	✓	✓	✓	✓
3	$P_{gross} \geq P_{net}$				✓	✓	
4	$P_{inverter} \geq P_{net}$				✓	✓	
5	Check that values match regex: unit ID, municipality ID, zip code	✓	✓	✓	✓	✓	✓
6	Check that P_{gross} / number modules is in accepted range				✓		
7	Check that P_{gross} / $P_{inverter}$ is in accepted range				✓	✓	
8	Check that P_{gross} / area is in accepted range (only ground-mounted PV)				✓		
9	Check that rated power matches with rotor diameter						✓
10	Check that balcony PV systems have a small installed capacity				✓		
11	Check that the installed power is reasonable	✓	✓	✓	✓	✓	✓
12	Check that the coordinates lie in the district	✓	✓	✓	✓	✓	✓
13	Check that the coordinates lie in the municipality	✓	✓	✓	✓	✓	✓
14	Check that the installation year is reasonable	✓	✓	✓	✓	✓	✓
15	Check that the hub height matches with rotor diameter						✓

Table 2. Table of the different data tests. The column *ID* is used to reference the tests in the text, *description* gives a short summary of the test. A ✓ in the technology columns denote that the test is applied for these technologies.

5 RESULTS

We ran the data pipeline and tests as described in Section 4 on the MaStR dataset obtained at 2024-03-12. All units that failed at least one test can be browsed online[2]. If the data is needed for further processing, it can be obtained from there as csv or json.

5.1 Basic Tests

As expected, most of the basic tests do not find errors. This shows that basic integrity of the MaStR dataset is granted. Basic expectations, like the existence of unique unit IDs are fulfilled. We only found two biomass units that do not have a municipality ID. Moreover, 31 hydropower units, roughly 30 solar and storage units that are not allocated at a municipality (and hence have no municipality ID and no zip code). We also found 6 storages where the inverter power was larger than the net power.

5.2 System size tests

The second error category is about system size. We compared values from the dataset that correlate with the system size and compared it to the rated power. In Fig. 3 a) we see that for 23,251 PV system, each represented by one dot in the scatter plot, the quotient of installed power of the PV modules and the power of the inverter does not lie in the allowed range of [1/20, 20]. This represents a share of roughly 0.6% of PV units. Especially the higher densities of points, where the units' power differ of a factor of 10^3 shows a typical error, where the physical quantities W, kW, and MW are mixed up. Since this factor of 20 is chosen rather arbitrary, we have plotted the share of failed tests depending on this factor in d). Here a unit fails the test if $1/a < P_{PV}/P_I < a$. For storages, we see that up to 8% of units fail the test if we only allow for thresholds a=3. However, keeping in mind that the inverter here are usually for both PV systems and storages, it is possible that some of those units are actually correct. For $a > 10$, we see that less then 2% of DSO approved units fail. Large PV systems fail more frequently, but since large PV systems usually are built with multiple inverters, one source of error could be that the power of the whole PV system is matched with the power of only one inverter.

In b) we compared the installed power of PV systems with the number of units, where we accepted a range of 50W - 700W per module. We selected these boundaries with an additional buffer to ensure that any raised errors are genuine. As result we see that for 43,733 (representing 1.12 %) of PV units the number of modules do not match with the installed PV power, where a large number of units are registered with only one module, possibly as a placeholder number. In c) we compared the installed power of ground-mounted PV systems with the area they needed. Knowing the area of ground-mounted PV systems is important for understanding the conflict between using arable area for agriculture or for electricity production. We see that the large share of 21.3% (3758 units) do not lie within our accepted range of 0.05MW/ha - 1.5MW/ha.

For wind turbines, the rated power scales quadratically with the rotor diameter $P \backsim r^2$. From [85] we took accepted ratios to be $160W/m^2$ - $700W/m^2$ and found 302 units (0.8%) that do not lie in this range. This is a rather small share, and we also saw that most erroneous wind turbines are small turbines with rated powers $\ll$ 1MW. For storages, we only found 770 storages where the inverter power and the power of the storage differ with a factor 20 or more.

[2] See marktstammdaten.kotthoff.dev

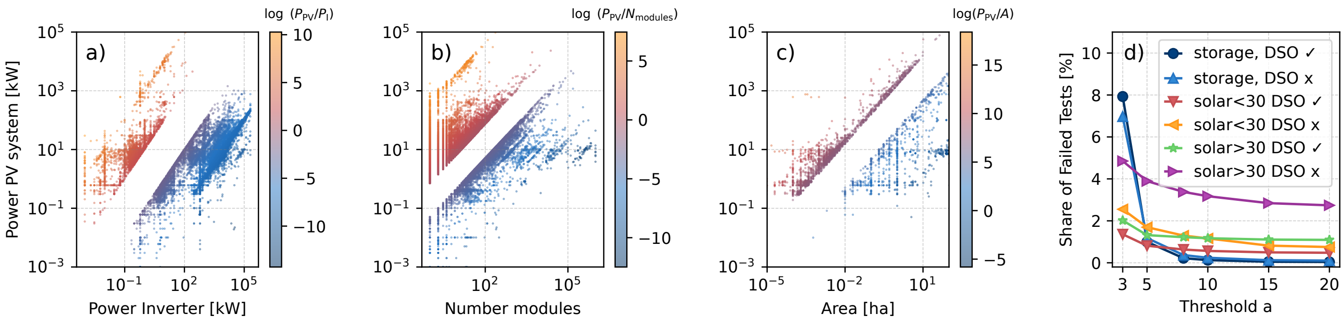


Fig. 3. Results of the System Size Tests. The white bands in the three plots represent the ranges of allowed values. In the scatter plots a)-c) each dot represents one PV system. In a) the power of the PV modules is compared to the power of the inverter, in b) the power of the PV modules is compared to the number of modules, and in c) the power of ground-mounted PV systems is compared to the area needed for the installation. Colors represent the log value of the fraction of y-axis and x-axis. In d) we plot the share of failed tests when we adapt the boundaries for the comparison of PV system and inverter power, where a unit fails a test if $1/a < P_{PV}/P_I < a$.

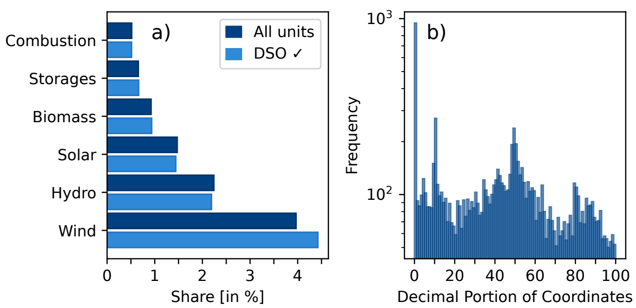


Fig. 4. Results of the Location Tests. In a) the share of units that have mismatching coordinates and districts is shown, where dark blue represents the fraction of wrong units divided by the total number of units. Light blue represents the share of erroneous units that were approved by DSOs, divided by the total number of DSO approved units. In b) a histogram of the decimal portion of coordinates from units with wrong locations is shown, where we see that a large number of erroneous units have coordinates with decimal portion .00, indicating that their coordinates lack any decimal portion.

When testing the rated power of units (Test ID 11), we did not find units lying outside of the accepted ranges.

Finally, we also tested the rated capacity of balcony-PV systems. In Germany, they fall under an own category of units with a simplified installation and registration, but they are only allowed to have a net power < 600W, which is planned to be increased to 800W within the year 2024. We first tested all units with unit type "balcony PV" to have a rated power of less than 1kW (with a small error tolerance of 200W). Then we also tested that units that have the word "balcony" in their name are small. Those units are not necessarily registered as balcony PV systems, but their name suggests that they are installed on a balcony, hence we set their upper limit to 5kW. We found about 9822 units failing one or both of these tests.

5.3 Location Tests

When testing if the coordinates of units lie within their specified districts, we saw that especially for wind turbines, a high share of roughly 4% of units have non matching coordinates and districts (see Fig. 4a). This corresponds to an accumulated installed power of more than 2 GW. Here we also see a concentration of location errors, where roughly half of the wind turbines with erroneous locations stem from only five districts. For all other technologies, the share of units with wrong location information is smaller (between 0.5% and 2%). For PV systems, we need to highlight that the publicly available dataset only contains coordinates of the 5% of systems with a power >30kW, hence this rather small share of PV units that fail the location tests has to be handled carefully and is not representative for the whole dataset. Additionally, we compared the amount of location information of the whole dataset with the subset of entries in the dataset that were already corrected by the DSO and find that the amount of location errors changes only slightly, but many location errors passed these DSO checks unnoticed. For wind turbines, we even saw that the share of errors for all units is slightly smaller then for DSO approved units.

One possible reason why those erroneous entries exist is the continuous change of geographic areas of districts and municipalities. In the MaStR, districts and municipalities occur even though they do not exist anymore because they were legally integrated into other districts or renamed. The entries within the MaStR are not automatically updated if above legal change occurs. However, not all erroneous entries can be traced back to this issue and some entries are simply wrong. We saw several wind farms with turbines allocated to different districts, even though next district borders were far away. In the appendix B, we further evaluate the distance between unit coordinates and their districts.

Two more reasons for wrong coordinates can be concluded from Fig. 4b). The figure shows a histogram of the decimal portion of wrong coordinates for units from all technologies where the district and coordinates do not match. If a unit fails the coordinate test and has coordinates (10.**13**4, 49.**30**7) we add the two decimal portions 13 and 30 to the histogram. We first see that a large amount of wrong units that fail the test have decimal portion of 0, indicating that those units are placed at integer latitude and longitudes. One reason could be that some unit owners do not want to share precise coordinates and delete this information in the registration form. The same reasoning could be applied to the peaks at .10 and .50. Moreover, it is evident that decimal values ranging from .60 to .99 are less frequent compared to those between .00 and .59. This discrepancy arises due to the conversion process of latitude and longitude from degrees, minutes, seconds (DMS) format to decimal degrees, where the minute values, which range from [0,59], are directly converted to decimal form.

We also tested coordinates against the specified municipalities of units and found about 10 GW of wind turbines, 0.3 GW of hydro power plants, 2.9 GW of solar power plants, 0.2 GW of biomass plants, 0.2 GW of storages, and 4.3 GW of combustion power plants that had wrong location information regarding their registered municipalities.

5.4 Technology specific Tests

For the installation years, we found 287 storages and 1169 PV systems to have installation dates earlier than 1980. For example, the MaStR contains battery storages with a rated power of more than 1MW for the installation year 1923. However, the accumulated power of all units with unplausible installation years is only of the order of several MW.

A rather unusual error comes up when testing the hub height of wind turbines. Here we found 4 turbines with a hub height smaller than the rotor radius, resulting in either broken turbine blades or a well-plowed field.

5.5 Updating results regularly

The MaStR dataset is updated once per day. Every update introduces new units that might contain erroneous data. Every update can also fix errors in existing entries. At the moment of reading this paper, the results section is already outdated. To overcome this issue, we decided to package the whole data pipeline and testing in one shell script, so that it can be triggered regularly. By using the framework datasette [103], it is then possible to visualize and publish the test results in an interactive website. All relevant plots of this results section are part of this dashboard and can be updated regularly. We host the current version of the dashboard online [3].

6 DISCUSSION

The MaStR is a rich and valuable dataset, both for the control and operation of the energy system as well as for research. This is underlined by the increasing number of publications that utilised the MaStR dataset, as we have shown in our literature review. We only considered publications in scientific journals. However, we assume that the utilization of the registry in applied science[4] is also of high relevance. Because of this, we're convinced that monitoring the data quality and highlighting possible shortcomings of this core dataset from the German energy system is an important contribution.

Regarding our results, we saw that the **Basic Tests** do not uncover crucial errors in the MaStR dataset, as most of them find none or few errors. Their main purpose comes when running the data and testing pipeline regularly to detect very fundamental mistakes (like non unique unit IDs). This is also true for the **Technology specific Tests**, especially when checking that rated powers lie in acceptable ranges. Testing hub heights of wind turbines or rated powers of balcony PV systems shows one important aspect of data tests: For some tests, domain knowledge is crucial. A collaborative approach that allows experts from various energy domains to easily define and implement data tests would significantly enhance the data quality of the MaStR, benefiting the entire community.

The **Location Tests** resulted in a mismatch between unit coordinates and assigned districts and municipalities for a substantial amount of units. Thus, researchers intending to use regional sub-samples of the registry should be cautious when selecting and aggregating data. It would be at least awkward to present large renewable energy plants in regional case studies which do not exist there.

The **System Size Tests** were an effective way of flagging erroneous entries. However they cannot be easily used to correct the data in the MaStR, as it is not clear which of the two properties is correct. For example, when seeing that an inverter and the modules power differ with a factor of 10^3, one of them must be wrong, but we cannot say which one. As a small contribution to also improve the data quality of the MaStR, we added a sub page [5] where system size errors can be obtained on a district level. We then disseminated this site over a Mail distribution list for the german energy domain. We hope that some DSOs are concerned about the data quality of the MaStR and try to correct the entries within their districts. As DSOs usually have the technical reports of both the installed modules and the inverter, this should be feasible.

To foster transparency and repeatability of results, we make our code openly available. The described dasboards can provide valuable insights into data quality of the MaStR. The pipeline and dashboard creation scripts are uploaded on github [51].

6.1 Limitations of our work

Our analysis is subject to limitations that we want to state clearly:

- To design elaborated test cases, domain knowledge is needed. Our knowledge helped us designing many test cases, but we are convinced that tests are missing. A process for different domain experts to collaboratively add data tests would be beneficial for the whole community.
- We only validate data from within the MaStR, thus our results only apply for the dataset itself. We also do not cross-check the content with other data sources. One further step could be to compare units registered in the MaStR with aerial images using deep learning based segmentation algorithms, as in [69, 83].
- The scope of the work is the monitoring of data quality of the whole MaStR dataset, hence the detailed investigation of specific errors and their reasons is missing out sometimes.

[3]See https://marktstammdaten.kotthoff.dev/-/dashboards/verify-marktstammdaten

[4]For example visualization tools like https://www.energy-charts.info/charts/power/chart.htm?l=en&c=DE or regional case studies like https://web.archive.org/web/20221108120815/https://opus.hs-osnabrueck.de/frontdoor/deliver/index/docId/3628/file/Energieversorgung.pdf

[5]Visit https://marktstammdaten.kotthoff.dev/-/dashboards/erronous-entries-district

7 CONCLUSION

As energy systems grow more complex, both energy research and operation needs to be based on detailed and reliable data to provide sound insights. Our contribution to this goal is a first analysis of the data quality of the MaStR as the central power and energy unit registry of Germany.

We start by emphasizing the urgency to validate the dataset by investigating its relevance in scientific literature. We find an increasing number of studies that use the MaStR data since its first release in January 2019.

In our opinion, the MaStR dataset is crucial for controlling and understanding the transforming energy system. It contains a huge amount of valuable and high-quality information, bearing in mind the identified issues with some dimensions of data quality. In the future, we recommend to increase the validation effort for the MaStR data with the help of automated techniques, together with scientific and public stakeholders.

In conclusion, we see this work as a first step towards a continuous validation process of the MaStR dataset. Validation efforts of the MaStR are most-likely already performed by many of its users. However, we think that results and methods should be shared to avoid redundant work and save resources which could be used better in producing valuable research of sustainable energy systems. Researchers using the MaStR dataset should be aware of the existing shortcomings and our identified errors. Wrong locations, for example, are of special relevance in the case of regional potential analyses, where it is important that the renewable power plants used in the studies were actually built within this region. With this paper and our data testing pipeline we have made a first contribution to an ongoing and open validation process.

ACKNOWLEDGMENTS

We would like to thank Annette Hammer and Susanne Weyand for fruitful discussions. DT and FK would like to thank the Bavarian State Ministry for Economic Affairs, Energy and Technology for their funding as part of the program "BayVFP Foerderlinie Digitalisierung - Informations- und Kommunikationstechnologie" (Förderkennzeichen DIK-2101-0005//DIK0240/01). FK would also like to thank the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy consortium - Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 501865131 within the German National Research Data Infrastructure (NFDI, https://www.nfdi.de/). CM thanks for the support by the Federal Ministry for Economic Affairs and Climate Action (BMWK).

A APPENDIX: COMPLETENESS OF THE MASTR DATASET

To give a clear idea of the content that we tested in the MaStR dataset, we provide table 3. It shows all columns that we have analyzed, together with an example entry and an indication of whether this column is apparent in the different technology tables. An x denotes that the column does not exist for this table, a number < 100 denotes the share of entries that are not null.

B APPENDIX: DISTANCE OF WRONGLY ALLOCATED UNITS

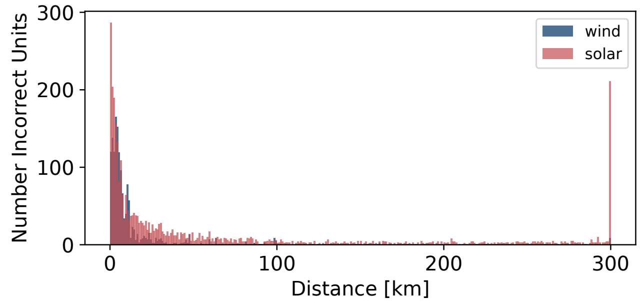


Fig. 5. Number of units where the coordinates do not lie within the district, plotted over the distance of the coordinates to the district boundaries. For wind turbines (blue) we see that most turbines are close to their district with distance < 40km. For PV systems, the distance to the district can be larger. The last bar is relatively large, since we used the aggregated value of all units that have a distance larger than 300km.

As described in section 5, the coordinates and assigned districts did not match for all units. Especially for wind and PV systems, we found erroneous entries. To see how far coordinates and districts lie apart, we calculated the distance between coordinates and district boundaries and plotted the results as a histogram in Fig. 5. We see that most of the units, where the coordinates and districts do not match, lie relatively close to the district boundaries. This is true for both wind turbines and PV systems. Only for units with a distance of several km we assume that the reason for the errors cannot be due to changes to the district boundaries over time.

REFERENCES

[1] Fabian Arnold, Samir Jeddi, and Amelie Sitzmann. 2022. How prices guide investment decisions under net purchasing — An empirical analysis on the impact of network tariffs on residential PV. *Energy Economics* 112 (aug 2022), 106177. https://doi.org/10.1016/j.eneco.2022.106177

[2] Miguel F. Astudillo, Karin Treyer, Christian Bauer, Pierre-Olivier Pineau, and Mourad Ben Amor. 2017. Life cycle inventories of electricity supply through the lens of data quality: exploring challenges and opportunities. *The International Journal of Life Cycle Assessment* 22, 3 (March 2017), 374–386. https://doi.org/10.1007/s11367-016-1163-0

[3] Keyu Bao, Louis Kalisch, Thunyathep Santhanavanich, Daniela Thrän, and Bastian Schröter. 2022. A bottom-up GIS-based method for simulation of ground-mounted PV potentials at regional scale. *Energy Reports* 8 (2022), 5053–5066. https://doi.org/10.1016/j.egyr.2022.03.187

[4] Julian Bartels, Christopher Varela, Timo Wassermann, Wided Medjroubi, and Edwin Zondervan. 2022. Integration of water electrolysis facilities in power grids: A case study in northern Germany. *Energy Conversion and Management: X* 14 (2022), 100209. https://doi.org/10.1016/j.ecmx.2022.100209

[5] Juliane Biehl, Leonard Missbach, Franziska Riedel, Ruben Stemmle, Julian Jüchter, Jessica Weber, Johanna Kucknat, Adrian Odenweller, Christian Nauck, Laura J. Lukassen, Matthias Zech, and Marie Grimm. 2023. Wicked facets of the German energy transition – examples from the electricity, heating, transport, and industry sectors. *International Journal of Sustainable Energy* 42, 1 (Sept. 2023), 1128–1181. https://doi.org/10.1080/14786451.2023.2244602

[6] Alexander J. Bogensperger, Yann Fabel, and Joachim Ferstl. 2022. Accelerating Energy-Economic Simulation Models via Machine Learning-Based Emulation and Time Series Aggregation. *Energies* 15, 3 (feb 2022), 1239. https://doi.org/10.3390/en15031239

[7] Meisam Booshehri, Lukas Emele, Simon Flügel, Hannah Förster, Johannes Frey, Ulrich Frey, Martin Glauer, Janna Hastings, Christian Hofmann, Carsten Hoyer-Klick, Ludwig Hülk, Anna Kleinau, Kevin Knosala, Leander Kotzur, Patrick

column	example	$C_{biomass}$	$C_{combustion}$	C_{hydro}	C_{solar}	$C_{storages}$	C_{wind}
unit owner mastr id	ABR989393706204	97	95	99	100	100	97
mastr id	SEE900002935310	100	100	100	100	100	100
operating status	In Betrieb	100	100	100	100	100	100
grid operator inspection	1	100	100	100	100	100	100
commissioning date	2001-12-21	99	99	100	98	99	91
planned commissioning date	2024-12-31	1	1	0	1	1	9
installation year	2017	100	100	100	100	100	100
download date	2024-03-12	100	100	100	100	100	100
zip code	17291	100	100	100	100	100	95
municipality	Bad Wünnenberg	100	100	100	100	100	95
district	Nordfriesland	100	100	100	100	100	95
coordinate	48.1748, 11.5961	96	26	55	5	0	97
power gross	5	x	x	x	100	100	x
power inverter	10	x	x	x	100	100	x
power net	5	x	x	x	100	100	x
power	2000	100	100	100	x	x	100
combustion technology	Verbrennungsmotor	100	x	x	x	x	x
fuel type	Gasförmige Biomasse	100	x	x	x	x	x
energy carrier	Erdgas	x	100	x	x	x	x
type of inflow	Flusskraftwerk	x	x	87	x	x	x
plant type	Laufwasseranlage	x	x	100	x	x	x
combination with storage	Kein Stromspeicher	x	x	x	98	x	x
number of modules	8	x	x	x	98	x	x
orientation	Süd	x	x	x	98	x	x
orientation secondary	West	x	x	x	21	x	x
unit type	Freifläche	x	x	x	100	x	x
storage capacity	10	x	x	x	x	100	x
battery technology	Lithium-Batterie	x	x	x	x	100	x
technology	Horizontalläufer	x	x	x	x	x	100
type description	E-70 E4	x	x	x	x	x	99
manufacturer	ENERCON GmbH	x	x	x	x	x	99
position	Windkraft an Land	x	x	x	x	x	100
hub height	65	x	x	x	x	x	98
rotor diameter	82	x	x	x	x	x	99

Table 3. Overview of selected columns with example and completeness for each table, where completeness is defined as fraction of number entries that are not null to the total number of entries (For example: 97% of units in the biomass table have a "unit owner mastr id"). An x marks columns that do not exist for this table.

Kuckertz, Till Mossakowski, Christoph Muschner, Fabian Neuhaus, Michaja Pehl, Martin Robinius, Vera Sehn, and Mirjam Stappel. 2021. Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis. *Energy and AI* 5 (sep 2021), 100074. https://doi.org/10.1016/j.egyai.2021.100074

[8] Bundesamt für Justiz. 2018. *Verordnung über das zentrale elektronische Verzeichnis energiewirtschaftlicher Daten (Marktstammdatenregisterverordnung - MaStRV) Anlage Im Marktstammregister zu erfassende Daten, Fundstelle: BGBl. I 2018, 1895 - 1904*. Bundesamt für Justiz. https://www.gesetze-im-internet.de/mastrv/anlage.html

[9] Bundesamt für Kartographie und Geodäsie. 2024. *Open Data - Verwaltungsgebiete*. Bundesamt für Kartographie und Geodäsie. https://gdz.bkg.bund.de/index.php/default/open-data.html

[10] Bundesnetzagentur. 2019. *Marktstammdatenregister*. https://www.marktstammdatenregister.de/MaStR © Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen | DL-DE-BY-2.0.

[11] Bundesnetzagentur für Elektrizität, Gas, and Telekommunikation, Post und Eisenbahnen. 2018. Das Marktstammdatenregister – Gesamtkonzept. (09 2018). https://www.marktstammdatenregister.de/MaStRHilfe/files/regHilfen/MaStR-Gesamtkonzept_Stand%20September%202018.pdf

[12] Bundesnetzagentur für Elektrizität, Gas, and Telekommunikation, Post und Eisenbahnen. 2022. Monitoringbericht 2022. (12 2022).

[13] Jonas Böhm, Thomas de Witte, and Coline Michaud. 2022. Land use Prior to Installation of Ground-mounted Photovoltaic in Germany—GIS-analysis Based on MaStR and Basis-DLM. *Zeitschrift für Energiewirtschaft* 46, 2 (jun 2022), 147–156. https://doi.org/10.1007/s12398-022-00325-4

[14] Clara Büttner, Katharina Esterl, Ilka Cußmann, Carlos Andrés Epia Realpe, Jonathan Amme, and Amélia Nadal. 2024. Influence of flexibility options on the German transmission grid — A sector-coupled mid-term scenario. *Renewable and Sustainable Energy Transition* 5 (Aug. 2024), 100082. https://doi.org/10.1016/j.rset.2024.100082

[15] Li Cai and Yangyong Zhu. 2015. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal* 14 (May 2015), 2–2. https://doi.org/10.5334/dsj-2015-002

[16] Stefano Carattini, Béla Figge, Alexander Gordan, and Andreas Löschel. 2024. Municipal building codes and the adoption of solar photovoltaics. *Journal of*

Environmental Economics and Management 124 (March 2024), 102937. https://doi.org/10.1016/j.jeem.2024.102937
[17] Wen Chen, Kaile Zhou, Shanlin Yang, and Cheng Wu. 2017. Data quality of electricity consumption data in a smart grid environment. *Renewable and Sustainable Energy Reviews* 75 (Aug. 2017), 98–105. https://doi.org/10.1016/j.rser.2016.10.054
[18] Nils Christiansen, Jeffrey R. Carpenter, Ute Daewel, Nobuhiro Suzuki, and Corinna Schrum. 2023. The large-scale impact of anthropogenic mixing by offshore wind turbine foundations in the shallow North Sea. *Frontiers in Marine Science* 10 (May 2023). https://doi.org/10.3389/fmars.2023.1178330
[19] dbt Labs. 2024. *Data Build Tool - dbt.* https://github.com/dbt-labs/dbt-core
[20] Elias Dörre, Sebastian Pfaffel, Alexander Dreher, Pedro Girón, Svenja Heising, and Kay Wiedemann. 2021. Flexibility Reserve of Self-Consumption Optimized Energy Systems in the Household Sector. *Energies* 14, 11 (2021), 3017. https://doi.org/10.3390/en14113017
[21] Thorsten Engelhorn and Thomas Möbius. 2022. On the Development of Wind Market Values and the Influence of Technology and Weather: a German Case Study. *Zeitschrift für Energiewirtschaft* 46, 1 (2022), 61–83. https://doi.org/10.1007/s12398-022-00319-2
[22] Danial Esmaeili Aliabadi, David Manske, Lena Seeger, Reinhold Lehneis, and Daniela Thrän. 2023. Integrating Knowledge Acquisition, Visualization, and Dissemination in Energy System Models: BENOPTex Study. *Energies* 16, 13 (July 2023), 5113. https://doi.org/10.3390/en16135113
[23] Max Eysholdt, Ralf Kunkel, Claus Rösemann, Frank Wendland, Tim Wolters, Maximilian Zinnbauer, and Roland Fuß. 2022. A model-based estimate of nitrate leaching in Germany for GHG reporting. *Journal of Plant Nutrition and Soil Science* (2022). https://doi.org/10.1002/jpln.202200119
[24] Federal Agency for Cartography and Geodasy. 2021. Administrative areas. https://gdz.bkg.bund.de/index.php/default/digitale-geodaten.html © GeoBasis-DE / BKG 2021.
[25] Jan Figgener, Peter Stenzel, Kai-Philipp Kairies, Jochen Linßen, David Haberschusz, Oliver Wessels, Georg Angenendt, Martin Robinius, Detlef Stolten, and Dirk Uwe Sauer. 2020. The development of stationary battery storage systems in Germany – A market review. *Journal of Energy Storage* 29 (jun 2020), 101153. https://doi.org/10.1016/j.est.2019.101153
[26] Simon Fink, Hendrik Teichgräber, and Mareike Wehling. 2021. Der Ausbau der deutschen Stromnetze: Kohärente Parteienideologie oder Sollbruchstelle entlang lokaler Interessen? *Zeitschrift für Vergleichende Politikwissenschaft* 15, 4 (nov 2021), 617–639. https://doi.org/10.1007/s12286-021-00494-5
[27] Martha Frysztacki and Tom Brown. 2020. Modeling Curtailment in Germany: How Spatial Resolution Impacts Line Congestion. In *2020 17th International Conference on the European Energy Market (EEM)*. IEEE, 1–7. https://doi.org/10.1109/EEM49802.2020.9221886
[28] Ray Galvin. 2022. Why German households won't cover their roofs in photovoltaic panels: And whether policy interventions, rebound effects and heat pumps might change their minds. *Renewable Energy Focus* 42 (sep 2022), 236–252. https://doi.org/10.1016/j.ref.2022.07.002
[29] Kirstin Ganz, Timo Kern, and Michael Hinterstocker. 2024. Systemic Evaluation of PV Self-Consumption Optimization Using Electric Vehicles. *World Electric Vehicle Journal* 15, 3 (March 2024), 98. https://doi.org/10.3390/wevj15030098
[30] Laura Gaßner and Joachim Ritter. 2023. Ground motion emissions due to wind turbines: observations, acoustic coupling, and attenuation relationships. *Solid Earth* 14, 7 (July 2023), 785–803. https://doi.org/10.5194/se-14-785-2023
[31] Abe Gong, James Campbell, and Great Expectations. [n. d.]. *Great Expectations*. https://github.com/great-expectations/great_expectations
[32] Fabian Gotzens, Heidi Heinrichs, Jonas Hörsch, and Fabian Hofmann. 2019. Performing energy modelling exercises in a transparent way - The issue of data quality in power plant databases. *Energy Strategy Reviews* 23 (Jan. 2019), 1–12. https://doi.org/10.1016/j.esr.2018.11.004
[33] Leonie Grau, Christopher Jung, and Dirk Schindler. 2021. Sounding out the repowering potential of wind energy–A scenario-based assessment from Germany. *Journal of Cleaner Production* 293 (2021), 126094. https://doi.org/10.1016/j.jclepro.2021.126094
[34] Thorsten Hack, Zheng Ma, and Bo Nørregaard Jørgensen. 2021. Digitalisation potentials in the electricity ecosystem: lesson learnt from the comparison between Germany and Denmark. *Energy Informatics* 4, S2 (sep 2021). https://doi.org/10.1186/s42162-021-00168-2
[35] Lion Hirth, Jonathan Mühlenpfordt, and Marisa Bulkeley. 2018. The ENTSO-E Transparency Platform – A review of Europe's most ambitious electricity data platform. *Applied Energy* 225 (Sept. 2018), 1054–1067. https://doi.org/10.1016/j.apenergy.2018.04.048
[36] Hannes Hobbie and Martin Lieberwirth. 2024. Compounding or Curative? Investigating the impact of electrolyzer deployment on congestion management in the German power grid. *Energy Policy* 185 (Feb. 2024), 113900. https://doi.org/10.1016/j.enpol.2023.113900
[37] Maximilian Hoffmann, Jan Priesmann, Lars Nolting, Aaron Praktiknjo, Leander Kotzur, and Detlef Stolten. 2021. Typical periods or typical time steps? A multi-model analysis to determine the optimal temporal aggregation for energy system models. *Applied Energy* 304 (dec 2021), 117825. https://doi.org/10.1016/j.apenergy.2021.117825
[38] Ludwig Hülk, Guido Pleßmann, Christoph Muschner, Florian Kotthoff, and Deniz Tepe. 2023. *open-MaStR.* https://github.com/OpenEnergyPlatform/open-MaStR/
[39] Ludwig Hülk, Guido Pleßmann, Christoph Muschner, Florian Kotthoff, and Deniz Tepe. 2023. *open-MaStR - Marktstammdatenregister.* https://doi.org/10.5281/zenodo.8225106
[40] Mateo Jesper, Felix Pag, Klaus Vajen, and Ulrike Jordan. 2021. Annual Industrial and Commercial Heat Load Profiles: Modeling Based on k-Means Clustering and Regression Analysis. *Energy Conversion and Management: X* 10 (jun 2021), 100085. https://doi.org/10.1016/j.ecmx.2021.100085
[41] Simon Johanning, Daniel Abitz, Emily Schulte, Fabian Scheller, and Thomas Bruckner. 2022. PVactVal: A Validation Approach for Agent-based Modeling of Residential Photovoltaic Adoption. In *2022 18th International Conference on the European Energy Market (EEM)*. IEEE, 1–6. https://doi.org/10.1109/eem54602.2022.9921039
[42] Christopher Jung and Dirk Schindler. 2022. Development of onshore wind turbine fleet counteracts climate change-induced reduction in global capacity factor. *Nature Energy* 7, 7 (2022), 608–619. https://doi.org/10.1038/s41560-022-01056-z
[43] Christopher Jung and Dirk Schindler. 2023. Reasons for the Recent Onshore Wind Capacity Factor Increase. *Energies* 16, 14 (July 2023), 5390. https://doi.org/10.3390/en16145390
[44] C. Jätz, B. Petters, N. Dult, A. Ahmadifar, and A. Monti. 2023. Balancing PV generation in low voltage grids with limited data. In *27th International Conference on Electricity Distribution (CIRED 2023)*. Institution of Engineering and Technology. https://doi.org/10.1049/icp.2023.0690
[45] Christoph Jörges, Hedwig Sophie Vidal, Tobias Hank, and Heike Bach. 2023. Detection of Solar Photovoltaic Power Plants Using Satellite and Airborne Hyperspectral Imaging. *Remote Sensing* 15, 13 (July 2023), 3403. https://doi.org/10.3390/rs15133403
[46] Steffen Karalus, Benedikt Köpfer, Philipp Guthke, Sven Killinger, and Elke Lorenz. 2023. Analysing Grid-Level Effects of Photovoltaic Self-Consumption Using a Stochastic Bottom-up Model of Prosumer Systems. *Energies* 16, 7 (March 2023), 3059. https://doi.org/10.3390/en16073059
[47] Maximilian Kilthau, Martin Asman, Alexandra Karmann, Ghayathri Suriyamoorthy, Jan-Philip Beck, Vincenz Regener, Christian Derksen, Nils Loose, Moritz Volkmann, Shashank Tripathi, Felix Gehlhoff, Kamil Korotkiewicz, Philippe Steinbusch, Volker Skwarek, Markus Zdrallek, and Alexander Fay. 2023. Integrating Peer-to-Peer Energy Trading and Flexibility Market With Self-Sovereign Identity for Decentralized Energy Dispatch and Congestion Management. *IEEE Access* 11 (2023), 145395–145420. https://doi.org/10.1109/access.2023.3344855
[48] Lena Kitzing, Muhammad Bilal Siddique, Ivan Nygaard, and Wikus Kruger. 2022. Worth the wait: How South Africa's renewable energy auctions perform compared to Europe's leading countries. *Energy Policy* 166 (2022), 112999. https://doi.org/10.1016/j.enpol.2022.112999
[49] Christina Kockel, Lars Nolting, Jan Priesmann, and Aaron Praktiknjo. 2022. Does renewable electricity supply match with energy demand?–A spatio-temporal analysis for the German case. *Applied Energy* 308 (2022), 118226. https://doi.org/10.1016/j.apenergy.2021.118226
[50] Hendrik Kondziella, Lucien Genge, and Thomas Bruckner. 2021. Status quo des technischen Potenzials von Flexibilitätsoptionen im Nordosten Deutschlands. *Zeitschrift für Energiewirtschaft* 45, 3 (apr 2021), 181–193. https://doi.org/10.1007/s12398-021-00300-5
[51] Florian Kotthoff and Christoph Muschner. 2024. *verify-marktstammdaten.* https://github.com/FlorianK13/verify-marktstammdaten
[52] Tobias Kraschewski, Tim Brauner, Maximilian Heumann, and Michael H. Breitner. 2023. Disentangle the price dispersion of residential solar photovoltaic systems: Evidence from Germany. *Energy Economics* 121 (May 2023), 106649. https://doi.org/10.1016/j.eneco.2023.106649
[53] Mats Kröger, Karsten Neuhoff, and Joern Constantin Richstein. 2022. Discriminatory Auction Design for Renewable Energy. (2022). https://doi.org/10.2139/ssrn.4178053
[54] Matthias Kühnbach, Stefan Pisula, Anke Bekk, and Anke Weidlich. 2020. How much energy autonomy can decentralised photovoltaic generation provide? A case study for Southern Germany. *Applied Energy* 280 (2020), 115947. https://doi.org/10.1016/j.apenergy.2020.115947
[55] Isaac Kyere, Thomas Astor, Rüdiger Graß, Thomas Fricke, and Michael Wachendorf. 2021. Spatio-temporal analysis of the effects of biogas production on agricultural lands. *Land Use Policy* 102 (mar 2021), 105240. https://doi.org/10.1016/j.landusepol.2020.105240
[56] Uwe Langenmayr and Manuel Ruppert. 2023. Renewable origin, additionality, temporal and geographical correlation – eFuels production in Germany under

the RED II regime. *Energy Policy* 183 (Dec. 2023), 113830. https://doi.org/10.1016/j.enpol.2023.113830
[57] Michael Lechl, Luis Schoppik, and Hermann de Meer. 2023. Aggregating multi-time-scale flexibility potentials of battery storages based on open data – a potential analysis. *Energy Informatics* 6, S1 (Oct. 2023). https://doi.org/10.1186/s42162-023-00273-4
[58] T. Lechner, J. Timmermann, S. Seifried, C. Bernecker-Castro, M. Finkel, R. Witzmann, G. Kerber, K. Chaarschmidt, and S. Herrmann. 2023. Frequency droop characteristic for grid forming battery inverters - operation in islanded grids with the infeed of distributed generation systems. In *27th International Conference on Electricity Distribution (CIRED 2023)*. Institution of Engineering and Technology. https://doi.org/10.1049/icp.2023.1015
[59] Yang Li, Przemysław Janik, and Harald Schwarz. 2023. Aggregated wind power characteristic curves and artificial intelligence for the regional wind power infeed estimation. *Electrical Engineering* 106, 1 (Sept. 2023), 655–671. https://doi.org/10.1007/s00202-023-02005-z
[60] Yang Li, Przemyslaw Janik, and Harald Schwarz. 2023. Prediction and aggregation of regional PV and wind generation based on neural computation and real measurements. *Sustainable Energy Technologies and Assessments* 57 (June 2023), 103314. https://doi.org/10.1016/j.seta.2023.103314
[61] Taimyra Batz Liñeiro and Felix Müsgens. 2021. Evaluating the German PV auction program: The secrets of individual bids revealed. *Energy Policy* 159 (dec 2021), 112618. https://doi.org/10.1016/j.enpol.2021.112618
[62] Taimyra Batz Liñeiro and Felix Müsgens. 2021. Lessons Learned from Photovoltaic Auctions in Germany. https://doi.org/10.48550/ARXIV.2104.07536
[63] Meike Löhr and Jannika Mattes. 2022. Facing transition phase two: Analysing actor strategies in a stagnating acceleration phase. *Technological Forecasting and Social Change* 174 (2022), 121221. https://doi.org/10.1016/j.techfore.2021.121221
[64] Lara Lück and Albert Moser. 2019. Artificial neural networks modeling photovoltaic power system allocation–Can artificial intelligence beat a fundamental approach?. In *2019 16th International Conference on the European Energy Market (EEM)*. IEEE, 1–6. https://doi.org/10.1109/eem.2019.8916440
[65] Lara Lück and Albert Moser. 2019. Wind Turbine Integration-Can Current Allocation Policies Curb Transmission Grid Congestions Caused by Renewables Feed-In?. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*. IEEE, 2161–2165. https://doi.org/10.1109/ispec48194.2019.8975277
[66] Alexandra Lüth, Jens Weibezahn, and Jan Martin Zepter. 2020. On distributional effects in local electricity market designs—Evidence from a german case study. *Energies* 13, 8 (2020), 1993. https://doi.org/10.3390/en13081993
[67] David Manske, Lukas Grosch, Julius Schmiedt, Nora Mittelstädt, and Daniela Thrän. 2022. Geo-Locations and System Data of Renewable Energy Installations in Germany. *Data* 7, 9 (2022), 128. https://doi.org/10.3390/data7090128
[68] David Matthäus, Sebastian Schwenen, and David Wozabal. 2021. Renewable auctions: Bidding for real options. *European Journal of Operational Research* 291, 3 (jun 2021), 1091–1105. https://doi.org/10.1016/j.ejor.2020.09.047
[69] Kevin Mayer, Benjamin Rausch, Marie-Louise Arlt, Gunther Gust, Zhecheng Wang, Dirk Neumann, and Ram Rajagopal. 2022. 3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D. *Applied Energy* 310 (mar 2022), 118469. https://doi.org/10.1016/j.apenergy.2021.118469
[70] Kevin Mayer, Zhecheng Wang, Marie-Louise Arlt, Dirk Neumann, and Ram Rajagopal. 2020. DeepSolar for Germany: A deep learning framework for PV system mapping from aerial imagery. In *2020 International Conference on Smart Energy Systems and Technologies (SEST)*. 1–6. https://doi.org/10.1109/SEST48500.2020.9203258
[71] Steffen Meinecke, Džanan Sarajlić, Simon Ruben Drauz, Annika Klettke, Lars-Peter Lauven, Christian Rehtanz, Albert Moser, and Martin Braun. 2020. Simbench—a benchmark dataset of electric power systems to compare innovative solutions based on power flow analysis. *Energies* 13, 12 (2020), 3290. https://doi.org/10.3390/en13123290
[72] Anica Mertins, Mathias Heiker, Sandra Rosenberger, and Tim Wawer. 2023. Competition in the conversion of the gas grid: Is the future of biogas biomethane or hydrogen? *International Journal of Hydrogen Energy* 48, 83 (Oct. 2023), 32469–32484. https://doi.org/10.1016/j.ijhydene.2023.04.270
[73] Lukas Mueller, Timothy Peter Marcroft, Constantin von Beck, Jan Pedro Zeiss, Valeria Jana Schwanitz, August Wierling, and Lars Holstenkamp. 2024. "First come, first served" or "the more, the merrier"? Organizational dynamics of citizen-led solar initiatives and the presence of photovoltaic installations in Germany. *Journal of Cleaner Production* 449 (April 2024), 141861. https://doi.org/10.1016/j.jclepro.2024.141861
[74] Mathias Müller, Yannic Blume, and Janis Reinhard. 2022. Impact of behind-the-meter optimised bidirectional electric vehicles on the distribution grid load. *Energy* 255 (2022), 124537. https://doi.org/10.1016/j.energy.2022.124537
[75] Renewables Now. 2023. *Mingyang presents 22-MW offshore wind turbine concept.* http://web.archive.org/web/20231104232701/https://renewablesnow.com/news/mingyang-presents-22-mw-offshore-wind-turbine-concept-837415/
[76] Nesrine Ouanes, Jan Kegel, Jan Wiesenthal, Clara Lenk, Hannes Bluhm, Julika Weiß, and Lukas Torliene. 2022. Prosuming–energy sufficiency and rebound effects: Climate impact of changing household consumption patterns in Germany. *TATuP-Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis* 31, 2 (2022), 18–24. https://doi.org/10.14512/tatup.31.2.18
[77] Yannik Pflugfelder, Hendrik Kramer, and Christoph Weber. 2024. A novel approach to generate bias-corrected regional wind infeed timeseries based on reanalysis data. *Applied Energy* 361 (May 2024), 122890. https://doi.org/10.1016/j.apenergy.2024.122890
[78] Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. 2002. Data quality assessment. *Commun. ACM* 45, 4 (April 2002), 211–218. https://doi.org/10.1145/505248.506010
[79] Benjamin Rausch, Kevin Mayer, Marie-Louise Arlt, Gunther Gust, Philipp Staudt, Christof Weinhardt, Dirk Neumann, and Ram Rajagopal. 2020. An Enriched Automated PV Registry: Combining Image Recognition and 3D Building Data. https://doi.org/10.48550/ARXIV.2012.03690
[80] Hannes Salomon, Martin Drechsler, and Felix Reutter. 2020. Minimum distances for wind turbines: A robustness analysis of policies for a sustainable wind power deployment. *Energy Policy* 140 (2020), 111431. https://doi.org/10.1016/j.enpol.2020.111431
[81] Tamara Schröter, André Richter, Jens Götze, André Naumann, Jenny Gronau, and Martin Wolter. 2020. Substation related forecasts of electrical energy storage systems: transmission system operator requirements. *Energies* 13, 23 (2020), 6207. https://doi.org/10.3390/en13236207
[82] Daniel Schröer and Uwe Latacz-Lohmann. 2024. Flexibilization or biomethane upgrading? Investment preference of German biogas plant operators for the follow-up of guaranteed feed-in tariffs. *GCB Bioenergy* 16, 2 (Jan. 2024). https://doi.org/10.1111/gcbb.13111
[83] Maximilian Schulz, Bilel Boughattas, and Frank Wendel. 2021. DetEEktor: Mask R-CNN based neural network for energy plant identification on aerial photographs. *Energy and AI* 5 (2021), 100069. https://doi.org/10.1016/j.egyai.2021.100069
[84] Anna-Lena Schönauer and Sabrina Glanz. 2023. Local conflicts and citizen participation in the German energy transition: Quantitative findings on the relationship between conflict and participation. *Energy Research & Social Science* 105 (Nov. 2023), 103267. https://doi.org/10.1016/j.erss.2023.103267
[85] Javier Serrano-González and Roberto Lacal-Arántegui. 2016. Technological evolution of onshore wind turbines—a market-based analysis. *Wind Energy* 19, 12 (Dec. 2016), 2171–2187. https://doi.org/10.1002/we.1974
[86] Muhammad Bilal Siddique and Jagruti Thakur. 2020. Assessment of curtailed wind energy potential for off-grid applications through mobile battery storage. *Energy* 201 (jun 2020), 117601. https://doi.org/10.1016/j.energy.2020.117601
[87] Chris Stetter, Henrik Wielert, and Michael H Breitner. 2022. Hidden repowering potential of non-repowerable onshore wind sites in Germany. 168 (2022), 113168. https://doi.org/10.1016/j.enpol.2022.113168
[88] Laura Stößel, Leila Poddie, Tobias Spratte, Ralf Schelenz, and Georg Jacobs. 2021. County Clustering with Bioenergy as Flexible Power Unit in a Renewable Energy System. *Energies* 14, 17 (2021), 5227. https://doi.org/10.3390/en14175227
[89] Judith Stute and Matthias Kühnbach. 2023. Dynamic pricing and the flexible consumer – Investigating grid and financial implications: A case study for Germany. *Energy Strategy Reviews* 45 (Jan. 2023), 100987. https://doi.org/10.1016/j.esr.2022.100987
[90] Daniela Thrän, Karen Deprie, Martin Dotzauer, Peter Kornatz, Michael Nelles, Kai Sven Radtke, and Harry Schindler. 2023. The potential contribution of biogas to the security of gas supply in Germany. *Energy, Sustainability and Society* 13, 1 (May 2023). https://doi.org/10.1186/s13705-023-00389-1
[91] Jacob Tran, Pascal Pfeifer, Christoph Wirtz, Dominik Wursthorn, Hendrik Vennegeerts, and Albert Moser. 2019. Modelling of synthetic power distribution systems in consideration of the local electricity supply task.
[92] Merlin Sebastian Triebs, Elisa Papadis, Hannes Cramer, and George Tsatsaronis. 2021. Landscape of district heating systems in Germany–Status quo and categorization. *Energy Conversion and Management: X* 9 (2021), 100068. https://doi.org/10.1016/j.ecmx.2020.100068
[93] Jan Frederick Unnewehr, Mirko Schäfer, and Anke Weidlich. 2022. The value of network resolution – A validation study of the European energy system model PyPSA-Eur. In *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*. 1–7. https://doi.org/10.1109/OSMSES54027.2022.9769123
[94] Jonas van Ouwerkerk, Karlo Hainsch, Soner Candas, Christoph Muschner, Stefanie Buchholz, Stephan Günther, Hendrik Huyskens, Sarah Berendes, Konstantin Löffler, Christian Bußar, et al. 2022. Comparing open source power system models-A case study focusing on fundamental modeling parameters for the German energy transition. *Renewable and Sustainable Energy Reviews* 161 (2022), 112331. https://doi.org/10.1016/j.rser.2022.112331
[95] Rebekka Volk, Christoph Stallkamp, Magnus Herbst, and Frank Schultmann. 2021. Regional rotor blade waste quantification in Germany until 2040. *Resources, Conservation and Recycling* 172 (sep 2021), 105667. https://doi.org/10.1016/j.

resconrec.2021.105667

[96] Julia Walgern, David Baumgärtner, Johannes Fricke, Niklas Requate, Athanasios Kolios, Martin Dörenkämper, Tobias Meyer, and Lukas Vollmer. 2023. Economic feasibility study for continued operation of German offshore wind farms. *Journal of Physics: Conference Series* 2626, 1 (Oct. 2023), 012031. https://doi.org/10.1088/1742-6596/2626/1/012031

[97] Felix Wege, Erdem Gümrükcü, Tristan Wortberg, Alexandra Mateeva, Roman Höller, Ferdinanda Ponci, and Antonello Monti. 2023. Impacts of Local Energy Exchange in Low Voltage Communities: A German Case Study. *IEEE Access* 11 (2023), 95768–95781. https://doi.org/10.1109/access.2023.3308815

[98] F. Weise, B. Koch, C. Voglstätter, T. Smolinka, and C. Hebling. 2023. Developing the GEO-techno-economic analysis of hydrogen ecosystems. In *7th International Hybrid Power Plants & Systems Workshop*. Institution of Engineering and Technology. https://doi.org/10.1049/icp.2023.1432

[99] Andreas Weiss, Florian Biedenbach, and Mathias Mueller. 2021. Simulation and analysis of future electric mobility load effects in urban distribution grids. In *ETG Congress 2021*. 1–6.

[100] August Wierling, Valeria Jana Schwanitz, Jan Pedro Zeiss, Constantin von Beck, Heather Arghandeh Paudler, Ingrid Knutsdotter Koren, Tobias Kraudzun, Timothy Marcroft, Lukas Müller, Zacharias Andreadakis, Chiara Candelise, Simon Dufner, Melake Getabecha, Grete Glaase, Wit Hubert, Veronica Lupi, Sona Majidi, Shirin Mohammadi, Negar Safara Nosar, Yann Robiou du Pont, Philippa Roots, Tadeusz Józef Rudek, Alessandro Sciullo, Gayatri Sehdev, Mehran Ziaabadi, and Nahid Zoubin. 2023. A Europe-wide inventory of citizen-led energy action with data from 29 countries and over 10000 initiatives. *Scientific Data* 10, 1 (Jan. 2023). https://doi.org/10.1038/s41597-022-01902-5

[101] August Wierling, Jan Pedro Zeiss, Veronica Lupi, Chiara Candelise, Alessandro Sciullo, and Valeria Jana Schwanitz. 2021. The Contribution of Energy Communities to the Upscaling of Photovoltaics in Germany and Italy. *Energies* 14, 8 (apr 2021), 2258. https://doi.org/10.3390/en14082258

[102] August Wierling, Jan Pedro Zeiss, Constantin von Beck, and Valeria Jana Schwanitz. 2022. Business models of energy cooperatives active in the PV sector—A statistical analysis for Germany. *PLOS Sustainability and Transformation* 1, 9 (2022), e0000029. https://doi.org/10.1371/journal.pstr.0000029

[103] Simon Willison. 2024. *datasette*. https://github.com/simonw/datasette

[104] C. Wirtz, M. Murglat, J. Tran, S. Krahl, and A. Moser. 2021. Modelling of synthetic high voltage networks based on open data and integration into a modular synthetic distribution grid generator. In *CIRED 2021 - The 26th International Conference and Exhibition on Electricity Distribution*, Vol. 2021. 2755–2759. https://doi.org/10.1049/icp.2021.1654

[105] Matthias Wrede. 2022. The influence of state politics on solar energy auction results. *Energy Policy* 168 (sep 2022), 113130. https://doi.org/10.1016/j.enpol.2022.113130

[106] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2015. Quality assessment for Linked Data: A Survey: A systematic literature review and conceptual framework. *Semantic Web* 7, 1 (March 2015), 63–93. https://doi.org/10.3233/SW-150175

Decentralised Privacy-preserving Power Flow Forecast on Smart Meters

SEVERIN NOWAK and FABIAN WIDMER, Lucerne University of Applied Sciences and Arts, Switzerland
JEREMY TAYLOR, Via Science, Canada
DANIEL RAIMUNDO and ANTONIOS PAPAEMMANOUIL, Lucerne University of Applied Sciences and Arts, Switzerland

The transition toward decarbonisation of energy systems necessitates significant adaptations in distribution networks, challenging the grid planning and operation for Distribution System Operators (DSOs). This paper explores how smart meters can improve DSO capabilities amidst evolving regulatory frameworks and technological landscapes. DSOs face the challenge of accommodating more distributed generation, electrification of loads, and the surge in electric vehicles, all while striving to maintain grid reliability and cost-effectiveness. To meet these demands without compromising data security and privacy, advanced data science techniques are explored, leveraging smart meter data in near real-time.

The paper proposes decentralised algorithms capable of near real-time computation and forecasting of network power flows, revealing congestion issues without centralising consumer data. Field trials conducted in Rolle VD provide concrete evidence of the feasibility and advantages of this approach. They not only offer insights into network conditions but also ensure data privacy. Moreover, the developed decentralised congestion calculation method has shown to deliver results comparable to commonly used centralised algorithms, even when confronted with real-world challenges such as connection issues. Utilizing smart meters with enhanced computing capabilities and Internet of Things (IoT) infrastructure, the proposed method presents a viable alternative to traditional approaches, enabling DSOs to adapt to the evolving energy landscape efficiently.

CCS Concepts: • **Computing methodologies → Distributed algorithms**; • **Hardware → Power networks**; **Smart grid**.

Additional Key Words and Phrases: Consumption Forecast, Congestion Calculation, Smart Grid, Distributed Power Flow, Decentralised Power Flow

1 INTRODUCTION

The decarbonisation of energy supply and the shift away from nuclear energy are catalysts for significant changes in the fundamental characteristics of distribution networks. As a result, the responsibilities of Distribution System Operators (DSOs) are evolving, influenced by legislative measures such as the Revised Act on Emission Reductions, the CO2 Ordinance, and the energy strategy for 2050 [9]. DSOs are required to provide high reliability electric power to end-customers while adapting to a higher level of distributed generation, the electrification of various loads such as heating systems, and the proliferation of new electrical loads like electric vehicles. The data collected from smart meters holds significant potential to aid DSOs in managing the evolving network landscape. Consequently, there is a growing focus on leveraging advanced data science techniques, primarily utilizing smart meter data. This exploration aims to assist DSOs in achieving their regulatory and corporate objectives related to enhancing grid reliability, resilience, and cost performance. Typically, such data science techniques are based on data centralisation, which introduces delays in data processing, data security, and privacy protection vulnerabilities. Moreover, centralisation of data may create regulatory challenges for DSOs given the legal constraints that exist in relation to accessing and analysing consumer load profiles.

This paper investigates data-driven approaches based on smart meter data that can improve the DSOs ability to adapt to decarbonisation and decentralisation in a way that addresses the privacy and security constraints described above while enabling near real-time insights into the present and future grid operation. Algorithms and models are deployed decentrally to analyze data without physically co-locating them, yet still yield results comparable to those obtained from applying the same algorithm or model to a centralised database. We further investigate how algorithms work in conjunction with smart meters at the grid edge and consider the practical integration of AI-based algorithms and federated analytics approaches in a real-world setting through simulations and field-based trials.

Within this paper, end-consumer electric load forecasting and decentralised power flow has been developed and implemented within a field trial. The field trials encompasses 30 end-customers in a sample distribution system in Rolle VD, Switzerland. The main contributions of the proposed approach are listed below:

(1) Enabling edge-based near real-time computation and forecasting of network power flows, including congestion levels. In contrast to existing smart metering infrastructure which only transmits 15-minute measurement data every 24 hours, network conditions can be accessed in near real-time.
(2) Ensuring data protection and avoids any centralisation of end-consumer data and hence a more efficient data processing.
(3) Presenting an innovative alternative to the traditional approach of installing additional measurement devices at the MV/LV transformer level to obtain congestion levels.
(4) Providing a prototype setting of the implementation using an IoT infrastructure in a real distribution system.

The proposed approach requires the use of smart meters equipped with enhanced computing capabilities to execute advanced decentralised algorithms. To allow testing on a real distribution gird, the implementation of Information and Communication Technology (ICT) at the Distribution System Operator (DSO) level with a direct Internet of Things (IoT) interface was developed using the Message Queuing Telemetry Transport (MQTT) protocol.

Authors' addresses: Severin Nowak, severin.nowak@hslu.ch; Fabian Widmer, fabian.widmer@hslu.ch, Lucerne University of Applied Sciences and Arts, Department of Electrical Engineering, Digital Energy and Electric Power, Technikumsstrasse 21, 6048, Horw, Lucerne, Switzerland; Jeremy Taylor, jtaylor@viascience.com, Via Science, 666 Sherbrooke Street West, Montreal, Québec, Canada, QC H3A 1E7; Daniel Raimundo, daniel.raimundo@hslu.ch; Antonios Papaemmanouil, antonios.papaemmanouil@hslu.ch, Lucerne University of Applied Sciences and Arts, Department of Electrical Engineering, Digital Energy and Electric Power, Technikumsstrasse 21, 6048, Horw, Lucerne, Switzerland.

1.1 Structure of this Paper

Within this paper, the development, implementation and test of a decentralised power flow forecasting algorithm is described. First, the state-of-the-art for the electric load forecasting and decentralised power flow methods is discussed (Sec. 2). Secondly, Sec. 3 introduced the developed methods within this paper. Thereafter, Sec. 4 describes the simulation-based results. Sec. 5 describes the test system and the outcome of the field trials. Lastly, Sec. 6 concludes the paper.

2 RELATED WORK

This section discusses previously published papers and experiments related to the topics of forecasting electrical load and decentralised power flow methods. It establishes the relevant state-of-the-art for the methodology and results of this paper.

2.1 Electric Load Forecasting

Electric load forecasting involves predicting the electricity demand of an individual end-customer or a group of end-customers within the grid. This process is essential to maintain the reliability and efficiency of the electric power system, as it allows utilities to anticipate the power required to achieve the equilibrium of supply and demand [1].

This forecasting process entails anticipating future load profiles by analyzing historical data and prevailing trends. These projections are essential for the efficient management of energy resources, the streamlining of grid operations, and the optimal use of the integration of renewable energy sources. [16].

For continuous datasets like electric load profiles, time series prediction algorithms prove highly effective. Recent research findings [3, 5, 7, 16] demonstrate that modern neural network architectures generally outperform traditional statistical methods in forecasting. Among these architectures, Convolutional Neural Networks (CNNs) have emerged as particularly superior, as highlighted by [2, 5, 7, 10, 16].

The Temporal Convolutional Neural Network (TCN) stands out in electric load forecasting. This network comprises stacked convolutional layers together with dilutions and residual connections to accommodate extensive historical data. The work in [2, 5, 10] has shown that TCN outperforms other architectures.

Similarly the DeepEnergy Network, introduced by [7], shares a similar architecture with TCN, comprising stacked convolutional networks alongside alternating pooling layers. Within their research [7] found that the DeepEnergy Network also provides superior performance compared to other state-of-the-art architectures.

2.2 Decentralised Power Flow

Efficient and robust computation of power flow solutions is at the core of many applications in distribution systems, such as the computation of network congestion and voltage levels in distribution networks. Furthermore, as distributed energy resources (DERs) become more prevalent, power flow studies at the distribution system level are growing in significance. Digitalization efforts are underway to facilitate more advanced grid analysis, extending down to the individual end-consumer level. Traditional power flow analysis techniques are typically rooted in centralised frameworks. These approaches facilitate modelling and analysis of (potentially bidirectional) power flows, voltage regulation, and system constraints inherent in distribution systems.

Recent research efforts have explored various methodologies to extend power flow analysis to decentralised settings, which offer the benefit of distributing the computational effort on multiple sub-area controllers, and preservation of data privacy (e.g., grid models or end-consumer load profiles) [11].

There are several approaches using consensus-based techniques to solve the non-linear power problem. These include the Alternating Direction Method of Multiplyers (ADMM) and the Augmented Lagrangian based Alternating Direction Inexact Newton method (ALADIN) to solve decentralised power flow problems [4, 11].

Additionally, decentralised solutions tailored to the *optimal* power flow problem can be adapted to solve the power flow problem by tightening the bounds on decision variables, such as the methods in i.e., [12, 13]. However, deploying distributed optimisation solvers for the power flow problem appears excessive, particularly for the modest computational capabilities of grid-edge devices (i.e. smart meters).

Alternatively, the direct power flow, or also known as backward-forward sweep power flow method may be leveraged to solve the radial nature of LV distribution systems [6, 14]. In this paper, it is demonstrated how the decomposability of the computation matrices lends itself to an iterative decentralised implementation. This enables safeguarding of data privacy and limited requirement of grid model for each sub-area controller. Only the electrical properties of grid connections (i.e., line impedances) to adjacent sub-area controllers are required.

3 METHOD

The objective of the approach in this paper is to compute past and future decentralised load and network power flows. It builds on two main building blocks: The first, is local electric load forecasting, and the second is a decentralised power flow approach. Within this section, both building blocks are presented.

3.1 Electric Load Forecasting

The electric load forecasting in this paper is based on the work of [2, 5, 7, 16]. In this work, a comparison of 11 types of deep neural network (DNN) architectures is made, applied to load profile forecasting, with the principal architectures being FNN, TCN, ERNN, LSTM, GRU and Seq2Seq. Individual and aggregate household data are used to train the models, and comparisons are made on various metrics. The work of [5] was repeated using the Low Carbon

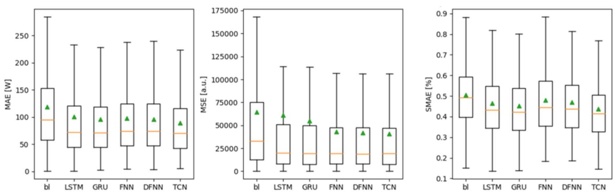


Fig. 1. Model performance of baseline and forecasting models, based on [5]

London dataset as a data source. An initial comparison of model performance identified the TCN architecture as the best performing. This is illustrated in Figure 1, where different mean average error (MAE), mean squared error (MSE) and symmetric mean absolute percentage error (SMAPE) values are plotted for different architectures from [5]. All perform better than the baseline (denoted by 'bl'), which is a simple forecast based on repetition of the previous day's load profile. Mean errors and standard deviations are shown for the models in consideration, with the TCN network providing the best results.

The chosen model architecture was subsequently optimised using a Hyper-BandForBOHB optimiser [8, 15]. The resulting TCN Model is comprised of eight dilated TCN layers, with each layer containing 64 filters for feature extraction, and a subsequent FNN backbone for feature selection. Overall, the optimised model consisted of 760,800 trainable parameters as presented in [15].

The implementation and testing of the developed load forecasting models are documented in Sec. 4.1.

3.2 Decentralised Power Flow

Traditional power flow algorithms, such as the Newton-Raphson or Gauss-Seidel algorithms require computationally challenging LU decomposition of the Y admittance matrix and inversion of the Jacobian matrix, respectively. To take advantage of the topological characteristics of distribution systems, [14] has proposed a direct power flow approach, which makes use of oriented bus-branch information.

The distribution system is modelled with N buses collected in the set $\mathcal{N} = \{1, \dots, N\}$, and $N-1$ branches collected in the set $\mathcal{B} = \{1, \dots, N-1\}$. It is assumed that bus 1 is the substation representing the slack bus. The substation bus voltage is fixed and known, and its power injection into the distribution feeder is modelled. Each line segment $(i, j) \in \mathcal{B}$ is represented by its equivalent series impedance $z_{ij} = r_{ij} + jx_{ij}$ where r_{ij} is the line resistance and x_{ij} the line reactance. The shunt admittance will be neglected for the short lines of the distribution system. Further the set $\mathcal{N}^- = \mathcal{N} \setminus \{1\}$ represents all buses expect for the slack bus. The complex voltage at bus $i \in \mathcal{N}$ is denoted V_i; and P_i and Q_i denote the active- and reactive-power injection (load - generation) at bus $i \in \mathcal{N}$. The complex power at bus i is:

$$S_i = P_i + jQ_i \tag{1}$$

The resulting equivalent current injection at bus i is

$$I_i = \left(\frac{P_i + jQ_i}{V_i}\right)^* \tag{2}$$

The direct power flow approach is developed based on the bus-injection to branch-current (BIBC) and the branch-current to bus-voltage (BCBV) matrix and the equivalent current injections from buses of the distribution network [14]. The branch currents are collected in $I_{\mathrm{Branch},ij}$ for all branches $(i, j) \in \mathcal{B}$ in vector $I_{\mathrm{Branch}} = [\{I_{\mathrm{Branch},ij}\}_{(i,j)\in\mathcal{B}}]^{\mathrm{T}}$, and the current injections I_i for all buses $i \in \mathcal{N}$ in vector $I = [\{I_i\}_{i\in\mathcal{N}}]^{\mathrm{T}}$. The BIBC matrix maps bus current injections to branch currents, such that

$$I_{\mathrm{Branch}} = \mathbf{BIBC}\, I. \tag{3}$$

Furthermore, the BCBV matrix maps branch currents to bus voltage differences, such that

$$\Delta V = \mathbf{BCBV}\, I_{\mathrm{Branch}}, \tag{4}$$

where $\Delta V = [\{V_1 - V_i\}_{i\in\mathcal{N}^-}]^{\mathrm{T}}$ collects the bus voltage differences at all the buses $i \in \mathcal{N}^-$ from the voltage at the substation V_1. A detailed derivation on how to build the BIBC and BCBV matrices is given in [14] based on the network topology and line impedances. The direct power flow approach is based on a backward/forward sweep algorithm, in which for each iteration ℓ the following Algorithm 1 executed.

Initialization:
Set initial condition $V_i^0 = 1.0$ p.u. $\forall\, i \in \mathcal{N}$;
Calculate the equivalent current injection at all buses using (2);
$\ell \leftarrow 0$;
while *Power flow algorithm not converged* **do**
  $\ell \leftarrow \ell + 1$;
  Backward sweep:
  Calculate all branch currents using (3);
  Forward sweep:
  Calculate all bus voltage differences using (4);
  Calculate bus voltage at all buses using $V_i^\ell = V_i^0 - \Delta V_i$;
  Check convergence:
  if $\max_{i=1\dots N}\left\{\left|V_i^\ell - V_i^{\ell-1}\right|\right\} < \varepsilon$ **then**
    Exit loop;
  end
end

Algorithm 1: Direct Power Flow Algorithm

Once the algorithm has converged, all bus voltages are obtained, branch currents are provided in B, and simple arithmetics can be applied in order to determine branch power flows.

Hereafter, we introduce a decentralised power flow method based on the direct power flow approach. We partition the distribution network into non-overlapping areas collected in the set $\mathcal{A}$. Within area $\alpha \in \mathcal{A}$, the network buses are collected in the set $\mathcal{E}^\alpha$. For distinct areas $\alpha, \gamma \in \mathcal{A}$, $\mathcal{E}^\alpha \cap \mathcal{E}^\gamma = \emptyset$. We denote the upstream bus of area $\alpha \in \mathcal{A}$ by $\mathcal{B}_{\text{up}}^{\alpha}$, and collect the downstream buses of area $\alpha \in \mathcal{A}$ in $\mathcal{B}_{\text{down}}^{\alpha}$. Furthermore, we define the area upstream of area α in $\mathcal{A}_{\text{up}}^{\alpha}$ and the areas downstream from area α in $\mathcal{A}_{\text{down}}^{\alpha}$.

For all areas $\alpha \in \mathcal{A}$, we collect the equivalent current injections at buses within the area in vector $I^{\alpha} = [I_{i \in \mathcal{E}^{\alpha}}]^{\mathrm{T}}$, and the branch currents upstream of buses $i \in \mathcal{E}^{\alpha}$ in vector $I_{\text{Branch}}^{\alpha} = [\{I_{\text{Branch},ij}\}_{(i,j)\in\mathcal{B}, i \in \mathcal{E}^{\alpha}}]^{\mathrm{T}}$. The sub-area BIBC matrix maps bus current injections to branch currents such that:

$$I_{\text{Branch}}^{\alpha} = \mathbf{BIBC}^{\alpha} I^{\alpha}. \tag{5}$$

In addition, for the bus upstream of area α which connects to the first bus within area α, we add the current of the first bus in area α ($I_{\text{Branch},0}^{\alpha}$) to the bus injection so that the load of downstream areas is considered, as expressed in:

$$I_i^{\gamma} = I_i^{\gamma} + I_{\text{Branch},0}^{\alpha} \tag{6}$$

$$\forall \alpha \in \mathcal{A}, \gamma \in \mathcal{A}_{\text{up}}^{\alpha}, i \in \mathcal{B}_{\text{down}}^{\gamma} \tag{7}$$

Furthermore, the sub-area BCBV matrix maps branch currents to bus voltage differences. For the first area connecting to the substation, the voltage drop due to loads is computed as follows:

$$\Delta V^{\alpha} = \mathbf{BCBV}^{\alpha} I_{\text{Branch}}^{\alpha}. \tag{8}$$

where $\Delta V^{\alpha} = [\{V_1 - V_i\}_{i \in \mathcal{E}^{\alpha}}]^{\mathrm{T}}$ collects the differences in bus voltage in all buses in the area α, $i \in \mathcal{E}^{\alpha}$ from the voltage at the substation V_1. In order to also consider the voltage drop due to upstream loads within area α, for all areas that are not connected to the substation, the voltage drop at the bus of the upstream area connecting area α needs to be considered:

$$\Delta V^{\alpha} = \mathbf{BCBV}^{\alpha} I_{\text{Branch}}^{\alpha} + \Delta V_{i, i \in \mathcal{B}_{\text{up}}^{\alpha}}. \tag{9}$$

Finally, convergence needs to be checked for each area according to:

$$\max_{i = \mathcal{E}^{\alpha}} \left\{ \left| V_i^{\ell} - V_i^{\ell-1} \right| \right\} < \varepsilon, \tag{10}$$

where ℓ indicates the iteration index. Once the voltage deviation in each area has converged to some tolerance ε, the algorithm has completed.

3.3 Implementation

The decentralised power flow method can be separated onto multiple edge devices (i.e. smart meters), where each edge device represents a sub-area and only needs to communicate with its upstream and downstream neighbor edge devices. Additionally, in case some meters do not provide computational resources, the proposed algorithm is flexible in such a way that the computation can be executed on neighbour devices, which perform computations for an area of smart meters in the network.

The approach is illustrated in Figure 2. The most important factors to consider are:

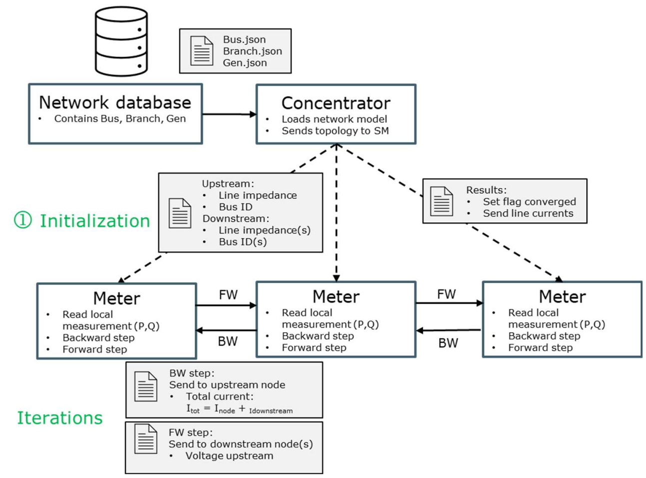
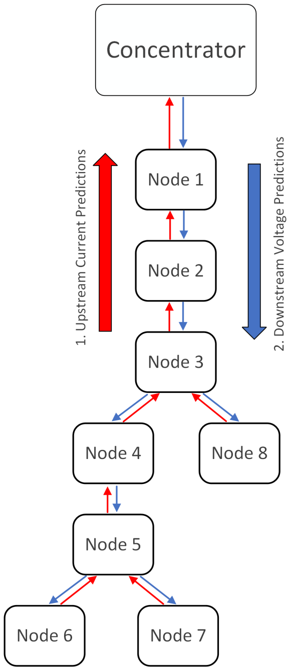


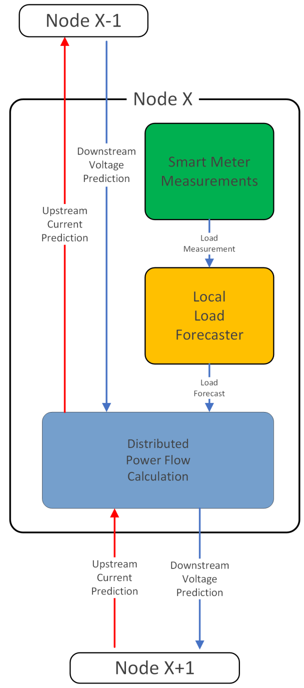
Fig. 2. Diagrammatic Illustration of Decentralised Power Flow Approach

1. Each participating edge-device (i.e. smart meter) requires a subset of the network topology information, such as upstream and downstream smart meter IDs as well as electrical connectivity information (i.e. line impedances that connect adjacent smart meters).
2. A concentrator loads a network topology from a data base, sends topology information to participating edge-devices, and initiates the forecasting and decentralised power flow algorithm on each edge-device.
3. The smart meter's memory are uniquely defined by the equipment installed and its connectivity: Thus this is known to the DSO and is not private and does not change unless the network topology is changed.
4. The system can be partitioned in arbitrary divisions: partitioning can be defined to suit the computational resources in the network.
5. Computation is decentralised: The updated elements are uniquely computed by the edge-device that is responsible for the area. No consensus is required.
6. Communication is a simple broadcast: So long as each processor can hold the whole current/voltage vector in memory, all that is required is to ensure that the neighbors receive the entire updated vector.
7. Privacy: Sensitive data from end customers (electrical load profiles), are kept within the edge-device, only aggregated information is exchanged with neighboring nodes.

Figure 3 shows an illustration of the decentralised load forecasting and power flow algorithm. Each node representing an area of the distribution system is connected to its neighbours and exchanges information with its upstream and downstream neighbours, as shown in Fig. 3a. Fig. 3b illustrates the processes taking place within each node of the decentralised system which are the extraction of smart meter measurements, the local electric load forecasting and the iterative decentralised power flow calculation.



(a) Decentralised power flow algorithm overview



(b) Processes within each node

Fig. 3. Illustration of decentralised load forecasting and power flow algorithm

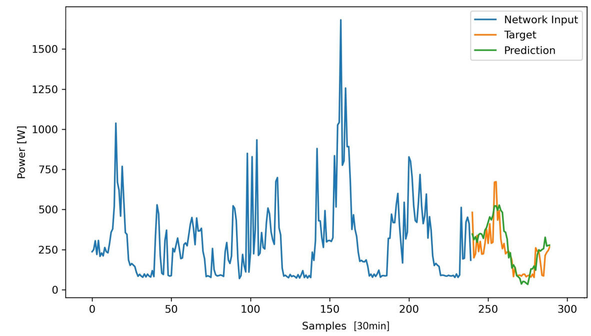


Fig. 4. Sample Prediction of Trained TCN Model [15]

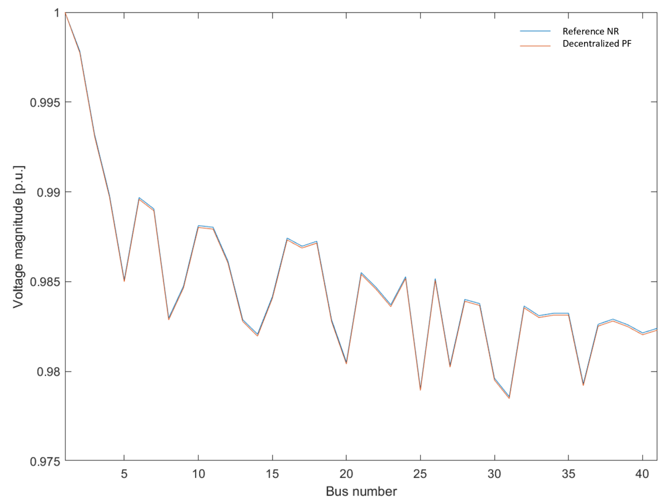


Fig. 5. Voltage profile obtained via the Newton-Raphson power flow algorithm compared to the decentralised power flow method

4 SIMULATION RESULTS

This section presents the results of the electric load forecasting method and the decentralised power flow.

4.1 Electric Load Forecasting

The developed and optimised load forecasting model underwent testing using unseen test data derived from the same household used to train the forecasting model. This approach ensured that the model was evaluated on unseen samples originating from a familiar data source. The forecasting results are visualized in Fig. 4, demonstrating the model's efficacy in capturing electric load patterns and generating reliable day-ahead electric load values.

A specific performance metric is intentionally omitted, acknowledging the substantial impact that variations in source data sets can have on the overall model performance, particularly those with significant randomness. This is known as the domain shift problem [15, 16], where more detailed results on the forecasting performance are provided.

4.2 Decentralised Power Flow

The decentralised power flow method was compared with a standard Newton-Raphson power flow solver to validate its performance against a conventional approach. Remarkably similar performance was observed in power flow computation, with the maximum error observed being 4.988e-7 p.u.

Figure 5 illustrates the voltage profile of the Romande Energie sample test network obtained using the PandaPower Newton-Raphson power flow solver compared to the decentralised power flow method. In fact, the results demonstrate that the two methods are very closely in match.

5 FIELD TRIAL

5.1 Test system description

The Field tests were carried out within the Romande Energie area in Rolle, using selected infrastructure from the project P+D Reel Demo – FURIES (SFOE contract SI/501523). The site at Rolle includes 70 phasor measurement units (PMU's), 100 GridEye units, 750 smart meters, a battery energy storage system, integrated PV, seven remotely controlled stations and a data management system, spread within 36 local LV systems. Its simulated topology is described in Fig. 6. There are 30 smart meters, each representing an area within

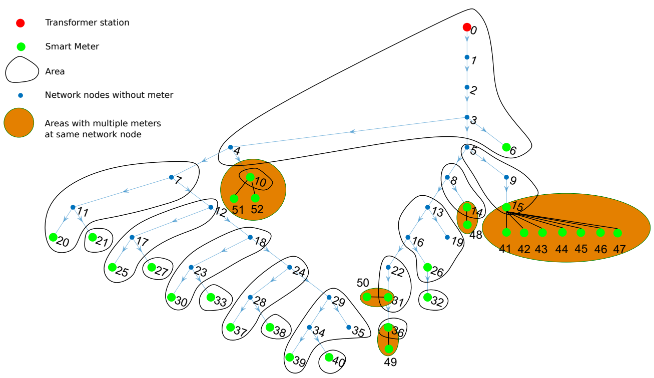


Fig. 6. Network topology of Romande Energie network with partitioning into areas

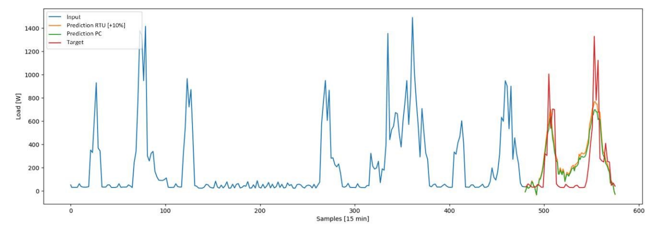


Fig. 7. Comparison of Deployed and Local Execution of the Electric Load Forecasting on Test Data

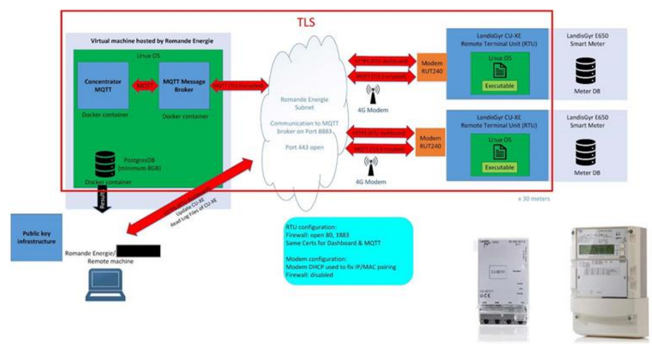


Fig. 8. ICT Concept of the Smart Meter Deployment in Role

the network. The areas with orange background illustrate smart meters, which are all connected to the same network node, for example a multi-family home.

A test feeder has been identified within the Rolle area with a mix of multi-family homes, single-family homes, and PV generation units. 30 metering points have been identified within a feeder that was previously equipped with a Depsys GridEye device. The meters in the feeder were replaced with Landis+Gyr S650 industrial smart meters and Landis+Gyr CU-XE remote terminal units (RTU) for the field trials in order to execute the algorithms developed.

The RTU is a communication unit provided by Landis+Gyr which offers a Linux environment to execute dedicated scripts for the field trials. This setup allows algorithms to be deployed and operated in customer premises. Customers were approached for an opt-in to participate in the field trials.

In addition to the deployment of smart meters and RTUs at the customer sites, an IT infrastructure has been established to enable safe and secure meter-to-meter communication.

5.2 Deployment of algorithms

To prepare the field trial, a load monitoring script, the load forecasting algorithm, and the decentralised power flow calculation were deployed onto the Remote Terminal Unit (RTU) that is connected to the S650 Landis+Gyr smart meters. The purpose of the load monitoring script is to write the current end-customer load values in a log file, which is kept only local to ensure data privacy.

The scripts deployed were tested within a controlled test environment to ensure successful execution and functionality. The test execution of the load forecasting script is depicted in Figure 7.

5.3 Smart Meter Communication Setup

The field trial setup uses a Romande Energie private metering sub-net via a Global System for Mobile Communication (GSM) modem to enable communication between the RTUs and the concentrator hosted on a virtual machine within a Kubernetes cluster by Romande Energie. A custom encrypted MQTT messaging service has been used to transfer relevant data between smart meters. All communication links have been secured via Transport Layer Security (TLS) encryption both on the MQTT protocol and on the Secure Hypertext Transfer Protocol (HTTPS). Romande Energie has managed the TLS certificates via their public key infrastructure. From the virtual machine hosted on a Kubernetes cluster, a concentrator as well as a messaging broker were implemented to orchestrate the algorithms. The aggregated power flow forecasts were written to a data base on the virtual machine, which enabled extraction of results from an operator via remote access to the virtual machine.

Finally, the communication has been configured in a way that HTTPS communication allowed remote access to the RTUs web server and enabling remote firmware updates of the RTUs. Figure 8 illustrates an overview of the IT-infrastructure established for the field trials. Figure 9 illustrates smart meters that have been deployed at a multi-family home with 5 end customers, including the GSM modem and a switch to enable the communication to the virtual machine.

5.4 Field Trial Outcomes

Figure 10 illustrates the field trial results for voltage profiles in all nodes (left) and currents transmission lines/transformers (middle) for an example 15-minute interval. On the (right), the prediction of

Fig. 9. Mounted Smart Meters in an End-Users Basement

the transformer load for the next 24 hours is illustrated. The load flow analysis for the past 24 hours using synthetic data aligns with the results obtained from the offline simulations.

5.4.1 Performance. The load flow analysis and forecasting are initiated every full hour, analysing 15-minute values from the past and future 24 hours of load flow across all segments and voltages in each node within the test network. This results in a substantial dataset, comprising approximately 20,000 data points per hour calculated across 192-time intervals, 50 nodes, and 49 edges. In the distribution network of Rolle with 30 end customers, the runtime for both forecasting and decentralised power flow is approximately 3.5 minutes. Figure 11 provides a histogram of the execution times of the algorithm for two weeks. It is important to note that the runtime is contingent upon the size and topology of the distribution grid due to the sequential nature of the algorithm. For this relatively limited size distribution grid (feeder) it is possible to execute the algorithm in about 3.5 minutes. This is well under 15 minutes, which enables execution in near real-time. The complexity lies in managing the timing and synchronisation of various message flows during algorithm execution, presenting a challenge to maintain efficiency and accuracy in the decentralised power flow and forecasting processes. This has been achieved by properly synchronising the individual edge-devices/smart-meters through the concentrator.

5.4.2 Deployment of smart meters. In the initial phase of the field test, the 30 customers in the test system were contacted through opt-in letters, with the aim of replacing the existing smart meters with new ones hosting the algorithms introduced above in their homes. Remarkably, 29 of the end-customers contacted accepted to participate in the trial, agreeing with the replacement of the smart meter. Some technical challenges were encountered during the installation process, as communication could not be established with four smart meters due to issues such as poor reception and incorrect manipulation. Subsequently, during the test phase, communication was unfortunately lost with three additional smart meters. Despite these setbacks, the final tests were conducted with 22 real smart meters, providing valuable insights in the performance and reliability of the approach. This comprehensive field test has not only highlighted areas for improvement but also affirmed the functionality of the majority of the deployed smart meters. Figure 12 illustrates the availability of smart meters for the field trials with 22 smart meters remaining for the field trials.

5.4.3 Field test limitations. Due to the shortcomings described above, 8 smart meters had to be virtualized, making live data integration impossible for these end-customers. While the algorithm is adaptable in the number of end-customers that are included in the calculations, simply dropping 8 smart meters would distort the final power flow results. Consequently, synthetic load profiles were generated to simulate the missing smart meter readings. The live version encountered further challenges as the reading of measured values on smart meters could only be achieved within the controlled environment of the laboratory test setup. Unfortunately, deploying the live version in the actual setup was impeded, requiring an essential software update that could technically be implemented remotely due to the flexibility of the IT infrastructure. Despite these efforts, the field test phase, which lasted only one month, imposed limitations on the opportunity for additional iterations of firmware updates. Consequently, the results obtained do not authentically reflect real physical conditions, making direct comparisons with transformer measurements challenging. While troubleshooting of these limitations would be straightforward, the constraints of the field test duration hinder a more comprehensive assessment and refinement of the field trial results via an additional firmware update.

6 CONCLUSION

The development and execution of decentralised power flow calculation and forecasting in near real-time is the main contribution of this paper. The proposed method facilitates computation and forecasting of network power flows and congestion levels at the edge on real smart meters. Unlike existing smart meter systems, which transmit data to grid operators only once every 24 hours, this approach enables near real-time availability of various data streams in 15-minute intervals. Moreover, the system prioritizes data protection and processes end-consumer data locally were it is collected, leading to more efficient data processing.

The solution serves as a viable alternative to the installation of supplementary measuring devices at the transformer circuit level, streamlining the integration process. This decentralised approach requires smart meters with increased computing capabilities for the additional functionality. The implementation of ICT in the DSO, which features a direct IoT interface using the MQTT secure protocol, introduces a novel way to interact with field devices for DSOs. Looking ahead, the potential for future industrialisation of such solutions holds promise for improved visibility into distribution grids, offering a pathway toward optimising grid planning and operation for increased efficiency and reliability.

ACKNOWLEDGMENTS

This work was supported by the research grant from the Swiss Federal Office of Energy, grant number SI/502076-01. The authors also acknowledge the support and collaboration with industry partners Landis+Gyr AG and Via Science and the distribution system operator Romande Energie.

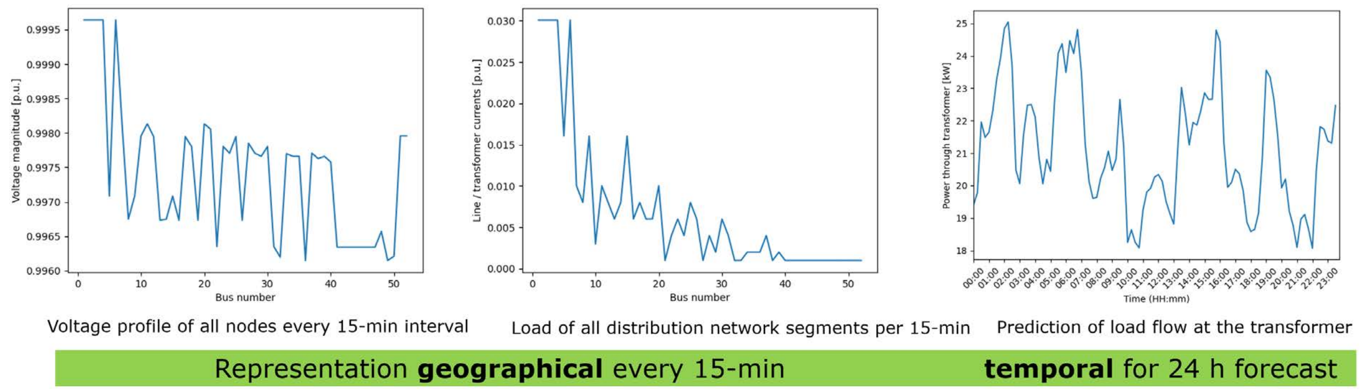


Fig. 10. Summary of power flow and forecasting results

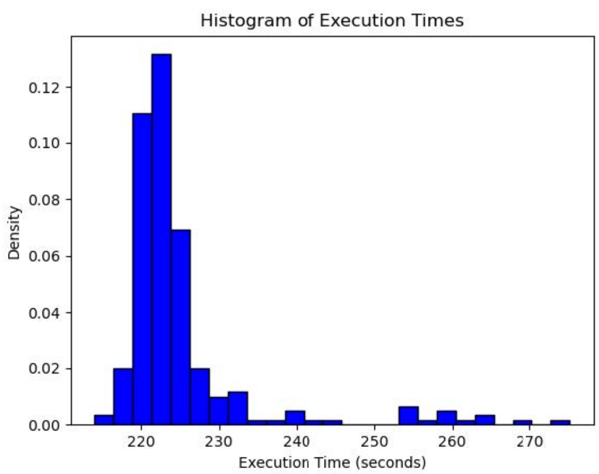


Fig. 11. Histogram Execution Times

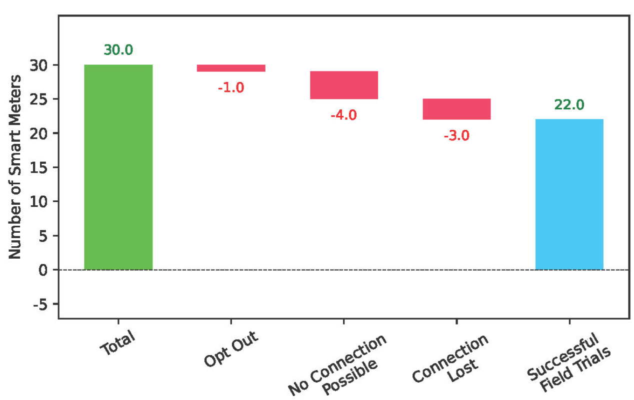


Fig. 12. Available Smart Meters during Field Trials

REFERENCES

[1] Hesham K. Alfares and Mohammad Nazeeruddin. 2002. Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science* 33, 1 (2002), 23–34. https://doi.org/10.1080/00207720110067421

[2] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271 [cs.LG]

[3] European Commission, Directorate-General for Energy, F Tounquet, M Linden, P Mandatova, J Knapp, J Van Tilburg, A Dadkhah, A Villar Lejarreta, S Yordanov, J Van Roy, C Alaton, M Goes, and P Gentili. 2020. *Supporting country fiches accompanying the report Benchmarking smart metering deployment in the EU-28.* Publications Office, European Commission Enterprise, Brussels, Belgium. https://doi.org/10.2833/728829

[4] Xinliang Dai, Yichen Cai, Yuning Jiang, and Veit Hagenmeyer. 2022. Rapid Scalable Distributed Power Flow with Open-Source Implementation. *IFAC-PapersOnLine* 55, 13 (2022), 145–150. https://doi.org/10.1016/j.ifacol.2022.07.250 9th IFAC Conference on Networked Systems NECSYS 2022.

[5] Alberto Gasparin, Slobodan Lukovic, and Cesare Alippi. 2022. Deep learning for time series forecasting: The electric load case. *CAAI Transactions on Intelligence Technology* 7, 1 (2022), 1–25.

[6] Andreas Kotsonias, Lenos Hadjidemetriou, and Elias Kyriakides. 2019. Power Flow for a Four-Wire Radial Low Voltage Distribution Grid with a Single Point Grounded Neutral. In *IEEE Xplore*. IEEE. https://doi.org/10.1109/ISGTEurope.2019.8905641

[7] Ping-Huan Kuo and Chiou-Jye Huang. 2018. A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting. *Energies* 11, 1 (2018), 100254. https://doi.org/10.3390/en11010213

[8] Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez, and Ion Stoica. 2018. Tune: A Research Platform for Distributed Model Selection and Training. *CoRR* abs/1807.05118 (2018), 10534. arXiv:1807.05118 http://arxiv.org/abs/1807.05118

[9] Swiss Federal Office of Energy. 2023. Energiestrategie 2050. online. https://www.bfe.admin.ch/bfe/de/home/politik/energiestrategie-2050.html/

[10] Elena Mocanu, Phuong H. Nguyen, Madeleine Gibescu, and Wil L. Kling. 2016. Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks* 6 (2016), 91–99. https://doi.org/10.1016/j.segan.2016.02.005

[11] Tillmann Mühlpfordt, Xinliang Dai, Alexander Engelmann, and Veit Hagenmeyer. 2021. Distributed power flow and distributed optimization—Formulation, solution, and open source implementation. *Sustainable Energy, Grids and Networks* 26 (2021), 100471. https://doi.org/10.1016/j.segan.2021.100471

[12] Severin Nowak, Yu Christine Chen, and Liwei Wang. 2022. Distributed Measurement-Based Optimal DER Dispatch With Estimated Sensitivity Models. *IEEE Transactions on Smart Grid* 13, 3 (2022), 2197–2208. https://doi.org/10.1109/TSG.2021.3139450

[13] Qiuyu Peng and Steven H. Low. 2018. Distributed Optimal Power Flow Algorithm for Radial Networks, I: Balanced Single Phase Case. *IEEE Transactions on Smart Grid* 9, 1 (2018), 111–121. https://doi.org/10.1109/TSG.2016.2546305

[14] Jen-Hao Teng. 2003. A direct approach for distribution system load flow solutions. *Power Delivery, IEEE Transactions on* 18 (08 2003), 882 – 887. https://doi.org/10.1109/TPWRD.2003.813818

[15] Fabian Widmer, Roman Loetscher, Daniel Raimundo, Severin Nowak, Peter Allenspach, and Antonios Papaemmanouil. 2024. Optimizing Energy Flexibilities in Local Energy Communities A Case Study on the Development of an Energy

Management System Incorporating Consumption and Production Forecasting Methods. In *Energy Informatics 2024*. AMC, Technikumsstrasse 21 6048 Horw switzerland, 9.

[16] Fabian Widmer, Severin Nowak, Benjamin Bowler, Patrick Huber, and Antonios Papaemmanouil. 2023. Data-driven comparison of federated learning and model personalization for electric load forecasting. *Energy and AI* 14 (2023), 100253. https://doi.org/10.1016/j.egyai.2023.100253

The plan4.energy Approach for Planning Support for Positive Energy Districts

YANNICK WIMMER*, AIT Austrian Institute of Technology, Austria
DANIEL SCHWABENEDER*, AIT Austrian Institute of Technology, Austria
MICHAEL NIEDERKOFLER†, Innovation Lab act4.energy, Austria
PATRIZIA-ILDA VALENTINI‡, Mobilize Financial Services, Austria

This paper introduces a software solution for planning support of Positive Energy Districts (PEDs), highlighting the challenges of interdisciplinary planning and stakeholder involvement. The development of the software tool, *plan4.energy*, addresses these challenges by providing stakeholders with accessible means to assess project variants and understand key planning levers. Based upon a database of over 300,000 simulation results and a user-friendly interface, *plan4.energy* facilitates informed decision-making for PED planning, enhancing the integration of renewable energy sources and optimizing system-level performance.

Additional Key Words and Phrases: Positive Energy Districts, Sustainable Urban Development, Integrated Energy Modeling

1 INTRODUCTION

Positive Energy Districts (PEDs) as defined in [12] represent a new approach to the implementation of sustainable system solutions that can contribute to the achievement of objectives within the framework of local climate protection strategies. This approach bundles previous technological developments and combines them at system level in order to achieve a positive effect on the energy system. As the planning of PEDs requires interdisciplinary approaches, the multiple interdependencies may be challenging to grasp, particularly for stakeholders who are not involved with the technical details on a daily basis (such as mayors, real estate developers, architects, planners and others). The project *plan4.energy* therefore aimed to develop a software tool designed to assist in the planning process of PEDs. Its main aim is to realistically visualize complex interdependencies and identify the key levers during the planning process. This enables stakeholders without a specialist background to evaluate project variants. This is achieved by combining a large database of pre-simulated data with an intuitive online frontend.

2 MOTIVATION

Tools that use simulations incorporating components from all mentioned energy sectors with detailed profiles are already available, as demonstrated in the overview study [19]. These solutions are, however, typically targeted towards users which are able to specify in detail the components and input parameters and thus not suitable for a quick assessment. Others lack key components such as an economic analysis. On the other hand simple decision support tools for stakeholders are also available [7, 11], these approaches, however, lack the necessary depth of analysis.

Authors' addresses: Yannick Wimmer, yannick.wimmer@ait.ac.at, AIT Austrian Institute of Technology, Giefinggasse 4, Vienna, Austria, 1210; Daniel Schwabeneder, daniel.schwabeneder@ait.ac.at, AIT Austrian Institute of Technology, Vienna, Austria; Michael Niederkofler, niederkofler@energiekompass.at, Innovation Lab act4.energy, Stegersbach, Austria; Patrizia-Ilda Valentini, patrizia-ilda.valentini@mobilize-fs.com, Mobilize Financial Services, Vienna, Austria.

3 SIMULATION FRAMEWORK AND DESIGN

The simulation framework upon which the *plan4.energy* tool is based contains many individual component models, which are integrated to form a comprehensive system for the simulation. The characteristic data needed for each of the components models or load profiles is given in the yellow boxes in Fig. 1. The following components and variable input profiles were taken into account in the simulation framework as part of *plan4.energy* (see also Fig. 1):

- Three building types
  - Single-family house (SFH)
  - Multi-family house (MFH)
  - Office building
- Building sizes (various sizes)
- Building efficiency classes (A, B, C, D)
- Photovoltaic system
- Heating systems
  - Air/water heat pump
  - Gas
  - Oil
  - Pellets
  - District heating
  - Water/water heat pump (energy recovery from wastewater)
- Floor heating
- Automatic shading
- Fossil energy demand for individual mobility
- Charging profiles for electric vehicles
- Cooling system
  - Air-conditioning
  - Cooling via heat pump
- Heating and cooling demand for different building types and efficiency classes
- Hot water demand
- Hot water storage
- Battery storage
- Grid electricity consumption (including various load profiles)

Apart from the detailed component models, the primary complexity lies in obtaining suitable input profiles and determining a reasonable operational strategy. One of the key strengths of the proposed approach is the ability to compare various combinations and sizes of the components in using different annual demand profiles (as shown in Fig. 1). This results in a database containing 258,048 pre-simulated combinations for single-family homes (SFH), 26,880 for multi-family homes (MFH) and 60,480 for office buildings. In the frontend (see section 4) these can be combined with PEDs, which allows for a quick assessment of countless combination possibilities of buildings forming such a PED. In Fig. 2 illustrates the simulation principle for a simulation run. Note that all results for the *plan4.energy* tool were calculated using profiles in 15 minutes resolution in order to holistically capture possible simultaneities in generation and load profiles, which is of utmost importance for the operation strategy of potential flexibilites. The operational strategy is determined by an in-house developed linear optimizer [22] which utilizes these flexibilites to establish a possible strategy for each subsequent 24-hour period. However, this linear optimizer

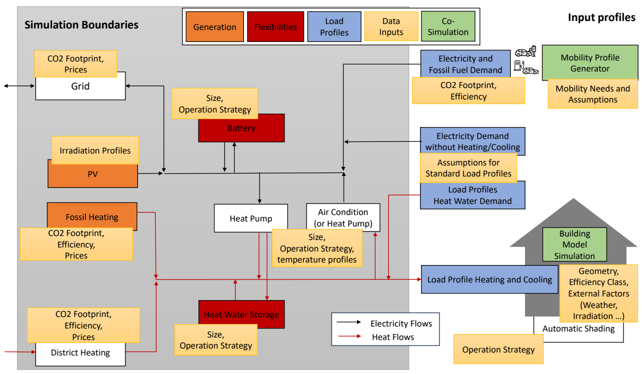


Fig. 1. plan4.energy simulation setup: Generation components are depicted in orange, flexibilities in red, load profiles in blue. Where co-simulation was used for profile generation, this is marked in green. The yellow boxes define the required data when modelling or simulating the specific components or profiles.

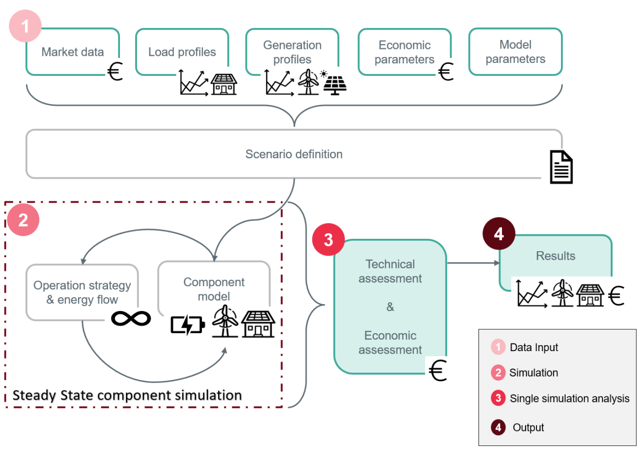


Fig. 2. plan4.energy Schematics of a typical simulation run. Note that all simulations are based on profiles with a 15 minutes resolution in order to properly account for simultaneities.

uses only linearized models, whereas the component model implemented in the simulation framework is considerably more detailed. Consequently, the schedules are subsequently evaluated using these more detailed simulation models, resulting in more accurate outcomes. This ensures both an optimized operational strategy and detailed component simulations, which would not be possible with linearized models alone. The quality of a simulation is, however, always heavily dependent on proper input data. Therefore, the used input data and profiles will be discussed as follows:

3.1 Photovoltaics

The sizes of the Photovoltaics (PV) systems and battery systems are based on the experience of the partners in the project (see Tab. 1). They are relatively larger for very small properties, where the ratio of roof area to living space is greater. The PV-sizes are based on the m^2 values, typical values of living space in m^2 are derived from the statistics for residential buildings: For the multi-family houses, the average apartment size results in around 73 m^2/unit [21] (in the project this was rounded to 75 m^2).

Table 1. PV sizes for different building types and sizes

Building Type and m^2	PV size (kWp)		
	Small	Medium	Large
SFH 50	4	6	8
SFH 150	5	10	15
SFH 400	10	20	30
MFH 300	4.8	9	18
MFH 1500	16	30	60
MFH 7500	80	150	300
Office 300	4.8	9	18
Office 5000	80	150	300
Office 25000	400	750	1500

3.2 Battery

The assumed battery sizes are also based on the experience of the individual partners. They are based on the ratio to the various PV-sizes (see above)

Table 2. Battery sizes for different building types and sizes

Building Type and m^2	Battery size (kWh)		
	Small	Medium	Large
SFH 50	3.2	4.8	6.4
SFH 150	4	8	12
SFH 400	8	16	24
MFH 300	2.4	4.5	9
MFH 1500	8	15	30
MFH 7500	40	75	150
Office 300	2.4	4.5	9
Office 5000	40	75	150
Office 25000	200	375	750

3.3 Load Profiles

One of the project partners provided the project with a large number of metered load profiles from the DACH region. The load profiles from Austria without a battery storage system were analyzed in detail to determine how they could be clustered in a meaningful way. The majority of the profiles were available from 2019, among other years. As the last fully available pre-pandemic year, this was therefore the logical choice as a reference year. The aim was initially to use averaged measured load profiles for the simulations. However, it proved difficult to identify which profiles may contain heat pumps or charging profiles of Battery Electric Vehicles (BEVs). By adding BEVs charging or heat pump profiles in addition to the average profiles, the contribution of those components would then be counted twice. The use of a very strongly averaged profile is also counterproductive, as the consumption and generation peaks are precisely the effects that determine the need for flexibility and any necessary storage sizes. As an alternative approach, an attempt was made to compare available synthetic load profiles [15] with the cluster results. Based on these results it was concluded that for the tool should use standard load profiles [5] for single-family houses which includes values for low, medium and high annual consumption based on the cluster results. For MFHs, five representative profiles were selected from [17] (002, 007, 027, 031, 060). Using these profiles results in a diversification across different types of profiles, as both couple households with and without children and single households of different ages are covered. The profiles used result in an average household size of 2.1 persons, exactly the average household size determined by Statistics Austria in its study [21]. The data sets [5] also include information for commercial buildings. An average annual consumption of 50 kWh/m² for office buildings can be derived from an ÖGUT report [25].

3.4 Hot Water Demand

The hot water demand in the *plan4.energy* tool is covered by discharging a hot water storage tank. The size of the storage tank is selected in such a way that it can cover the demand if necessary, and at the same time it also serves as storage (flexibility), which is taken into account by the optimizer. The hot water storage tank was modeled in the simulation framework according to [24]. Unfortunately, to the knowledge of the authors, there are hardly any hot water demand profiles available in literature. Synthetic hot water demand profiles for many different households are included in [17] and were therefore also used for MFHs. As part of the project, some profiles were available from the measurement data of several SFHs. Here too, it became apparent that individual demand profiles vary greatly. However, the standard load profiles from [5] only address electricity usage, such as charging the storage tank, and do not cover hot water usage. Therefore, it is crucial to account for hot water tapping to operate the storage tank effectively as a flexible system. Synthetic hot water demand profiles 060 from [17] based on [15] was used for the SFH hot water profile, representing a family with one toddler one at work, one at home. For the multi-family houses the same profiles as above (002, 007, 027, 031, 060) where used for consistency. An important key figure is the ratio of the electricity energy demand to the hot water energy demand of a household. An average value from the measured values of SFHs described above comes to the conclusion that hot water demand for these single-family houses accounts for approx. one third (34%) of electricity demand. The hot water profile used was scaled with these findings in order to obtain the correct ratios for the respective annual requirements. This ratio is significantly lower for office buildings. A 2013 study by ÖGUT shows a yearly consumption of 5.9 kWh/m^2 [6] (which corresponds to a ratio of 6.7% of electricity demand, given the assumptions above). Due to a lack of available hot water tapping profiles for office buildings, a profile based on the combined multi-family house profile was used. However, the profile was adapted so that no tapping takes place outside working hours or on weekends and public holidays, and the profile was subsequently smoothed.

3.5 Mobility

The assumptions for the number of kilometers driven by motor vehicles are based on the Österreich Unterwegs 2013/14 study [9]. It is also assumed that the average annual mileage is independent of the building type, but that the charging profiles for electric mobility show different charateristics for SFHs and MFHs compared to Offices (see below). The EVVP tool [20] developed at the AIT was used to create such charging profiles, which is based on the data from the Österreich Unterwegs study [9].

3.5.1 Single Family Home. This study results in an average mileage of around 30km/day or approx. 11,000 km/year. The values for "low" and "high" are set at half and double this value respectively. For single-family homes, the tool offers the option of selecting both a "fossil" mileage (none, low, medium, high) to represent fossil fuel vehicles and a parallel electric mobility (BEVs) mileage (none, low, medium, high) in order to cover a variety of fleet options (e.g. second vehicle).

3.5.2 Multi-Family Home. For the multi-family house, the assumption of an "averaged" vehicle fleet applies, with the assumption that the average mileage is 11,000 km/year per vehicle. However, the size of the vehicle fleet is allowed to vary (no cars, 0.8 cars per household or 1 car per household) and different degrees of electrification of the vehicle fleet (fossil fuel vehicles only, 15% BEVs, 40% BEVs). The resulting annual mileage with fossil fuel vehicles and BEVs is calculated from these assumptions and the respective charging profiles are also calculated using the EVVP tool. A value of 40% electrification can be used as an optimistic assumption based on the Austrian Ministries Mobility Master Plan 2030 for Austria [8].

3.5.3 Office Building. The fleet of office buildings in the *plan4.energy* tool is composed according to a very similar logic to that of MFHs. The assumed average mileage here is also 11,000 km/year per vehicle, as are the degrees of electrification (0%, 15%, 40% BEVs). The number of vehicles is influenced by the number of employees in the building. In the *plan4.energy* tool, it was assumed that there are 25 m² per person, though this can vary significantly depending on the company. This value is in line with the EERAnet project (Sunex), which assumed 21.3m^2 [13]. This number of people per m^2 is also relevant for calculating the heating and cooling load profiles and is therefore also shown separately in Table 4. Regarding individual mobility, the Österreich Unterwegs study suggests that approximately 72% of commuters use cars, while a lower percentage is around 37%. Alternatively, it is also possible to exclude a vehicle fleet from consideration. Using the above figures, the annual mileage and the number of vehicles with either fossil fuel or electric propulsion can also be calculated. Similar to MFHs, the BEV charging profiles were calculated using the EVPP tool. The EVVP tool takes into account the varying arrival times of BEVs, which differ significantly between trips to and from the workplace [9], leading to substantial differences in profiles compared to SFHs and MFHs.

3.6 Heating and cooling load profiles

A key lever for potential energy savings is the building itself. For the *plan4.energy* tool, a special focus was therefore given to the simulation of heating/cooling requirements for different building types, building sizes and building quality classes according to [1]. The open-source software EnergyPlus [10] was used to simulate the buildings. One main advantage is that the building physics including interaction with the environment (outside temperatures, solar radiation, ...) are taken into account for the calculations. As can be seen in Tables 1,2 and 4, there are three sizes for each category: a particularly small building, a medium building, and a particularly large building. Thus, for simulations that differ only in the building size parameter, it is possible to approximate different building sizes in between by simple interpolation and is used by the *plan4.energy* tool. The values for m^2 per inhabitant are based on statistical data from Statistics Austria [21].

Further information on the modeling of the various building models can be found in [16]. To enhance comparability, simple building shapes with flat roofs were used for all cases. The calibration of the individual models for different building efficiency classes was carried out by adjusting the insulation material thickness and window quality in the building model in order to arrive at average heating demand values for different building efficiency classes according to [1]. For the different classes A to D this corresponds to an annual heating demand of A: 12.5; B: 37.5; C: 75 and D: 125 kWh/m^2 respectively. Simulations were carried out with EnergyPlus using the weather file provided by the Meteonorm [2] software for the Vienna-Schwechat location for the year 2019. The building temperature is kept within a certain corridor (in this case 21.5 to 25.5°C). In order to maintain this corridor, thermal energy is added or withdrawn in the EnergyPlus simulation as required. This energy is therefore the required heating or cooling profile for the respective building of a certain size and building efficiency class.

3.7 Heating systems, cooling systems and floor heating

As described above, one can choose between several different heating systems in the *plan4.energy* tool. A rough distinction is made between the non-electricity-based variants (gas, oil, district heating, pellets) and, in contrast, the two heat pump options (air-to-water and the water-to-water heat pump system for energy recovery from wastewater). For the non-heat pump systems, efficiencies from literature were assumed and are given in Table 5. The efficiencies of the heat pumps are taken from [24]. The strength of the heat pump is that it can potentially provide several units of heat energy for a single unit of electricity, depending on the current COP value. The COP depends on the two temperature levels between which the heat pump must operate (and a certain temperature surplus for the heat exchangers involved). In the case of the air/water heat pump, for example, this would be the outside temperature and the flow temperature of the heating system, adding a temperature surplus for both sides respectively. The COP is therefore a key variable in the simulation. The COP can be derived using the temperatures from the Meteonorm Weather files described above. It should be noted that the *plan4.energy* tool also provides results for heating the building using wastewater heat recovery (water-water heat pump). A temperature profile for a wastewater sewer was provided by one of the project partners. The flow temperatures as input to the heating system represent the higher temperature level for the COP calculation. A lower flow temperature ensures higher COP - and is therefore more efficient. However, the flow temperature is a variable that typically changes continuously for a defined heating system as it is dependent on the outside temperature via the heating curve. To avoid this unnecessary complexity, the approach chosen for the *plan4.energy* tool is to determine an "average weighted flow temperature" for each building quality class. Firstly, a weighted average outdoor temperature is calculated as $\sum_t (heatingoutput(t) * T_{out}(t))/totalheatingoutput$. This weighted outdoor temperature can then be converted to the weighted average flow temperature (see Tab. 3) according to a typical heating curve. Depending on whether floor heating is installed or not, the associated heating curve is significantly flatter and therefore leads to lower flow temperatures.

Table 3. Average weighted flow temperatures with and without floor heating for different building quality classes

Efficiency Class	Average Weighted Flow Temperature (°C)	
	Floor Heating	No Floor Heating
A	27.64	34.62
B	30.08	36.78
C	31.24	39.13
D	32.58	41.82

In most cases, a typical air conditioning system is assumed for the cooling system (this corresponds to an air/air heat pump modeled according to [24]). For combinations of heating systems with heat pumps and floor heating, however, the *plan4.energy* tool also offers the option of using air/water or water/water heat pumps for cooling. This results in higher efficiencies for the respective cooling process, but is only possible for systems with floor heating.

Table 4. Assumed dimensioning of hot water storage, heat pump, and air conditioning

Building Type and m^2	Parameters			
	m^2/Person	Hot Water Storage (Liters)	Heat Pump Nominal Power (kW)	Air Conditioning Nominal Power (kW)
SFH_50	25	150	2	2
SFH_150	42.86	250	5	5
SFH_400	66.67	350	12.5	12.5
MFH_300	36.9	450	12	10
MFH_1500	36.9	700	60	45
MFH_7500	36.9	2500	200	200
Office_300	25	250	6	5
Office_5000	25	700	110	90
Office_25000	25	2500	400	400

3.8 Automatic shading

Automatic shading is a subsystem whose influence on energy saving potential is often underestimated. The exact technical implementation of such a control system can vary greatly [14, 18], but the effect is similar in each case. It achieves a reduction in the direct solar input through the window surfaces, typically when the room temperature exceeds a certain value. For the simulations in this project, this was implemented using a function in EnergyPlus. If a certain room temperature is exceeded (in this case 23.5°C), the automatic shading is simulated in such a way that the transmissivity of the window surfaces is reduced to just 40%. This leads to a significantly reduced heating of the building and therefore a lower cooling requirement. Interestingly, it can be seen that automatic shading also slightly increases the annual heating requirement. This is due to the fact that the building itself is also acting as a heat storage and the shading kicks in earlier than the AC-cooling. Particularly in spring and fall, this leads to slightly less heat stored in the building, which can lead to slightly more heating being required (on specific days). However, energy savings from AC are significantly higher than this additional energy requirement.

3.9 Assumptions on efficiencies and conversion factors

For comparing different systems running on different energy carriers, different sources provide slightly different efficiencies. In Table 5 the assumptions (and conversion factors) for this work are summarized.

Table 5. Various Used Efficiencies and Conversions including Sources

Parameter	Value	Source
Average Gas Heater Efficiency	0.97	[3]
Average Oil Heater Efficiency	0.92	[3]
Average Pellet Heater Efficiency	0.82	[3]
Vehicle Fuel Consumption / km	0.077	[23]
Conversion kWh/l Oil	0.0965	[4]
Conversion kWh/kg Pellets	0.20	[4]
Conversion kWh/l Vehicle Fuel	0.1053	[4]

3.10 Time resolution of profiles

Depending on the source of the input profiles, the typical resolution is either 15 minutes or, in some cases like weather data, 1 hour, as summarized in Table 6. Profiles having initially a 1 hour resolution were interpolated to the 15 minutes time steps used for the simulations in this work. This also holds true for sewer temperatures needed for the water/water heat pumps for heat recovery from waste water, which only were available on a daily basis. In Fig. 3 an example for these load profiles used for simulation is given for a typical period in winter and summer.

Table 6. Time resolution of the sources of the main input profiles

Profile and source	Resolution
PV generation / irradiation	1h
el. load profiles	15min
EV charging profiles	15min
heating and cooling loads	1h
Air temperature	1h
Waste water temperature	1day

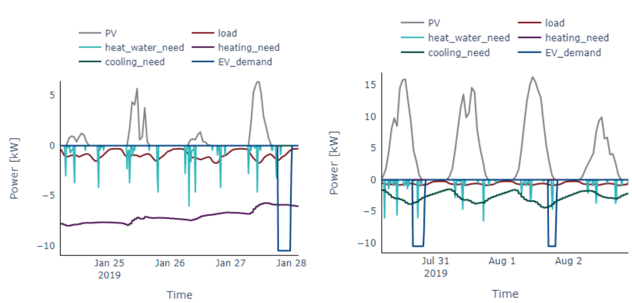


Fig. 3. Example of electric and heat load profiles for a 400m^2 SFH.

3.11 Future work

In the present work results were calculated using Austrian weather data and are therefore not suitable for different climate zones. This however can easily be adapted using different weather input data and creating additional databases. It should further be mentioned, that the selected approach of the simulation database offering countless combinations of buildings in a PED does have the shortcoming that load shifts between different buildings could not be incorporated. The back-end of the *plan4.energy* approach does, however, provide the tool chain to be able to conduct such simulations. However, since every simulation typically takes a couple of minutes in computation time for every single change in configuration, this may be undesirable when using the tool without pre-simulated results for consulting purposes.

4 FRONTEND, RESULTS AND DISCUSSION

In the following, the frontend is described in more detail. It should be noted that the tool is currently targeted toward Austrian stakeholders and is therefore in German language. The typical setting in which it would be used is to help an expert (e.g. a planner) to quickly show different variants of a project, thereby identifying reasonable levers to improve the overall energy balance of a PED. The assessment process is roughly divided into two stages: 1) The assessment of the current status of an existing or planned project ("Ist") and 2) the several variants ("Variante"). The selection of the different buildings is at the very beginning of the "Ist" assessment. Each building type is defined by different parameters, such as size, heating type, size of the PV system, etc., which clearly defines the building (and therefore the pre-simulated results assigned to it) (see Fig. 4). In order to create the district, several buildings can be combined.

Variations of the "Ist" survey can be generated by creating several variants. Examples of this include building renovation (higher building quality class), installation of PV and/or battery storage, change of heating system, etc.

As such changes are usually associated with investments and possibly also with changes in additional annual running costs, the menu items "Investments and subsidies" and "Cost calculation" have been created. These costs can be recorded here (Fig. 4 bottom) and are used to provide and evaluation of the cost/benefit or amortization calculation of certain variants in comparison to the actual status.

The underlying primary energy costs can also be set in one of the submenus, as experience has shown that these can change quickly (see Fig. 5). Depending on the variant, this may have a decisive influence on the cost/benefit analysis or amortization time. However, one could also imagine two variants differing only in the price assumptions in order to be able to draw comparisons. The same is of course also possible for the investment/promotion calculation. One example is a comparison of variants with and without subsidies.

Since it is impossible to present the benefits of this dynamic tool using static figures in the following a simple example of how the *plan4.energy* tool can be used to interpret complex interdependencies is shown. The results presented in Fig. 4 to 6, using the example of PED consisting of two SFH (efficency class C) one MFH (B) and a office building (A). The baseline "Ist" is considered with fossil gas

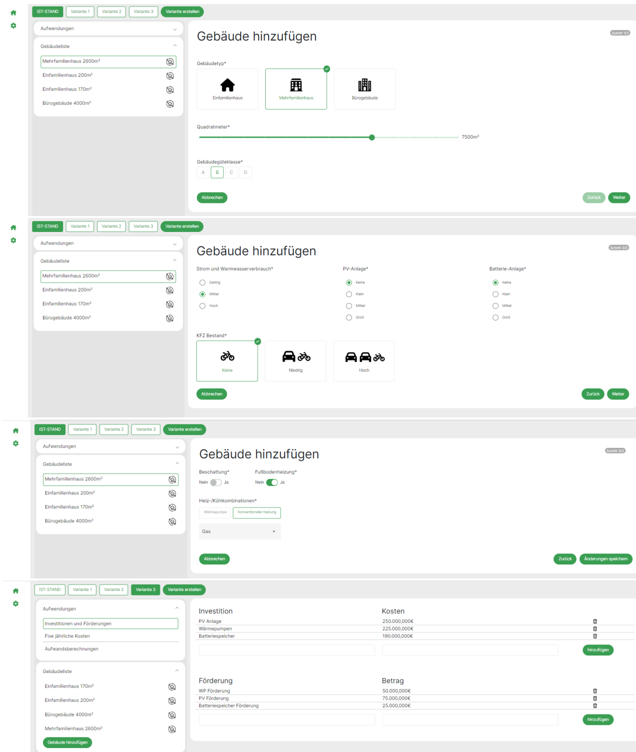


Fig. 4. Sequential input screens for the assessment of different variants.

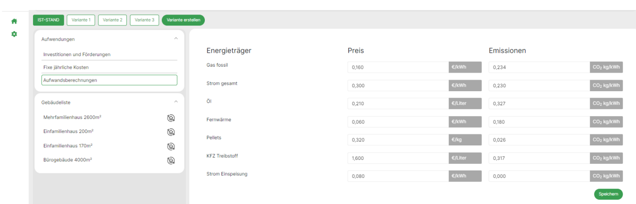


Fig. 5. Submenu for defining primary energy costs and CO_2 footprints

heating and a medium sized PV (see Tab. 1) on the office building only. In the variant 1 the heating system is changed to heatpumps, in variant 2, PVs are added and in variant 3, medium sized battery storages are included. From the pie chart in Fig. 6 it becomes clear that the situation concerning emissions constantly improves. The step from IST to variant 1 and subsequently variant 2 show the largest improvements. Amortization times of the investment costs in Fig. 6 top should be highlighted. It is highest for variant 1, but lowest for variant 2 (due to the valuable interplay between heatpumps and PV). Battery storage in variant 3 does pay off, however at a slower rate than variant 2 without. This is one example illustrating the many possibilities where the tool can be used to show complex interlinks between different components. In the typical use case of a planner consulting on these matters, these interlinks could then be assessed by adding other variants using different variations of the heat pump (air/water, water/water and with or without floor heating). Furthermore, mobility could be considered and included in a further assessment. For example results show that the battery storage would perform better in the above example in combination with EV charging.

In conclusion, the strength of the *plan4.energy* tool lies in its foundation of a database containing more than 300,000 simulation results for different configurations. These simulations results were created using detailed component modelling and optimization makings use of sector coupling across the different technologies. The use of profiles with a resolution of 15 minutes is essential to capture possible simultaneities, which can then be stored and/or shifted via flexibilities (as determined by a linear optimizer). Finally, an appealing graphical user interface which allows for easily compilation of the many different possibilities (and allows to display the main results and key performance indicators in an detailed, but easily understandable manner) is essential for a tool as the one described in this work.

ACKNOWLEDGMENTS

This work was funded by the project program "Stadt der Zukunft - 7. Ausschreibung" by the Austrian Research Promotion Agency FFG.

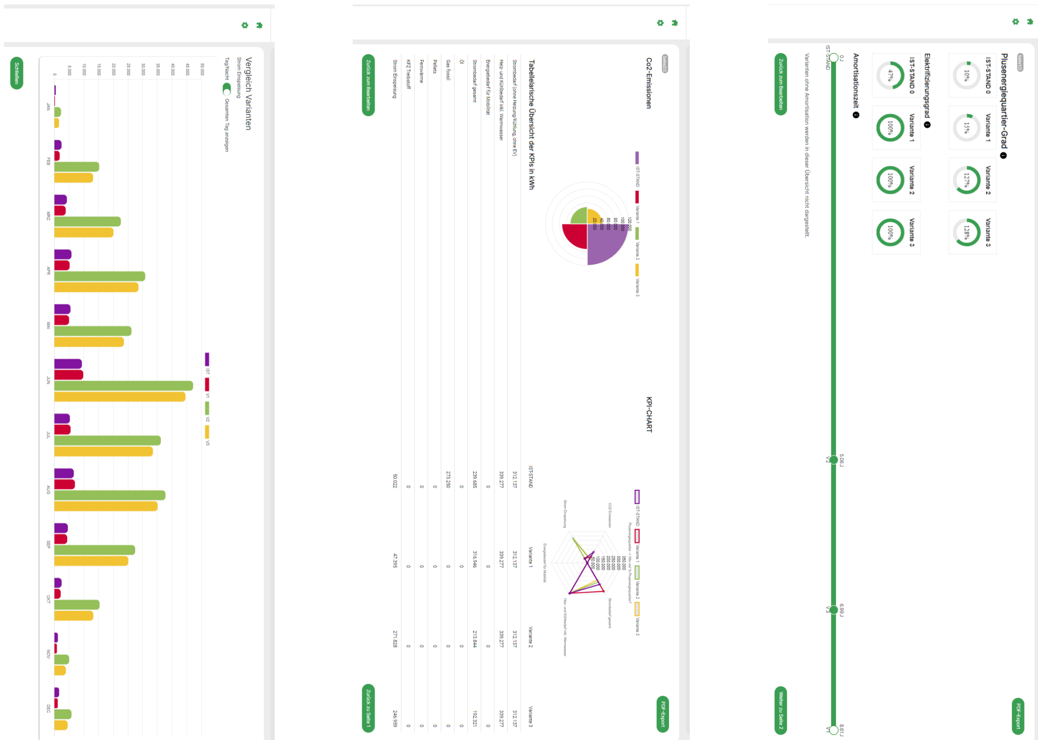


Fig. 6. Example of the several screens of the results presentation

REFERENCES

[1] 2010. Richtlinie 2010/31/EU des Europäischen Parlaments und des Rates vom 19. Mai 2010 über die Gesamtenergieeffizienz von Gebäuden (Neufassung): COD 2008/0223. Online. http://data.europa.eu/eli/dir/2010/31/oj

[2] 2020. *Meteonorm version 8.* https://meteonorm.com/assets/publications/5bv.3.8_pvsec_2020_mn8.pdf

[3] 2021. *Technology Data for Individual Heating Plants.* https://ens.dk/en/our-services/projections-and-models/technology-data/technology-data-individual-heating-plants

[4] 2024. *Energy Density.* https://en.wikipedia.org/wiki/Energy_density

[5] APCS Power Clearing and Settlement AG. 2024. *Synthetische Lastprofile.* https://www.apcs.at/de/clearing/technisches-clearing/lastprofile

[6] G. Bayer, T. Sturm, and S. Hinterseer. 2011. *Kennzahlen zum Energieverbrauch in Dienstleis-tungsgebäuden.* https://www.oegut.at/downloads/pdf/e_kennzahlen-ev-dlg_zb.pdf

[7] BuildingSync. 2024. BuildingSync: A Common Schema for Energy Audit Data. https://buildingsync.net/ Accessed: 2024-07-26.

[8] Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie. 2021. *Mobilitätsmasterplan 2030 für Österreich: Der neue Klimaschutz-Rahmen für den Verkehrs-sektor Nachhaltig – resilient – digital.* Wien.

[9] Bundesministerium für Verkehr, Innovation und Technologie. 2016. *Ergebnisbericht zur österreichweiten Mobilitätserhebung „Österreich unterwegs 2013/2014".* www.oesterreich-unterwegs.at

[10] D. B. Crawley and et al. 2001. EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings* 33, 4 (2001), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6

[11] Energie- und Umweltagentur des Landes Niederösterreich. 2024. Klimarelevanz-tool. https://www.energie-noe.at/klimarelevanztool Accessed: 2024-07-26.

[12] JPI Urban Europe and SET Plan Action. 2020. White paper on PED reference framework for positive energy districts and neighbourhoods. *Online* 30 (2020), 2021.

[13] A. Hainoun and W. Loibl. 2021. Analyses of the Long-Term Energy Demand of Vienna City and Mod-elling Related-Key Food-Water-Energy Nexus Effects. In *Advances in Science, Technology & Innovation, SUSTAINABLE ENERGY-WATER-ENVIRONMENT NEXUS IN DESERT CLIMATE*, E. Heggy, V. Bermudez, and M. Vermeersch (Eds.). SPRINGER NATURE, 457–462.

[14] R. Lionar, D. Kroll, V. Soebarto, E. Sharifi, and M. Aburas. 2024. A review of research on self-shading façades in warm climates. *Energy and Buildings* (2024), 114203. https://doi.org/10.1016/j.enbuild.2024.114203

[15] N. Pflugradt, P. Stenzel, L. Kotzur, and D. Stolten. 2022. LoadProfileGenerator: An Agent-Based Behavior Simulation for Generating Residential Load Profiles. *JOSS* 7, 71 (2022), 3574. https://doi.org/10.21105/joss.03574

[16] David Reihs, Aurelien Bres, Karim Rezk, Jerik Catal, and Yannick Wimmer. 2024. Reduced complexity building and heat pump model to facilitate large scale deployment evaluations. (2024). accepted.

[17] Daniel Schwabeneder. 2023. *GeneratedProfiles: gitLab project.* https://codeberg.org/daschw-lab/GeneratedProfiles.jl

[18] I. Shah, B. Soh, C. Lim, S.-K. Lau, and A. Ghahramani. 2023. Thermal transfer and temperature reductions from shading systems on opaque facades: Quantifying the impacts of influential factors. *Energy and Buildings* 278 (2023), 112604. https://doi.org/10.1016/j.enbuild.2022.112604

[19] Sunanda Sinha and S.S. Chandel. 2014. Review of software tools for hybrid renewable energy systems. *Renewable and Sustainable Energy Reviews* 32 (2014), 192–205. https://doi.org/10.1016/j.rser.2014.01.035

[20] D. Stahleder, S. Übermasser, D. Reihs, S. Ledinger, and S. Lehfuss. 2021. ELECTRIC VEHICLE CAR PARK CHARGING SIMULTANEITY AND GRID CONNECTION POWER REQUIREMENT ANALYSIS. In *IET Digital Library.* https://digital-library.theiet.org/content/conferences/10.1049/icp.2021.1565

[21] Statistik Austria. 2023. *Wohnsituation.* https://www.statistik.at/statistiken/bevoelkerung-und-soziales/wohnen/wohnsituation

[22] S. Strömer, D. Schwabeneder, and contributors. 2024. Integrated Energy System Optimization (IESopt). https://github.com/ait-energy/IESopt. (accessed May 1, 2024).

[23] Umweltbundesamt D. 2024. *Energieverbrauch und Kraftstoffe.* https://www.umweltbundesamt.de/daten/verkehr/endenergieverbrauch-energieeffizienz-des-verkehrs

[24] ÖNORM H 5056-1. 2019. Gesamtenergieeffizienz von Gebäuden: Teil 1: Heiztechnikenergiebedarf. Austrian Standards International.

[25] Österreichische Gesellschaft für Umwelt und Technik ÖGUT. 2014. Energieflüsse in Bürogebäuden - (NEWID-IST): Eine Studie im Auftrag der MA 20 - Energieplanung. https://www.oegut.at/downloads/pdf/e_newid_praesentation.pdf

Enabling Moral Agency in Distributed Energy Management: An Ethics Score for Negotiations in Multi-Agent Systems

MALTE STOMBERG and MARTIN TRÖSCHEL, OFFIS Institut für Informatik, Germany

Intelligent and autonomous agents have gained increased recognition as a solution for the efficient and reliable operation of distributed and digitalised power systems. In multi-agent systems, agents use negotiation to coordinate their behaviour, letting them exchange information and make informed control decisions according to a shared goal. However, as power systems are socio-technical systems, too, the agents' autonomous decisions directly or indirectly impact the welfare of human beings. Thus, the question of moral agency arises: How can the agents' decision-making, and the resulting behaviour of the overall system, be aligned with moral values and ethical principles? Unfortunately, this important question is often overlooked in the current state of the art. In this paper, we therefore introduce an ethics score to be used within agent-based power systems in order to enable explicit moral agency in distributed power management. As a showcase, we extend the multi-agent system `WINZENT` in such a way that all agents are able to include the ethics score as a notion of goodness and fairness in their decision-making without impacting the overall efficiency and effectiveness of the agent-based control. We demonstrate the feasibility of the concept in a case study of a future urban distribution grid modelled after the city of Bremerhaven, Germany, and discuss strengths and weaknesses of the proposed approach.

Additional Key Words and Phrases: multi-agent systems, ethics in agent systems, ethical decision making

INTRODUCTION

Motivation

Distributed control plays a pivotal role in the robust and efficient operation of complex distributed systems such as the power system of the future. Intelligent and autonomous agents [Dipti Srinivasan 2010] have gained increased recognition as a solution for the implementation of distributed control in digitalised power systems. Here, agents often represent individual distributed energy resources (DER), monitoring and controlling their DER's operational state [Merabet et al. 2014].

In a multi-agent system (MAS), where multiple autonomous agents operate concurrently to achieve common goals, the importance of decentralized autonomous decision-making and coordination cannot be overstated. Distributed control mechanisms enable these agents to make local decisions based on local information, fostering adaptability, scalability, and robustness within the system. This approach not only enhances the overall performance and responsiveness of the system but also mitigates the risks associated with single points of failure [Wooldridge 2009].

The autonomous decision-making of agents controlling a power system can have profound implications for human welfare, as it introduces the potential for actions that may be ethically problematic or even detrimental. In scenarios where these agents lack a sense of moral agency, their decisions might not align with human values, potentially leading to unintended consequences or ethical conflicts. As an illustrative example of such a conflict, consider a power grid congestion that requires load shedding [Hu et al. 2015; Lim et al. 2014]. When deciding which load to disconnect from the grid in order to solve the congestion, a hospital might be the best choice from an efficiency standpoint as it is a consumer with high load. However, from an ethical point of view, depending on the agreement among stakeholders of different groups in society, the decision to disconnected might collide with agreed upon ethical rules of prioritising the continued powering of critical infrastructure. With regard to this and all other examples, we would like to make it clear that these are purely illustrative in nature and do not imply an ethical standpoint on our part.

As MAS become more integrated into power systems, the need to ensure that autonomous agents possess a form of moral agency to make decision with respect to the current ethical set of rules becomes increasingly important. This is also a crucial point with respect to the acceptance of distributed and autonomous decision-making [Karnouskos and Leitao 2016]. However, the current state of the art does not adequately address the problem of moral agency in agent-based distributed energy management.

Authors' address: Malte Stomberg, malte.stomberg@offis.de; Martin Tröschel, martin.troeschel@offis.de, OFFIS Institut für Informatik, Oldenburg, Lower Saxony, Germany.

Related Work and Research Gap

In the last two decades, several different applications of intelligent agents and MAS to solve control and optimisation problems in power systems have been discussed [Roche et al. 2010]. In a survey, Mahela et al. [Mahela et al. 2020] differentiate two major areas of application for MAS in digitalised energy systems: a) energy management, i.e. the control of distributed generators, consumers and storage systems in order to achieve a supply-demand equilibrium, and b) power grid control, i.e. all tasks related to the security, efficiency and resilience of the operation of the power grid and its assets such as lines and transformers.

Both energy management and power grid control are relevant with regard to ethical decision-making: In energy management, agents embedded in households or commercial buildings directly interact with end users and their preferences, while in power grid control the agents' decision may directly (and negatively) impact the welfare of human beings, e.g. in case of load shedding or other grid stabilisation measures.

In the remainder of this paper, we focus on agent-based energy management for two reasons: First, energy management use cases such as supply-demand matching are easier to discuss in that they require less technical knowledge about the specifics of power grids. Second, in energy management including distributed energy resources (DER), agents directly interact with human users. Focusing on agent-based energy management thus allows us to discuss the ethical dimension and the specifics of our contribution more clearly. However, this does not limit the application of our contribution to energy management use cases only.

Agent-based Energy Management. In agent-based energy management, each agent typically represents a consumer, generator or storage in the power system. In their survey on agent-based control in microgrids [Kantamneni et al. 2015], Kantamneni et al. discuss centralised, hierarchical and distributed architectures for the implementation of agent-based control systems. While centralised solutions are closely related to classical supervisory control and data acquisition (SCADA) schemes and thus possibly more suitable for integration into existing control centres [Abbas 2014], the strengths of intelligent agents capable of autonomous decision-making are especially suited for hierarchical and distributed systems. In the following, we discuss two exemplary MAS, one hierarchical, the other fully distributed. For a more thorough overview on MAS in power systems we refer to the survey papers of Mahela et al. [Mahela et al. 2020], Roche et al. [Roche et al. 2010], Merabet et al. [Merabet et al. 2014], Kantamneni et al. [Kantamneni et al. 2015] and Dorri et al. [Dorri et al. 2018].

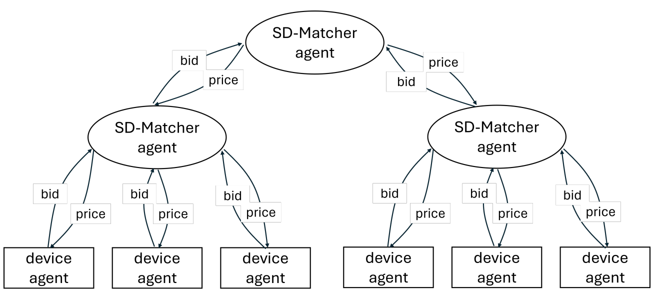


Fig. 1. Hierarchically organised supply-demand matching in the PowerMatcher-MAS. [Kok et al. 2005]. Each agent representing a device sends bids to supply-demand-matching agents (SD-Matchers) in order to settle its device's power demand. The bids may be aggregated by SD-Matchers and submitted to higher-level SD-Matchers in order to find an power equilibrium, resulting in an equilibrium price that is then communicated back down to device agents.

`PowerMatcher` [Kok et al. 2005] is one of the first MAS to be used in the field of smart grids and demand-side management. The key idea behind `PowerMatcher` is to implement a market-based system for flexible and real-time matching of energy supply and demand within a local energy system. As shown in fig. 1, loads and DER are represented by device agents which supervise and control their respective devices. Device agents place bids for their expected consumption or generation with SD-Matcher agents, which in turn place aggregated bids with hierarchically superordinate SD-Matcher agents until an equilibrium between supply and demand is reached and an equilibrium price can be calculated. Here, agents follow self-interested bidding strategies with the goal of economical optimisation [Kok et al. 2005]. The agents' bidding strategies incorporate knowledge about the type of device that they represent, e.g. shiftable operation or freely-controllable devices. However, agents only take economical goals into account and neglect both users' preferences (aside, obviously, the preference for cheap energy) and ethical considerations such as a preferred usage of renewable energy or a minimisation of carbon dioxide emissions.

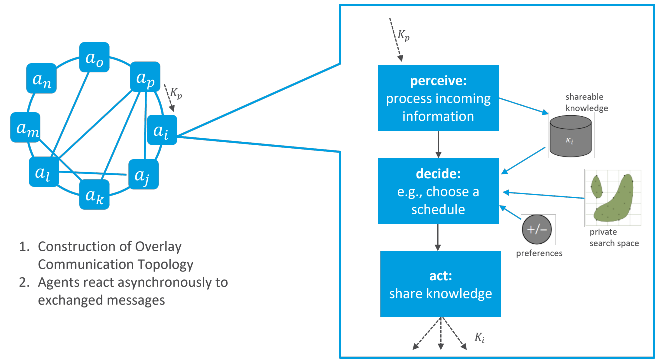


Fig. 2. Fully distributed information exchange and information processing steps in `COHDA`. Each agent (1) perceives new information from its environment and from other agents' messages, (2) decides on its local contribution to solving the overall optimisation problem, and (3) acts by informing all neighbouring agents about its decision. Image curtesy of Frauke Oest and Jörg Bremer, Carl von Ossietzky University of Oldenburg.

`COHDA` [Hinrichs et al. 2014] is a fully distributed optimisation heuristic that is able to solve a combinatorial problem, without being dependent on centrally or hierarchically gathered information. As in `PowerMatcher`, a `COHDA` agent supervises and controlles a single load or DER. In contrast to the hierarchical design of `PowerMatcher`, `COHDA` is a fully distributed system inspired by swarm intelligence and gossiping mechanisms [Hinrichs et al. 2014]. Here, agents are connected in an communication overlay that defines with which neighbouring agents a `COHDA` agent shares information in a multi-step optimisation process. All `COHDA` agents try to achieve a shared goal, e.g. to balance supply and demand in a given power system. For this, each agent first perceives incoming information from its neighbouring agents; second, it decides on its contribution to solving the shared goal based on a local search space and some user preferences; and third, it shares its newly gained knowledge with all of its neighbours. Thus, information regarding all agents' contributions and the overall performance with respect to the shared goal is eventually exchanged within the `COHDA` agent system. A reference implementation of `COHDA` is both available open source as standalone software[1] and as part of the Python mango-agents library[2]. `COHDA` has also been successfully applied to multi-objective optimisation problems in power systems [Stark et al. 2024] taking carbon dioxide emissions into account; however, no further explicit ethical considerations have been integrated into the agents' decision-making processes.

Machine Ethics and Artificial Moral Agency. Ethical concerns regarding autonomous decision making in digitalised energy systems are not limited to emergency situations, with human lives immediately at stake. Fairness, solidarity, and privacy are among the many values that should be reflected in an autonomous agent's decisions, and especially in its behaviour towards human beings and human welfare [Anderson and Anderson 2011]. Machine ethics is the young field of interdisciplinary research concerned with designing autonomous

[1]https://github.com/ambimanus/cohda
[2]https://github.com/OFFIS-DAI/mango-library/tree/main/mango_library/negotiation/cohda

systems "whose decisions and actions might be considered good" [Allen et al. 2006].

The notion of artificial moral agency refers to the idea that artificial entities, such as machines or systems, can possess a level of moral responsibility and decision-making capability analogous to that of human beings [Moor 2006]. Artificial moral agency implies that multi-agent systems can not only perform tasks and make decisions based on predefined rules or learning algorithms but also have the capacity to understand, reason about, and act in accordance with moral principles. This goes beyond simple automation and involves imbuing multi-agent systems with a sense of ethical reasoning, allowing them to navigate situations that involve moral dilemmas and make choices that align with societal values.

There have been numerous attempts in the last two decades to implement explicit ethical agents. Anderson et al. developed the expert systems “Jeremy” [Anderson and Anderson 2007] and “W.D.” [Anderson et al. 2005] based on the Hedonistic Act Utilitarianism from Jeremy Bentham and the notion of prima facie duties of William D. Ross. Both systems rely on machine learning algorithms, training the agents to resolve ethical dilemmas. For this, inductive-logic programming is used in the training process to model the relationships among the duties and values involved in a particular dilemma. Both “Jeremy” and “W.D.” defer the final decision making in a given situation to a human operator who was supported with information generated by the expert systems.

“MedEthEx”, also developed by Anderson et al. [Anderson et al. 2006], is an ethical healthcare agent whose underlying architecture consists of three components: (1) a knowledge-based human-machine interface to produce guidance in selecting duty intensities for a particular case, (2) an advisor module to determine the correct action for a particular case based on learned knowledge, and (3) a learning module that abstracts the guiding principles from particular cases supplied by a biomedical ethicist acting as trainer. Similar to “Jeremy” and “W.D.”, “MedEthEx” is an expert system, deferring final decisions to a human expert.

With “MoralDM”, Deghani et al. [Dehghani et al. 2008] designed a decision-making process for explicit ethical agents that tries to imitate human moral decision making. It makes use of natural language processing to produce formal representations of ethical dilemmas given in textual form, includes a qualitative reasoning algorithm for modelling and measuring the impacts of differently prioritised values, and uses an analogical reasoning algorithm to determine consequences and utilities when making moral judgments. “MoralDM” is capable of exhibiting both deontological and utilitarian behaviour depending on the problem faced.

Franklin et al. [Franklin et al. 2013; Wallach et al. 2020] proposed "LIDA", an advanced cognitive decision-making system for physical agents (robots). The ethical decision-making model, which is a computational model based on neuroscience, considers four evaluations: (1) primary evaluation, (2) evaluation of reward, (3) evaluation of punishment, and (4) evaluation based on ethical norms. The primary evaluation results in a level of pleasure calculated for persons or things affected by each likely action.

Berreby et al.’s [Berreby 2020] developed a generic ethical assessment model for moral agents. The model involves four modules: action, causal motor, ’good’, and ’right’. The action and causal motor modules provide a morally neutral understanding of an agent’s environment and action space, while ’good’ and ’right’ guide decision-making and evaluation. The action module receives the action, producing an event trace. The causal motor adds causation, creating a causal trace. This trace, along with a ’goodness’ assessment, is fed into the ’right’ model, which, based on deontological principles, provides an ethical assessment of the action.

Similar to that, Cointe et al.’s [Cointe 2016] approach involves a four step process for decision-making. The awareness process assesses the environment by generating beliefs and desires. The Evaluation Process categorizes possible actions into subsets of desirable, feasible, and morally good actions. The Rightness Process combines these subsets to create a set of rightful actions which the user can then choose from.

Chaput et al. [Chaput et al. 2021] present a hybrid approach that combines machine learning and symbolic reasoning in order to build artificial moral agents that learn value-aligned behaviors under the guidance of human-supported judgements. Judging agents are associated with single moral values and are equipped with a set of rules to determine if an action presented by a learning agent supports or defeats this moral value. Here, rules can target consequences of actions (consequentialism) or actions themselves (deontological ethics). The feedback from judging agents is then used to continually improve the performance of learning agents regarding the moral valuation of their selected actions. The concept is evaluated in a simulation of a small microgrid comprising households, offices and schools. Each building is represented by a learning agent that has to optimise the comfort of the buildings’ occupants with respect to the energy demand. To our knowledge, this is the first work relating artificial moral agency to power systems and energy management. However, while the learning agents learn to adapt their building’s energy demand to a given supply, there is no explicit communication between the agents and no coordination of the agents’ behaviours in the sense of a MAS [Ferber and Weiss 1999].

In summary, on the one hand agent-based distributed energy management systems currently do not take ethical considerations or moral values into account, while on the other hand the concepts from artificial moral agency have not been successfully applied to distributed control in power systems yet. The work of Chaput et al. is a promising first step in this direction; however, the learning agents act individually and not in a coordinated multi-agent system.

Contribution

This paper introduces and discusses an ethics score that can be used in negotiations of autonomous agents in distributed energy management. For this, the ethics score attributes a specific, time-dynamic value to every agent that correlates with a user-defined ethical importance of the consumer, producer or storage system the agent represents. This score can then be used to induce preferred treatment to agents with higher ethics scores. For showcasing the feasibility of our concept, we extend the negotiations of the existing agent system `WINZENT` [winzent 2021] with our ethics score and run experiments on supply-demand matching in a simulation model of the power system of Bremerhaven, a German city of medium size. Our experiments demonstrate the usage of the ethics score for a prioritised consumption of renewable energy and for a preferred treatment of industrial consumers over private households in case of supply shortages. The source code for the WINZENT extension and the showcase including all necessary data itself is open source and freely available in a Zenodo repository [ZE2 2024].

The rest of this paper is structured as follows: In the next section, we introduce some theoretical background on the agent system WINZENT. We briefly discuss both the negotiation protocol for information exchange as well as the individual decision function of WINZENT agents. Both have to be extended in order to include an ethical valuation of different consumers, producers and storage systems represented by the agents. After that, we introduce the concept of our ethics score for negotiations in MAS-based distributed power management. It takes three aspects into account: a user-defined basic priority score, an amount of balancing failures and the recency of balancing failures, allowing for a time-dynamic change of priority. Then, we showcase the usage of our ethics score in a simulation of distributed energy management in an urban power grid. The results highlight that our ethics score can be integrated into existing agent-based energy management systems, allowing them to take locally optimal decisions while taking ethical and moral considerations into account. Finally, we discuss weaknesses of our approach and give an outlook to future work.

THEORETICAL BACKGROUND

In the following, we briefly introduce the basic concepts of the agent system WINZENT and the Lightweight Power Exchange Protocol (LPEP) that we use in our showcase. For a more detailed discussion that extends the scope of this paper, we refer to [Veith 2017] and [Frost et al. 2020].

Distributed Energy Management with Winzent

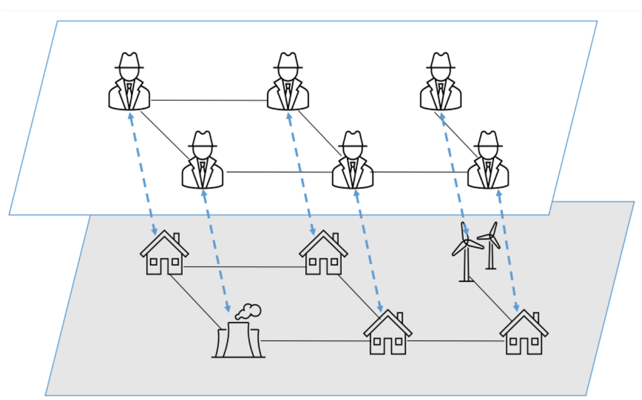


Fig. 3. Overlay topology including communication paths used by agents in WINZENT. The bottom layer represents the power system components; the top layer shows the agents getting assigned to a component to manage it. Furthermore, the connections between the agents represent their communication overlay, which in this case resembles the connections between the grid components in the real grid [ASC 2022, 107].

WINZENT is a distributed, autonomous agent system designed to balance supply and demand at distribution system level in order to achieve an efficient and robust operation of a power grid [Veith 2017]. As with PowerMatcher and COHDA, WINZENT agents represent individual loads and generators and negotiate power surplus or needs with each other. For the negotiation, WINZENT agents use a lightweight messaging protocol [Frost et al. 2020; Veith et al. 2013] that has been developed as an extension of the FIPA Request Interaction Protocol for Intelligent Agents [3]. In order to efficiently calculate power equilibrium states based on the information exchanged between the agents, a Boolean logic solver is used [Veith and Steinbach 2017]. Thus, WINZENT agents are able to coordinate supply and demand in real-time and to take power grid states such as node voltages or power line loading into account. Here, the economical optimisation of individual loads or generators is subordinate to an efficient and robust operation of the overall power system including the power grid. As with PowerMatcher and COHDA, WINZENT agents do not take ethical considerations into account.

As shown in fig. 3, each agent represents a physical component in the power grid. WINZENT agents communicate with their neighbors via connections that are established when creating the WINZENT grid. The agents always follow a protocol, sending predefined messages about local power imbalances to each other and cooperatively attempt to balance production and consumption. Offers to balance electricity can be obtained from other agents and confirmed by both sides using this protocol. Confirmation is reached once a 4-way-handshake is fulfilled. This includes an initial demand or offer request sent to all agents of the network, a reply of the opposite type where available flexibility is offered, an acceptance reply that confirms the offered flexibility and an acceptance acknowledgement reply that successfully closes the negotiation for both agents. Overall, this process is called negotiation and protocol used is called Lightweight Power Exchange Protocol (LEPL).

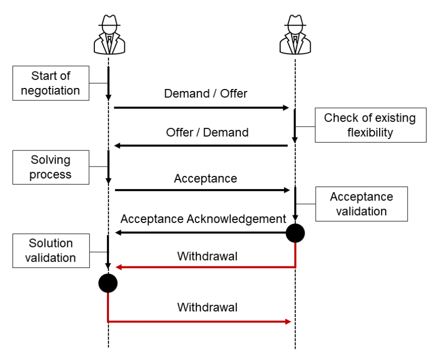


Fig. 4. The negotiation procedure between two WINZENT agents. After an agent sends a demand or offer request to another one, the recipient will always answer with the counterpart of either an offer or a demand if he can at least partly fulfill the request. Red arrows represent the unsuccessful termination of the negotiation with one the agents withdrawing his offer.

A schematic negotiation to illustrate the used protocol is shown in fig. 4. If no more acceptance acknowledgements are pending, the negotiation ends and the final solution for the agent is confirmed. If

[3] http://www.fipa.org/specs/fipa00026/SC00026H.html

however the required acknowledgements do not arrive on time, the negotiation is aborted, withdrawals are sent to all agents with pending acknowledgements and the incomplete solution is confirmed as a final solution.

In practical application, `WINZENT` usually communicates with more than one other agent. To produce a valid solution from the solution process noted in fig. 4, the agent has to handle many different offers from other agents with varying values and time frames. An example of a solution created from such a complex environment is illustrated in fig. 5. In this example, the agent who started the negotiation received offers by four different agents, illustrated by the associated numbers of the received flexibilities. All of these flexibilities are valid for a different portion of the negotiated time frame. The initiating agent tries to find a solution by placing them so that they cover the most space possible, but in this case, it is not enough to satisfy the power demand since there is still uncovered space. Here, the initiating agent has no other option than to accept this power shortage.

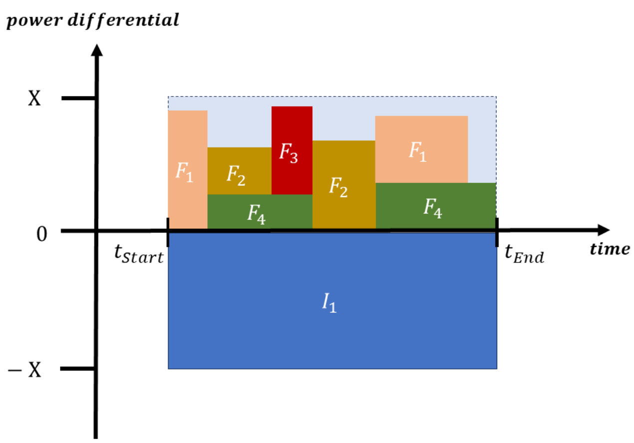


Fig. 5. An example of a realistic `WINZENT` solution for Δt. The block denoted with I_1 is the imbalance the agent tries to negate. The blocks denoted with F_x are the flexibilities offered by corresponding agents. They are arranged by a boolean algorithm to cover the largest possible area of the deficiency.

In order to always find the solution that offers the most coverage for the agent's imbalance, a particular boolean algorithm is used that is mathematically guaranteed to always find exactly that solution. Since this algorithm is not altered in any way during the creation of the ethics module, a summary of it in the context of this paper is omitted. We refer interested readers to [Veith 2017][153] for further details.

Ethical concepts

'Goodness', 'rightness' and 'fairness' can be identified to be three fundamental ethical concepts that help lay the foundations for directed, non-random decision-making in agent systems, as already laid out in the chapter 'Machine Ethics and Artificial Moral Agency'. They also form the ethical backbone for the components of the ethics score, explained in the upcoming chapters. In this chapter however, our understanding of these three concepts is explained and connections are made between the concepts.

The 'goodness' concept defines a set of rules that the agent understands as good. These set the ground work for basic judgements [Cointe 2020, 3]. Using the topic of this paper, an example could be: *Avoiding emission of CO2 is good*, which would lead to the statement *Preferring renewable energies over conventional ones is good*, a statement that would then be enforced in the decision making of an agent. If possible, the set of rules defined by the 'goodness' concept should at least be partly applicable to every situation the agent could find himself in. If that is not the case, the agent's decisions might either become less predictable or the agent is not able to make a decision at all.

The 'rightness' concept broadly describes what to do when there is no clear *good* option available. It allows the agent to make compromises between its various *good*-principles, even if the chosen option ultimately breaks with some of these principles. An example for this, using the aforementioned example rules in the 'goodness' section, could be to forego a renewable energy offer to guarantee security of supply with conventional energy that fits the demand better. The agent might break with the *Avoiding emission of CO2 is good*-principle but it secured grid stability which , in this case, might be valued more highly. The principles behind the *theory of the right*, as the concept is phrased in [Cointe 2020, 3], fall back on Kant's Categorical Imperative and Thomas Aquinas' Doctrine of Double Effect and have to be implemented individually depending on the use case of the particular agent system (see [Cointe 2020, 3]), effectively expressing that there is no one *theory of the right* applicable for all use cases involving agent system.

'Fairness' is another ethical concept to be taken into account. While the term itself generally describes the equal distribution of a certain resource among a group of entities, it is not as trivial to define as one might think since '…'fairness' has multiple context-dependent,and sometimes even conflicting, theoretical understandings …'[Wallach and Jacobs 2021]. Therefore, the definition of fairness used in this paper is inspired by the definition of individual 'fairness' in [Wallach and Jacobs 2021][9]: '…similar people be treated similarly …'. In the case of this paper, the 'people' mentioned in the citation are the agents that govern similar power grid entities. They should be treated similarly, as long as the power grid entities they administer are of comparable priority within the power grid. In a case where a decision between two agents of similar ethical importance needs to be made and equal treatment is impossible without hurting grid stability, the neglected agent should be preferred next time around to ensure long term fairness.

THE CONCEPT OF THE ETHICS SCORE IN MULTI-AGENT SYSTEMS

In this part of the paper, the general concept and design of the ethics score with all its collaborating components are discussed.

General concept

The main focus of the idea is to assign an ethics score to every agent that represents its ethical value within the context of the power grid. This score is higher, the more ethically valuable the power grid entity behind an agent is. The ethics score changes dynamically based on the supply-and-demand situation an agent faces during the course of the simulation. Agents communicate their current ethics score to other agents by appending it onto the negotiation messages they regularly send out. Users are able to freely change the ethics order of the power grid components and its impact on the MAS solution finding process.

To contextualize and exemplify this idea within a power grid environment, a simple power surplus situation shown in fig. 6. It shows two power producers, a wind power plant and a coal power plant, competing for a single power consumer, a household. They could match the household's demand exactly, but the ethics score is higher for the wind power plant. With the additional ethical value

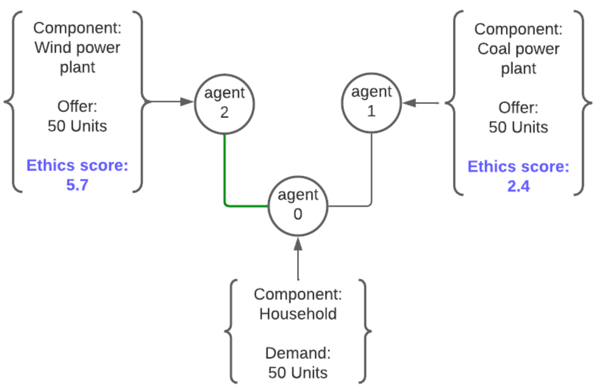


Fig. 6. Three agents in a power surplus scenario. Agent 0 has sent a demand request over 50 units to both agents and both are able to answer that request. Ethics scores of the offering agents are marked blue. The decision of the demanding agent to take agent 2 because of the higher ethics score is marked green. Figure created by us.

as a parameter considered by agent 0 managing the household, it becomes possible to make a confident decision regarding whether to accept the power offered by the wind power plant managed by agent 2.

The same principle can be applied to power shortage scenarios. With such a situation being presented in fig. 7, a decision towards providing the hospital with power can now be made on the basis of the ethics value.

Design

The final ethics score is comprised of three components:

- Priority score (PRI)
- Amount of balancing failures (ABF)
- Recency of balancing failures (RBF)

While the PRI is a static component, the ABF and the RBF are calculated dynamically. The PRI defines a range of possible values and restricts the ethics score to that range. The value of that component is defined by the ethical group the agent is allocated to. This ethical

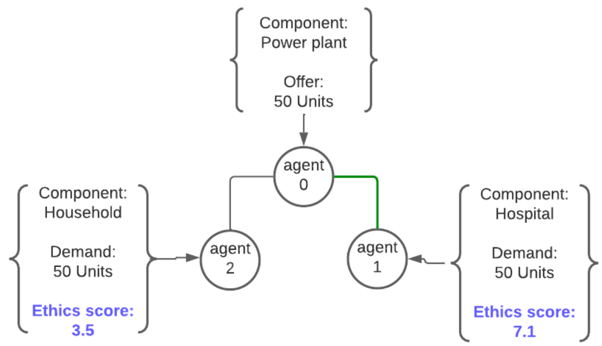


Fig. 7. Three agents in a power shortage scenario. Agent 0 has sent an offer request over 50 units to both agents and both are able to answer that request. Ethics scores of the offering agents are marked blue. The decision of the demanding agent to take agent 1 because of the higher ethics score is marked green

group could, for example, be 'hospital' or 'household'. The restriction ensures that an ethical group ranked lower than another can never overtake the other during runtime. If, for example, 'household' is ranked lower than 'hospital', it will never be possible for an agent managing a 'household'-component become more ethically valuable than an agent managing a 'hospital'-component. The ABF and RBF let the corresponding MAS incorporate events that happen during runtime to either bump up or reduce an agent's ethics score. If, on the one hand, an agent was not able to successfully secure his power supply through a negotiation, its ethics score is raised to increase the likelihood of being supplied during the next step. On the other hand, if an agent gets supplied regularly after an initial supply failure, its ethics score slowly decreases again.

A balancing failure, as specified in the names of the ABF and RBF, is characterized as an unsuccessful negotiation by an agent seeking power balance. This situation may arise when either a producer aims to dispose of excess energy and fails to do so or a consumer seeks and fails to meet its energy demand.
The ethics score is calculated by putting those three components into a quantifiable perspective using the functions *APO_calc* and *RPO_calc* and adding all of them together:

$$PRI + ABF\_calc(ABF) + RBF\_calc(RBF) = Ethics\ score$$

Following, the calculation and meaning of all three of these factors will be explained and exemplified.

The *Priority score* is a basic numeric positive integer greater than zero given to each agent at the start of the simulation. This number is user-modifiable and depends on the type of plant that is represented by the agent. The higher the number, the higher the priority when a decision regarding energy distribution is made.

In order to prevent lower priority entities from gaining the priority over higher ranked entities, entity types are sorted into specific tiers that define the priority value.

An example table filled with priority range values can be seen in table 1. The priority value is the most significant influence on

Priority Score	Grid Components
3	wind power plants, solar power plants, hospitals, critical traffic infrastructure
2	gas power plants, waste incineration power plants, industrial loads
1	public institutions, households, coal power plants

Table 1. An example priority table displaying the priority scores and the corresponding power grid entities. The higher the score, the higher the priority of that particular power grid component in negotiations.
This priority table in particular is used as the default priority score loadout for the ethics module. It favors renewable energy sources and critical public infrastructure as well as public healthcare.

the resulting ethics score. The other two factors are tuned in such a way that an entity can never escape its original priority score tier to make sure that the higher ranked entities will always have priority over lower ranked ones. An example of this can be created using the table in figure 4. According to the table and the rules of the ethics score, an industrial load will never be preferred over a hospital, no matter the supply history of both entities.

The priority score represents and realises the ethical 'goodness' principle by assigning a value to a grid component and thus classifying its 'goodness' without any justification from within the system itself.

The second-most important factor in determining the ethics score is the amount of outages a power entity has experienced during the ongoing simulation, called *Amount of balancing failures (ABF)*. For this to be quantifiable, the aforementioned priority range is divided by the total amount of steps of the upcoming simulation. The value created represents the sub tier range within the selected priority tier. The amount of sub tiers within a priority tier is equal to the amount of simulation steps of the simulation to allow a change of the ethics score regarding the amount of power outages in every simulated time step. Every time a client is denied its needs, be it supply or demand, it climbs up the ethics score ranking by one tier. This concept is exemplified in fig. 8, where a simulation run consisting of five steps leads to five different sub tiers.

As with the priority score, an agent can never leave its sub tier unless another balancing failure occurs. Furthermore, it can only ever go up a tier, but never climb down.

Conclusively, *ABF_calc* is defined as:

$$ABF\_calc(ABF) = ABF \times sub\_tier\_size \tag{1}$$

To distinguish the ethics score of two agents with the same amount of balancing failure, a function is implemented that includes the *Recency of last balancing failure (RBF)* into the equation. It decreases the ethics score linearly depending on the amount of time the corresponding agent has not experienced a balancing failure. The gradient, called *decay rate*, determining the decrease is user-modifiable but cannot transfer the agent into a lower sub-tier to guarantee that an agent with more balancing failure will always be favored over an agent with fewer.

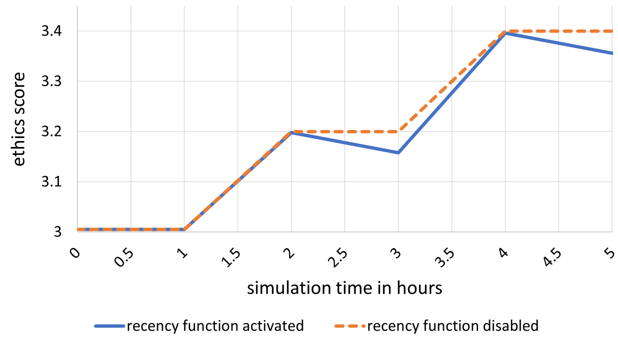


Fig. 8. An example of the effects of the recency function and the power outages combined on the ethics score of an agent in a 5-step-simulation. (1): The first power outage of the agent leading to an increase of the ethics score by one sub tier for the next time step. (2): Another power outage to increase the ethics score by another sub tier from the next time step onwards. Before this incident, the recency function caused the ethics score to decrease slowly.

The default configuration sets the *decay rate* at

$$decay\_rate = (-1) \times \frac{sub\_tier\_range}{total\_simulation\_steps} \tag{2}$$

to allow maximum granularity. With this configuration, one balancing failure at any point of the simulation will have an influence on the overall ethics score for the rest of the simulation. Finally, the *RBF_calc* function is defined as follows:

$$\begin{aligned} RBF\_calc = \max\big(&Current\_Sub\_Tier\_Minimum, \\ &Sub\_Tier\_Maximum\big) + decay\_rate \quad \times \\ &Amount\_Of\_Steps\_Since\_Last\_Balancing\_Failure \end{aligned} \tag{3}$$

The scores for ABF and RBF represent the 'fairness' component of the ethics score. They make sure that agents in the same 'goodness' category receive equal treatment regarding the power supply they require. The 'rightness' concept of the ethics module is represented by the interaction of the three values. It quantifies the ethical priorities, changes dynamically and allows the agent to make the right decision depending on the state of the initial priority scores and the state of the simulation.

SHOWCASE: ETHICAL DISTRIBUTED ENERGY MANAGEMENT IN A FUTURE URBAN POWER GRID

In this chapter, the implementation of the aforementioned ethics score concept is tested and compared against the original `WINZENT` version using a complex test scenario to confirm that the concept is functional and does not significantly impact the performance of the MAS.

Scenario Introduction

To test the limits of the ethics module and the base version of `WINZENT` itself, a simulation of a large, existing power grid is used to create power shortage and power surplus situations that both `WINZENT` versions then have to perform in. Due to availability, the power grid of the city of Bremerhaven was chosen for this task.

Since it is virtually impossible to compare the WINZENT ethics module to the base version in all possible configurations, a representative key scenario is identified and tested that puts an emphasis on the decision making capabilities of the two WINZENT versions in both power shortage and power surplus situations. The term 'scenario', in the context of this paper, describes a set of experiments that are connected by at least one configuration parameter that is slightly modified for each new experiment.

Scenario description

The scenario aims to gradually increase the available wind power by increasing the flexibility of all wind parks in 10 % -steps, starting from 0 %. Modifying the wind production capacity is expected have the most noticeable influence on the power supply situation of the grid compared to all other options.

When running this scenario on the grid without any WINZENT interference, the situation changes from a significant power shortage situation to a clear power surplus one, as can be seen in fig. 9. When

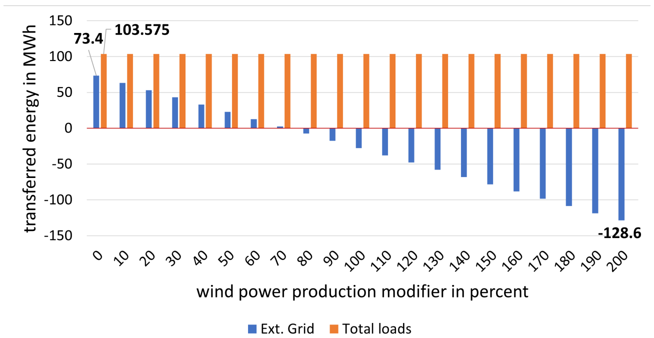


Fig. 9. Overview of the feed-in rate of the external grid over the course of changing wind power production modifiers as well as the power consumption of all loads in the grid. The external grid's purpose changes from a power producer to a power consumer as
It is to be noted that for this figure, the maximum possible power production for all power generators without WINZENT interference was assumed. Figure created by us.

looking at the feed-in rate of the external grid at 0 %, it can be noted that it supplies three quarter of the loads with power when wind is taken away. This means that the WINZENT agents only have around a quarter of the demanded power available for negotiation and need to distribute that power efficiently. Conversely, when looking at the values at 200 %, it can be observed that the external grid consumes more power than all loads of the grid combined, a clear power surplus scenario where the WINZENT agents get to choose their preferred power source.

Simulation Setup

The Bremerhaven grid is simulated using MIDAS, which in turn uses Pandapower for the power flow calculations of the individual grid components. In fig. 10, an overview of the grid is shown and all relevant actors are marked. Pandapower is a python library that combines the two libraries pandas and PYPOWER to create a python

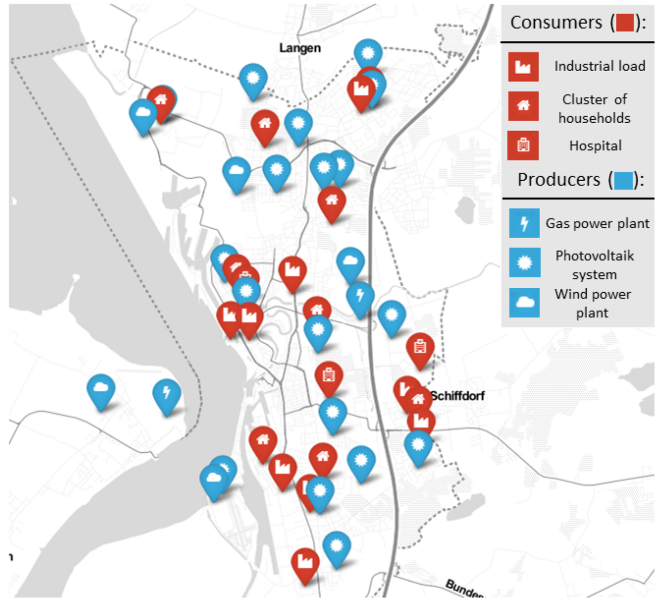


Fig. 10. Map of the Bremerhaven power grid with all the relevant producers and consumers. Connections between the grid components have been left out and households as well as solar power sources have been clustered due to visual clarity. Figure created by us with data from [ARL 2021].

module that can create, simulate and analyse custom power grid structures. It allows for different power grid components such as busses, lines, transformators, generators and loads to be added. After the component configuration, it executes a power flow calculation to determine the state of the power grid [Thurner et al. 2018].

The Multi-Domain test Scenario (MIDAS) allows for pandapower power flow calculations to be put into a simulation context introducing simulation steps and dynamically setting parameters based on user configurations for each of those steps. Furthermore, it introduces the usage of modules which give access to load profiles of different power grid actors such as power plants, households or commercial loads. These functionalities result in the potential to realistically simulate a power grid for a specified amount of time and test it in different, user-defined scenarios [MIDAS 2022].

When using WINZENT in combination with the Bremerhaven grid, distinct advantages arise regarding the simulation complexity and the authenticity of the results. Alas, there are also a few restrictions regarding the grid usage that have to be considered. With the Bremerhaven grid providing an image of a real power grid, results gained by using WINZENT and its ethics module serve as both a confirmation for the functioning of the module itself and for the general usability of the module in complex grids modelled after reality. Furthermore, the grid provides the foundation for a variety of possible dilemmas with its season-specific power schedules, its different grid components that can be manipulated or turned off entirely and its size allowing for an intricate testing of the ethics score functionality. Lastly, it can be combined with the experiment scheduler created by the project group ASC (see [ASC 2022]), allowing for experiment queueing and and result data visualisation using Grafana.

Restrictions

The Pandapower framework the grid is implemented in is not fit to simulate a standalone microgrid without access to an external grid, as the stationary power flow calculation requires a reference node both defining the basic voltage values and providing (theoretically) unlimited power for all loads if necessary. This means that dynamic effects with respect to e.g. voltage stability or frequency can not be simulated. However, we focus on the impact of ethical values on the decision-making and optimisation of autonomous agents, and not on dynamic stability of a future microgrid.

Additionally, with using MIDAS as simulation tool, it is only possible to simulate one time step at a time, step by step. This means that all offers an agent considers always have the same time slot, effectively trivialising the solution finding process of WINZENT. This restriction is referred to in this paper as single time slot solving.

With these restrictions in mind, separate test cases have been created to showcase the functioning of the complete ethical multi time slot solving process. These can be found in the WINZENT repository ([OFFIS Energy 2024]).

Definition of the comparison metrics

In order to identify scenarios ideal for a meaningful comparison, the metrics to be compared between the two WINZENT versions need to be defined first.

As the first comparison metric, the 'Average Relative Negotiated Volume' (ARNV) describes the average amount of successfully negotiated volume relative to the total volume that needed to be negotiated over the course of the experiment. In this context, volume references the total amount of negotiated power. While this value is highly dependent on the type of scenario used, it is a useful comparative value to measure the overall performance of the WINZENT module relative to one another. It is measured in percent. The average negotiation success rate is typically low in scenarios where a large demand for electricity cannot be matched by a comparatively smaller production of electricity and vice versa. On the contrary, the value nears its maximum when demand and production match or the production facilities have the flexibility to decrease their power production in a timely manner.

The second comparison metric, the 'Total Amount Of Balancing Failures', counts all occurrences of insufficiently supplied power grid components across all priority score tiers. This does not only help to evaluate the efficacy of the ethics module regarding high priority grid components, it also quantifies the impact of the priority score on components with a lower ethics score. Furthermore, it can be combined with the aforementioned ARNV to provide information about the types of grid components that get preferred in the negotiation process. If, for example, the ARNV is high and so are the total amount of balancing failures, it can be concluded that large-scale consumers have the priority in the ongoing negotiation process.

The third metric used is the 'Electricity Mix' (EM) of the grid. Depending on which power producer gets to negotiate his production capacity away, the overall electricity mix of the grid changes. The ethics module aims to adjust the electricity mix based on prioritized preferences, thereby mitigating balancing failures associated with specific ethically preferred components.

Results

In this chapter, the results of the comparison between the ethics module implemented in WINZENT and the WINZENT base version, both deployed in the same grid environment, are presented. Specifically, two different priority score configurations are used for a a series of experiments to highlight the differences when changing the priority score.
The code and the data can be found in a public Zenodo repository [ZE2 2024].

Results: Renewable-focused ethics score

This configuration, displayed in Table 2, is focused on prioritising renewable energy production over conventional power sources whenever possible. When looking at power consumers, hospitals are prioritised over everything else. Households rank second and public institutions like malls, libraries or stadiums possess the lowest priority. In fig. 11, the influence of the ethics module on the type of balancing failures allowed can be observed. While balancing failures for low priority targets possessing the priority value 1 is significantly higher for the ethics module, the opposite is true for grid components with the priority score 2 where the base agent produces more balancing failures. Focusing on the priority score 3 high-priority targets, it's notable that the base version had a higher balancing failure occurrence, reaching a peak of 60 recorded denials when set up with a wind power production coefficient of 10%. In contrast, the ethics module did not register any balancing failures. It consistently satisfied all needs for all high priority targets for every step in every experiment of the scenario. When taking a look at the total balancing failure rate displayed in fig. 11, it might look like the WINZENT version using the ethics module allowed more balancing failures in total, trading supply security of high priority targets for a higher occurrence of balancing failures in low priority targets. However, his does not mean that is has been less effective at distributing flexibility across the grid. When taking a look at fig. 12, we can see that the overall volume that was negotiated is nearly identical between the two versions. This implies that the ethics module used the available flexibility to supply a small amount of high-volume targets rather that many low-volume ones, following the directions given by the ethics score.

Priority Score	Grid components
3	wind power plants, solar power plants, hospitals
2	households, waste incineration power plant
1	gas power plant, public institutions, industrial plants

Table 2. The priority score loadout that is used for this series of experiments. It possesses three tiers and favors renewable energy as well as health care institutions. Compared to table 1, it is missing the critical infrastructure components as they are not present in our version of the simulated Bremerhaven grid.

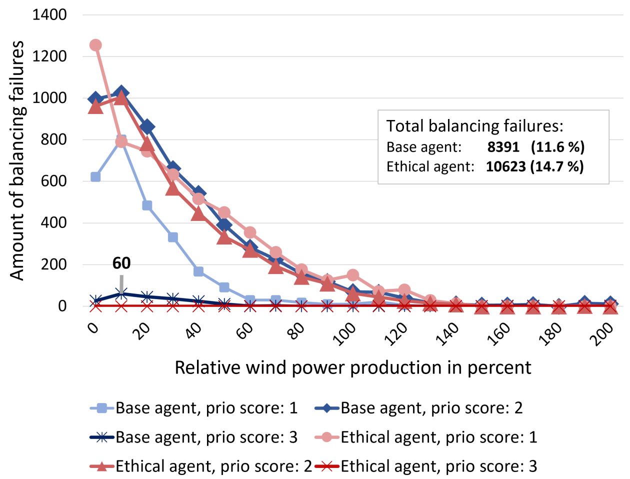


Fig. 11. The amount of balancing failures denials recorded over the incrementally increased wind power production. The value of the balancing failure count for high priority targets of the base agent is labeled as the difference to the ethics module regarding high value targets is biggest here. Figure created by us.

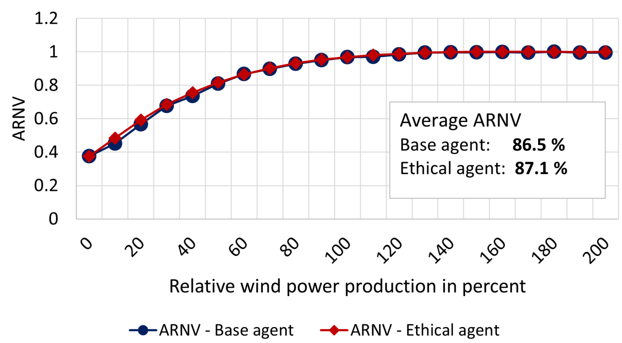


Fig. 12. The average relative negotiated volume over the over the incrementally increased wind power production. Additionally, the overall mean value for both WINZENT systems is given. Figure created by us.

When looking at the electricity mix of the scenario illustrated in fig. 13, it can be observed that the ethics module incorporated more wind power into its electricity mix as the external grid loses its relevance due to sufficient wind power being available.

To further quantify the increased incorporation of wind power by the ethics module, the electricity mix for both MAS is displayed in fig. 14. The difference of around 6.5% in wind power is created by reducing both the usage of waste power by 3.9% and gas power by 2%. Furthermore, we can see that solar power only plays a minor role in the scenario and thus cannot be used to highlight the differences between both versions.

Results: Industry-focused ethics score

For this series of experiments, the underlying priority score has been changed from the default configuration (see Table 2) to the

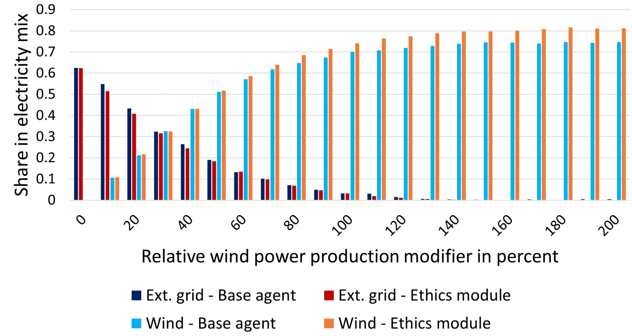


Fig. 13. The shares in the electricity mix of the external grid and the wind power plants over the incrementally increased wind power production For both the WINZENT base version and the ethics module. Figure created by us.

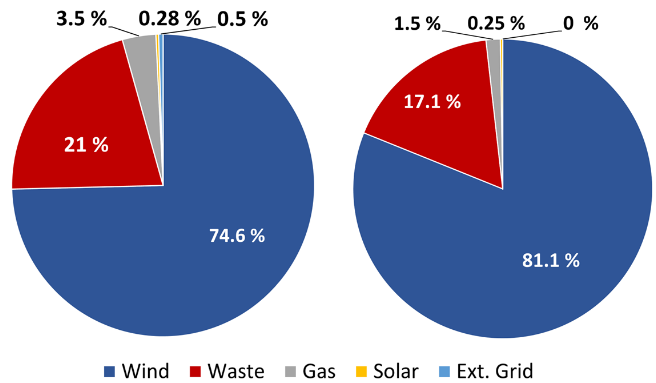


Fig. 14. The electricity mix for both the ethical version (left) and the base version (right) when using a wind power production modifier of 200% and the renewable-focused priority score loadout displayed in table 2. In the top part of both pie charts, the three percentages shown from left to right are: Gas, Solar, Ext. Grid. Figure created by us.

configuration shown in Table 3. The goal of this change is twofold.

Priority Score	Grid components
4	gas power plant, industrial plants, waste incineration power plant
3	households
2	hospitals, public institutions
1	wind power plants, solar power plants

Table 3. The industrial priority score setup.

Firstly, it aims to increase the challenge of ethical distribution of power by tripling the amount of consumers in the highest ethics tier from 3 hospitals to 9 industrial plants. Secondly, it aims to show a clear power distinction in the composition of the electricity mix compared to the default priority score assignment.

As for results, the amount of balancing failures for this new ethics configuration is displayed in fig. 15. The power distribution shown

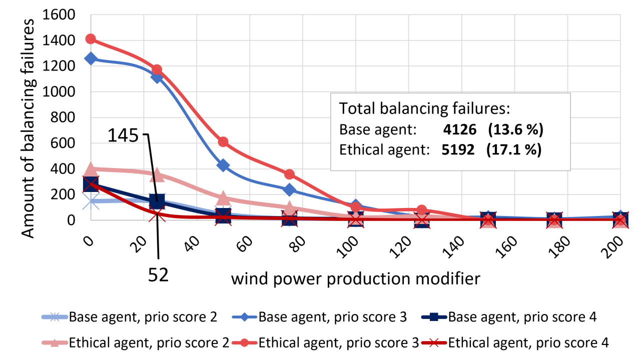


Fig. 15. The amount of balancing failures over the wind power production modifier with an industrial-focused custom priority score configuration.

in the data is similar to fig. 11 with the ethical agent recording higher total balancing failures, but supplying high priority targets more reliably once wind energy production increases.

When focusing on the data point at 0 on the x-axis, it can be observed that the amount of balancing failures for the highest priority tier is exactly similar. That is because the amount of available power in the grid is so low that the ethical agent cannot even fully supply the high priority industrial plants. As the wind power modifier increases, the ethics module proves to be faster in stabilizing the high priority targets. At a modifier of 25%, the ethics module counted 52 balancing failures of high priority targets which is one third of the 145 balancing failures that the `WINZENT` base version caused. As the wind power increases, this advantages diminishes.

The ARNV for this scenario is once again similar for both versions, as seen in fig. 16. As with the default priority loadout, the ethics module produced more total balancing failures (see fig. 15), but distributed the available power just as well as the base `WINZENT` version. From a producer perspective, the electricity mix of the ethics

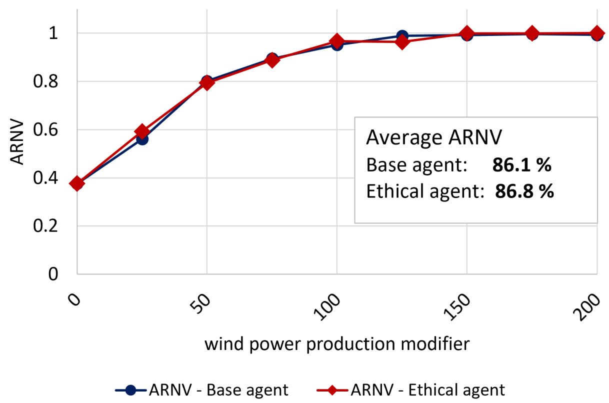


Fig. 16. The ARNV over the wind power production modifier with an industrial-focused priority score configuration for both the base agent and the ethics module. Figure created by us.

module depicted in fig. 17 showed around 9 % less wind power when compared to fig. 14, as was expected due to a lower prioritization. When comparing the performance of both priority score loadouts, a

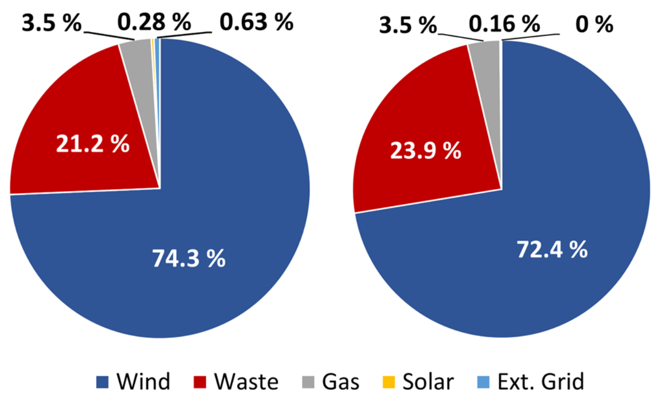


Fig. 17. The electricity mix for both the ethical version (left) and the base version (right) when using a wind power production modifier of 200% and the industry-focused priority score loadout displayed in table 3. In the top part of both pie charts, the three percentages shown from left to right are: Gas, Solar, Ext. Grid. Figure created by us.

stark difference between the default and the industrial ethics setup can be observed as the available wind power ramps up. As seen in fig. 18, once the grid is supplied with enough power to let the agents choose which power source to go for, they prioritize the power source with the highest ethics score available. As power produced

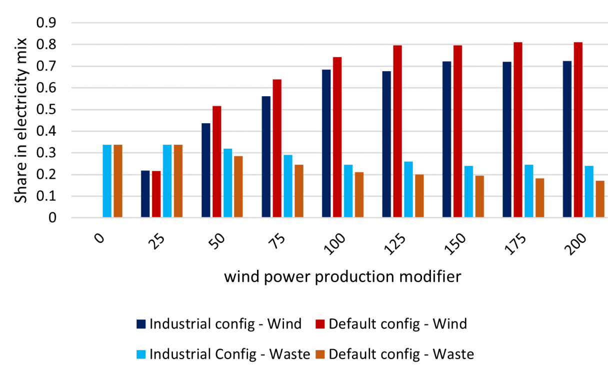


Fig. 18. The share of power produced by wind power plants and the waste incineration plant of both the industry-focused priority score loadout and the renewable-focused one. Figure created by us.

by the waste incineration plant is less prevalent than wind power, the difference stops growing once the grid is fully supplied and waste power is maximized.

Result summary

Both test cases, one being the default priority score loadout and the other being the industry-focused priority score loadout, highlight the differences in outcome when comparing the ethics module with the base `WINZENT` version.

From a consumer perspective, we observed that the highest priority score group in both cases got supplied significantly better by the ethics module than the base WINZENT version. This effect is amplified in scenarios where a insufficient amount of power is available and thus certain consumers need to be prioritised (see fig. 11 and fig. 15). From a producer perspective, this prioritisation worked as well as can be seen in the electricity mix of both scenarios. The driving factor of power production, the wind power, was either significantly higher for the default priority score loadout (see fig. 14) or lower for the fossil-power-focused loadout with fossil power plants getting a bigger share (see fig. 17) when compared to the base WINZENT version.

Lastly, the usage of the ethics score did not harm WINZENT's ability to distribute power in any way. While the type of consumer supplied differs from the base version (see fig. 11 and fig. 15), the total power supplied is nearly identical to the base version (see fig. 12 and fig. 16).

LIMITATIONS

The current implementation of the ethics score in our model presents several limitations that constrain its applicability and accuracy. One major limitation is the simplified assumption that agents will always deliver 100% of their negotiated commitments. This assumption, while convenient for modeling purposes, is unlikely to hold true in real-world scenarios, especially when considering the variability inherent in renewable energy sources.

Moreover, the ethics score in its present form leaves out significant factors such as costs, emissions, and other ethical values. This limits the meaningfulness of the ethical assessment, as these factors play a critical role in evaluating the overall ethical impact of actions within the network. Additionally, the model assumes complete trustworthiness among all agents in the network, disregarding the variability in trust that can arise from factors like past behavior or dependency on external conditions.

Another important limitation is the ethics score's reliance on a purely quantitative approach. While this allows for seamless integration into optimization algorithms, it fails to account for the broader philosophical aspects of ethics. The complexities of ethical decision-making may require more nuanced approaches, such as rule-based or hybrid systems, which the current model does not use.

Lastly, as mentioned in the 'Restrictions' section of the show case chapter, while there has been exhaustive testing for the functionalities of the ethics module in single time slot cases, the same cannot be said for multiple time slots, where the existing proof-of-concept cases ([ZE2 2024]) cannot cover all possible scenarios in a way that testing on a large grid can.

CONCLUSION AND FUTURE WORK

In this paper, we presented the theoretical conception and the testing of an ethics score usable by negotiation-based MAS for a more ethical energy distribution in power grids. This novel approach to ethical decision making was implemented and tested with the MAS WINZENT.

The core idea, the ethics score, is a dynamically-changing, numerical value that is assigned to every grid component the chosen MAS manages. This number then defines the ethical value of the grid component in a MAS negotiation and helps it make decisions on which component to supply in critical grid situations.

As a test environment, a configurable digital version of the Bremerhaven power grid was chosen that could put the WINZENT agents in dilemma situations. These situations were then used to find out how WINZENT would handle these situations with and, as a comparison, without using the ethics score.

We showed that using the ethics module does have a significant impact on the outcome of the negotiations. Grid components possessing a high ethical value are better covered by the ethics module in energy shortage situations. To add to that, the share of highly valued power production sources in the electricity mix in surplus situations is notably higher as well compared to the base WINZENT version. This difference could also be observed when comparing two different ethical rule sets against one another.

An interesting direction for future research is the examination of the impact of the ethics score on multi time slot negotiations in combination with the predictability and reliability of weather-reliant energy sources in complex grid environments, a relevant topic in current research [Brand et al. 2023]. This would not only tackle the unrealistic limitation of agents always being flawless suppliers of the power that was negotiated, it would also allow for the classification of more and less reliable agents. This, when put into quantifiable metrics, could add another layer to the ethics score and might lead to interesting scenarios such as high priority agents negotiating for more than exactly the needed amount of power to ensure a steady supply or refusing to negotiate with non-reliable agents.

Improvements could also be made to the restrictive nature of the ethics score in its current iteration. The ethics score currently stands as a purely quantitative measure. This offers the advantage of seamless integration into optimization calculations. However, there is a philosophical discourse suggesting that not all pertinent aspects of ethics can be quantified, and that rule-based or hybrid implementations of artificial moral agency are better suited [Tolmeijer et al. 2020]. In such cases, agents require the capacity to operate with logical rules and engage in reasoning to determine which of their actions is most ethically appropriate.

The introduction of the ethics score represents a small step towards fostering the acceptance of agent-based control of power grids and for the deployment of autonomous agents in a real-life power system environment, be it in an assisting or an operating role.

ACKNOWLEDGEMENTS

We want to thank the reviewers for their constructive criticism that allowed us to improve this paper and Frauke Oest for proofreading the first draft. Furthermore, we would like to thank Eduard Malcev and Michael Krah for their impulses on ethical decision making in MAS.

REFERENCES

2024. *Data and software for the WINZENT ethics module simulation.* https://zenodo.org/records/11066637

Hosny A Abbas. 2014. Future SCADA challenges and the promising solution: the agent-based SCADA. *International journal of critical infrastructures* 10, 3-4 (2014), 307–333.
Colin Allen, Wendell Wallach, and Iva Smit. 2006. Why machine ethics? *IEEE Intelligent Systems* 21, 4 (2006), 12–17.
Michael Anderson, Susan Anderson, and Chris Armen. 2005. Towards machine ethics: Implementing two action-based ethical theories. In *Proceedings of the AAAI 2005 fall symposium on machine ethics*, Vol. 11.
Michael Anderson and Susan Leigh Anderson. 2007. Machine ethics: Creating an ethical intelligent agent. *AI magazine* 28, 4 (2007), 15–15.
Michael Anderson and Susan Leigh Anderson. 2011. *Machine ethics*. Cambridge University Press.
Michael Anderson, Susan Leigh Anderson, and Chris Armen. 2006. MedEthEx: a prototype medical ethics advisor. In *Proceedings of the national conference on artificial intelligence*, Vol. 21. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 1759.
PG ARL. 2021. Projektdokumentation - Adversarial Resilience Learning: Einsatz von künstlicher Intelligenz zum Angriff sicherheitskritischer CPS. (2021).
PG ASC. 2022. Projektdokumentation - Agent System Competition. (2022).
Ganascia Berreby, Bourgne. 2020. A Declarative Modular Framework for Representing and Applying Ethical Principles. (2020).
Michael Brand, Anand Narayan, and Sebastian Lehnhoff. 2023. Applying Trust for Operational States of ICT-Enabled Power Grid Services. *ACM Transactions on Autonomous and Adaptive Systems* (2023).
Rémy Chaput, Jérémy Duval, Olivier Boissier, Mathieu Guillermin, and Salima Hassas. 2021. A multi-agent approach to combine reasoning and learning for an ethical behavior. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 13–23.
Boissier Cointe, Bonnet. 2016. Ethical Judgment of Agents' Behaviors in Multi-Agent Systems. (2016).
Nicolas Cointe. 2020. A Ethics based Cooperation in Multi-Agent Systems. (2020).
Morteza Dehghani, Emmett Tomai, Kenneth D Forbus, and Matthew Klenk. 2008. An Integrated Reasoning Approach to Moral Decision-Making.. In *AAAI*. 1280–1286.
Lakhmi C. Jain Dipti Srinivasan. 2010. An Introduction to Multi-Agent Systems. In *Innovations in Multi-Agent Systems and Application – 1*. Springer International Publishing, Chapter 1, 1–27.
Ali Dorri, Salil S Kanhere, and Raja Jurdak. 2018. Multi-agent systems: A survey. *Ieee Access* 6 (2018), 28573–28593.
Jacques Ferber and Gerhard Weiss. 1999. *Multi-agent systems: an introduction to distributed artificial intelligence*. Vol. 1. Addison-wesley Reading.
Stan Franklin, Tamas Madl, Sidney D'mello, and Javier Snaider. 2013. LIDA: A systems-level architecture for cognition, emotion, and learning. *IEEE Transactions on Autonomous Mental Development* 6, 1 (2013), 19–41.
Emilie Frost, Eric MSP Veith, and Lars Fischer. 2020. Robust and deterministic scheduling of power grid actors. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, Vol. 1. IEEE, 100–105.
Christian Hinrichs, Sebastian Lehnhoff, and Michael Sonnenschein. 2014. COHDA: A combinatorial optimization heuristic for distributed agents. In *Agents and Artificial Intelligence: 5th International Conference, ICAART 2013, Barcelona, Spain, February 15-18, 2013. Revised Selected Papers 5*. Springer, 23–39.
Junjie Hu, Arshad Saleem, Shi You, Lars Nordström, Morten Lind, and Jacob Østergaard. 2015. A multi-agent system for distribution grid congestion management with electric vehicles. *Engineering Applications of Artificial Intelligence* 38 (2015), 45–58.
Abhilash Kantamneni, Laura E Brown, Gordon Parker, and Wayne W Weaver. 2015. Survey of multi-agent systems for microgrid control. *Engineering applications of artificial intelligence* 45 (2015), 192–203.
Stamatis Karnouskos and Paulo Leitao. 2016. Key contributing factors to the acceptance of agents in industrial environments. *IEEE Transactions on Industrial Informatics* 13, 2 (2016), 696–703.
J. K. Kok, C. J. Warmer, and I. G. Kamphuis. 2005. PowerMatcher: Multiagent Control in the Electricity Infrastructure. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (The Netherlands) *(AAMAS '05)*. Association for Computing Machinery, New York, NY, USA, 75–82. https://doi.org/10.1145/1082473.1082807
Yujin Lim, Hak-Man Kim, and Tetsuo Kinoshita. 2014. Distributed load-shedding system for agent-based autonomous microgrid operations. *Energies* 7, 1 (2014), 385–401.
Om Prakash Mahela, Mahdi Khosravy, Neeraj Gupta, Baseem Khan, Hassan Haes Alhelou, Rajendra Mahla, Nilesh Patel, and Pierluigi Siano. 2020. Comprehensive overview of multi-agent systems for controlling smart grids. *CSEE Journal of Power and Energy Systems* 8, 1 (2020), 115–131.
Ghezlane Halhoul Merabet, Mohammed Essaaidi, Hanaa Talei, Mohamed Riduan Abid, Nacer Khalil, Mohcine Madkour, and Driss Benhaddou. 2014. Applications of multi-agent systems in smart grids: A survey. In *2014 International conference on multimedia computing and systems (ICMCS)*. IEEE, 1088–1094.
MIDAS. 2022. MultI-DomAin test Scenario (MIDAS). https://gitlab.com/midas-mosaik/midas. [Online; zuletzt aufgerufen 20.12.2022].
James H Moor. 2006. The nature, importance, and difficulty of machine ethics. *IEEE intelligent systems* 21, 4 (2006), 18–21.
OFFIS Energy. 2024. *Winzent ethics module repository*. https://github.com/OFFIS-DAI/mango-library/tree/Winzent-ethics_module_refactor/tests
Robin Roche, Benjamin Blunier, Abdellatif Miraoui, Vincent Hilaire, and Abder Koukam. 2010. Multi-agent systems for grid energy management: A short review. In *IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society*. IEEE, 3341–3346.
Sanja Stark, Emilie Frost, and Marvin Nebel-Wenner. 2024. Distributed Multi-objective Optimization in Cyber-Physical Energy Systems. In *ACM Energy Informatics Review April Issue 2024*. ACM.
L. Thurner, A. Scheidler, F. Schäfer, J. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun. 2018. pandapower — An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. *IEEE Transactions on Power Systems* 33, 6 (11 2018), 6510–6521. https://doi.org/10.1109/TPWRS.2018.2829021
Suzanne Tolmeijer, Markus Kneer, Cristina Sarasua, Markus Christen, and Abraham Bernstein. 2020. Implementations in machine ethics: A survey. *ACM Computing Surveys (CSUR)* 53, 6 (2020), 1–38.
Eric Veith. 2017. *Universal smart grid agent for distributed power generation management*. Logos Verlag Berlin.
EM Veith, Bernd Steinbach, and Johannes Windeln. 2013. A lightweight messaging protocol for Smart Grids. In *Proceedings of the Fifth International Conference on Emerging Network Intelligence (EMERGING 2013)*. IARIA XPS Press, 6–12.
Eric MSP Veith and Bernd Steinbach. 2017. Agent-based power equilibrium in a smart grid with XBOOLE. In *2017 International Conference on Information and Digital Technologies (IDT)*. IEEE, 406–416.
Hanna Wallach and Abigail Z. Jacobs. 2021. Measurement and Fairness. (2021), 11.
Wendell Wallach, Colin Allen, and Stan Franklin. 2020. Consciousness and ethics: artificially conscious moral agents. In *Machine Ethics and Robot Ethics*. Routledge, 299–314.
winzent. 2021. Modular Python Agent Framework (winzent). https://gitlab.com/mango-agents/mango-library/-/tree/Winzent-Integration/mango_library/negotiation/winzent. [Online; zuletzt aufgerufen 06.11.2021].
Michael Wooldridge. 2009. *An introduction to multiagent systems*. John wiley & sons.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009

The impact of redispatch on grid and storage expansion planning in the German energy system

CLARA BÜTTNER, Flensburg University of Applied Sciences, Germany
ULF MÜLLER, Flensburg University of Applied Sciences, Germany

In light of the necessity for a cost-effective transition to a new energy system, it has become increasingly important to consider the integration of redispatch measures as a potential option in long-term grid planning. This could help to avoid the need for significant expansion of the grid and storage facilities. In this paper, we present a novel method for the consecutive optimization of the sector-coupled electricity system of the year 2035 at bidding zone level, as well as on a 300 bus model of the German extra high and high voltage grid. Therefore, redispatch can be explicitly co-optimized in conjunction with grid and storage expansion. The results demonstrate that redispatch measures can markedly reduce the necessity for grid and storage expansion, ranging from 76 % to 78 % in the German grid. By employing only 21 % of the optimal annual redispatch volume, 77 % of the potential investment cost savings can already be captured.

CCS Concepts: • **Applied computing → Engineering**.

Additional Key Words and Phrases: Grid Expansion Planning, Storage Expansion Planning, Redispatch, Open Source, Open Data

1 INTRODUCTION

The speed of transmission grid expansion is currently one challenge for the transition towards a 100% renewable energy system in Germany [Heering and Gustafson 2021]. German grid development plans (GDPs) by the German Transmission System Operators (TSOs) identify a significant need for additional lines [German TSOs 2021, 2023]. However, the implementation of these measures is taking longer than anticipated, as documented in [Heering and Gustafson 2021].

Until the year 2023, these expansion plans were designed to enlarge the grid capacities to be able to transfer as much electricity as needed by the market. Redispatch or curtailment was employed only to a limited extent and was not a viable alternative to grid expansion. This results in an overinvestment in transmission capacity and a decline in welfare according to [Kemfert et al. 2016]. Since 2023, there has been a notable shift in the planning paradigm of the TSOs. As stated by the TSOs, redispatch measures are regarded as a potential alternative solution if they demonstrate to be more cost-effective [German TSOs 2023]. Nevertheless, the necessity for grid expansion remains significant and is not diminished to any considerable extent compared to the preceding plans.

Scientific studies such as [Brown et al. 2018b; Büttner et al. 2024; Senkpiel et al. 2018], employ methods for identifying grid expansion needs with a particular focus on the transmission grid, while also considering the current mechanisms of the energy market in a limited and rather idealized way. Generation dispatch and grid expansion are calculated in one optimization problem, which implicitly includes an optimal redispatch without any compensation for power plants participating in the redispatch market apart from marginal costs for additional generation. It can be assumed that the efficient utilization of flexibility is not reflected by the current market design. Furthermore, it is not possible to calculate redispatch explicitly. In contrast, other studies (e.g. [Hirth and Schlecht 2018; Trepper et al. 2013]) focus on modeling future energy markets in Germany. The transmission grid is not considered, redispatch is calculated based on market conditions.

An interface between the energy market and the electrical grid is analyzed by [Nappu and Arief 2015]. The study employs a model of economic redispatch that considers both transmission congestion and network losses. The proposed method is applied to a small exemplary system of three nodes, which lacks the scalability to be applied to large-scale real-world systems. The impact of cost-based redispatch as a flexibility to reduce grid expansion needs in Germany is analyzed in [Staudt and Oren 2021] through the use of a five-zone model for the transmission grid. In [Grimm et al. 2021], a methodology for optimizing generation expansion, spot-market trading, transmission expansion, and cost-based redispatch is presented. This method is applied to the German electricity system with a spatial resolution of 16 nodes to analyze the impact of different market designs on transmission grid expansion. Within the dissertation of [Nüßler 2012] a model to analyze the impact of the electricity market on the German transmission grid is developed. The German transmission grid is aggregated to 31 zones. The results of these studies collectively indicate that the incorporation of redispatch considerations into the planning process for transmission grid expansion in Germany has the potential to reduce the necessity for such expansion and to enhance economic welfare.

The German transmission grid is highly aggregated within these studies. The quantification of bottlenecks and grid expansion needs is feasible only between the five and 31 clusters. In contrast, the actual extra high voltage (eHV) grid in Germany is far more complex, comprising over 500 substations. Consequently, the high aggregation may result in overlooked grid expansion needs. This study aims to quantify the impact of redispatch as a flexibility option in grid expansion planning, with a particular focus on the German extra high voltage and high voltage (eHV/HV) grid at a high spatial resolution (300 nodes). The following research questions will be addressed:

- How can redispatch explicitly be co-optimized in national power system planning on eHV/HV level at a high spatial resolution of 300 buses?
- To what extent can different degrees of redispatch reduce the need for grid and storage expansion in the German eHV/HV grid?

The developed method, which co-optimizes redispatch and the expansion of grid and storage, is described in section 2. Results of the application on the German transmission grid in the year 2035 are presented in Section 3. These are discussed in Section 4. Section

Authors' addresses: Clara Büttner, clara.buettner@hs-flensburg.de, Flensburg University of Applied Sciences, Kanzleistraße 91-93, Flensburg, Germany, 24943; Ulf Müller, ulf.p.mueller@hs-flensburg.de, Flensburg University of Applied Sciences, Kanzleistraße 91-93, Flensburg, Germany, 24943.

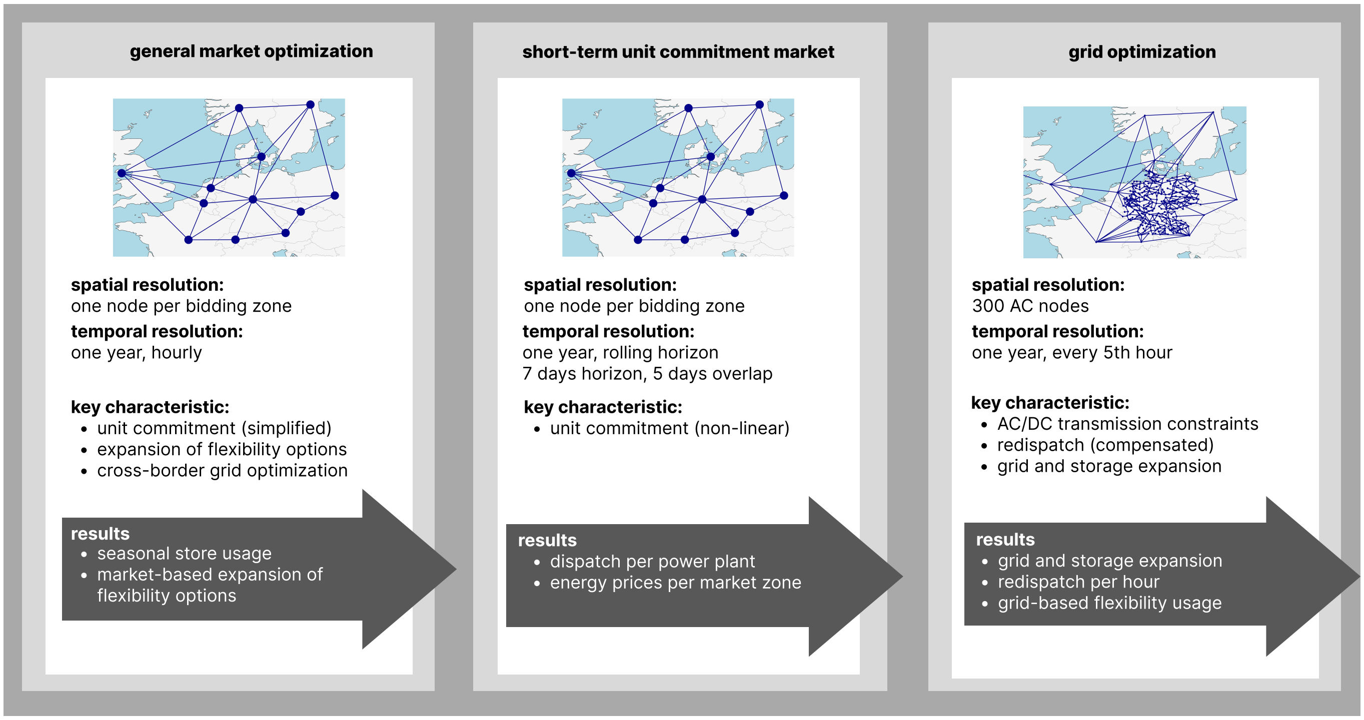


Fig. 1. Method steps and key characteristics

5 summarizes the main findings and gives an overview of potential improvements.

2 METHOD

To address the research questions of this paper, a consecutive method was developed that separates the market model from the grid model (e.g. as in [German TSOs 2023; Staudt and Oren 2021]). The subsequent sections describe the corresponding optimization steps and their interfaces. Furthermore, a comprehensive overview is provided in Figure 1.

The developed method is implemented in the python tool *Electricity Transmission Grid Optimization* (*eTraGo*)[1] mainly using the python package *Python for Power System Analysis* (*PyPSA*)[2][Brown et al. 2018a]. The fundamental methodology used in this paper is the Linear Optimal Power Flow (LOPF), which minimizes system costs restricted by several technical constraints. For each optimization step, the objective function and the main constraints are summarized based on [Brown et al. 2018a]. More details can be found in the original source.

The used input data and the general parametrization are characterized by the scenario *eGon2035_lowflex* which is described in [Büttner et al. 2024]. This midterm scenario for the year 2035 is based on the GDP[German TSOs 2021]. In particular, the GDP sets the general exogenous definition concerning future demands, generation capacities, vehicle charging and power-to-heat infrastructure [Büttner et al. 2024]. The *lowflex* scenario is characterized by a relatively conservative set of assumptions that do not account for sophisticated flexibility options such as demand-side management (DSM). Instead, the optimization focuses on the grid, hydrogen store, battery expansion and the flexible dispatch of power plants. While this paper focuses on the novel method, a more detailed explanation of the data and parametrization of the underlying scenario can be found in [Büttner et al. 2024].

[1]https://github.com/openego/eTraGo
[2]https://github.com/PyPSA/PyPSA/

2.1 Market optimization

The market model is subdivided into two optimization steps. Initially, seasonal storage behavior and market-based expansions of flexibility options are identified through a simplified yet annual calculation (see Section 2.1.1). The second step aims to come up with a realistic dispatch for each power plant and hour, taking into account non-linear Unit Commitment (UC) constraints and a short-term planning horizon (see Section 2.1.2). In both steps, the spatial resolution is defined by one node per current bidding zone (as defined in [ENTSO-E 2021]).

2.1.1 General market optimization. As short-term planning horizons are often unsuitable for mid- and long-term planning [Krüger et al. 2015; Étienne Cuisinier et al. 2022], we designed a general market model as one optimization problem for 8,760 hours of an entire year. The objective function is defined as in Equation 1. A random noise on the marginal cost of generators $o_{n,r}$ is added, in order to differentiate between different generators with the same carrier and marginal cost [Müller 2021]. Within this first optimization step, the

NOMENCLATURE

Indices

n	Bus labels
lf	Free of charge redispatch-able link labels
s	Store labels
r	Generator labels
t	Snapshot labels
ℓ	Passive branch labels
rp	Redispatch-able generator labels
rf	Free of charge redispatch-able generator labels
l	Link labels
lp	Redispatch-able link labels

Variables

$g_{n,r,t}$	Market dispatch of generator r at node n at time t
$g_{n,rf,t}$	Dispatch of generator rf at node n at time t
$g^{\text{down}}_{n,rp,t}$	Ramp down of redispatch generator rp at node n at time t
$g^{\text{up}}_{n,rp,t}$	Ramp up of redispatch generator rp at node n at time t
$f_{l,t}$	Market dispatch of link l at time t
$f_{lf,t}$	Dispatch of link lf at time t
$f^{\text{down}}_{lp,t}$	Ramp down of redispatch link lp at time t
$f^{\text{up}}_{lp,t}$	Ramp up of redispatch link lp at time t
$f_{\ell,t}$	Loading of branch ℓ at time t
$h_{n,s,t}$	dispatch of store s at node n at time t
$e_{n,s,t}$	state of charge of store s at node n at time t
F_ℓ	Capacity of passive branch ℓ
F_l	Capacity of link l
$H_{n,s}$	Capacity of store s at node n

Coefficients and bounds

c_ℓ	CAPEX passive branch ℓ
c_l	CAPEX link l
$c_{n,s}$	CAPEX store s at node n
$o_{n,r}$	OPEX of generator r at node n
$o^{\text{down}}_{n,rp,t}$	Ramp-down cost of generator rp at node n at time t
$o^{\text{up}}_{n,rp,t}$	Ramp-up cost of generator rp at node n at time t
$o_{n,rf}$	OPEX of generator rf at node n
o_l	OPEX of link l
o_{lf}	OPEX of link lf
$o_{n,s}$	OPEX of store s at node n
w_t	Weight coefficient at time t
$\eta^{\omega_t}_{n,s,0}$	Standing loss of store s at node n
$\eta_{n,s,+}$	Charging efficiency of store s at node n
$\eta_{n,s,-}$	Discharging efficiency of store s at node n
x_ℓ	Series inductive reactance of branch ℓ
$\lambda_{m,t}$	Market price in zone m at time t
$\bar{g}_{n,r,t}$	Per-unit power availability of generator r at node n at time t
$G_{n,r}$	Installed capacity of generator r at node n
$o^{\text{add}}_{n,r}$	Management fee for redispatch r
$\alpha_{l,n,t}$	-1 if l starts at node n, 1 if l ends at node n, $\eta_{l,t}$ if l is a link with time-dependent efficiency (e.g. heat pump)
$\alpha_{\ell,n,t}$	-1 if ℓ starts at node n, 1 if ℓ ends at node n
$C_{\ell,c}$	Matrix including circles c as a combination of branches ℓ
$u_{n,r,t}$	On/off binary status for generator unit commitment
$T^{\text{min\_up}}$	Minimum up time of generator r
$T^{\text{min\_down}}$	Minimum down time of generator r
$suc_{n,r(,t)}$	Start up cost of generator r

dispatch of seasonal storage units (i.e. hydrogen tank stores and primarily cavern-based methane stores) can be identified. The dispatch of these stores is defined as a subset of $h_{n,s,t}$. Moreover, long-term investment decisions are made concerning the expansion of cross-market-zone links (F_l) and the expansion of storage ($H_{n,s}$). These resulting values of these optimization variables become exogenous input values in the subsequent short-term UC optimization problem (explained in Section 2.1.2). While all the other variables are also of significance within the context of general market optimization, they are not utilized as inputs in any subsequent optimization step.

Given the high temporal complexity of the problem, it is necessary to address the computational burden associated with the short-term planning variables. Therefore, the UC constraints were simplified using a designated PyPSA functionality[3] following the relaxation approach described in [Hua et al. 2018]. As a consequence of the linearization of the UC problem, the start-up and shut-down costs are excluded from the objective function.

$$\min_{F_l,H_{n,s},g_{n,r,t},f_{l,t},h_{n,s,t}} \Bigg[\sum_l c_l \cdot F_l + \sum_{n,s} c_{n,s} \cdot H_{n,s} + \sum_{n,r,t} o_{n,r} \cdot g_{n,r,t} + \sum_{n,l,t} o_{n,l} \cdot f_{l,t} + \sum_{n,s,t} o_{n,s} \cdot [h_{n,s,t}]^+ \Bigg] \tag{1}$$

Constraints following Equation 2 ensure that the inelastic demands at each node are met at each time t by the energy of generators, stores or by the flows $f_{l,t}$ from the branches l. The bidding zones are interconnected via link components, thus $\sum_\ell \alpha_{\ell,n,t} \cdot f_{\ell,t}$ is zero in the market optimization.

$$\sum_r g_{n,r,t} + \sum_s h_{n,s,t} + \sum_l \alpha_{l,n,t} \cdot f_{l,t} + \sum_\ell \alpha_{\ell,n,t} \cdot f_{\ell,t} = d_{n,t} \tag{2}$$

The dispatch of generators is constrained by their capacities and, in the case of fluctuating renewables, time-dependent availability (Equation 3). The same equation can be easily adapted for link and store components, thus their dispatch is limited analogously. The state of charge of energy stores is given by Equation 4.

$$0 \leq g_{n,r,t} \leq \bar{g}_{n,r,t} \cdot G_{n,r} \tag{3}$$

$$e_{n,s,t} = \eta^{\omega_t}_{n,s,0} e_{n,s,t-1} + \eta_{n,s,+} \cdot \omega_t [h_{n,s,t}]^+ - \eta_{n,s,-} \cdot \omega_t [h_{n,s,t}]^- \tag{4}$$

[3] https://github.com/PyPSA/PyPSA/blob/cf17345dfd0c94546cdc44a631e0ab9d09565183/pypsa/optimization/optimize.py#L512

2.1.2 Short-term Unit Commitment. The second optimization step develops a more realistic short-term planning of the hourly power plant dispatch. While the constraints from the general market optimization (Equations 2 - 4) remain, most importantly the non-linear UC constraints of conventional power plants are additionally considered. The mathematical definition of these constraints based on [Brown et al. 2018a] is given in Equations 5 - 8. The carrier-specific parameters are taken from the tool *pypsa-eur*[4].

$$u_{n,r,t} \cdot \tilde{g}_{n,r,t} \cdot \bar{g}_{n,r} \leq g_{n,r,t} \leq u_{n,r,t} \cdot \bar{g}_{n,r,t} \cdot \bar{g}_{n,r} \tag{5}$$

$$\sum_{t'=t}^{t+T^{\text{min\_up}}} u_{n,r,t'} \geq T^{\text{min\_up}}(u_{n,r,t} - u_{n,r,t-1}) \tag{6}$$

$$\sum_{t'=t}^{t+T^{\text{min\_down}}} (1 - u_{n,r,t'}) \geq T^{\text{min\_down}}(u_{r,s,t-1} - u_{n,r,t}) \tag{7}$$

$$suc_{n,r,t} \geq suc_{n,r}(u_{n,r,t} - u_{n,r,t-1}) \tag{8}$$

With regard to particular operational expenses, while shutdown costs remain unconsidered, additional startup costs ($suc_{n,r,t}$) for conventional generators are regarded as stated in the objective function as defined in Equation 9. These start-up costs $suc_{n,r,t}$ are also parametrized as in *pypsa-eur*[5].

$$\begin{aligned}\min_{g_{n,r,t}, f_{l,t}, h_{n,s,t}} \Bigg[&\sum_{n,r,t} (o_{n,r} \cdot g_{n,r,t} + suc_{n,r,t}) \\ &+ \sum_{l,t} o_l \cdot f_{l,t} + \sum_{n,s,t} o_{n,s} \cdot [h_{n,s,t}]^+ \Bigg]\end{aligned} \tag{9}$$

In addition to the introduction of non-linear UC constraints, the temporal planning horizon is significantly shortened in this second optimization step, while maintaining an hourly resolution. In a rolling horizon approach, the optimization is divided into weekly problems with an overlap of five days resulting in a look-ahead period of two days. This approach is motivated by two aspects: First, the rolling short-term planning horizon seems to be a more realistic model for short-term dispatch decisions [Bødal et al. 2022]. In the case of weather-dependent renewable power plants, it is reasonable to assume a limited look-ahead period. Second, due to the higher complexity of the model induced by the non-linear UC constraints, it is necessary to reduce the computational burden. The explained rolling horizon approach successfully generates tractable optimization problems.

In contrast, as stated in the previous Section, the used rolling horizon approach is not suitable for mid- and long-term decisions. Consequently, no investment decisions are optimized (see Equation 9) but instead exogenously defined by the previous general market optimization. The general market optimization also sets the initial state of charge as well as the state of charge at the end of each planning horizon for the seasonal storage units. A maximum deviation of ±1% is permitted.

[4]https://github.com/PyPSA/pypsa-eur/blob/03fd9bfb528a4148614865837977f28a10ea7fc7/data/unit_commitment.csv?plain=1#L1-L7
[5]https://github.com/PyPSA/pypsa-eur/blob/03fd9bfb528a4148614865837977f28a10ea7fc7/data/unit_commitment.csv?plain=1#L8

2.2 Grid optimization

In this final optimization phase, it can be evaluated how well the market dispatch aligns with the power grid and its associated technical constraints. Consequently, the spatial resolution is significantly enhanced. Given the computational limitations associated with performing annual optimization runs on the original grid complexity, a representative power system with 300 AC buses is considered (as in [Büttner et al. 2024]). This resolution is still comparably high and has been used in many previous scientific works [Büttner et al. 2024; Müller 2021; Müller et al. 2019b; Wienholt et al. 2018].

The mathematical definitions of the constraints stated in Section 2.1.1 also apply for this optimization step but lead to more constraints as the number of buses n increases.

The inner-German AC lines are modeled in this optimization in addition to the active cross-border flows which are already part of the previous optimization steps. Consequently, the linear power flow equations (Equation 10) ensure the physical behavior of AC-lines. Due to linearization, only the predominant inductive reactance of the lines is considered.

$$\sum_{\ell} C_{\ell c} \cdot x_{\ell} \cdot f_{\ell,t} = 0 \tag{10}$$

Furthermore, as stated in Equation 11, AC line flows are limited to the specific line capacities (including voltage-level specific sweeping (n-1) security margins $\bar{f}_{\ell}$ as defined in [Müller et al. 2019b]).

$$|f_{\ell,t}| \leq F_{\ell} \cdot \bar{f}_{\ell} \tag{11}$$

At a relatively decentralized level, long-term investment decisions and hourly redispatch options are optimized in one optimization problem, as stated in Equation 12. The integrated nature of the optimization problem motivates an annual planning horizon and an hourly resolution. Due to the high computational burden, only every fifth time step is considered and weighted with $w_t = 5$ (see [Büttner et al. 2024]). In order to consider changing reactances due to line expansion, four iterations of the LOPF are performed updating the reactances after each iteration (as in [Büttner et al. 2024]).

To prevent the generation of implausible outcomes, the expansion of transmission lines is constrained to a maximum of four circuits per eHV and two circuits per HV line. The assumed investment costs for battery and grid expansion are listed in Table 1.

Table 1. Investment costs for storage and grid expansion

type	**value**	**source**
battery expansion cost	838.0 €/kW	[Danish Energy Agency 2020]
grid expansion cost HV	0.23 €/kVA/km*	[Agricola et al. 2012]
grid expansion cost eHV	0.70 €/kVA/km*	[German TSOs 2021]
discount rate	5 %	[Victoria et al. 2021]

* part of sensitivity analysis

$$\min_{F_\ell, F_l, H_{n,s}, g^{\text{down}}_{n,rp,t}, g^{\text{up}}_{n,rp,t}, g_{n,rf,t}, f^{\text{down}}_{lp,t}, f^{\text{up}}_{lp,t}, f_{lf,t}, h_{n,s,t}} \Bigg[\sum_\ell c_\ell \cdot F_\ell + \sum_l c_l \cdot F_l + \sum_{n,s} c_{n,s} \cdot H_{n,s} + \sum_{n,rp,t} w_t \left(o^{\text{down}}_{n,rp,t} \cdot g^{\text{down}}_{n,rp,t} + o^{\text{up}}_{n,rp,t} \cdot g^{\text{up}}_{n,rp,t} \right) + \sum_{n,rf,t} w_t \cdot o_{n,rf} \cdot g_{n,rf,t} + \sum_{n,lp,t} w_t \left(o^{\text{down}}_{n,lp,t} \cdot f^{\text{down}}_{lp,t} + o^{\text{up}}_{n,lp,t} \cdot f^{\text{up}}_{lp,t} \right) + \sum_{lf,t} w_t \cdot o_{lf} \cdot f_{lf,t} + \sum_{n,s,t} w_t \cdot o_{n,s} \cdot [h_{n,s,t}]^+ \Bigg] \quad (12)$$

Investment decisions on links (F_l), AC lines (F_ℓ) and storage units ($H_{n,s}$) are made without any direct constraining linkage to the previous market optimization. Therefore, cross-border line expansion is re-evaluated. Furthermore, storage and grid expansion can be determined locally without having to meet the bidding zone results from the general market optimization at the aggregated level.

In contrast, the power plants that participate in the redispatch market and are therefore compensated for its activation (see Table 2) depend on the result of the previous UC optimization. Their market dispatch ($g_{n,r,t}$ or $f_{l,t}$) has to be met and is therefore not part of the objective function. Instead, the remaining potential redispatch flexibility for ramping up or down is constrained as in Equation 13 and 14. These redispatch variables $g^{\text{down}}_{n,rp,t}$ and $g^{\text{up}}_{n,rp,t}$ as well as $f^{\text{down}}_{lp,t}$ and $f^{\text{up}}_{lp,t}$ multiplied by their specific compensation costs are subject to the minimization of the overall objective function. The specific compensation costs depend on the market prices $\lambda_{m,t}$ from the UC market optimization as well as the carrier-specific operational costs. In particular, the specific ramp up and ramp down costs are defined in Equation 15 and 16. The management fee $o^{\text{add}}_{n,r}$ is set to zero in the base setting. However, in the sensitivity analysis this value is set differently (see Section 3.5). Equation 13-16 can be directly adapted for the link components with the index lp. Hence, the constraints and redispatch compensations for these open cycle gas turbine (OCGT) are analogous to those of the other generators.

$$0 \leq g^{\text{down}}_{n,rp,t} \leq g_{n,r,t} \quad (13)$$

$$g_{n,r,t} \leq g^{\text{up}}_{n,rp,t} \leq \bar{g}_{n,r,t} G_{n,r} \quad (14)$$

$$o^{down}_{n,rp,t} = \lambda_{m,t} - o_{n,r} + o^{\text{add}}_{n,r} \quad (15)$$

$$o^{\text{up}}_{n,rp,t} = \begin{cases} \lambda_{m,t}, & \lambda_{m,t} > o_{n,r} \\ o_{n,r}, & \lambda_{m,t} < o_{n,r} \end{cases} + o^{\text{add}}_{n,r} \quad (16)$$

Furthermore, Table 2 shows, that in addition to the energy carriers that provide redispatch flexibility in a structured and therefore compensated manner, all the others except combined heat and power plant (CHP) are also obliged to flexibly produce energy in a system-optimal way but without receiving any compensation. CHP do not provide any redispatch flexibility. Conversely, their dispatch is fixed in accordance with the parameters established by the UC market optimization.

Table 2. Categorization of generator and link carriers with respect to their redispatch availability

Redispatch availability	**carrier**	**label**
Compensated	coal, lignite, nuclear, oil, others, reservoir, run of river, solar (ground-mounted and rooftop), wind (on- and offshore), biomass, OCGT	rp for generators, lp for links i.e. OCGT
Free of charge	geo thermal, solar thermal, heat pumps (central and rural), central gas boiler, central resistive heater, electrolysis, fuel cells, natural gas feed-in, steam methane reforming (SMR)	rf for generators, lf for links
None	CHP	chp

2.3 Different degrees of redispatch

The question of whether and to what extent long-term power system planning should rely on short-term redispatch measures is open to debate. In order to address this question, we have generated results for several possible system designs, which allow for different degrees of redispatch. Therefore, in certain calculations the curtailment (i.e. ramp down) of power plants within the overall system is limited to a varying defined value R. The additional constraints of the optimization problem are defined according to Equation 17.

$$\sum_{n,rp,t} w_t \cdot g^{\text{down}}_{n,rp,t} + \sum_{n,lp,t} w_t \cdot f^{\text{down}}_{lp,t} \leq R \quad (17)$$

The model is first optimized without any national upper limit of redispatch usage ($R = \infty$). Subsequently, calculations are performed with different redispatch limits ranging from zero ($R = 0$) to the resulting redispatch observed in the calculations with unlimited redispatch.

3 RESULTS

This section presents the findings of market optimization and grid optimization studies, employing different degrees of redispatch. Moreover, a sensitivity analysis was conducted to evaluate the influence of selected, presumed sensitive parameters.

3.1 Market optimization

The electricity generation in the market optimization is predominantly based on renewable energy. In Germany and its neighboring countries, 74 % of the total generation is derived from renewable sources. Wind onshore, reservoir and solar power plants account for the largest share. The potential of wind and solar power generation is almost fully utilized, with only a few percent being not used (wind onshore: 5 %, solar: 0.4 %). Over 50 % of the electricity generated in Germany is produced by wind turbines (118 TWh offshore, 275 TWh onshore). Solar power plants feed in 136 TWh of electricity. The

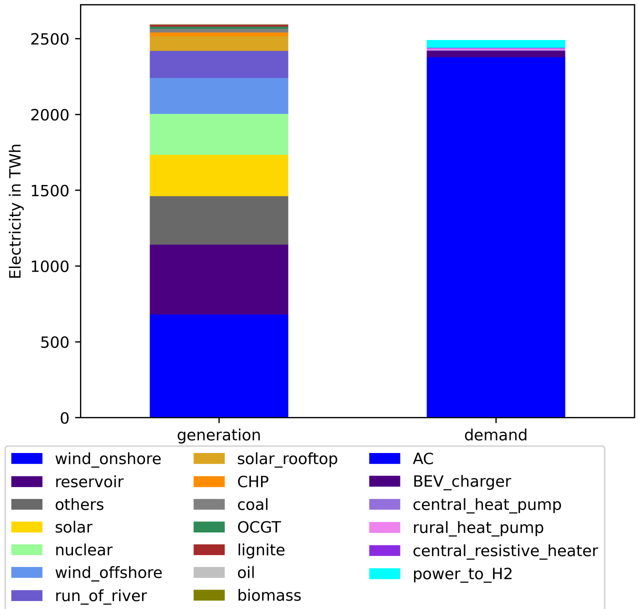


Fig. 2. Results of market optimization

dispatch of conventional power plants in Germany is comparably small. The annual dispatch per carrier and demand category for the overall system is shown in Figure 2.

It can be observed that the majority of the electricity is required to meet the exogenous electricity demand, amounting to over 2.000 TWh. It should be noted that electricity demands due to sector-coupling are modeled separately and are not included in this demand. The inflexible charging of battery electric vehicle (BEV) consumes 41 TWh. Power-to-heat dispatch competes with conventional, gas-based heat supply in district heating grids. In the market optimization, 20 TWh electricity are utilized for the production of heat. In addition, electrolyzers with an installed capacity of 9.1 GW are built in Germany. They consume in total 52 TWh of electricity. The primary objective of hydrogen production is to reduce the reliance on natural gas. Although the utilization of hydrogen as a seasonal store is theoretically possible through the construction of hydrogen stores and fuel cells, the relatively small size of fuel cells and the frequent charging and discharging of hydrogen stores suggest that hydrogen is not widely employed as a seasonal store.

All subsequent grid optimizations are based on these market results.

3.2 Optimal redispatch

In the absence of constraints on redispatch, about 30.7 TWh of electricity is curtailed and 37.0 TWh is ramped up in the overall model region. These numbers include only compensated redispatch (see Table 2). The discrepancy between the compensated curtailment and compensated ramp-up can be attributed primarily to the free-of-charge redispatch of electrolysis.

Although the installed electrolysis capacity is similar in the market (9.1 GW) and grid optimization (9.0 GW), there is a 4 TWh increase in dispatch in the grid optimization compared to the market optimization. This redispatch is free of charge (see Table 2) and is not included in the given redispatch amounts. Less natural gas is used to fulfill the hydrogen demands in exchange. In addition, a smaller shift (1.8 TWh) from conventional, gas-based heating supply to power-to-heat can be seen. Furthermore, the increased utilization of energy storage in the grid optimization leads to a higher rate of energy loss compared to the market optimization.

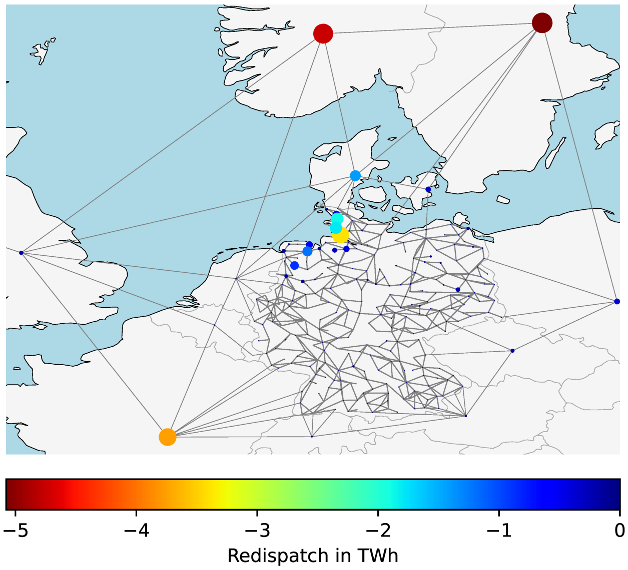
The spatial distribution of redispatch throughout the year is visualized in Figure 3. The majority of the total curtailment occurs in the German system (14.5 TWh), since in other countries eHV/HV grid bottlenecks are barely considered. Mainly wind turbines offshore (9.4 TWh) and onshore (12.9 TWh) are curtailed in all countries. In neighboring countries, conventional power plants such as coal, lignite and nuclear are curtailed. Solar power plants are curtailed to a smaller amount (0.65 TWh).

Notably high levels of curtailment are observed at offshore wind grid connection points along the German North Sea coast, particularly near Hamburg. At the bus near Hamburg, 3.4 TWh of wind offshore energy is curtailed, which accounts for 28 % of its market-optimal dispatch. Although energy feed-in from wind offshore plants is curtailed throughout the whole year, in autumn and winter the effect is particularly high (see Figure 4). The curtailment of solar power plants is less prevalent at numerous nodes within the grid. The temporal distribution demonstrates that curtailment is confined to specific hours that align with the pattern of solar daytime potential.

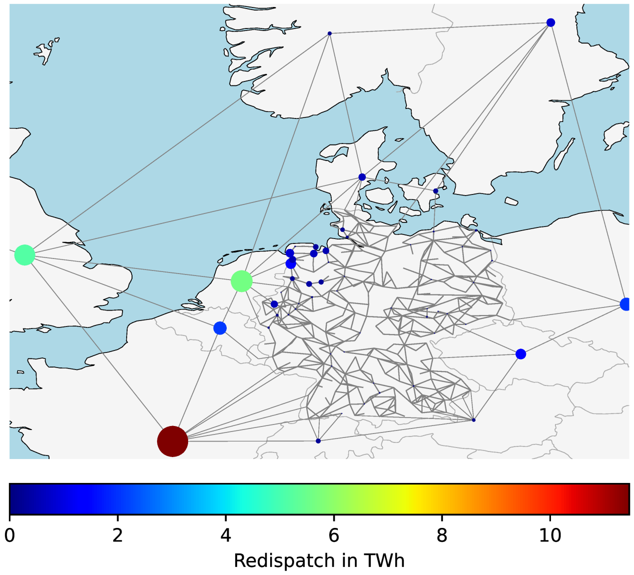
Positive redispatch is provided by nuclear (10.9 TWh), wind (onshore 4.2 TWh, offshore 17.1 TWh), OCGT (1.2 TWh) and coal (1.6) power plants. In Germany, the largest shares are provided by wind onshore plants and OCGT. OCGT were not dispatched in the market simulation due to their high marginal costs, but can provide flexibility when the grid topology is considered, especially in grid regions with high energy demand and low renewable production. Curtailment and ramp-up of power plants often do not coincide, which is possible using battery storage. Such measures permit the redispatch to be reallocated to periods when market prices are lower, thereby reducing the costs associated with redispatch.

Transmission lines are highly expanded in the north-west of Germany (see Figure 5a). Some lines have reached the maximum feasible extent of expansion. The expansion of these lines is primarily driven by the need to transfer wind energy to regions with limited renewable energy resources and higher energy demand in the center and south of Germany. High transmission capacities are added between the North and the center of Germany. Lines in central and Southern Germany are expanded less. In addition, transmission lines on two routes from east to west are significantly expanded by about 5 GVA each.

The capacity of batteries is expanded to 11.2 GW in the overall system, of which 1.6 GW are located in Germany. The largest battery capacity is expanded in the United Kingdom. The majority of battery capacities in Germany are situated in the southern region. Additional larger batteries are built near Frankfurt and at the west coast of Schleswig-Holstein. The expanded batteries in Germany



(a) Ramp down (curtailment)



(b) Ramp up

Fig. 3. Spatial distribution of redispatch. Colors of buses indicate the amount of redispatch over the year, their sizes increase proportionally to the amount of redispatch to improve visibility.

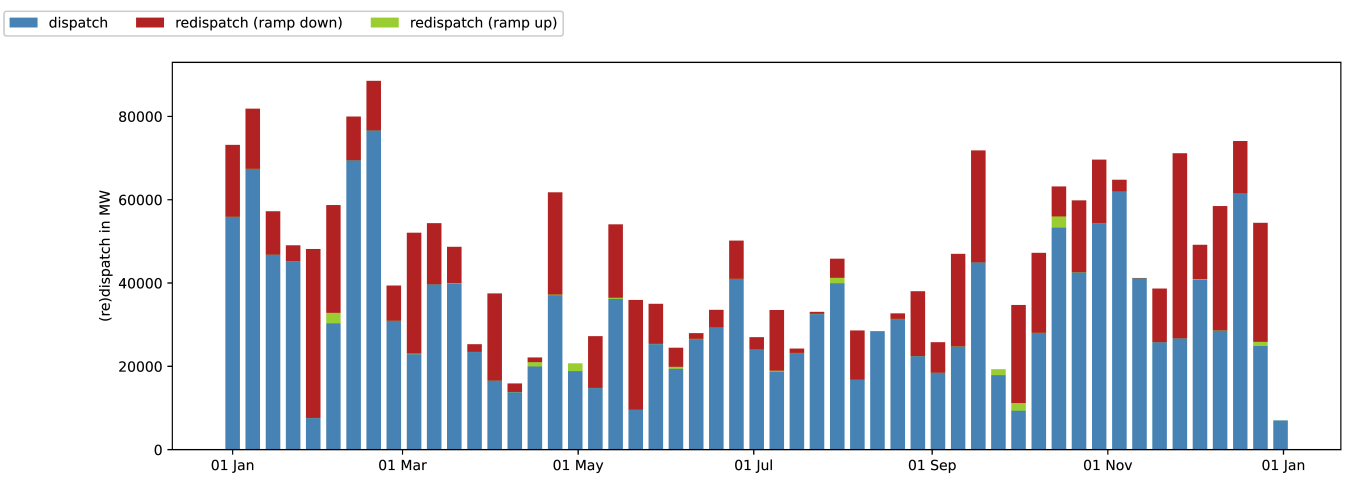
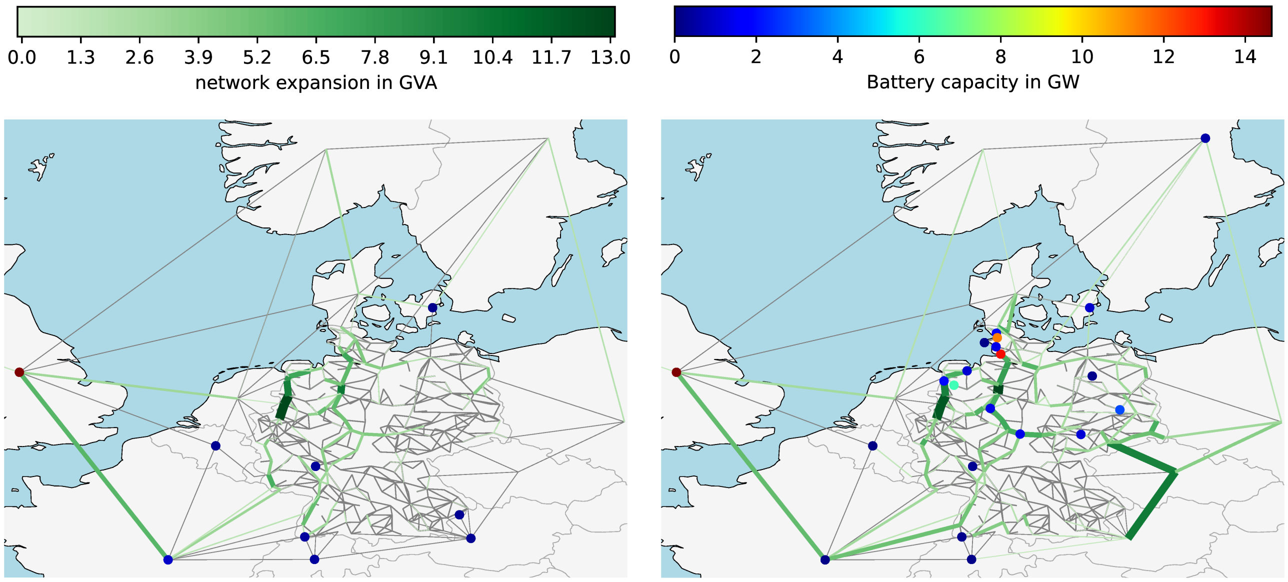


Fig. 4. Temporal distribution of dispatch and redispatch of wind offshore power plant near Hamburg

reach installed capacities of up to 217 MW each (corresponding to a storage capacity of 1.3 GWh). Batteries are charged and discharged frequently over the year, following the pattern of electricity prices.

Hydrogen stores are extendable to allow storing electricity on a seasonal basis. The capacity of hydrogen stores in Germany is expanded to 66 GWh_{H2}. Electrolysis and fuel cells can be built in addition to the use of hydrogen stores in the electricity system. Given the limited capacity of fuel cells and the dispatch time series of hydrogen stores, it can be inferred that these stores are not utilized to serve seasonal flexibility for the electricity system. This is likely due to high investment costs.

In a system with unlimited redispatch potential, system costs for all considered countries are summed up to 166 billion €/a. These costs include dispatch cost (163 billion), redispatch cost (0.6 billion) as well as annualized investment costs for grid expansion (1.3 billion), stores (2.0 billion) and sector-coupling technologies (0.4 billion).



(a) Optimal redispatch

(b) Without redispatch

Fig. 5. Spatial distribution of grid and battery expansion. Colors and widths of lines indicate grid expansion. Colors of buses show battery expansion, their sizes increase proportionally to the expansion to improve visibility. Both legends refer to both figures.

3.3 Without redispatch

System costs increase when redispatch is not an allowed option and the generator dispatch from the market optimization must be strictly met in the grid optimization. The overall system costs sum up to 170 billion €/a (+3.0 billion / +2.4 % compared to optimal redispatch). This increase is driven by rising investment costs (+4 billion €/a / +108 % compared to optimal redispatch). Notably, the costs associated with storage investments are higher (5.4 billion €/a / +170 %), while grid investments also increase and sum up to 2.5 billion €/a (+92 %).

Similar to the results in the previous Section 3.2, the spatial distribution of grid expansion (see Figure 5b) indicates that expansion is required primarily in the northwestern region of Germany. The expanded transmission routes are consistent with the results of the system with optimal redispatch, but the additional capacity is higher in some line sections.

In this calculation, the expansion of batteries is significantly more pronounced. In the overall system, a total of 65 GW (391 TWh) of capacity is installed, of which 45 GW (271 TWh) is connected to the German grid. In contrast to the calculation in Section 3.2, batteries are being expanded primarily in the north of Germany. The installed storage capacities are much higher (up to >10 GW at one grid node). Additional locations for larger batteries are identified in regions in the east and south of Germany. The frequent usage of battery storage units follows electricity prices.

The capacity of hydrogen stores in Germany is expanded to 19 GWh_{H2}, about 75 % less than in a system with optimal redispatch. Charging and discharging of hydrogen stores do not have a seasonal pattern but change frequently during the year. While there are a few more electrolysis capacities compared to the results with optimal redispatch (+1 %), their dispatch decreases by 4 %. The capacity of fuel cells is increasing (+248 %), but remains comparatively small (81 MW), the dispatched energy is similar. The increased battery capacities may decrease the need for hydrogen stores since batteries store electricity more efficiently and flexibly.

3.4 Limited redispatch

Between the evaluated two extreme cases ($R = 0$ and $R = 31\,TWh$) 16 steps were selected as R in Equation 17. As the limitation approaches $R = 0$, the step size decreases from 5 TWh to 0.25 TWh.

The investments in grid and storage expansion are increasing when redispatch is limited. Figure 6 shows the additional investment costs and saved redispatch costs for the different redispatch limits compared to the results with unlimited redispatch. It indicates an exponential correlation between allowed curtailment and investment needs. Costs are markedly reduced when a minimal quantity of curtailment is permitted. Already 52 % of the savings of system costs are possible when allowing 0.8 % of the optimal redispatch (0.25 TWh) in all considered countries. Savings in grid and storage investment costs reach 42 % of the possible savings (see Figure 7). The curtailment of approximately 20 % of the optimal result (i.e. 6 TWh) reduces the overall system costs by 93 % of the total savings that can be achieved by redispatch. With regard to investment costs (grid and storage) 89 % of the savings are then achieved in all considered countries. Investment cost reductions increase by only six percentage points when the allowed redispatch is doubled (12 TWh). The costs associated with redispatch can be reduced through the implementation of limited curtailment. Nevertheless, the savings are

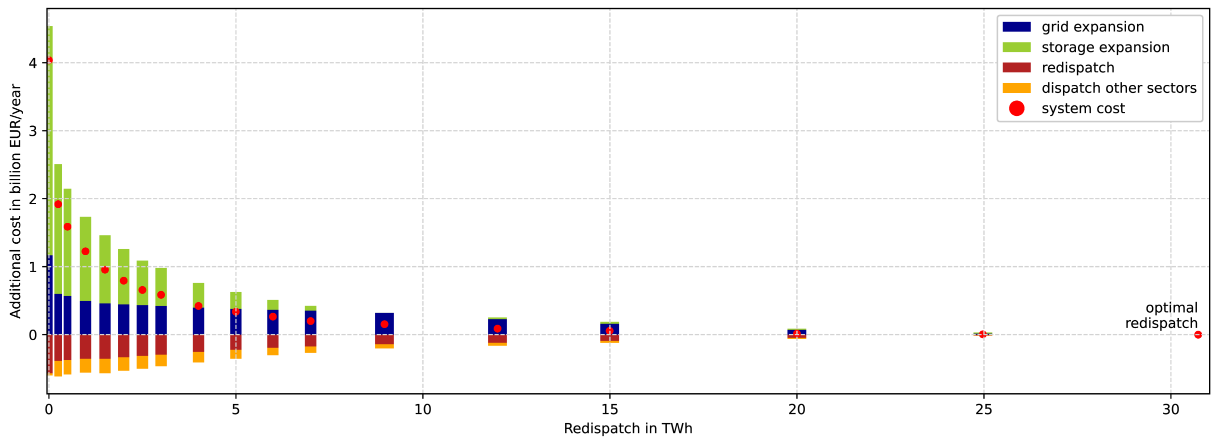


Fig. 6. Differences in system costs when limiting redispatch compared to results with unlimited redispatch

relatively insignificant in comparison to the augmented investment costs.

Furthermore, the potential for savings in system costs in Germany is dependent upon the level of redispatch that is permitted. But the possible savings with a limited amount of redispatch are smaller than in the overall system. When redispatch in the overall system is limited, a disproportionate amount of redispatch is allocated to the German grid. Thus, the shares of redispatch are higher. It is possible to reduce investments in grid and storage infrastructure in Germany by 30 % with 0.2 TWh redispatch in Germany (1.8 % of the optimal redispatch). About 21 % of the optimal redispatch is necessary to reduce the investment costs by 77 % of the possible savings. The results indicate that allowed redispatch and savings in expansion costs in Germany follow the *Pareto priciple* [Nisonger 2008]. Approximately 80 % of the savings can be achieved with 20 % of the optimal redispatch.

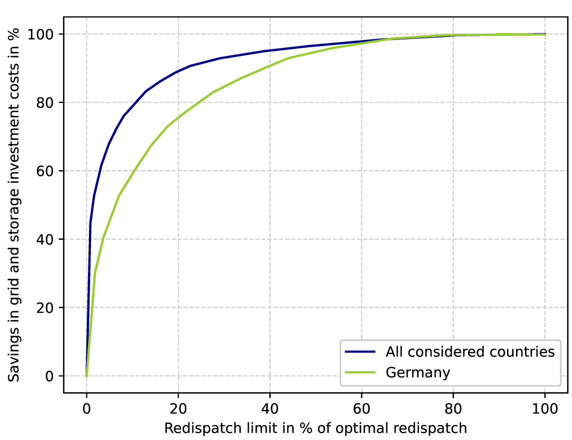


Fig. 7. Relative savings in grid and storage investment costs with limited redispatch

In general, Figure 7 shows an exponential correlation between investment savings and permitted redispatch, with a more pronounced gradient observed in the overall system compared to the one in the German grid.

3.5 Sensitivity analysis

The sensitivity analysis includes parameter variations on specific redispatch cost, specific grid expansion cost, and grid expansion limit. These parameters are assumed to be essential and therefore a significant influence on the results is expected.

Specific redispatch costs are increased by adding a management fee of 4 €/MWh for fluctuating renewables and 2 €/MWh for flexible generation, adapting the latest (but outdated) explicit incentive scheme for fluctuating renewable energy to participate in the German power market [EEG 2019]. A reduction in the specific redispatch cost is considered by reducing the compensation paid for positive and negative redispatch (as defined in Equations 15 and 16) to 70 % and 90 %. The specific grid expansion cost are scaled to 75 %, 150 % and 200 % of the assumed cost. Furthermore, model runs were conducted without upper limits on grid expansion. For each sensitivity setting, six calculations were executed considering unlimited redispatch, no redispatch and redispatch limited to 0.5 TWh, 3 TWh, 15 TWh and 31 TWh (optimum in base setting). The results are summarized in Figure 8, which plots regression curves of differences in system costs compared to the base setting with unlimited redispatch for each sensitivity.

In optimizations without upper limits on grid expansions, redispatch demand is decreased. The maximal total ramp down is optimized to about 17 TWh. Investments in grid and storage units are reduced to 90 % of the base setting. System costs are still higher without upper limits on capacities when redispatch is impossible. This suggests that redispatch can not be fully substituted by grid and storage expansion. An allowed curtailment of 0.25 TWh achieves savings in system costs of 95 %.

Compared to the upper limits variations, the results are less sensitive to changing specific grid expansion costs. Higher specific grid expansion costs result in higher optimal redispatch (+47 % with

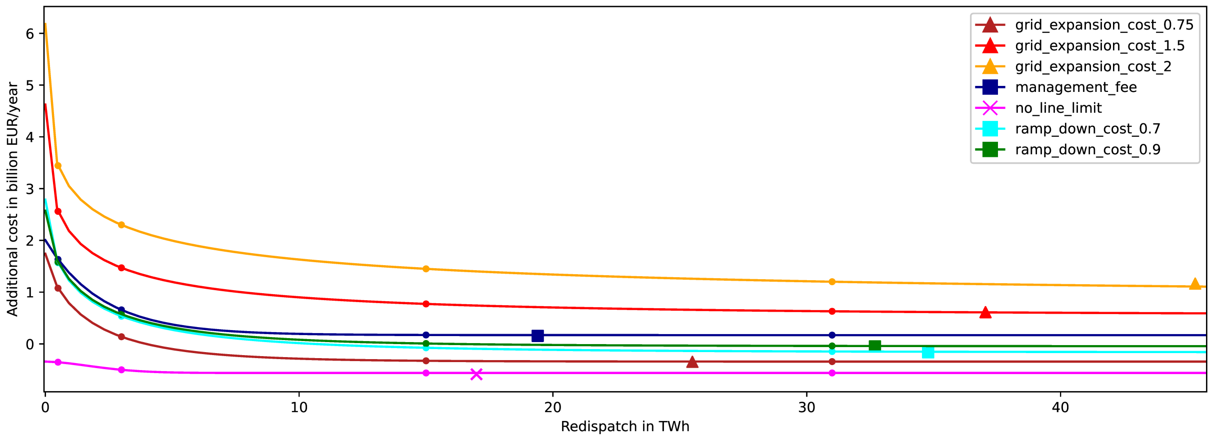


Fig. 8. Differences in system cost for sensitivities compared to the base setting without redispatch limit (regression curves). Points show the results which were the basis for exponential regression, larger markers indicate optimal redispatch.

200 % grid expansion cost) and less grid expansion. Vice versa, the maximum redispatch is decreasing (-17 %) when grid expansion costs are reduced to 75 %. The slope of the regression curve is higher when grid expansion costs are increased. This indicates that the impact of redispatch limitations is bigger when grid expansion costs are increasing. When specific redispatch costs are reduced, the optimized redispatch is up to 13 % higher. In contrast, the introduction of an additional management fee for redispatch results in a reduction of the optimal redispatch by 37 %. Investment costs change in the opposite direction. Redispatch constraints reduce the impact of varying redispatch costs. The potential for changes in investment costs achievable with unlimited redispatch ranges from 46 % to 58 % when specific redispatch or grid expansion costs change. In the German grid, savings in investment costs of 76 % to 78 % are possible with unlimited redispatch.

4 DISCUSSION

This section presents a discussion of the results in comparison to other papers and to the official grid expansion planning in Germany. Furthermore, the limitations of the developed methods and their impact on the results are critically addressed.

The data model was developed based on the results of the GDP created in the year 2021 [German TSOs 2021]. However, the grid expansion needs identified in this paper are less than those published by the German TSOs in [German TSOs 2021]. Several reasons for this discrepancy have been discussed in [Büttner et al. 2024]. It is noteworthy that the GDP 2021 did not consider redispatch measurements as a potential alternative for grid expansion [German TSOs 2021]. Nevertheless, in comparison to the scenarios that do not include or only include a minimal amount of redispatch, the grid expansion needs remain relatively low. The discrepancy is likely attributable to a significant divergence in the assessment of storage expansion. In this paper, storage expansion is optimized in a combined problem with grid expansion and redispatch. This resulted in large battery capacities when redispatch was severely restricted. In contrast, the German TSOs defined significantly smaller storage capacities exogenously [German TSOs 2021]. Thus, there are fewer alternatives to expanding the grid.

Recently, TSOs consider redispatch as an alternative when planning grid expansion. The identified redispatch in Germany varies from up to 2.5 TWh for the year 2037 and 2.8 TWh to 5.5 TWh for the year 2045 in different scenarios [German TSOs 2023]. Quantitative comparisons to the findings of this study have to be conducted carefully due to the different target years and changing shares of renewable power plants. Nevertheless, the results of this paper demonstrate that optimal curtailment in Germany is higher, despite the reduced installed capacity of renewable power plants and new demand resulting from sector coupling. One potential explanation for this discrepancy is the use of disparate modeling approaches and parameterization of redispatch. However, as the TSOs method is only described in a cursory manner in publicly accessible sources and cost parameters are not provided, it is not possible to quantify this effect. The grid expansion needs in [German TSOs 2023] are higher than in our results. These needs are lower in scenarios with more redispatch, which is plausible and in line with our findings. Battery storage units are still defined exogenously, but the capacity is significantly increased compared to the previous GDP. Their dispatch is, in general, oriented towards the market, but can also cover the needs for redispatch to a certain extent [German TSOs 2023]. It can be assumed that this results in additional costs. This is not the case in our optimizations which assume that batteries have to be used grid-supportive without any compensation. In consequence, batteries are more valuable for the reduction of redispatch and grid expansion needs in the results presented here.

The impact of redispatch on long-term planning is also the subject of various scientific studies. The necessity for grid expansion in Germany in the year 2035 can be reduced by up to 45 % when dispatch and expansion are co-optimized according to [Kemfert et al. 2016]. The authors assume that market mechanisms will change in the future. They propose a nodal-pricing approach that implicitly incorporates redispatch at no additional cost. In our results, redispatch resulted in additional cost although certain changes in the market mechanisms were assumed. The results of our analysis indicate a reduction in grid expansion investments in Germany by 40 %. When both grid and storage expansion are taken into account, savings of 78 % are achievable.

In [Franken et al. 2018], a method is developed for classifying grid expansion measures based on their potential to reduce redispatch. The method is applied to a model of the German transmission system, taking into account the expansion needs identified by the German TSOs. The results show that a scenario with less grid expansion leads to three to four times higher redispatch demand. Increased congestion management due to delayed grid expansion is also analyzed in [Müller et al. 2019a]. For the year 2035, delayed grid expansion causes about 20 TWh of redispatch in Germany which can be reduced to 5 TWh when grid expansion is implemented as planned by the TSOs. In this work, the amount of redispatch in the German transmission grid is optimized resulting in 14.5 TWh redispatch, which is more than twice the best-case scenario presented in [Müller et al. 2019a]. In contrast, the grid expansion needs are lower. However, the applied methodologies differ as redispatch and grid investments are not co-optimized but analyzed using different settings of grid expansion [Franken et al. 2018; Müller et al. 2019a]. As a result, the results are not directly comparable.

In this work, storage units, particularly battery storage, are expanded considerably to reduce redispatch. The potential of batteries to reduce redispatch is also shown in [Eickmann et al. 2014]. The potential reduction of redispatch and grid expansion needs by storage units in the German transmission grid is also identified by [Göke et al. 2022]. In contrast to this work, long-term storage options are built more than short-term storage units (batteries).

The model used in this study considers the eHV/HV grid in Germany. In contrast, in grid planning by the TSOs and in scientific papers such as [Franken et al. 2018; Göke et al. 2022; Müller et al. 2019a], only the eHV grid is considered. According to [Hoffrichter et al. 2018], considering the HV grid results in a reduction of over 30 % in the required redispatch. Therefore, the reduction in redispatch or grid expansion needs observed in the results of this study may be caused by considering the HV grid.

The scenario used in this study considers battery storage units and electrolyzers as the only technologies shifting demands in time. DSM from industrial processes, heat pumps or flexible charging of BEV is not considered, although it could potentially reduce redispatch needs, as previously analyzed in [Staudt et al. 2018].

The developed method aims for a co-optimization of redispatch measurements as well as grid and storage investment needs. The way redispatch is modeled and parameterized includes aspects of the current market situation in Germany following *Redispatch 2.0* and aspects of future updates, some following the approach of *Redispatch 3.0* [E-Bridge Consulting GmbH 2022]. Power plants can provide positive and negative redispatch regardless of the installed capacities which does not represent the current situation but follows the trend to increase the number of power plants contributing to redispatch that can be observed in *Redispatch 2.0* and *Redispatch 3.0*. In our calculations, demand-side flexibilities (such as electrolyzers) contribute to redispatch without additional compensation. In contrast, demand-side flexibilities cannot contribute to redispatch in the current system. The upcoming *Redispatch 3.0* rule may include this option [E-Bridge Consulting GmbH 2022]. The manner in which they will receive compensation has not yet been determined. From a macroeconomic point of view, the compensation should be lower than the one for power plants. However, the assumption that they redispatch for free is rather ideal and thus questionable and affects the results. In addition, the amounts of redispatch reported in this paper exclude their free redispatch. This must be taken into account when comparing the results of this study with those of other studies.

Simplifications were necessary to enable the co-optimization of redispatch and investments in a spatial and temporal high resolution. The electricity market is represented by a general market and a short-term market including UC. This does not exactly represent the current energy markets in Germany. The selected planning horizon for the short-term market optimization of one week with an overlap of five days does not accurately reflect the existing day-ahead or intra-day market. A clear definition of the planning horizons for a short-term market model could not be derived from other studies or official sources from the grid planning by the TSOs[6]. The literature reveals a variety of approaches and trends, which were employed in defining the planning horizon in this study. The impact of this parametrization on the results is not quantified.

The market model does not include the behavior of market participants, as is typical in agent-based models like [Fatras et al. 2022; Harder et al. 2023]. A random noise is used instead to select which generators with the same marginal cost are dispatching in the market optimization (see Section 2.1). According to [Torralba-Díaz et al. 2020] this results in an efficiency gap, given that market behavior does not always reflect optimal solutions.

The parameterization of redispatch costs in the grid optimization also includes assumptions regarding future developments. A sensitivity analysis indicates that variations in these costs within a reasonable range do not greatly affect the overall results. However, notable shifts are possible at specific locations or times. The sensitivity analysis included assumptions regarding grid expansion costs. Such costs were observed to be more sensitive than specific redispatch costs. Nevertheless, comparisons of the impact of variations in specific grid expansion and redispatch costs are complex due to different cost units and scopes. The total cost of grid expansion is in the order of magnitude of billions, whereas the cost of redispatch is in the order of magnitude of millions. This implies that the specific costs associated with grid investment are likely to exert a more pronounced influence on the overall results. In addition, the cost variations were not aligned. Grid expansion costs were increased up to 200 % whereas the increase in redispatch cost is limited to additional management fees. This makes the different sensitivities less comparable. The increase in grid investment costs was influenced by rising costs published by the German TSOs [German TSOs 2023]. In contrast, costs for redispatch above the market price (apart from a management fee) seem to be not realistic, since a decrease of costs for redispatch can be observed in new regulations.

[6]A bilateral exchange with a TSO member influenced the assumptions.

UC constraints were excluded in the grid optimization due to the high computational efforts. The (n-1) security of the transmission grid was modeled by rather sweeping limitations of the available transmission capacity. Hence, a (n-1) contingency analysis by e.g. using a security-constrained LOPF was not part of the analysis as proposed e.g. in [Leite da Silva et al. 2010; Ronellenfitsch et al. 2017]. The hydrogen grid was not modeled, assuming that the hydrogen backbone grid will be implemented by the year 2035 and will not cause any bottlenecks in the energy system. A number of additional assumptions regarding parameters were required to enable the modeling of the German transmission grid, the influence of which was not determined in this study. The main results are therefore derived from comparing different model runs that all include these parameters and assumptions.

Despite these limitations, the novel method is able to co-optimize grid and storage expansion and explicit redispatch in a large sector-coupled energy system. The results are in general in line with the findings of other studies. Further developments and tests of the presented method are planned. The development process followed open-source and open-data principles. All data models and implemented methods are publicly available. Consequently, the results are transparent, reproducible, and external users can further develop and apply the developed software tools.

5 CONCLUSION AND OUTLOOK

This paper presents a novel method to co-optimize grid and storage expansion as well as redispatch using a consecutive market and grid optimization. The method is designed for use in large-scale grid systems with high spatial and temporal resolution, and it has been applied to the German eHV/HV grid system.

The impact of redispatch as a flexibility in expansion planning is analyzed by several model runs with different degrees of redispatch. The results indicate that grid and storage expansion costs in Germany decrease up to 78 % when redispatch can be used as a flexibility option in long-term planning. When the scope for redispatch is limited, the necessity for expansion of both storage and the grid is increased. Even relatively small amounts of redispatch result in significant cost savings. About 77 % of the potential savings in grid and storage investment costs in Germany can be achieved with 21 % of the optimal redispatch. However, investments can not level out all savings that can be achieved by redispatch, system costs are always higher when redispatch is not an option. The results demonstrate that redispatch significantly reduces the necessity for expansion in all scenarios, although the extent of this reduction varies.

The principal findings are consistent with the current state of the art. The design of the selected model and scenarios affects the results and causes differences to other studies. As an example, the selected scenario excludes flexibility options such as flexible charging and demand side management, which have the potential to reduce the necessity for redispatch. This impact will be analyzed and quantified in future research.

In future studies, the method will be further validated and developed, for example, to encompass multiple market zones within Germany. Furthermore, it will be interesting to analyze how the value of redispatch as a flexibility in long-term planning evolves with increasing shares of renewable energy towards a 100 % renewable energy system.

ACKNOWLEDGMENTS

The authors thank the Federal Ministry for Economic Affairs and Climate Action for funding the research project *PoWerD* (grant number: 03EI1042C).

REFERENCES

Annegret-Cl. Agricola, Bernd Höflich, Philipp Richard, Jakob Völker, Christian Rehtanz, Marco Greve, Björn Gwisdorf, Jan Kays, Theresa Noll, Johannes Schwippe, André Seack, Jan Teuwsen, Gert Brunekreeft, Roland Meyer, and Vanessa Liebert. 2012. Ausbau- und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030 (kurz: dena-Verteilnetzstudie): Endbericht. https://www.dena.de/fileadmin/dena/Dokumente/Pdf/9100_dena-Verteilnetzstudie_Abschlussbericht.pdf

T. Brown, J. Hörsch, and D. Schlachtberger. 2018a. PyPSA: Python for Power System Analysis. *Journal of Open Research Software* 6, 4 (2018). Issue 1. https://doi.org/10.5334/jors.188 arXiv:1707.09913

T. Brown, D. Schlachtberger, A. Kies, S. Schramm, and M. Greiner. 2018b. Synergies of sector coupling and transmission reinforcement in a cost-optimised, highly renewable European energy system. *Energy* 160 (Oct. 2018), 720–739. https://doi.org/10.1016/j.energy.2018.06.222 arXiv:1801.05290 [physics].

Espen Flo Bødal, Audun Botterud, and Magnus Korpås. 2022. Capacity Expansion Planning with Stochastic Rolling Horizon Dispatch. *Electric Power Systems Research* 205 (2022), 107729. https://doi.org/10.1016/j.epsr.2021.107729

Clara Büttner, Katharina Esterl, Ilka Cußmann, Carlos Andrés Epia Realpe, Jonathan Amme, and Amélia Nadal. 2024. Influence of flexibility options on the German transmission grid — A sector-coupled mid-term scenario. *Renewable and Sustainable Energy Transition* 5 (Aug. 2024), 100082. https://doi.org/10.1016/j.rset.2024.100082

Danish Energy Agency. 2020. Technology Data - Energy Plants for Electricity and District heating generation. https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_el_and_dh.pdf

E-Bridge Consulting GmbH. 2022. Redispatch 3.0: Regulatorischer Rahmen, Markt- und Produktdesign.

EEG. 2014 (last changes 2019). *Gesetz für den Ausbau erneuerbarer Energien (Erneuerbare-Energien-Gesetz - EEG 2017).* Technical Report. Bundesgesetzblatt 1066. https://www.bgbl.de/xaver/bgbl/start.xav#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl114s1066.pdf%27%5D__1578835066303

Jonas Eickmann, Tim Drees, Jens D. Sprey, and Albert Moser. 2014. Optimizing Storages for Transmission System Operation. *Energy Procedia* 46 (Jan. 2014), 13–21. https://doi.org/10.1016/j.egypro.2014.01.153

ENTSO-E. 2021. ENTSO-E Bidding Zone Configuration - Technical Report 2021. https://eepublicdownloads.azureedge.net/clean-documents/mc-documents/211209_ENTSO-E%20Bidding%20Zone%20Configuration%20Technical%20Report%202021.pdf

Nicolas Fatras, Zheng Ma, and Bo Nørregaard Jørgensen. 2022. An agent-based modelling framework for the simulation of large-scale consumer participation in electricity market ecosystems. *Energy Informatics* 5, 4 (Dec. 2022), 47. https://doi.org/10.1186/s42162-022-00229-0

Marco Franken, Alexander B. Schrief, Hans Barrios, and Armin Schnettler. 2018. Network Reinforcement Applying Redispatch-based Indicators. In *2018 53rd International Universities Power Engineering Conference (UPEC).* 1–6. https://doi.org/10.1109/UPEC.2018.8542104

German TSOs. 2021. Grid Development Plan Electricity 2035, Version 2021, second draft.

German TSOs. 2023. Grid Development Plan Electricity 2035, Version 2023, second draft.

Veronika Grimm, Bastian Rückel, Christian Sölch, and Gregor Zöttl. 2021. The impact of market design on transmission and generation investment in electricity markets. *Energy Economics* 93 (Jan. 2021), 104934. https://doi.org/10.1016/j.eneco.2020.104934

Leonard Göke, Mario Kendziorski, Claudia Kemfert, and Christian von Hirschhausen. 2022. Accounting for spatiality of renewables and storage in transmission planning. *Energy Economics* 113 (Sept. 2022), 106190. https://doi.org/10.1016/j.eneco.2022.106190

Nick Harder, Ramiz Qussous, and Anke Weidlich. 2023. Fit for purpose: Modeling wholesale electricity markets realistically with multi-agent deep reinforcement learning. *Energy and AI* 14 (Oct. 2023), 100295. https://doi.org/10.1016/j.egyai.2023.100295

Jonas Heering and Thane Gustafson. 2021. Germany's Energiewende at a Crossroads. (June 2021), 47–69. https://doi.org/10.3167/gps.2021.390203 Num Pages: 47-69 Publisher: Berghahn Books, Inc. Section: Articles.

Lion Hirth and Ingmar Schlecht. 2018. Market-Based Redispatch in Zonal Electricity Markets. https://doi.org/10.2139/ssrn.3286798

André Hoffrichter, Hans Barrios, Janek Massmann, Bhavasagar Venkataramanachar, and Armin Schnettler. 2018. Impact of Considering 110 kV Grid Structures on the Congestion Management in the German Transmission Grid. *Journal of Physics: Conference Series* 977, 1 (Feb. 2018), 012004. https://doi.org/10.1088/1742-6596/977/1/012004 Publisher: IOP Publishing.

Bowen Hua, Ross Baldick, and Jianhui Wang. 2018. Representing Operational Flexibility in Generation Expansion Planning Through Convex Relaxation of Unit Commitment. *IEEE Transactions on Power Systems* 33, 2 (2018), 2272–2281. https://doi.org/10.1109/TPWRS.2017.2735026

Claudia Kemfert, Friedrich Kunz, and Juan Rosellón. 2016. A welfare analysis of electricity transmission planning in Germany. *Energy Policy* 94 (July 2016), 446–452. https://doi.org/10.1016/j.enpol.2016.04.011

Christine Krüger, Mathis Buddeke, Frank Merten, and Arjuna Nebel. 2015. Modelling the interdependencies of storage, DSM and grid-extension for Europe. In *2015 12th International Conference on the European Energy Market (EEM)*. 1–5. https://doi.org/10.1109/EEM.2015.7216669

Armando Leite da Silva, Leandro S. Rezende, Luiz A. F. Manso, and George Anders. 2010. Transmission expansion planning: A discussion on reliability and N-1 security criteria. *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2010*, 244 – 251. https://doi.org/10.1109/PMAPS.2010.5528652

C. Müller, A. Hoffrichter, L. Wyrwoll, C. Schmitt, M. Trageser, T. Kulms, D. Beulertz, M. Metzger, M. Duckheim, M. Huber, M. Küppers, D. Most, S. Paulus, H. J. Heger, and A. Schnettler. 2019a. Modeling framework for planning and operation of multi-modal energy systems in the case of Germany. *Applied Energy* 250 (Sept. 2019), 1132–1146. https://doi.org/10.1016/j.apenergy.2019.05.094

Ulf Philipp Müller. 2021. *Combined optimization of grid and storage expansion in the german power system.* Ph. D. Dissertation. Europa-Universität Flensburg. https://d-nb.info/125765599X/34

Ulf Philipp Müller, Birgit Schachler, Malte Scharf, Wolf-Dieter Bunke, Stephan Günther, Julian Bartels, and Guido Pleßmann. 2019b. Integrated Techno-Economic Power System Planning of Transmission and Distribution Grids. *Energies* 12, 11 (2019). https://doi.org/10.3390/en12112091

Muhammad Bachtiar Nappu and Ardiaty Arief. 2015. Economic redispatch considering transmission congestion for optimal energy price in a deregulated power system. In *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*. 573–578. https://doi.org/10.1109/ICEEI.2015.7352565 ISSN: 2155-6830.

Thomas Nisonger. 2008. The "80/20 Rule" and Core Journals. *Serials Librarian - SERIALS LIBR* 55 (July 2008), 62–84. https://doi.org/10.1080/03615260801970774

Ariette Nüßler. 2012. *Congestion and Redispatch in Germany. A model-based analysis of the development of redispatch.* text.thesis.doctoral. Universität zu Köln. https://kups.ub.uni-koeln.de/4652/1/DisNuessler.pdf

Henrik Ronellenfitsch, Debsankha Manik, Jonas Hörsch, Tom Brown, and Dirk Witthaut. 2017. Dual Theory of Transmission Line Outages. *IEEE Transactions on Power Systems* 32, 5 (2017), 4060–4068. https://doi.org/10.1109/TPWRS.2017.2658022

Charlotte Senkpiel, Wolfgang Biener, Shivenes Shammugam, and Sven Längle. 2018. Evaluation of load flow and grid expansion in a unit-commitment and expansion optimization model SciGRID International Conference on Power Grid Modelling. *Journal of Physics: Conference Series* 977, 1 (Feb. 2018), 012008. https://doi.org/10.1088/1742-6596/977/1/012008 Publisher: IOP Publishing.

Philipp Staudt and Shmuel S. Oren. 2021. Merchant transmission in single-price electricity markets with cost-based redispatch. *Energy Economics* 104 (Dec. 2021), 105610. https://doi.org/10.1016/j.eneco.2021.105610

Philipp Staudt, Marc Schmidt, Johannes Gärttner, and Christof Weinhardt. 2018. Using vehicle-to-grid concepts to balance redispatch needs: A case study in Germany. In *Proceedings of the Ninth International Conference on Future Energy Systems (e-Energy '18)*. Association for Computing Machinery, New York, NY, USA, 80–84. https://doi.org/10.1145/3208903.3208926

Laura Torralba-Díaz, Christoph Schimeczek, Matthias Reeg, Georgios Savvidis, Marc Deissenroth-Uhrig, Felix Guthoff, Benjamin Fleischer, and Kai Hufendiek. 2020. Identification of the Efficiency Gap by Coupling a Fundamental Electricity Market Model and an Agent-Based Simulation Model. *Energies* 13, 15 (Jan. 2020), 3920. https://doi.org/10.3390/en13153920 Number: 15 Publisher: Multidisciplinary Digital Publishing Institute.

Katrin Trepper, Michael Bucksteeg, and Christoph Weber. 2013. An Integrated Approach to Model Redispatch and to Assess Potential Benefits from Market Splitting in Germany. https://doi.org/10.2139/ssrn.2359328

Marta Victoria, Kun Zhu, Elisabeth Zeyen, and Tom Brown. 2021. PyPSA Technology data, v0.3.0. https://github.com/PyPSA/technology-data

Lukas Wienholt, Ulf Philipp Müller, and Julian Bartels. 2018. Optimal Sizing and Spatial Allocation of Storage Units in a High-Resolution Power System Model. *Energies* 11, 12 (2018). https://doi.org/10.3390/en11123365

Étienne Cuisinier, Pierre Lemaire, Bernard Penz, Alain Ruby, and Cyril Bourasseau. 2022. New rolling horizon optimization approaches to balance short-term and long-term decisions: An application to energy planning. *Energy* 245 (2022), 122773. https://doi.org/10.1016/j.energy.2021.122773

A ACRONYMS

BEV battery electric vehicle
CHP combined heat and power plant
DSM demand-side management
eHV extra high voltage
eHV/HV extra high voltage and high voltage
eTraGo *Electricity Transmission Grid Optimization*
GDP German grid development plan
HV high voltage
LOPF Linear Optimal Power Flow
OCGT open cycle gas turbine
PyPSA *Python for Power System Analysis*
SMR steam methane reforming
TSO Transmission System Operator
UC Unit Commitment

Towards a State Explanation Framework in Cyber-Physical Systems

KATRIN SCHREIBERHUBER, Wirtschaftsuniversität Wien, Austria
FAJAR J. EKAPUTRA, Wirtschaftsuniversität Wien, Austria
MARTA SABOU, Wirtschaftsuniversität Wien, Austria
DANIEL HAUER, Siemens AG Österreich, Austria
KONRAD DIWOLD, Siemens AG Österreich, Austria
THOMAS FRÜHWIRTH, TU Wien, Austria
GERNOT STEINDL, TU Wien, Austria
TOBIAS SCHWARZINGER, TU Wien, Austria

The concept of Explainable Cyber-Physical Systems (ExpCPSs) aims to provide clear explanations for system decisions and actions, which is crucial for understanding, managing, and controlling high-risk Cyber-Physical Systems (CPSs), such as intelligent energy distribution networks or Smart Grids (SGs). Existing approaches for ExpCPSs, however, are typically designed only for specific application areas, limiting their impact. This paper proposes a domain-independent framework for automatically generating system state explanations in a CPS through (i) a unified data model and (ii) novel explanation algorithms to improve systems transparency, interpretability, and trustworthiness. The proposed framework facilitates root cause analysis on CPSs independently of their domains, enabling users to query system states and receive comprehensive explanations. We evaluated our solution in a real-world Electric Vehicle (EV) charging garage use case, highlighting its feasibility and effectiveness.

CCS Concepts: • **Computer systems organization → Embedded and cyber-physical systems**; • **Information systems → Resource Description Framework (RDF)**; • **Hardware → Smart grid**; **Sensor devices and platforms**; • **Computing methodologies** → *Causal reasoning and diagnostics.*

Additional Key Words and Phrases: System explainability, Domain knowledge representation

1 INTRODUCTION

The increasing integration of digital technologies into industrial systems has given rise to CPSs [Müller 2017], where the physical and digital worlds converge. A prominent example for such adoption of CPSs is on the energy distribution systems in form of Smart Grids (SGs).

In this domain, the Energiewende [Jacobs 2012] (i.e., the European goal to reach carbon neutrality) has lead to an increased integration of renewable energy resources and complex grid installations into the European distribution systems, including PV-battery systems, smart buildings, and community batteries.

They are becoming more complex, with power generation now occurring at voltage levels traditionally reserved for distribution. Additionally, new intelligent grid entities (e.g., smart building with PV-battery system) are to a certain extent decoupled from demand and supply [Qadrdan et al. 2018] thus making their forecasting difficult [Bittner et al. 2023].

Traditional distribution system operation strategies (which were essentially a passive operation of the low voltage level [Ghadi et al. 2019]) are no longer suitable to handle this complexity, prompting the development of new monitoring and control schemes known as SG technologies [Moreno Escobar et al. 2021; Tuballa and Abundo 2016]. These technologies leverage a combination of information and communication technologies (ICT) and monitoring and control strategies tailored specifically for distribution system operation.

An SG is designed to detect and respond to local changes in energy usage by controlling physical grid entities using digital communications technologies. It enables the management of localized energy production and consumption within distribution systems, contributing to a stable grid operation. While SGs are a big asset for achieving carbon neutrality, anomalies may occur during the operation of an SG as a result of unusual phenomena (e.g., extreme weather) or limited capacities of facilities (e.g., transformer overloads). The inherent complexity of an SG makes it increasingly difficult for engineers and operators to understand the system behaviour and identify root causes of anomalies.

Being able to provide explanations for these anomalies –or system states in general– therefore plays a crucial role in understanding, managing, and controlling SGs and other CPSs. To achieve this, the ExpCPS concept was proposed [Blumreiter et al. 2019], which aims for deriving and providing clear explanations for the decisions and actions taken by the system. In the context of SG operation ExpCPSs help power systems engineers, operators, and consumers to understand the reasons behind unusual system behavior and implement effective countermeasures, while ensuring understandability and trust even in increasingly complex systems [Jha 2022].

Earlier ExpCPS approaches have been proposed in several application domains [Aryan et al. 2021; Chhokra et al. 2020; Ploennigs et al. 2014]. However, these solutions are limited to their specific application areas. Furthermore, as CPSs cover diverse domains, the integration of heterogeneous data sources as well as the generation of meaningful and understandable explanations are crucial challenges to be addressed in such frameworks [Blumreiter et al. 2019].

In this paper, we propose a framework to automatically generate such an explanation in any CPS. The framework facilitates an interface for users of a system, where they can ask explanations about states in the system. As a response, the framework performs a root cause analysis to return a tree of potential root causes of the state. It aims to facilitate the management of ExpCPSs, providing an adaptable framework for CPSs in various domains, such as SGs or smart buildings. The framework provides increased transparency, interpretability and trustworthiness for any user ranging from system expert to lay user. Furthermore, while our evaluation is focusing on a local smart grid example, the framework can be applied to any

CPS, e.g., smart factories, smart buildings and complex smart grid applications (e.g., local energy communities).

We aim to answer the following research questions:

RQ1 Which minimal components are needed to provide all capabilities needed for an ExpCPS?

RQ2 How can relevant knowledge be represented effectively to enable (semi-)automated explanation generation?

RQ3 What methods can automate the generation of system explanations at runtime?

To answer these questions, we first define the problem of explainability in CPS in Sec. 2, introducing a motivating example for reference in future sections. We present our generic ExpCPS framework in Sec. 3 that enables the automatic generation of system explanations. As a basis of our explanation module, we describe a unified data model in Sec. 4 to facilitate a seamless integration of heterogeneous data sources relevant for explanation derivation in the system. The core explanation module is described in Sec. 5, which processes data to create causal paths between system states. These paths can then be queried through user interactions and are provided as explanations.

We evaluate the proposed solution in a feasibility study based on a real-life use case of an EV-charging garage reported in Sec. 6. Afterwards, we discuss related work on ExpCPS research and explainability in general within Sec. 7. Finally, we conclude our work, highlight its contributions, limitations as well as planned future work in Sec. 8.

2 PROBLEM DEFINITION

An ExpCPS is defined as "the capability of both the system and its engineering tools to explain certain aspects of interest about the system, both in a human-comprehensible and machine-processable format" [Greenyer et al. 2019].

A CPS may present numerous *aspects of interest*, as mentioned in the definition above. This paper focuses on providing explanations for *system states*. A state is a phenomenon occurring in the system, which happens over a period of time. Current event detection methods focus on the detection of changes in the system at a single time point. Therefore, we propose the following definition of events and states in this paper:

- An *event* is a phenomenon, which happens at a certain point in time and place and was detected by an event detection procedure.
- In contrast, a *state* is a phenomenon, which happens over a period of time. A state transition is triggered by an event.

An *explanation* in the context of ExpCPS is an output that is aimed at making a system state more understandable, appropriable and exploitable by its intended users and decision makers. In short, we regard explanations as answers to why-questions about system states, which is adapted from the eXplainable AI (XAI) domain [Cabitza et al. 2023]. An explanation is structured in three components:

- *explanandum*: the state to be explained
- *explanans*: information that explains
- *explanatory relationship*: links the explanandum and explanans

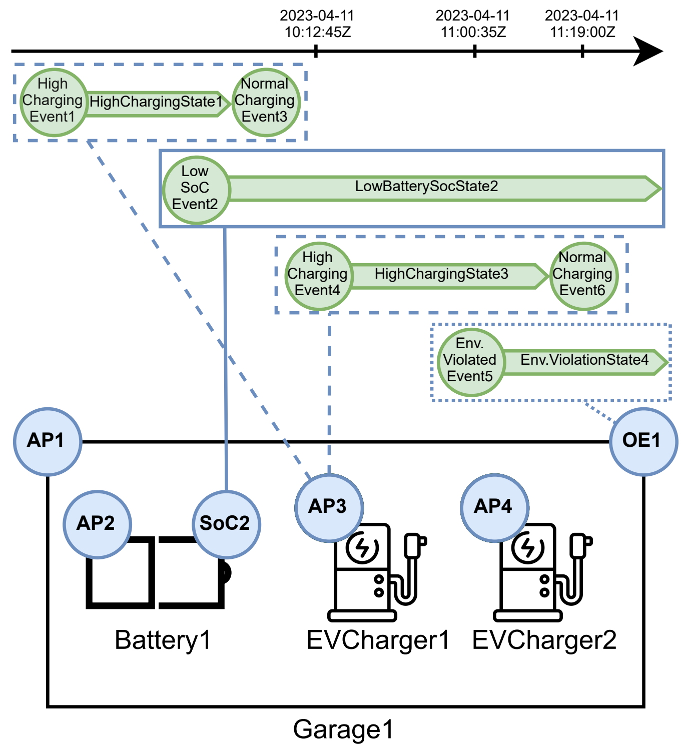


Fig. 1. EV garage as motivating example, showing Active Power (AP), State of Charge (SoC), Operating Envelope (OE) sensors in blue, events and states connected to the sensors in green.

Each explanandum can have multiple explanantia [Cabitza et al. 2023]. In the context of our research, this means that for a state requiring an explanation (i.e. the explanandum), there can multiple root causes to explain "why" this state occurred (i.e. the explanans/-explanantia). The explanatory relationship between the explanandum and explanans is the reasoning behind the explanation (i.e. the general causal knowledge the explanation is based on).

Considering the definitions above, in this paper, we aim to address the pressing issue of the lack of methodologies and techniques to provide clear, meaningful explanations for states of interest in a CPS independent of its domains.

Furthermore, we describe a motivating example of an ExpCPS from the area of SGs to clarify our problem context. In this example, we introduce a public EV charging garage, which is directly connected to a local transformer. As the garage is a major energy consumer in the area, an operating envelope is imposed on the facility. This means, the facility operator needs to make sure that the electricity demand of the facility from the local grid stays within an agreed limit (within the operating envelope). Fig. 1 depicts the setup of the example use case. The garage has multiple chargepoints, where an EV can charge. It also has a battery installed, which can be used for peak shaving in times of high consumption. Multiple sensors are installed at various devices (indicated as gray circles), which take measurements of various observable properties, such as Active Power (AP), State of Charge (SoC), Operating Envelope (OE).

During the operation of the garage, no envelope violation should occur, as the garage is supposed to regulate the use of its components autonomously, avoiding any violation of the imposed operating envelope. However, there are situations, where violations still happen. If they do, the facility operators as well as the distribution system operator (DSO) will request an explanation of why the violation has happened. An example explanation, which can be derived from the use case in Fig. 1 would be "`EnvelopeViolationState4` occurred due to `HighChargingState3` at `EVCharger1`, which could not be balanced out by the Battery as it was empty (`LowBatterySocState2`). High Charging Activity for multiple hours preceding the Violation has led to an empty battery (`HighChargingState1`)".

3 EXPCPS FRAMEWORK

We hereby introduce a technology-agnostic explainability framework for the development of an ExpCPS. The framework serves as the general architecture to enable the operation of the explanation module, which is the focus of this paper. Starting from timeseries and system data provided by the CPS, the information is processed to be stored based on a unified structure defined by the integrated data model (see Section 4) to build explanations for system phenomena in the explanation generation module (see Section 5).

The common ExpCPS framework (see Fig. 2) identifies all necessary modules to create an explainable CPS. The modules are structured based on the interoperability layers defined by the Smart Grid Reference Architecture (SGAM) [Group 2012]. However, we aim for a common architecture, which is not limited to SG applications to ensure its compatibility with various reference models (e.g., Reference Architecture Model Industry 4.0 [Hankel and Rexroth 2015]).

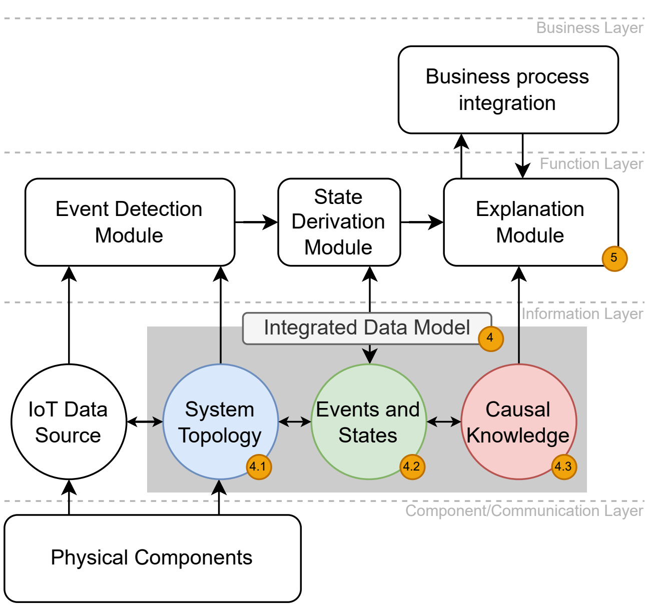


Fig. 2. ExpCPS framework - detailed explanations of the key aspects of this framework can be found in sections according to small orange circles

3.1 Component/Communication Layer

The component layer involves the physical components of the system (e.g., batteries, EV chargers, PV systems and transformers in a smart grid). These components are physically located at a certain geolocation, but they are also part of logical groups inside a system (e.g., components belonging to one household). All of these components are equipped with various sensors. The communication layer contains the communication between components as well as the integration of the physical components for further data processing. Sensor measurements are sent to and stored in a database according to the integrated data model described in Section 4.

This paper does not delve into the specifics of this integration. Our assumption is an existing data flow from physical entities to the information layer. There are various ways to ensure communication of sensor data to a database, such as OPC Unified Architecture (OPC UA) servers or MQTT brokers [Steindl et al. 2024].

3.2 Information Layer

The information layer contains multiple entities to store and process heterogeneous data of a CPS and its phenomena occurring at runtime. All these entities are connected to each other, forming an integrated database that contains all relevant data for an ExpCPS.

The *IoT Data Source* provides a reliable storage solution for sensor measurements. Sensor measurements are time-dependent values describing a feature in the system (e.g., AP consumed by an EV charger). These measurements are recommended to be stored in a time-series database (e.g., influxDB[1]) to facilitate easy access to current and historical values for further analysis. The IoT data is not part of the integrated data model. The location of the data as well as its relevance in the CPS are instead stored in the system topology as attributes of a sensor.

The *System Topology* contains static information about the CPS setup. It provides context to the sensors, sensor measurements and the events occurring in the system as well as platforms in the system and how they are connected (e.g., `APSensor3` is located at `Charger1`, a platform hosted by `Garage1`).

The *Events and States* data module contains events and states that are detected by the *Event Detection Module* and the *State Derivation Module*. It also contains a mapping between event and state types. In the mapping, each state type has at least one event type which can start the state and one which can end the state (e.g., `EnvelopeViolationState` is started by `EnvelopeExceededEvent` and ended by `EnvelopeNormalEvent`). The event-to-state type mapping serves as an input to the *state derivation module*, which is responsible for state creation.

Causal Knowledge is knowledge about causal relations within the system. These abstract causal relations are defined a priori by domain experts directly or through workshops between knowledge engineers and domain experts. Causal knowledge in this context is defined as general knowledge about connections between types of states, not between specific instances of states (e.g., `HighChargingState` causes `EnvelopeViolationState`).

A detailed description of each part of the integrated data model is provided in Section 4.

[1]https://www.influxdata.com/

3.3 Function Layer

The function layer contains components that add advanced functionalities to the system, enriching the data sources in the information layer with additional information and deriving new knowledge.

The *Event Detection Module* is responsible for analysing the IoT data source at runtime and for extracting events that are of interest. Event detection can be performed by simple rules imposed on the data (e.g., SoC is lower than a certain threshold), or through more complex methods, such as pattern recognition models applied on the data. The event detection module works not only on raw timeseries data as its input, but also incorporates the topology data of the system and the connection between components for an informed and integrated event detection.

The *State Derivation Module* examines the events detected by the event detection module and derives the corresponding states according to the event-to-state type mapping that is defined a priori and is stored in the events and states data store. The state derivation module is responsible for mapping detected events to a specific state at a sensor in the system.

The *Explanation Module* is responsible for finding causal relations between states occurring in the system and enabling the creation of causal paths (i.e. explanans). These causal paths can be used to find the root cause of a state that is relevant for the user of a system (i.e. explanandum) that is interested in an explanation. The explanation engine, as part of the module, integrates states, system topology data as well as causal and domain knowledge (i.e. explanatory relationship) to create these paths. The *explanation module* is the focus of this paper and will be explained in more detail in Section 5.

3.4 Business Layer

The business layer holds the connection of the framework to business processes.

The *business process integration* module represents the integration of explanations generated by the framework into everyday business processes. Explanations of events or faults in the system can help find root causes of unexpected behaviour more efficiently, which enables even non-expert users to understand why some behaviours of the system have occurred. The integration of these explanations into business processes can be implemented through various modes, such as chatbots, dashboards or multi-modal explanations. This implementation is very use-case specific, depending on the system as well as intended users of the ExpCPS system. Therefore, we will not go into details about business process integration in this paper.

4 INTEGRATED DATA MODEL

We propose an integrated data model as the foundation of an ExpCPS framework that is applicable across various CPS domains. The data model plays a key role in our approach as a means to facilitate the integration of diverse data sources for comprehensive system understanding and to enable root cause analysis. While different technologies can be used to implement the data model, we use semantic web technologies to exemplify the implementation of the model and to visualise its structure. A semantic model serves as a structured representation of knowledge or information, designed to improve understanding by both humans and machines. It relies on an ontology, which acts as a schema for representing a specific domain [Ehrlinger and Wöß 2016]. Instantiating an ontology with data points results in the creation of a Knowledge Graph (KG), where concepts and entities are uniquely identified using Unified Resource Identifiers (URIs) to ensure reusability. The semantic model builds upon the Resource Description Framework (RDF) [Tomaszuk and Hyland-Wood 2020] for creating a graph-based data structure. In the prototype, data points are published to the data model in RDF serialized in the Turtle format.

The data model consists of three parts, which focus on different aspects of the ExpCPS framework (i) System Topology (ii) Events and States (iii) Causal Knowledge. In Fig. 3, the data model is shown as an ontology. Each concept or entity is identified by a URI. Prefixes are defined to abbreviate long URI strings (in Fig. 3, relevant prefixes are defined in the top left corner). The ontology implementation of the integrated data model is documented and published online[2].

4.1 System Topology

The *System Topology* contains information about the system setup, all the devices in the system as well as the sensors, which provide a continuous data stream of system measurements at runtime. The topology information specifies the positions of the devices and their physical and logical connections. To represent the topology of a system, we use three core concepts: platform, sensor and observable property. The terms are derived from the SOSA Ontology[3], a "lightweight general-purpose ontology to represent the interaction between entities" in CPSs [Janowicz et al. 2019].

A **Platform** is an entity that hosts other entities, particularly sensors, and other platforms (e.g., EV charger, battery, garage). Platforms are devices or facilities within a system, where sensors are located.

A **Sensor** is a device or agent, which collects data measurements of an observable property (e.g., AP). Each sensor has a sensor type, which defines the type of measurements it can take and what types of states are possible at the sensor. Sensors are hosted by a platform, which means a sensor is located at/within a Platform and its measurements correspond to this platform (e.g., `EVCharger1` hosts `APSensor3` in Fig. 1). Each sensor stores information on how to access its sensor measurements (e.g., access token to a timeseries database).

An **Observable Property** is an observable quality (property, characteristic) in the system. Each sensor observes an observable property. (e.g., `APSensor3` observes `ActivePower`).

By using these three concepts (Platform, Sensor, Observable Property), the system setup can be represented in sufficient detail to facilitate event explainability. It enables the description of which sensors are located at which platform, what a sensor measures, where to find measurement data and how platforms are connected to each other.

[2] http://w3id.org/explainability/sense#
[3] https://www.w3.org/TR/vocab-ssn/

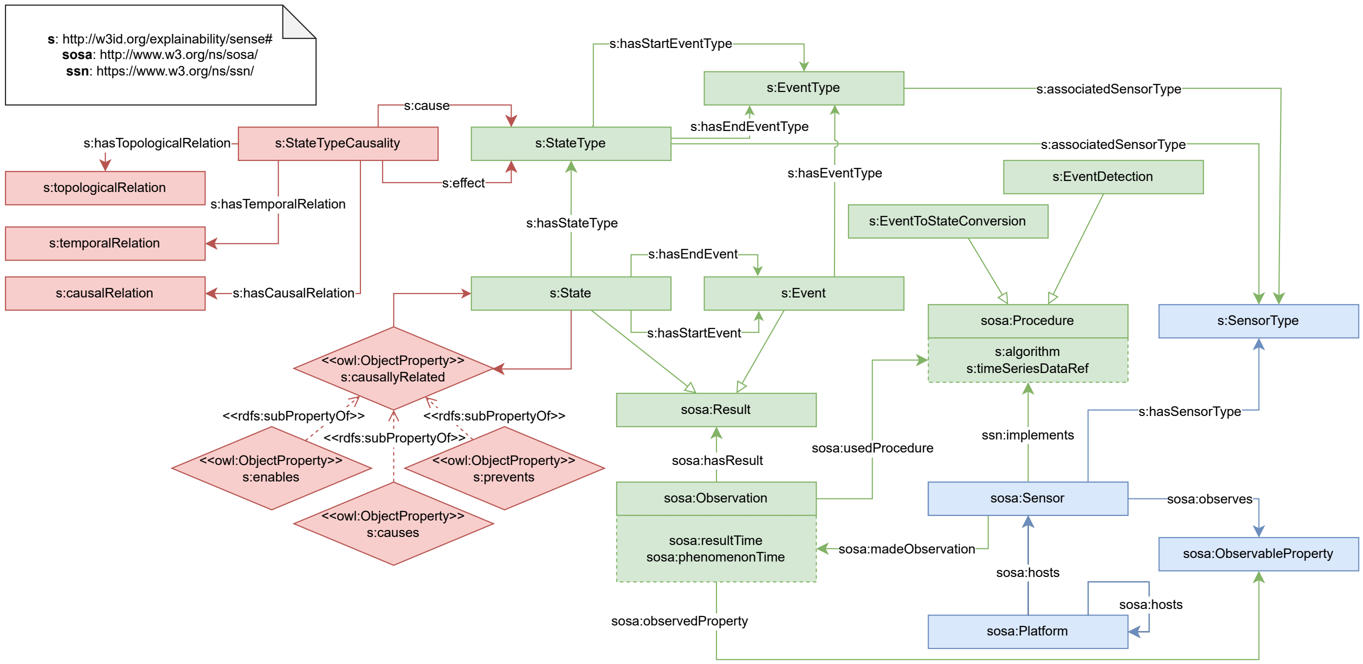


Fig. 3. Integrated Data Model represented as an ontology ◯ System Topology, ◯ Events and States, ◯ Causal Knowledge

4.2 Events and States

The *Events and States* data contains any information connected to the event detection and state derivation process. It contains event detection methods for each event type that are implemented to detect events. Additionally, it stores a priori information about the event-to-state type mapping between event types and state types. At runtime, detected events as well as states that are derived from these events are collected and stored in this module. As an example, `HighChargingEvent4` is detected at `"2023-04-11 10:12:45Z"` by `APSensor3` hosted by `Charger1` in the motivating example from Fig. 1.

Event: Each event (e.g., `HighChargingEvent4`) is stored in the data base, including added semantics about the event:

- *eventType* of the event that was detected (e.g., `HighChargingEvent`)
- *procedure* that is responsible for the detection of the event (e.g., `HighChargingDetectionQuery`). A procedure can be a query, a model, or any function that is responsible for the event detection.
- *sensor* where the observation was made (e.g., `APSensor3`)
- *timestamp* of the event (e.g., `"2023-04-11 10:12:45Z"`)
- *observableProperty* that was observed to detect the event (e.g., `ActivePower`)

State: The *State Derivation Module* is responsible for deriving states from detected events using the event-to-state mapping. A new state (e.g., `HighChargingState3` in Fig. 1) is stored in the database including the following semantics:

- *stateType* of the state (e.g., `HighChargingState`)
- *startEvent* that has started the state (e.g., `HighChargingEvent4`)
- *endEvent* that has ended the state (e.g., `NormalChargingEvent6`); if the state has not been ended yet, it is also possible that there is no endEvent related to the state (e.g., see `EnvelopeViolationState4` has no endeEvent (yet))
- *observableProperty* that was responsible to detect the event (e.g., `ActivePower`)

Users can query for potential root causes of specific system states. More information on the identification of a root cause for a trigger state is provided in Sections 5.2 and 5.3.

Event-to-State mapping: Each event type represents a state transition between state types, which is defined a priori (e.g., `HighChargingEvent` starts `HighChargingState`, and `NormalChargingEvent` ends `HighChargingState`). The state derivation module uses this information combined with topology data to identify which state has been changed at which sensor by a certain event (e.g., `HighChargingEvent4` at `APSensor3` starts `HighChargingState3` at `APSensor3` as a result of the event-to-state mapping indicating `HighChargingEvent` starts `HighChargingState`).

4.3 Causal Knowledge

The *causal knowledge* data is a description of causal relations between state types. Each causality relation consists of a cause state type and an effect state type as well as three parameters (causal, temporal, topological) that give detailed information on the nature of the causal relation. We rely on the definition in the Causal and Temporal Relation Scheme (CaTeRS) [Mostafazadeh et al. 2016] for causal and temporal relations. An example of such a relation is:

`HighChargingState` (**cause state type**)
`causes` (**causal**) an
`overlapping` (**temporal**)
`EnvelopeViolationState` (**effect state type**) at
`parentPlatform` (**topological**)

Each part of the causal relation is described in more details below.

cause state type and effect state type: A state type is the concept of a state. It has a description of what characteristics this type of state exhibits (e.g., `HighChargingState` is a state when an `EVCharger` charges more than 50kW). A state type is not related to a specific time or place. It is a concept of a state that can occur at runtime.

causal: The causal relation defines a more detailed view of the type of causality as compared to a simple "caused by" relation. There are three types of causal relations that can occur between two states ($stateA \rightarrow stateB$):

- *cause:* If stateA occurs, stateB most probably occurs as a result.
- *enable:* If stateA does not occur, stateB most probably does not occur (not enabled to occur).
- *prevent:* If stateA occurs, stateB most probably does not occur as a result.

By providing a more detailed description of a causal relation, intricate relations between state types can be represented, such as "`LowBatterySoCState` prevents `BatteryDischargeState`, while `BatteryUnusedState` (i.e. the non-existence of `BatteryDischargeState`) enables `EnvelopeViolationState`".

temporal: The temporal relation captures the temporal aspects of a relation between two states. There are two options for temporal relations ($stateA \rightarrow stateB$) that showed to be relevant for representing causal knowledge:

- *before:* stateA starts and ends before stateB.
- *overlaps:* stateA starts before stateB, but ends after stateB has started.

topological: The topological relation considers the logical/hierarchical relation between two states as a constraint for the causal relation to hold. We consider a hierarchical representation of platforms and their sensors. A state is always associated with one or more sensors, hosted by a platform. In Fig 4, the topology representation of the motivating example from Fig. 1 is depicted. Platforms are shown in light blue, while sensors are shown in dark blue. Three types of topological relations are defined as causal constraints. each of the relations is described below and exemplified in red in Fig. 4:

- *samePlatform:* stateA and stateB are associated with a sensor on the same platform (e.g.,`SoCSensor2` and `APSensor2` are associated to the same platform `Battery1`).
- *parentPlatform:* stateA is associated with a sensor on a platform that is hosted by the platform of stateB (e.g., `APSensor1` is associated with `Garage1`, which is the parent platform of `EVCharger1`. `APSensor3` is associated with `EVCharger1`, forming a `parentPlatform` relation from `APSensor3` to `APSensor1`). Note that this relation is the only non-symmetric topological relation.
- *siblingPlatform:* stateA is associated with a sensor on a platform that is hosted by the same platform as the platform associated with stateB (e.g.,`APSensor3` is associated with `EVCharger1` and `APSensor4` is associated with `EVCharger2`. Both platforms are hosted by `Garage1`, so the two sensors are hosted by `siblingPlatforms`).

The definition of these relations can be extended if needed for specific domains or use cases.

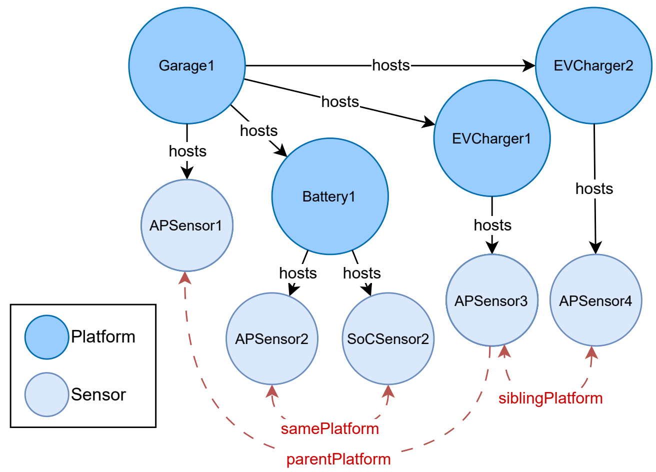


Fig. 4. Topology representation of the example from Fig 1

A thoroughly crafted causal knowledge dataset is crucial for the explanation engine to provide meaningful and correct explanations of states. This knowledge not only provides information about causality, but it also provides information on which types of states are relevant. Therefore, it serves as an indication to create corresponding event detection procedures in the event detection module of the ExpCPS framework.

5 EXPLANATION MODULE

The explanation module presented in this section provides a causal exploration method that can leverage domain knowledge about causal relations of a system from domain experts, while being domain-independent at its core. Fig. 5 shows the internal structure of the module, including its interfaces to other components of the ExpCPS framework. The module is dependent on the integrated data model and its components, as defined in Section 4. There are two main components of the explanation module: (i) *Actual Causality Exploration*, where causal relations between states are explored based on information from multiple data sources and (ii) *Causal Path Query*, where the causal path from an event to its root cause is queried from the system through a user request.

5.1 Actual Causality Exploration

The *actual causality exploration module* contains the extraction of causal relations between states (actual causality) from causal knowledge between state types (type causality). Actual causality is defined as the relation between two instances of states that happen at

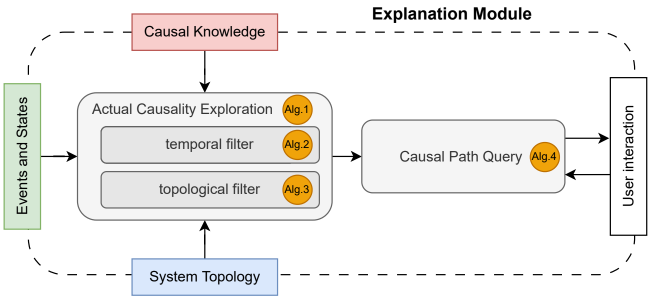


Fig. 5. Details of the Explanation Module

runtime (e.g., `HighChargingState1` causes `EnvelopeViolationState2`). On the other hand, type causality represents general relations between entities in the system (between State Types, e.g., `HighChargingState` causes `EnvelopeViolationState`). The distinction between type causality and actual causality comes from the definition of Halpern-Pearl [Halpern 2016] as a means to formalize causal reasoning. While type causality deals with more abstract or general patterns of causality that hold across different instances or types of states, actual causality is concerned with the causal relationships between specific instances of states.

In the causality exploration module, actual causality is derived from type causality through the actual causality exploration algorithm. Its main functionality is for each new state, to check the system for relations between state types that are defined in the causal knowledge data. For relations between two states that match the description in the causal knowledge, the causal relation between these two states is added to the database as an actual causal relation. The pseudo-code of the actual causality derivation is stated in Algorithms 1-3.

In *Algorithm 1*, the function `getActualCausalRelations` takes a new state (e.g., `EnvelopeViolationState75`) as its input and checks for new actual causal relations that can be detected between this new state and previous states in the system. It analyses the state type and the causal knowledge about this state type. If there is causal knowledge about this state type, all states of the cause state type in the system are filtered according to the causality constraints that are defined in the causal knowledge data (topological, temporal, cause state type). It returns a list of triples in the form `(causeState, causalRelation, newState)` (e.g., `(HighChargingState1, caused, EnvelopeViolationState2)`). Algorithms 2 and 3 are functions, which are called in the process.

In *Algorithm 2*, details of the filter on topological relations are defined, applying different filters depending on the stated relation (samePlatform, parentPlatform or siblingPlatform). If there is a causal relation defined as "samePlatform", only states that are associated with sensors at the same platform are considered potential causes. If it is defined as "parentPlatform", only states associated with sensors at a childPlatform of the effect sensor are considered potential causes (parentPlatform defined as effect at parentPlatform of cause). If it is defined as "siblingPlatform", only states associated with sensors at a siblingPlatform of the effect sensor are considered potential causes.

Algorithm 1 Get Actual Causal Relations

```
allStates ← all states recorded in the system
function GetActualCausalRelations(state)
    actualCauses ← [ ]
    for causalRelation in causalKnowledge do
        if state.stateType = causalRelation.effectStateType then
            causeStates ← TopologicalFilter(state, topological, allStates)
            causeStates ← TemporalFilter(state, temporal, causeStates)
            for cause in causeStates do
                actualCauses.append(cause, causalRelation.causal, state)
            end for
        end if
    end for
    return actualCauses
end function
```

Algorithm 2 Topological Filter

```
function TopologicalFilter(effect, topological, states)
    filteredStates ← [ ]
    effectPlatform ← effect.Sensor.hostedByPlatform
    effectParentPlatform ← effectPlatform.hostedByPlatform
    for cause in states do
        causePlatform ← cause.Sensor.hostedByPlatform
        causeParentPlatform ← causePlatform.hostedByPlatform
        if topological = "samePlatform" then
            if effectPlatform = causePlatform then
                filteredStates.append(cause)
            end if
        end if
        if topological = "parentPlatform" then
            if effectPlatform = causeParentPlatform then
                filteredStates.append(cause)
            end if
        end if
        if topological = "siblingPlatform" then
            if effectParentPlatform = causeParentPlatform then
                filteredStates.append(cause)
            end if
        end if
    end for
    return filteredStates
end function
```

In *Algorithm 3*, details of the temporal filter based on the temporal relation from the causal knowledge (before, overlaps) are shown. Depending on the temporal relation that is defined in the causal knowledge, only states are considered potential causes that fulfill the temporal constraints. If the causal relation is defined as "overlaps", only states that overlap with the effect state are considered. If the causal relation is "before", a potential cause only needs to be preceding the effect state.

Finally, any causes that remain after filtering on topological and temporal constraints are added to the database as relevant causal relations. Importantly, the type of causality needs to be saved as well (causes vs enables vs prevents) to ensure a comprehensive explanation path when querying the system. Therefore, *getActualCausalRelations(STATE)* returns a list of actual causalities as triples in the form `(causeState, causalRelation, STATE)`

5.2 Causal Path Query

The *causal path query module* queries the system data to find a causal path from a state to its potential root cause. A state for which an explanation is requested is called a trigger state. The causal path query module is triggered by entering specific states, so-called

Algorithm 3 Temporal Filter

```
function TemporalFilter(effect, temporal, states)
    filteredStates ← [ ]
    effectStart ← effect.startEvent.timestamp
    for cause in states do
        causeStart ← cause.startEvent.timestamp
        causeEnd ← cause.endEvent.timestamp
        if temporal = "before" then
            if causeStart ≤ effectStart then
                filteredStates.append(cause)
            end if
        end if
        if temporal = "overlaps" then
            if (causeStart ≤ effectStart) and (effectStart ≤ causeEnd) then
                filteredStates.append(cause)
            end if
        end if
    end for
    return filteredStates
end function
```

trigger states. A trigger state can either be set by the system through a predefined rule (e.g., each time an envelope violation happens, an explanation should be generated and sent to the facility operator), or by the user (e.g., a user realizes that the operating envelope is violated and requires an explanation).

Algorithm 4 shows the pseudo code to extract the causal path. The path identification is a recursive function to query all causal relations in a tree-like structure until there is no more causal relation to be added to the path.

Algorithm 4 Causal Path Identification

```
function FindCause(triggerState)
    allCauses ← [ ]
    Query (causeState, causalRelation, triggerState) as immediateCauses
    for each cause in immediateCauses do
        if cause not in allCauses then
            allCauses.append(cause)
        end if
        allCauses.append(FindCause(cause.causeState))
    end for
end function
```

5.3 User Interaction

The user interaction interface converts the user request for an explanation of a trigger state to a causal path identification query. The user interaction can have various implementations. Either an explanation request is only triggered directly by a user manually (e.g., via a user interface), or by a predefined notification rule, such as a weekly report or a notification that is sent every time a new state of a certain state type occurs (e.g., send a notification each time an envelope violation occurs). The implementation also depends on the business process integration of the framework, as chatbots or interactive dashboards require more interactive features and two-way communication than static reports. To allow for effective user interaction, user-centered explanations can address a user's needs and context individually. The SmartEx framework [Sadeghi et al. 2024] proposes one solution to create user-centered explanations in a smart building CPS use case.

6 USE CASE-BASED EVALUATION

This section introduces a prototype implementation of the explanation module within the ExpCPS framework, as discussed in Sections 3 - 5. In the scope of this study, the implementation serves two primary objectives. Firstly, it aims to clarify the underlying concepts by illustrating the architecture with specific technologies. Secondly, by applying the prototype to a practical use case, it aims to showcase the feasibility of the proposed approach.

The proof of concept implementation for EV charging was exclusively developed using open-source software tools and frameworks. It serves as a demonstration of one potential implementation and its capabilities. The source code of the prototype is accessible on GitHub[4] for reference.

6.1 Use Case Description

The proposed ExpCPS framework including the data model and the explainability module is implemented in a real-world use case of a smart charging garage for electric vehicles. The use case data consists of one month sensor measurements in a public EV charging garage located in an urban area. There was no restriction on charging power implemented, which means that charging peaks can only be mitigated by using the battery at the facility for peak shaving.

In Fig 6, a schematic overview of the garage and its components is shown. Similarly to the motivating example presented in the problem definition of the paper, the garage contains 8 EV charging stations, where one of them is a fast charger (potential charging power is 150kW). It has two batteries installed, which help to regulate peak consumption times, as well as a PV system, which can be used directly for EV charging, or to charge the battery. An operating envelope is imposed on the facility operator, which should not be violated at any time. Most of the system components have a sensor installed, which is indicated by a blue circle (AP = Active Power, SoC = State of Charge, OE = Operating Envelope).

As shown in Fig 6, sensors are located at most points of interest. Some measurements were calculated by aggregating data from other sensors, which are called *virtual sensors* (e.g, *AP* of *Garage1* is the sum of *AP* of all sensors hosted within *Garage1*, indicated as a dotted border around the sensor).

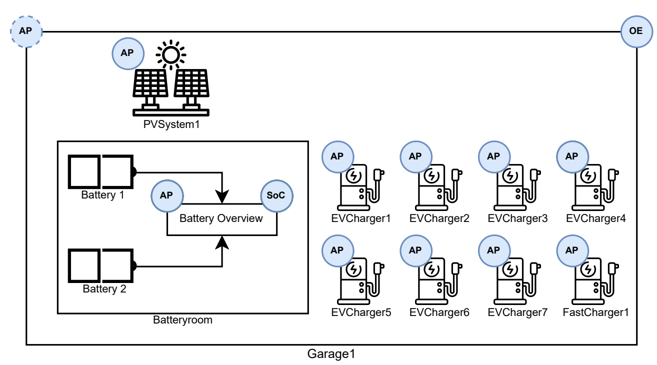


Fig. 6. EV charging garage use case with its platforms and sensors

[4]https://github.com/semanticsystems/semantic-explanation-module

The use case seeks to detect an envelope violation, and create an explanation of why it occurred. We will consider the following states as potentially interesting for the explanation (associated sensor type of a state is defined in brackets):

- *Envelope Violation(AP_Garage)*: active power consumed/produced within the platform exceeds the envelope limits imposed by the operating envelope of the platform.
- *High Charging(AP_EVCharger)*: active power consumed by the EV charger exceeds the operating envelope of the garage.
- *High Charging Difference(AP_EVCharger)*: active power consumed by the EVcharger is increasing very fast.
- *Low Battery SoC(SoC_Battery)*: the battery is depleted (SoC is close to 0)
- *Battery Discharge Loading(AP_Battery)*: active power provided by the battery is increasing, but has not reached its maximum discharging power limit
- *Battery Dip(AP_Battery)*: active power provided by the battery has dropped for a short time (i.e. only a single recorded time instance).
- *Battery Unused(AP_Battery)*: battery is neither charging nor discharging

For these states, corresponding event detection procedures and event-to-state-mappings were set up. Details about the event detection module are beyond the scope of this paper. For a potential implementation of the event detection module, we refer the reader to [Steindl et al. 2024].

The causal knowledge about this use case was developed in a cooperation between domain experts and engineers, constructing a list of type causalities between state types of the system. The resulting causal knowledge is shown in Table 1. Each causal relation consists of a cause state type and effect state type, with a combination of causal, temporal and topological restrictions/features. For example, a `HighChargingState` `causes` an `EnvelopeViolationState`, if the effect state occurs at a `parentPlatform` of the cause state and their timeframes `overlap` (shown in the first entry of Table 1).

Cause	causal	temporal	topol.	Effect
High Charging	causes	overlaps	parent	Envelope Violation
High Charging	enables	overlaps	same	High Charging Difference
High Charging Difference	enables	overlaps	sibling	Battery Discharge Loading
Battery Discharge Loading	causes	overlaps	parent	Envelope Violation
High Charging	causes	before	sibling	Low Battery SOC
Low Battery SOC	causes	before	parent	Envelope Violation
Battery Dip	causes	overlaps	parent	Envelope Violation

Table 1. State Type Causal Knowledge

6.2 Data Model Implementation

Data Model: The SENSE (Semantics-based ExplaNation of cyber-physical SystEms) Ontology[5] was developed to enable the ExpCPS framework implementation, defining the integrated data structure of the framework (see Fig. 3). The ontology is developed to ensure a unified and consistent representation of relevant knowledge for automated explanation generation (RQ2). It provides the generic data model required to represent event and state information, causal knowledge and topology data in a generic CPS (see Section 4.1 / Fig. 3) as the minimal components to provide explainability capabilities in a CPS (RQ1). For the implementation of the semantic model and its instantiation, GraphDB[6], a triple store, was employed, enabling the association of events with semantic context.

```
1  :Garage1 a sosa:Platform;
2      sosa:hosts Garage1_OperatingEnvelopeSensor1;
3      sosa:hosts Garage1_EnvelopeViolationVirtualSensor1;
4      sosa:hosts :EVCharger1;
5      sosa:hosts :BatteryOverview.
6
7  :EVCharger1 a sosa:Platform;
8    sosa:hosts :EVCharger1_APSensor1.
9
10 :BatteryOverview a sosa:Platform;
11   sosa:hosts :BatteryOverview_APSensor2;
12   sosa:hosts :BatteryOverview_SoCSensor1;
13   sosa:hosts :Battery1;
14   sosa:hosts :Battery2.
15
16 :Garage1_EnvelopeViolationVirtualSensor1 a sosa:Sensor;
17   sosa:observes :EnvelopeViolationProperty;
18   sosa:implements :DetectEnvelopeViolationQuery.
```

Listing 1. Instantiation excerpt of the use case in the Semantic Model

Data Instantiation: We instantiate the SENSE Ontology using RDF to model the EV charging use case. For future use cases, engineers can apply these capabilities to other domains by either referring to the generic ontology directly or by mapping domain-specific ontologies to the SENSE Ontology (e.g., Brick Ontology for smart building domain [Balaji et al. 2016]).

Listing 1 shows an excerpt of the data instantiation to represent the system topology according to Fig 6. The example defines a garage (`Garage1`), which has two sensors installed, `Garage1_OperatingEnvelopeSensor1` and `Garage1_EnvelopeViolationVirtualSensor1` (line 1-3). The garage hosts devices, which are logically located inside the garage - `EVCharger1` and `BatteryOverview` (line 4-5). Each of the platforms hosted by `Garage1` hosts more sensors and platforms (line 7-14). Lastly, `Garage1_EnvelopeViolationVirtualSensor1` is defined, which is a virtual sensor that is used for detecting envelope violations (line 16-18).

This example shows how the system topology of a smart grid system can be modeled in an integrated data model using semantic technologies, specifically RDF. A full example instantiation can be found in the GitHub repository.

[5] http://w3id.org/explainability/sense#

[6] https://www.ontotext.com/products/graphdb/

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX s: <http://w3id.org/explainability/sense#>
PREFIX sosa: <http://www.w3.org/ns/sosa/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?causeState ?causalRelation ?effectState
where {
    # effect State
    VALUES ?effectState { STATE }
    ?effectState s:hasStateType ?effectStateType .
    ?effectObservation s:observedState ?effectState ;
    ?effectObservation s:startTime ?effectStart ;
    ?effectObservation s:endTime ?effectEnd .
    ?effectSensor sosa:madeObservation ?effectObservation
        .
    ?effectPlatform sosa:hosts ?effectSensor .
    ?effectParentPlatform sosa:hosts ?effectPlatform .
    # cause State
    ?causeState a s:State .
    ?causeState s:hasStateType ?causeStateType .
    ?causeObservation s:observedState ?causeState .
    ?causeObservation s:startTime ?causeStart .
    ?causeObservation s:endTime ?causeEnd .
    ?causeSensor sosa:madeObservation ?causeObservation .
    ?causePlatform sosa:hosts ?causeSensor .
    ?causeParentPlatform sosa:hosts ?causePlatform .
    {
        select ?causeStateType ?effectStateType ?
            causalRelation
        ?topologicalRelation ?temporalRelation
                where {
            ?stc a s:StateTypeCausality .
            ?stc s:cause ?causeStateType .
            ?stc s:effect ?effectStateType .
            ?stc s:hasCausalRelation ?causalRelation .
            ?stc s:hasTemporalRelation ?temporalRelation
                .
            ?stc s:hasTopologicalRelation ?
                topologicalRelation .
        }
    }
    BIND (
        IF(?topologicalRelation = s:samePlatform,
            ?causePlatform = ?effectPlatform,
        IF(?topologicalRelation = s:parentPlatform,
            ?causeParentPlatform = ?effectPlatform,
        IF(?topologicalRelation = s:siblingPlatform,
            ?causeParentPlatform = ?effectParentPlatform
                &&
            ?causePlatform != ?effectPlatform, true)))
            as ?Platform_filter )
    BIND (
        IF(?temporalRelation= s:overlaps,
            ?causeStart <= ?effectStart &&
            ?effectStart <= ?causeEnd,
        IF(?temporalRelation = s:before,
            ?causeStart <= ?effectStart, true))
            as ?temporal_filter )
    FILTER(?Platform_filter && ?temporal_filter)
}
```

Listing 2. SPARQL Query implementing Algorithms 1-3

6.3 Algorithm Implementation

Actual State Causality Exploration: The actual state causality exploration is implemented as a SPARQL query. It applies the logic of algorithms 1- 3 in a single query to apply causal knowledge from Table 1 in order to create causal connections between instances of states in the graph database. The query takes a state as its input (a newly detected state, for which causal relations still need to be explored) and derives all states that have a direct causal relation to the new state.

Listing 2 shows the full SPARQL query that takes a STATE to derive triples for causal relations in the form of `(causeState, causalRelation, STATE)`. From the causal knowledge data , it derives relevant causalities to filter existing states for relevant causes of the input data based on temporal and topological constraints.

Causal Path Query: The causal path query is also implemented as a SPARQL query, exploring the KG exhaustively to find the full causal tree of the root cause exploration (see Listing 3). It takes a *TRIGGERSTATE* as its input, which is the state for which the full root cause analysis is performed and returned to the user. The causal path query is executed as requested by the user in the user interaction interface. In combination with the *Actual State Causality Exploration*, the query can be used for automated explanation generation of system states at runtime (RQ3).

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX s: <http://w3id.org/explainability/sense#>

SELECT ?cause ?causalRelation ?effect
where {
    VALUES ?triggerstate { TRIGGERSTATE }
  ?cause ?causalRelation ?effect .
  ?causalRelation rdfs:subPropertyOf s:causallyRelated .
  ?effect (s:causallyRelated)* ?triggerstate .
}
```

Listing 3. SPARQL Query implementing Algorithm 4

6.4 Initial Results

Results of the query are visualised in Fig 7. An explanation was requested for the trigger state `EnvelopeViolation25`, which occurred for approximately half an hour on April 8, 2023 at `Garage1_EnvelopeViolationSensor1`. By querying the semantic model, two causal paths could be identified as potential causes of the Envelope Violation:

Path1 `HighCharging22` state occurred at `EVCharger1`, which caused a `BatterySoCLow` state at the SoC Sensor of BatteryOverview. The depleted battery enabled `EnvelopeViolation25`.

Path2 `EnvelopeViolation25` was caused by `HighCharging24` directly. The temporal and topological proximity are the reasons for the actual causal relation between the two states.

Overall, the two causal paths show the user that the Envelope Violation was caused by a combination of two incidents: (i) the battery was depleted due to extended High Charging in the past, which made the operating envelope vulnerable to violations; and (ii) current high charging directly caused the violation.

In the full prototype implementation, our system detected 136 Envelope Violations over the course of one month of sensor measurements. Out of these, 100 Envelope Violations could be correctly explained using our framework. For the remaining 36 violations, there was a lack of causal knowledge in the prototype implementation to identify their full causes. The causal knowledge base is a key factor of the implementation and is a major factor for the success of the explainability framework. As this knowledge currently has to be defined by domain experts, creating a sound causal knowledge base is currently a bottleneck of the framework. In future work, new approaches should be explored in the direction of assisted causal

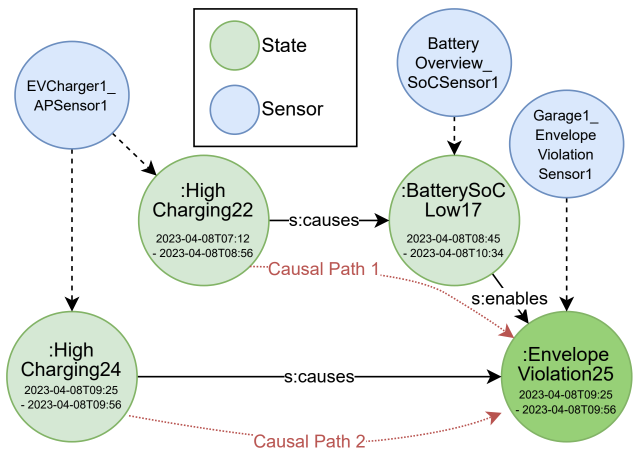


Fig. 7. Causal Path Query result

knowledge base creation by experts, automated causality detection using deep learning methods, or a combination of the two.

7 RELATED WORK

Our work contributes to the emerging field of ExpCPS research, which is focused on providing explanation capabilities to CPSs to make these increasingly complex systems more understandable and usable for humans. This area has gained prominence as a key factor contributing to the European Commission's vision of Industry 5.0, creating human-centric industrial systems that aim beyond efficiency and productivity [Commission et al. 2021]. Explanations are a key resource to make industrial systems more user-friendly.

A significant body of scientific research investigates XAI for SG applications. [Machlev et al. 2022] provides a comprehensive overview of the general concepts and existing works in this field. The authors outline critical challenges XAI techniques face in SG settings: (1) current methods generate explanations tailored for AI experts, not power system engineers, operators, or end users. (2) Machine learning models, often regarded as "black boxes", struggle to gain widespread acceptance in SG applications due to the high level of accountability required in this domain. The authors distinguish between intrinsic explainability (baked into the ML model) and post-hoc explainability (independent of the model) as potential solutions, which are general XAI concepts initially defined in [Das and Rad 2020]. Our proposed approach enables considerable improvements for both challenges: it leverages semantic web technologies, allowing for straightforward mapping to domain-specific concepts and terminology. This feature is integrated into the process of deriving the system topology, enabling the use of domain-specific language for event explanations. Moreover, the decisions and actions of ML applications can be considered events within the SG as a CPS. Consequently, explanations can be provided using causal knowledge in a post-hoc approach.

Research approaches in ExpCPS extend beyond the explanation capabilities of AI systems; they encompass the explainability of complex software systems in their entirety. These approaches are centered around explaining context-dependent decisions that are independent of machine learning algorithms and their decisions, including environmental factors that are needed to explain system states [Jha et al. 2021]. Initial research on ExpCPSs exist for various domains, such as smart grids [Aryan et al. 2020, 2021; Chhokra et al. 2020; Cordova et al. 2018; Larsson et al. 2007] and smart buildings [Lork et al. 2019; Ploennigs et al. 2014]. However, current solutions face several research challenges: (1) ExpCPS research is constrained by boundaries of individual disciplines [Greenyer et al. 2019], limiting the development of universal problem definitions and associated solution frameworks and tools. (2) As explainability is reliant on various, heterogeneous data sources, acquiring and meaningfully integrating relevant CPS data is challenging. (3) The exploration of explanations requires complex analytical processing founded in domain knowledge. We offer an improvement for all of these challenges by (1) proposing a common, domain-independent ExpCPS framework, which can be implemented for various use cases. (2) We define an integrated data model using semantic technologies, which can be adapted to different application domains, offering a common structure for further processing of CPS data. (3) An explanation module is proposed, which processes system events and general domain knowledge to derive explanations.

Anomaly detection and subsequent root-cause analysis [Lorenzo et al. 2008; Rooney and Heuvel 2004] are crucial topics in industrial CPS research. Current methods, such as Failure Mode and Effect Analysis (FMEA) [Ben-Daya 2009] and Fault Tree Analysis (FTA) [Ericson 2011], as well as functional modelling [Lind and Zhang 2014] require specification of possible anomalies and their causes by experts familiar with the system, which can be labor-intensive and reliant on fragmented domain knowledge. Moreover, these methods require a very detailed representation of systems to be able to derive root causes. This knowledge is predominantly documented using natural language descriptions, resulting in ambiguity and incompleteness [Dittmann et al. 2004]. Semantic methods, including ontology-based representations, have been proposed to mitigate these shortcomings [Rehman and Kifor 2016; Zhou et al. 2015]. Our framework builds upon these principles by integrating heterogeneous data sources into a unified data model, enabling complex analytical processing to derive meaningful explanations.

In summary, our ExpCPS framework brings novel developments to the field, by offering a holistic approach to explaining CPS states through the integration of causal knowledge representation, semantic web technologies, and advanced explanation algorithms. By addressing key challenges and providing a unified solution, we pave the way for enhanced transparency, interpretability, and usability of CPS across various domains.

8 CONCLUSION

The paper proposes a domain-independent ExpCPS framework, identifying minimal components needed to provide explanation capabilities in a CPS (RQ1). We focus on the definition of the integrated data model as a basis for generating explanations (RQ2) and an explainability module to automatically generate system state explanations at runtime according to causal domain knowledge from experts (RQ3). We demonstrated the feasibility of our approach in a use case of a public EV charging facility, where our approach is able to detect and explain violations of an operating envelope imposed on the charging facility. This approach aims to help in the realization of the energy transition by contributing to more explainable systems thus making systems of the future saver and more understandable for all stakeholders.

Despite its success, the current approach is highly reliant on well-curated causal knowledge from domain experts, which poses the main bottleneck for a wider adoption. To address this issue, in the future, we aim to develop methods and tools to assist experts in the definition of causal knowledge by proposing possible causal relations derived from data through machine learning and heuristic algorithms. Additionally, we are planning to further evaluate the scalability and generalizability of the approach in a larger smart grid scenario based on simulated energy community data and smart building scenarios. Finally, iterative refinements of the prototype based on feedback and empirical testing in different application domains is essential to validate its effectiveness, identify potential refinements, and ensure its suitability for different environments.

ACKNOWLEDGMENTS

This work was supported by the FFG research project SENSE (project Nr. FO999894802)

REFERENCES

Peb R Aryan, Fajar J Ekaputra, Marta Sabou, Daniel Hauer, Ralf Mosshammer, Alfred Einfalt, Tomasz Miksa, and Andreas Rauber. 2020. Simulation support for explainable cyber-physical energy systems. In *2020 8th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems*. IEEE, 1–6.

Peb Ruswono Aryan, Fajar Juang Ekaputra, Marta Sabou, Daniel Hauer, Ralf Mosshammer, Alfred Einfalt, Tomasz Miksa, and Andreas Rauber. 2021. Explainable cyber-physical energy systems based on knowledge graph. In *Proceedings of the 9th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems*. 1–6.

Bharathan Balaji, Arka Bhattacharya, Gabriel Fierro, Jingkun Gao, Joshua Gluck, Dezhi Hong, Aslak Johansen, Jason Koh, Joern Ploennigs, Yuvraj Agarwal, Mario Berges, David Culler, Rajesh Gupta, Mikkel Baun Kjærgaard, Mani Srivastava, and Kamin Whitehouse. 2016. Brick: Towards a Unified Metadata Schema For Buildings. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, Palo Alto CA USA, 41–50. https://doi.org/10.1145/2993422.2993577

Mohamed Ben-Daya. 2009. Failure mode and effect analysis. In *Handbook of maintenance management and engineering*. Springer, 75–90.

Matthias Bittner, Daniel Hauer, Christian Stippel, Katharina Scheucher, Robin Sudhoff, and Axel Jantsch. 2023. Forecasting Critical Overloads based on Heterogeneous Smart Grid Simulation. In *2023 International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 339–346.

Mathias Blumreiter, Joel Greenyer, Francisco Javier Chiyah Garcia, Verena Klos, Maike Schwammberger, Christoph Sommer, Andreas Vogelsang, and Andreas Wortmann. 2019. Towards Self-Explainable Cyber-Physical Systems. In *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. IEEE, Munich, Germany, 543–548. https://doi.org/10.1109/MODELS-C.2019.00084

Federico Cabitza, Andrea Campagner, Gianclaudio Malgieri, Chiara Natali, David Schneeberger, Karl Stoeger, and Andreas Holzinger. 2023. Quod erat demonstrandum? - Towards a typology of the concept of explanation for the design of explainable AI. *Expert Systems with Applications* 213 (March 2023), 118888. https://doi.org/10.1016/j.eswa.2022.118888

Ajay Chhokra, Nagabhushan Mahadevan, Abhishek Dubey, and Gabor Karsai. 2020. Qualitative fault modeling in safety critical cyber physical systems. In *Proceedings of the 12th System Analysis and Modelling Conference*. 128–137.

European Commission, Directorate-General for Research, Innovation, A Renda, S Schwaag Serger, D Tataj, A Morlet, D Isaksson, F Martins, M Mir Roca, C Hidalgo, A Huang, S Dixson-Declève, P Balland, F Bria, C Charveriat, K Dunlop, and E Giovannini. 2021. *Industry 5.0, a transformative vision for Europe – Governing systemic transformations towards a sustainable industry*. Publications Office of the European Union. https://doi.org/doi/10.2777/17322

Jose Cordova, Lalitha Madhavi Konila Sriram, Ayberk Kocatepe, Yuxun Zhou, Eren E Ozguven, and Reza Arghandeh. 2018. Combined electricity and traffic short-term load forecasting using bundled causality engine. *IEEE Transactions on Intelligent Transportation Systems* 20, 9 (2018), 3448–3458.

Arun Das and Paul Rad. 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* (2020).

Lars Dittmann, Tim Rademacher, and Stephan Zelewski. 2004. Performing FMEA using ontologies. In *18th International Workshop on Qualitative Reasoning. Evanston USA*. 209–216.

Lisa Ehrlinger and Wolfram Wöß. 2016. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)* 48, 1-4 (2016), 2.

Clifton A Ericson. 2011. *Fault tree analysis primer*. CreateSpace Incorporated.

M Jabbari Ghadi, Sahand Ghavidel, Amin Rajabi, Ali Azizivahed, Li Li, and Jiangfeng Zhang. 2019. A review on economic and technical operation of active distribution systems. *Renewable and Sustainable Energy Reviews* 104 (2019), 38–53.

Joel Greenyer, Malte Lochau, and Thomas Vogel. 2019. Explainable Software for Cyber-Physical Systems (ES4CPS). In *GI Dagstuhl Seminar*, Vol. 19023.

CEN-CENELEC-ETSI Smart Grid Coordination Group. 2012. *Smart Grid Coordination Group (SGCG) Reference Architecture Smart Grid*. Technical Report.

Joseph Y Halpern. 2016. *Actual causality*. MiT Press.

Martin Hankel and Bosch Rexroth. 2015. The reference architectural model industrie 4.0 (rami 4.0). *Zvei* 2, 2 (2015), 4–9.

David Jacobs. 2012. The German Energiewende–history, targets, policies and challenges. *Renewable Energy Law and Policy Review* (2012), 223–233.

Krzysztof Janowicz, Armin Haller, Simon JD Cox, Danh Le Phuoc, and Maxime Lefrançois. 2019. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics* 56 (2019), 1–10.

Sanjiv Subodhnarayan Jha. 2022. An Overview on the Explainability of Cyber-Physical Systems. *The International FLAIRS Conference Proceedings* 35 (2022). https://doi.org/10.32473/flairs.v35i.130646

Sanjiv S Jha, Simon Mayer, and Kimberly García. 2021. Poster: Towards Explaining the Effects of Contextual Influences on Cyber-Physical Systems. In *Proceedings of the 11th International Conference on the Internet of Things*. 203–206.

Jan Eric Larsson, Bengt Öhman, and Antonio Calzada. 2007. Real-time root cause analysis for power grids. In *Security and Reliability of Electric Power Systems, CIGRE Regional Meeting, Tallinn, Estonia*. Citeseer.

Morten Lind and Xinxin Zhang. 2014. Functional modelling for fault diagnosis and its application for NPP. *Nuclear Engineering and Technology* 46 (Dec. 2014), 753–772. https://doi.org/10.5516/NET.04.2014.721

Donald K Lorenzo, Laura O Jackson, et al. 2008. *Root cause analysis handbook: A guide to efficient and effective incident investigation*. Rothstein Publishing.

Clement Lork, Vishal Choudhary, Naveed Ul Hassan, Wayes Tushar, Chau Yuen, Benny Kai Kiat Ng, Xinyu Wang, and Xiang Liu. 2019. An ontology-based framework for building energy management with IoT. *Electronics* 8, 5 (2019), 485.

R Machlev, L Heistrene, M Perl, KY Levy, J Belikov, S Mannor, and Y Levron. 2022. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI* 9 (2022), 100169.

Jesús Jaime Moreno Escobar, Oswaldo Morales Matamoros, Ricardo Tejeida Padilla, Ixchel Lina Reyes, and Hugo Quintana Espinosa. 2021. A comprehensive review on smart grids: Challenges and opportunities. *Sensors* 21, 21 (2021), 6978.

Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen, and Lucy Vanderwende. 2016. CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures. In *Proceedings of the Fourth Workshop on Events*. Association for Computational Linguistics, San Diego, California, 51–61. https://doi.org/10.18653/v1/W16-1007

Hausi A Müller. 2017. The Rise of Intelligent Cyber-Physical Systems. *Computer* 50, 12 (2017), 7–9.

Joern Ploennigs, Anika Schumann, and Freddy Lécué. 2014. Adapting semantic sensor networks for smart building diagnosis. In *The Semantic Web–ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II 13*. Springer, 308–323.

Meysam Qadrdan, Nick Jenkins, and Jianzhong Wu. 2018. Smart grid and energy storage. In *McEvoy's Handbook of Photovoltaics*. Elsevier, 915–928.

Zobia Rehman and Claudiu Vasile Kifor. 2016. An ontology to support semantic management of FMEA knowledge. *International Journal of Computers Communications & Control* 11, 4 (2016), 507–521.

James J Rooney and Lee N Vanden Heuvel. 2004. Root cause analysis for beginners. *Quality progress* 37, 7 (2004), 45–56.

Mersedeh Sadeghi, Lars Herbold, Max Unterbusch, and Andreas Vogelsang. 2024. SmartEx: A Framework for Generating User-Centric Explanations in Smart Environments. *arXiv preprint arXiv:2402.13024* (2024).

Gernot Steindl, Tobias Schwarzinger, Katrin Schreiberhuber, and Fajar Ekaputra. 2024. Towards Semantic Event-handling for building Explainable Cyber-physical Systems. *IEEE Open Journal of the Industrial Electronics Society* (2024). under review.

Dominik Tomaszuk and David Hyland-Wood. 2020. RDF 1.1: Knowledge Representation and Data Integration Language for the Web. *Symmetry* 12, 1 (2020). https://doi.org/10.3390/sym12010084

Maria Lorena Tuballa and Michael Lochinvar Abundo. 2016. A review of the development of Smart Grid technologies. *Renewable and Sustainable Energy Reviews* 59 (2016), 710–725.
Anmei Zhou, Dejie Yu, and Wenyi Zhang. 2015. A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA. *Advanced Engineering Informatics* 29, 1 (2015), 115–125.

Received 10 May 2024

Analysing and Predicting Extreme Frequency Deviations: A Case Study in the Balearic Power Grid

ZAKARIEA SHARFEDDINE*, SEBASTIAN PÜTZ*, VEDANG TAMHANE, VEIT HAGENMEYER, and BENJAMIN SCHÄFER, Karlsruhe Institute of Technology, Germany

Maintaining the power grid stability requires a balance between the generation and demand across various timescales. The power grid frequency reflects this balance and hence plays an important role in stabilizing and controlling power grids. Large imbalances and hence frequency deviations represent a threat to power grid stability. Therefore, understanding such extreme deviations and having a prior warning would be very valuable for power grid operation. In this paper we analyze extreme frequency deviations and develop a decision-support tool to predict and mitigate such events, demonstrated using data from the Balearic power grid. We inspect the frequency and durations of extreme deviations and identify specific fast ramps as precursors. We develop an interpretable machine learning model that is intended to warn before extreme frequency deviations and provide insights into its reasoning by using SHapely Additive exPlanations (SHAP).

CCS Concepts: • **Computing methodologies → Machine learning**; • **Applied computing →** *Engineering*; • **Hardware → Energy distribution**.

Additional Key Words and Phrases: frequency stability, machine learning, explainable AI, decision support

1 INTRODUCTION

Balancing generation and demand at all timescales is essential for the reliable operation of power grids. The central observable to maintain this balance on short timescales is the power grid frequency. Deviations of the power grid frequency represent an imbalance between active power generation and demand [Machowski et al. 2020]. An excess of demand effectively slows down the synchronous generators connected to the grid and results in a decreasing frequency. Conversely, a surplus of generation entails a frequency increase. Large deviations of frequency from the rated value represent a threat to the stability of the grid by leading to power outages. Therefore, frequency control is a main challenge in maintaining the stable operation of power grids.

Since power generation transitions towards inverter-based renewable energy sources, like wind and solar power, the system inertia decreases while fluctuations increase [Milano et al. 2018; Ulbig et al. 2014]. Smaller grids, like those on islands, typically have lower inertia and show larger frequency fluctuations [Onsaker et al. 2023; Rydin Gorjão et al. 2020]. To manage these fluctuations and maintain stability, accurate and fast frequency control is essential.

If an island grid is connected to another synchronous system via a High Voltage Direct Current (HVDC) link, that link can be utilized to provide control power and thus benefit power system stability. HVDC links allow for the efficient transmission of power over long distances, independent of the frequency in either connected AC grid while providing precise control of the power flow [Alassi et al. 2019]. However, the flow of the HVDC link, if used for frequency control, could negatively affect the frequency stability of the other synchronous system [Pütz et al. 2022].

Modern machine learning (ML) techniques could support the transition to a sustainable power system [Rolnick et al. 2022]. In particular, ML models are effective at analyzing the ever growing data sets and complex features that are critical to managing grid stability. Therefore such methods can aid in processing otherwise humanly unmanageable data [Kummerow et al. 2018] and thereby support decision-making for Transmission System Operators (TSOs). However, the complexity of ML models often obscures why a model decision is made, posing challenges in critical areas such as the power system [Kruse et al. 2021b; Machlev et al. 2022]. Explainable Artificial Intelligence (XAI) methods address this limitation by making ML decisions transparent. We highlight Shapley Additive Explanations (SHAP) values [Lundberg and Lee 2017] as they explain the influence of each feature on model predictions and are particularly suitable for models such as gradient-boosted trees [Lundberg et al. 2020]. Lundberg et al. [2018] show how XAI can support human decision-making in a medical application. Inspired by this we implement a prototypical decision support tool for the prediction of power grid frequency deviations.

There have already been a few related studies investigating power grid frequency time series with ML: Yurdakul et al. [2020] applied Recurrent Neural Network architectures to forecast frequency with one-minute granularity for one-step-ahead predictions and Chettibi et al. [2021] aimed forecast grid frequency with horizons ranging from 183 ms up to 2 min (in 1 min increments). The study by Taylor and Roberts [2016] focused on the forecasting of minutely power grid frequency-corrected demand in Great Britain. A recent study centered on predicting hourly frequency trajectories using transformers [Pütz et al. 2024]. However, the above studies do not specifically focus on forecasting rare extreme deviations, which inherently exhibit low support in data and are therefore often overlooked by general-purpose forecasting models.

The contributions of this paper are twofold: First, we analyze extreme frequency deviations, which are mitigated via HVDC link frequency control and identify under which circumstances these events appear more likely. Second, we build a prototypical decision-support tool to predict and warn power system operations of large frequency deviations before they occur so they can prepare for counter-measures. We demonstrate both contributions on data from the Balearic power grid as a case study.

The paper is organized as follows. We start with the description and analysis of extreme deviations in the Balearic power grid in Section 2.2. After a short description of the Balearic power system followed by an analysis of extreme deviations, we search for precursors of such events. In Section 3, we investigate the development of a decision-support tool that predicts if an extreme deviation will

*Both authors contributed equally to this research.

Authors' address: Zakariea Sharfeddine, zakariea.sharfeddine@student.kit.edu; Sebastian Pütz, sebastian.puetz@kit.edu; Vedang Tamhane;
Veit Hagenmeyer, veit.hagenmeyer@kit.edu; Benjamin Schäfer, benjamin.schaefer@kit.edu, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, Germany, 76344.

occur within the next five minutes. Section 4 discusses our results before we close the paper with a conclusion and outlook in Section 5.

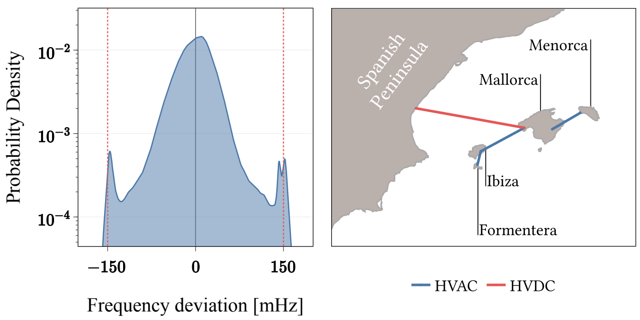


Fig. 1. (Left) Probability density (kernel density estimation) of the frequency time series shows two distinct peaks at ±150 mHz (Right) Map of the Balearic Power System. The islands Mallorca, Menorca, Ibiza, and Formentera are connected via High Voltage Alternating Current (HVAC). An HVDC link connects Mallorca to mainland Spain.

2 EXTREME FREQUENCY DEVIATIONS IN THE BALEARIC POWER GRID

2.1 The Balearic Power Grid

In the present paper, we use the Balearic grid as a case study. Its data availability in terms of frequency, generation and demand data, as well as its unique grid frequency characteristics, make the Balearic grid an exemplary candidate for our study.

The Balearic Islands are located in the Mediterranean Sea, off the east coast of Spain. Their synchronous power grid spans the four main islands: Mallorca, Menorca, Ibiza, and Formentera (see Fig. 1 (Right)). The Cometa HVDC link runs between Mallorca and the Spanish peninsula and thereby connects the Balearic to the Continental European power grid. With a maximum transmission capacity of 400 MW it is large compared to the daily peak load in the Balearic grid of under 1 GW [Red Eléctrica de España 2024]. For comparison the daily peak load of the Continental European system is around 300 GW [ENTSO-E 2024].

Throughout this paper, we use 1 s-resolution frequency time series recorded in Palma de Mallorca using an electrical data recorder (EDR) [Jumar et al. 2020; Rydin Gorjão et al. 2020]. Moreover, we utilize 5 min-resolution generation and demand data that is publicly provided by the Spanish transmission system operator [Red Eléctrica de España 2024]. We use data from 29 September 2019 through 30 June 2023. For more information on our data sources refer to Appendix A.1.

The probability density of the frequency deviation time series (see Fig. 1 (Left)) shows two distinct peaks at ±150 mHz. These peaks arise as a consequence of strict threshold-like frequency control [Martínez-Barbeito et al. 2021], which we presume is provided via the HVDC link. An example of a large frequency deviation from the 11th November 2022 is shown in Fig. 2. At around 19:45 the frequency sharply drops more than 300 mHz from the reference frequency before being stabilized at −150 mHz. The 5 min-resolution HVDC import data indicates a mitigating role in this event. Shortly

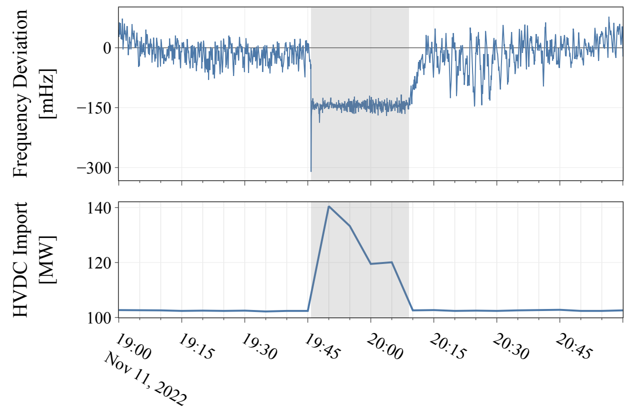


Fig. 2. Example of a large negative frequency deviation in the Balearic Power grid that seems to be mitigated via HVDC power imports. The link is appears to be used for frequency containment rather than restoration.

after the drop, the link ramps up almost 40 MW to stabilize the frequency and returns to its previous Import rate when the frequency returns to normal behaviour. We observe power adjustments of HVDC imports where the frequency deviates by 150 mHz in either direction [Red Eléctrica de España 2024]. However, from the available 5-minute data, we can neither conclusively infer if these adjustments are made for all 150 mHz deviations nor if the HVDC link is the only entity providing this kind of control.

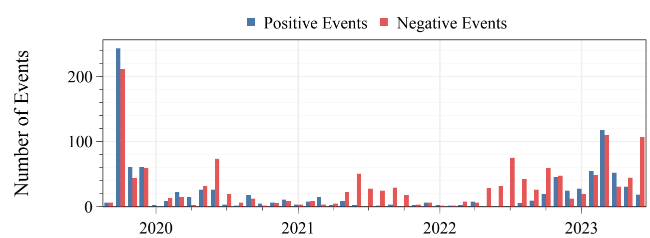


Fig. 3. Number of large positive and negative frequency deviations in the Balearic grid per month.

2.2 Analysing Extreme Deviations

After observing these extreme deviations, we are interested in how often they occur and how long the frequency stays that far from its nominal value. In the following, we define extreme events as instances where the frequency falls below the lower threshold of 49.85 Hz or exceeds the upper threshold of 50.15 Hz. We define a five-minute time frame during which the frequency can fall below the aforementioned thresholds and exceed them again. In these cases, we consider this to still belong to the same event. Should this duration exceed five minutes, the occurrences would be considered distinct events. Events exceeding the upper threshold are classified as positive events, while those exceeding the lower threshold are classified as negative events.

We observe a total of 1390 negative directed and 969 positive directed extreme frequency deviation events between 29 September 2019 and 30 June 2023. The month with the most events is October 2019 (see Fig. 3). Notably, this was before the only coal power plant in Mallorca went out of operation at the end of 2019 [Martínez-Barbeito et al. 2021]. Furthermore, there is a discernible seasonal trend, with a higher number of negative events occurring during the summer months.

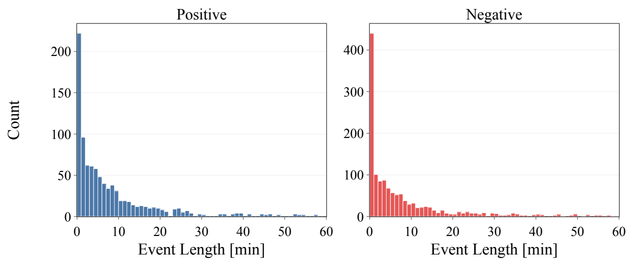


Fig. 4. The length distribution of positive and negative events shorter than one hour. Each bar represents one minute.

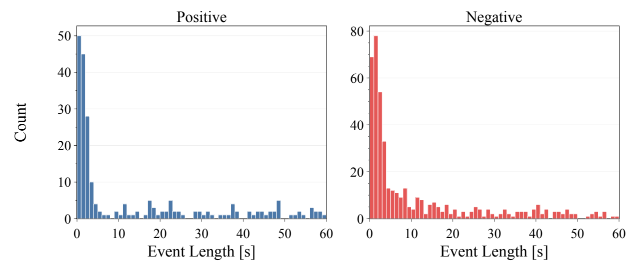


Fig. 5. The event duration distribution of positive and negative events shorter than one minute. Each bar represents one second.

The events vary drastically in length. The longest event lasted even more than six hours and in total 73 events took longer than 1 hour. However, most events are shorter. In Fig. 4 we show the distribution of event length of positive and negative events that are shorter than one hour. It is clearly visible that most of the events are shorter than one minute and a large part is shorter than 10 minutes. We observe a similar distribution for negative and positive events.

Most events are very short. We plot the distribution of event lengths restricted to events shorter than 1 min in Fig. 5. The majority of events in both directions are shorter than 10 seconds.

2.3 Searching for Precursors

To find precursors of the observed events, we examine the demand and generation ramps just before an event. Fast ramps can result in mismatched demand and generation and thereby entail large frequency deviations. In particular, the step-like market-based generation ramping is known to drive the so-called deterministic frequency deviations at the start of trading intervals [Kruse et al. 2021a; Schafer et al. 2018; Weissbach and Welfonder 2009]. Therefore, we apply k-means clustering to the generation ramp data (difference in generation between two timesteps) collected in the 5-minute windows preceding each event. By examining the characteristics of these ramps, such as their magnitude and rate, the algorithm clusters events into categories that represent similar operational scenarios. K-means clustering is a widely used and simple clustering method K-means is chosen for its efficiency and ease of understanding. However, it requires the number of clusters to be specified in advance. We chose the number of clusters in a way that balances clustering performance metrics and interpretability (see [Sharfeddine 2024] for a more detailed explanation). In Fig. 6 we show the centroids of the $k = 5$ clusters and the standard deviation of their included points for positive.

The largest clusters of both positive and negative events (cluster 1 in both cases) display the largest uncertainties but also hint towards an important role of the demand for frequency stability: Positive events are associated with a negative demand ramp and negative events with a positive demand ramp. This demand-dependency is consistent with the observation that more (particularly negative) events occur during summer when the overall demand and hence the potential demand ramps are increased on the islands.

Clusters 2 and 5 in Fig. 6 for positive events show a similar pattern. The HVDC link import power is ramping up fast. Cluster 2 shows HVDC ramping faster than demand while the fifth cluster shows the HVDC ramping up faster than combined cycle generation is ramped down. In contrast to its previously discussed mitigation potential, in these cases, the fast HVDC ramps seem to harm frequency stability. Due to this fast-ramping behaviour of HVDC links ramping speed restrictions have already been imposed on other links e.g. in the Nordic power grid [Energinet et al. 2017]. Cluster 3 highlights an increase in renewables, while the demand is slightly dropping. This can possibly be explained via unforeseen solar energy generation. Cluster 5, the combined cycle generation ramps down while the HVDC link power ramps up. The ramp-up of the HVDC link power is larger, potentially causing a frequency deviation. Cluster 3 for the negative deviations is similar in the opposite direction.

In the small clusters 2, 4 and 5 for negative deviations, we observe downramps of combined cycle or coal that are larger than HVDC link upramps. Notably, these ramp-downs are large (up to $20\,\mathrm{MW\,min^{-1}}$)and could point to failures that are tried to be mitigated via HVDC imports. This would also be coherent with their small cluster size.

3 DECISION-SUPPORT TOOL

3.1 Setup

Inspired by the work of Lundberg et al. [2018], we develop a decision-support tool that predicts the odds of extreme frequency deviations within the next five minutes in the Balearic grid as a case study (see Fig. 8). The authors of Lundberg et al. [2018] developed a machine-learning-based approach to prevent hypoxaemia during surgery. They trained a model that not only predicts whether a patient will suffer from hypoxaemia within the next five minutes, but using SHAP, also provides an explanation to the anaesthesiologist as to why it is predicting an increased risk. Similarly, we strive to investigate the design of a prototypical model that not only predicts

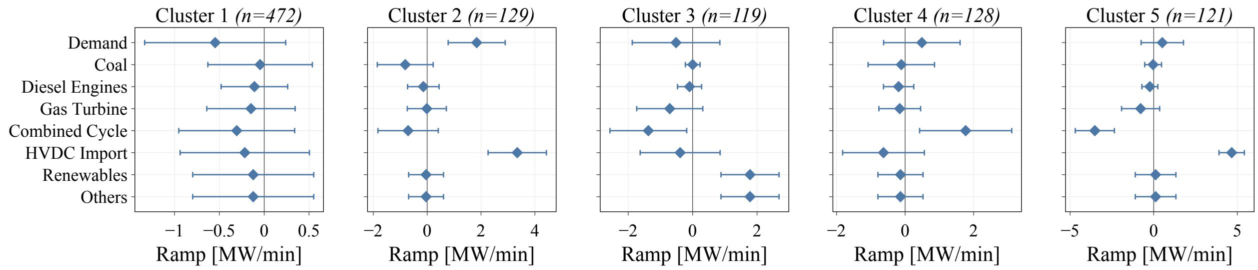


Fig. 6. Centroids and standard deviation per feature of five clusters before positive deviations.

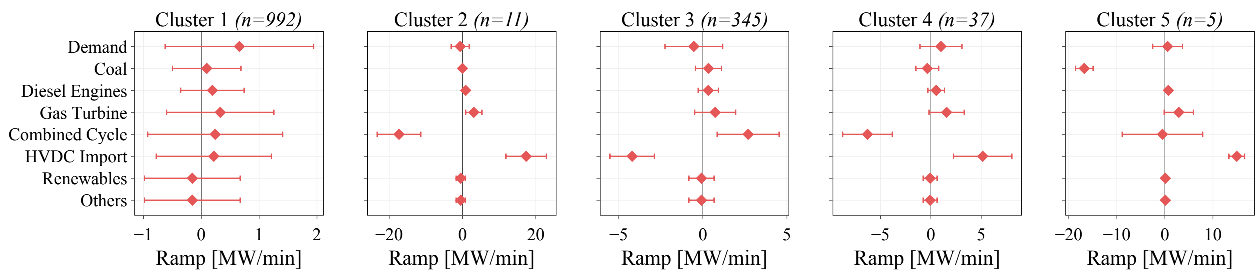


Fig. 7. Centroids and standard deviation per feature of five clusters before negative deviations.

extreme frequency events but is also able to inform grid operators about the reasoning of its prediction.

Although other types of frequency control are already active in the Balearic network below deviations of 150 mHz [Martínez-Barbeito et al. 2021], we will focus here only on the extreme cases that necessitate the activation of the previously described drastic threshold-like frequency control to keep the frequency within the statutory limits. We therefore train a classifier that distinguishes between three classes ("positive event", "negative event" and "no event"). These three classes refer to the same definition of extreme events given in Section 2.2. By treating this as a classification rather than a regression problem we ensure that the model is specifically trained on identifying extreme cases, rather than learning day-to-day behaviour.

Gradient-boosted tree models, as also used in the hypoxaemia model in Lundberg et al. [2018], show great performance on tabular data and allow for fast and efficient computation of SHAP values [Lundberg et al. 2020, 2018]. For validation purposes, we compare the performance of the tree-based model with a logistic regression baseline. Since our frequency data is in second resolution while our techno-economic features and hence ML model use a 5-minute time resolution, we make use of exponentially decaying weighted averages (exponential moving average (EMA)) and variances (exponential moving variance (EMV)) to capture information of the recent frequency trajectory. Similarly, we again compute features like ramps from the five-minute resolution generation and demand data. For more information on feature engineering refer to Appendix A.2. In total, we extracted 47 features from the two datasets.

In Table 1 we observe that the support for both types of extreme events is very small. Moreover, from Fig. 3 it is apparent, that the extreme events are not uniformly distributed over time. To still train a useful classifier while minimizing the effects of a look-ahead bias, we apply a stratified split of weekly chunks on our dataset to obtain train, validation and test set (for more details refer to [Sharfeddine 2024]). To validate our model we compare it to a simple linear logistic regression model as a baseline.

We aim to build a model that not only alerts a user in case of extreme deviations but also explains its prediction. Computing SHAP values provides the user with valuable insights to make informed decisions [Lundberg and Lee 2017]. The SHAP values are based on the game-theoretic Shapley values [Shapley 1953]. While Shapley values are fair attributions for players in cooperative games, SHAP values attribute the effects of individual features on the model output in the context of machine learning models. An important property of SHAP values is their efficiency, which guarantees that the SHAP values of all input features of a single sample together with the average prediction add up to the model output for the given sample. In the case of our classification model, the SHAP values sum up to the direct model output, which can be transformed into probabilities by applying the softmax function. SHAP values are inherently local explanations, i.e. they are calculated for each sample separately. While this is exactly what is needed for explaining a single model output to a user as in our case, SHAP values can offer additional insights on a global level [Lundberg et al. 2020]. By calculating the mean of the absolute SHAP values for multiple samples we find the most important features. By plotting SHAP values against their respective feature values we explore the nonlinear dependencies of the trained model. Such global insights are valuable to understand and potentially improve ML models.

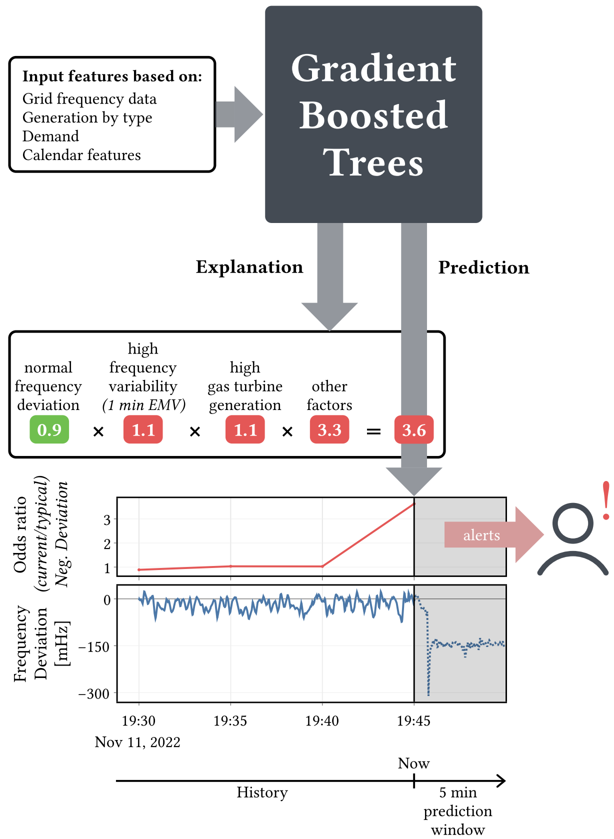


Fig. 8. Overview of the explainable ML model. The model predicts the odds of extreme events occurring within the next five-minute window and alerts the operator if it anticipates an extreme deviation. Here we show the exponentiated model outputs. Thereby the sum of SHAP values becomes a product.

Table 1. Classification Report of the trained model on the test set.

Class	Precision	Recall	F1-score	Support
Negative Event	0.12	0.73	0.21	267
No Event	1.00	0.97	0.98	65151
Positive Event	0.09	0.70	0.16	113

3.2 Performance

The gradient-boosted tree model predicts 70% of all positive events and 73% of all negative events in our test set (see Table 1). However, with 9% for positive events and 12% for negative events, the precision is rather low. For every true prediction, we have about nine false alarms. Notably, we are only using publicly available data within this model. The precision-recall curves in Fig. 9 show that the gradient-boosted tree model outperforms the logistic regression baseline, while still having room for improvement.

The false alarms predicted by our model are often precursors of real events. To quantify this, we plot the time until the next corresponding event in Fig. 10. Notably, for about a third of the false alarms, an extreme event actually occurred within the next hour. For more than 10% an actual extreme event happened already within 15 minutes after the prediction. This indicates that the model is indeed able to recognize high-risk periods, but cannot always identify the specific timing of these events.

3.3 Feature analysis

To obtain further insights into the model's predictions and how it utilizes the techno-economic features, we compute SHAP values on the test set. These insights could be very valuable to a user to make informed decisions mitigating extreme frequency deviations if such a model is applied in practice. Moreover, these insights help us to understand what features are important to the model and how to potentially improve the model. To find the most important features of the model, for each class, we calculate the average of the absolute SHAP values and normalize them by dividing by the highest value.

Analysing the relative feature importances in Fig. 11, we observe that the current frequency deviation and other features derived from the frequency data play an important role in the model output. Apart from frequency-derived features we also find the two main generation types, combined-cycle generation and gas turbine generation, as well as the HVDC import among the most important features.

We proceed by analysing how these important features influence the model predictions. By plotting the SHAP value for a specific feature against the respective feature values, we can get more detailed insights. Exemplary, we plot SHAP dependence plots for the current frequency deviation and the gas turbine generation in Fig. 12. Note that the SHAP values sum up to the direct model output before the softmax function is applied. The dispersion in the y-direction reflects the interaction effects with other features. We observe that frequencies below the nominal grid frequency increase the risk for extreme negative deviations while frequencies above the nominal grid frequency have a similar effect for extreme positive deviations. Moreover, frequencies in opposite directions seem to alleviate the risk slightly. This makes sense intuitively: if the frequency is already below or above its nominal value and thus a generation mismatch exists, a large deviation is much more likely, as only a smaller additional mismatch is then required.

The SHAP dependence plot for gas turbine generation in Fig. 12 *(bottom)* shows further interesting characteristics. For both, positive and negative extreme deviations, the predicted probabilities decrease if gas turbine generation is low or nonexistent. This can probably be linked to its role in driving these deviations in both directions with fast ramps (see Section 2.3).

4 DISCUSSION

From our analysis, we observe that the amount of extreme frequency deviations has not decreased over time (see Fig. 3). This observation

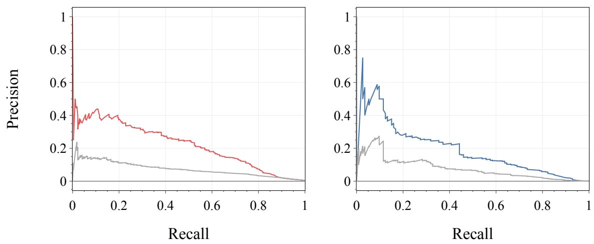


Fig. 9. Precision-recall curves for negative deviations *(left)* and positive deviations *(right)*. The grey lines show the performance of a logistic regression baseline model.

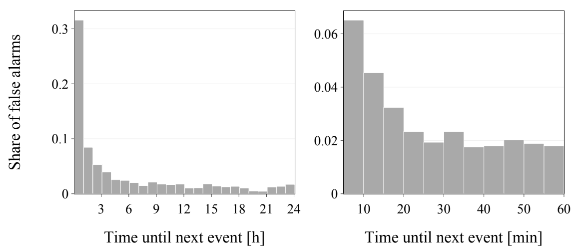


Fig. 10. Temporal distance to the next positive or negative event in case of false alarms. In more than 30% an event occurs within the next hour *(left)*. In more than 15% an event occurs within the next 15 minutes *(right)*.

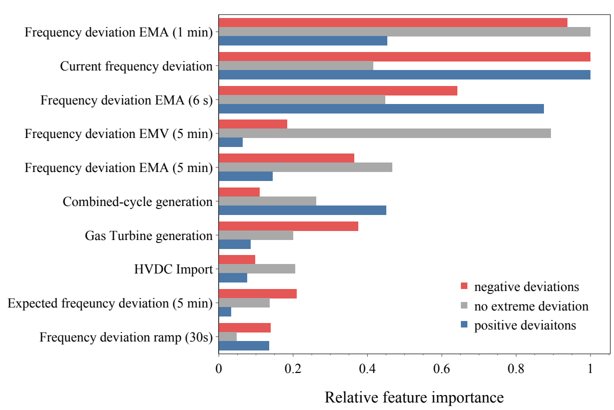


Fig. 11. The union of the seven most important features for each class. The features are sorted by their combined mean absolute SHAP values.

raises the question as to whether in the Balearic grid, this is even perceived as a problem, or if one rather relies on fast control, such as via the HVDC link, as a guard rail of the system. In Fig. 4 we observe, that while most events are short, multiple events are events lasting minutes or hours, which should usually offer sufficient time for frequency restoration. The reasons for these sustained discrepancies between the frequency and its nominal value remain unclear.

The lack of sufficient public data makes it difficult to identify any causal relationship related to frequency deviations. In the event analysis section of our results, we aimed to identify potential precursors to events using the k-means clustering algorithm. This approach was used to examine load and demand ramps preceding extreme frequency deviations (see Fig. 6 and Fig. 7). The underlying hypothesis is that load ramps in different directions and sudden changes in demand or generation technology can trigger frequency deviations. Our analysis revealed instances of load ramps moving in different directions that we are able to interpret. However, access to higher-resolution power generational data could further improve our analysis. As the generation power is averaged over 5 minutes, it is not possible to detect fast ramps which are driving frequency deviations. Such averaging effectively smooths out any abrupt peaks in load, blending them with other data points within the same time frame.

Our explainable ML model predicts extreme frequency deviations within a 5-minute prediction window. Despite the limits of the available data and the rarity of such events, our model predicts more than 70% of them. However, only with a precision of approximately 10%, as detailed in Table 1. The ML model categorizes the predictions into three outcomes: a positive event, no event or a negative event is expected in the next 5 minutes. These decisions are paired with interpretable explanations. Using SHAP values we can explain which features played a role and to what extent. In particular, we identify nonlinear dependencies and the interaction between features. Overall, this allows valuable insights into the model's reasoning. As shown in Table 1, the model shows better recall than precision. A low precision implies that we have a high number of false alarms. However, for ensuring frequency stability high recall is more important than high precision. The Transmission System Operator (TSO) should be alerted to potential frequency deviations. It is more valuable to miss fewer extreme events than to have false alarms. In the case of our model, we additionally found, that many false alarms are followed by an extreme deviation, indicating that the model can detect time periods of higher risk.

This also poses the question of how a modified approach would perform, that does not solely focus on the extreme events that violate the statutory limit but also includes less extreme cases. These cases are still interesting for grid operators. Shifting to a regression problem, models still emphasizing rare extreme cases are particularly interesting [Ding et al. 2019]. Moreover, adapting our approach in such a way could make such an approach more valuable in that application to other grids, as this observed threshold-like control seems to be a peculiarity of the Balearic grid.

We suspect that incorporating higher-resolution generation data or more knowledge of individual power plant scheduling and generation could facilitate the development of a similar model with improved predictive capabilities. Fig. 11 reveals that the key features influencing our predictions are frequency-related. The frequency data is in 1 s-resolution. However, to suit our tree-based ML model, it currently operates derived features in 5-minute resolution. By including more of the actual frequency deviation time series, e.g. by adapting to model architectures dedicated to time-series data, we could potentially improve our predictive performance. In order to

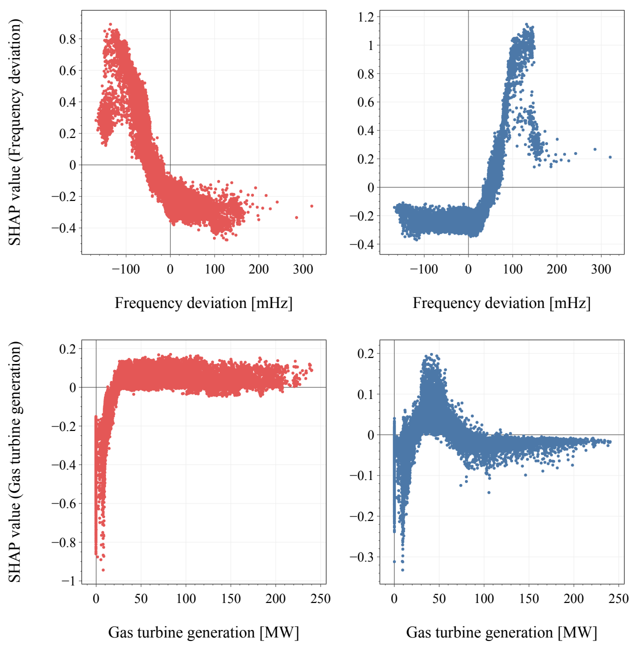


Fig. 12. Dependence between the SHAP values for the current frequency deviation *(top)* and gas turbine generation *(bottom)* and their respective feature values. We show the SHAP values for negative events *(left)* and positive events *(right)*.

still be able to provide insights into the model, future work could explore explainability techniques specifically designed for time series data [Enguehard 2023].

5 CONCLUSION & OUTLOOK

In our research on frequency deviations in the Balearic power grid, we observe a threshold-like frequency control that is frequently used to mitigate extreme frequency deviations and to keep the frequency within the range of 49.85 Hz and 50.15 Hz. While most of the time the nominal frequency is restored after a few seconds, sometimes it takes even several hours. We identify fast HVDC and combined cycle generation ramps, demand ramps and failures as typical precursors for extreme frequency deviations. We developed an ML model that could prove itself as a prototype of a decision support tool. Our model has demonstrated the capability to predict over 70% of the extreme frequency deviation events while using only publicly available data sources. Combined with SHAP, we can interpret the classifier's decisions.

Concluding, we have made two contributions. First, we analyzed extreme frequency deviations in the Balearic power grid and identified precursors of these events. Second, we built a prototypical decision support tool that is trained to predict if these extreme frequency deviations will occur within a 5-minute time window. The model is paired with interpretable explanations. Understanding the dynamics of extreme frequency deviations through XAI can help stakeholders make informed decisions and therefore facilitate transition to sustainable energy sources while ensuring grid stability.

Future research will explore highly resolved data, different island systems and alternative ML architectures for predictions.

ACKNOWLEDGMENTS

This project is funded by the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI. V Tamhane received funding via the Aspirant Grant (Doc) of the KHYS, KIT.

REFERENCES

Abdulrahman Alassi, Santiago Bañales, Omar Ellabban, Grain Adam, and Callum MacIver. 2019. HVDC transmission: Technology review, market trends and future outlook. *Renewable and Sustainable Energy Reviews* 112 (2019), 530–554.

Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

N. Chettibi, A. Massi Pavan, A. Mellit, A. J. Forsyth, and R. Todd. 2021. Real-Time Prediction of Grid Voltage and Frequency Using Artificial Neural Networks: An Experimental Validation. *Sustainable Energy, Grids and Networks* 27 (2021), 100502. https://doi.org/10.1016/j.segan.2021.100502

Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, and Xiangnan He. 2019. Modeling Extreme Events in Time Series Prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. Association for Computing Machinery, New York, NY, USA, 1114–1122. https://doi.org/10.1145/3292500.3330896

Energinet, Fingrid Oyj, Kraftnät Åland AB, Svenska kraftnät, and Statnett SF. 2017. Nordic Synchronous Area Proposal for Ramping Restrictions for Active Power Output in Accordance with Article 137(3) and (4) of the Commission Regulation (EU) 2017/1485 of 2 August 2017 Establishing a Guideline on Electricity Transmission System Operation. (2017). https://consultations.entsoe.eu/system-operations/nordic-tsos-proposal-on-ramping-restrictions/supporting_documents/210119%20Ramping%20restrictions%20for%20active%20power%20output%20amended%20vs%203.0%20for%20public%20consultation.pdf

Joseph Enguehard. 2023. Time Interpret: A Unified Model Interpretability Library for Time Series. https://doi.org/10.48550/arXiv.2306.02968 arXiv:2306.02968 [cs]

ENTSO-E. 2024. ENTSO-E Transparency Platform. https://transparency.entsoe.eu/.

Richard Jumar, Heiko Maass, Benjamin Schäfer, Leonardo Rydin Gorjão, Veit Hagenmeyer, and Ellen Förstner. 2020. Power Grid Frequency Data Base. (April 2020). https://doi.org/10.17605/OSF.IO/BY5HU

Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2021a. Exploring Deterministic Frequency Deviations with Explainable AI. In *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 133–139. https://doi.org/10.1109/SmartGridComm51999.2021.9632335

Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2021b. Revealing Drivers and Risks for Power Grid Frequency Stability with Explainable AI. *Patterns* 2, 11 (Nov. 2021), 100365. https://doi.org/10.1016/j.patter.2021.100365

André Kummerow, Steffen Nicolai, and Peter Bretschneider. 2018. Spatial and Temporal PMU Data Compression for Efficient Data Archiving in Modern Control Centres. In *2018 IEEE International Energy Conference (ENERGYCON)*. 1–6. https://doi.org/10.1109/ENERGYCON.2018.8398809

Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence* 2, 1 (Jan. 2020), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc.

Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. 2018. Explainable Machine-Learning Predictions for the Prevention of Hypoxaemia during Surgery. *Nature Biomedical Engineering* 2, 10 (Oct. 2018), 749–760. https://doi.org/10.1038/s41551-018-0304-0

R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron. 2022. Explainable Artificial Intelligence (XAI) Techniques for Energy and Power Systems: Review, Challenges and Opportunities. *Energy and AI* 9 (Aug. 2022), 100169. https://doi.org/10.1016/j.egyai.2022.100169

Jan Machowski, Zbigniew Lubosny, Janusz W Bialek, and James R Bumby. 2020. *Power System Dynamics: Stability and Control*. John Wiley & Sons.

María Martínez-Barbeito, Damià Gomila, and Pere Colet. 2021. Data Analysis of Frequency Fluctuations in the Balearic Grid Before and After Coal Closure. In *ENERGY 2021, The Eleventh International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies*. 13–18.

Federico Milano, Florian Dörfler, Gabriela Hug, David J Hill, and Gregor Verbič. 2018. Foundations and challenges of low-inertia systems. In *2018 power systems computation conference (PSCC)*. IEEE, 1–25.

Thorbjørn Lund Onsaker, Heidi S. Nygård, Damiá Gomila, Pere Colet, Ralf Mikut, Richard Jumar, Heiko Maass, Uwe Kühnapfel, Veit Hagenmeyer, and Benjamin Schäfer. 2023. Predicting the Power Grid Frequency of European Islands. *Journal of Physics: Complexity* 4, 1 (March 2023), 015012. https://doi.org/10.1088/2632-072X/acbd7f

Sebastian Pütz, Hadeer El Ashhab, Matthias Hertel, Ralf Mikut, Markus Götz, Veit Hagenmeyer, and Benjamin Schäfer. 2024. Feasibility of Forecasting Highly Resolved Power Grid Frequency Utilizing Temporal Fusion Transformers. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '24)*. Association for Computing Machinery, New York, NY, USA, 447–453. https://doi.org/10.1145/3632775.3661963

Sebastian Pütz, Benjamin Schäfer, Dirk Witthaut, and Johannes Kruse. 2022. Revealing Interactions between HVDC Cross-Area Flows and Frequency Stability with Explainable AI. *Energy Informatics* 5, 4 (Dec. 2022), 46. https://doi.org/10.1186/s42162-022-00241-4

Red Eléctrica de España. 2024. Visiona Project Home Page. https://demanda.ree.es/visiona/home.

David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. 2022. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–96.

Leonardo Rydin Gorjão, Richard Jumar, Heiko Maass, Veit Hagenmeyer, G Cigdem Yalcin, Johannes Kruse, Marc Timme, Christian Beck, Dirk Witthaut, and Benjamin Schäfer. 2020. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. *Nature communications* 11, 1 (2020), 6362.

Benjamin Schafer, Marc Timme, and Dirk Witthaut. 2018. Isolating the impact of trading on grid frequency fluctuations. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 1–5.

Lloyd S. Shapley. 1953. A Value for N-Person Games. *Contributions to the Theory of Games* 2, 28 (1953), 307–317.

Zakariea Sharfeddine. 2024. Github repository. https://github.com/KIT-IAI-DRACOS/extreme-frequency-balearic

James W. Taylor and Matthew B. Roberts. 2016. Forecasting Frequency-Corrected Electricity Demand to Support Frequency Control. *IEEE Transactions on Power Systems* 31, 3 (2016), 1925–1932. https://doi.org/10.1109/TPWRS.2015.2444665

Andreas Ulbig, Theodor S. Borsche, and Göran Andersson. 2014. Impact of Low Rotational Inertia on Power System Stability and Operation. *IFAC Proceedings Volumes* 47, 3 (Jan. 2014), 7290–7297. https://doi.org/10.3182/20140824-6-ZA-1003.02615

T. Weissbach and E. Welfonder. 2009. High Frequency Deviations within the European Power System: Origins and Proposals for Improvement. In *2009 IEEE/PES Power Systems Conference and Exposition*. 1–6. https://doi.org/10.1109/PSCE.2009.4840180

Ogun Yurdakul, Fatih Eser, Fikret Sivrikaya, and Sahin Albayrak. 2020. Very Short-Term Power System Frequency Forecasting. *IEEE Access* 8 (2020), 141234–141245. https://doi.org/10.1109/ACCESS.2020.3013165

A METHODS

A.1 Data

We study frequency, generation, and demand data of the Balearic grid. We have obtained the frequency data from an openly accessible database [Jumar et al. 2020; Rydin Gorjão et al. 2020]. This dataset includes measurements from 29 September 2019 through 30 June 2023. Within this period, we excluded multiple data gaps summing to a total of 131 missing days from the data. In addition, we include data from the official website of Red Eléctrica de España (REE), the Spanish transmission system operator (TSO). REE provides detailed information on aggregate demand and electricity production with 5-minute resolution [Red Eléctrica de España 2024]. The data also includes the power transmitted to the Balearic Islands via the HVDC line from mainland Spain and the generation data specified by generation technology. We sum this up as "techno-economic features". For the clustering in Section 2.3, we aggregated Solar, Wind and Other renewables as “Renewables”. Waste, Cogeneration and Auxiliary are grouped in “Others”. Albeit its small percentage coal generation is itself included, due to its importance before the closure of the power plant [Martínez-Barbeito et al. 2021].

Table 2. Percentage of total generation by generation technology rounded to one decimal place.

Generation Technology	Percentage of Total Generation (%)
Combined Cycle	55.9%
HVDC Import	18.9%
Diesel Engines	6.3%
Gas Turbine	5.7%
Waste	4.6%
Coal	4.3%
Solar	3.5%
Cogeneration	0.7%
Auxiliary	0.2%
Wind	0.0%
Other Renewables	0.0%

A.2 Features & Training details

Our dataset for training the model consists of 47 features, including the output of each generation technology, the HVDC link, and their respective ramps over a time range of 5 to 15 minutes. Ramps are defined as the difference between the current power and the one before 5, 10 or 15 minutes. We include actual, scheduled, and forecast demand, with day-ahead data also used for scheduled and forecast demand. We also calculate the difference between actual demand and both forecasted/scheduled demand. Additionally, we include calendar information such as the current time, day, month, day of the year, and day of the month. We use exponentially decaying weighted average and variance with 6 seconds, 1 minute and 5 minutes half-lifes for the initially 1-second-resolved frequency measurements. Moreover, we used the daily average frequency profile of the next timesteps as input. To eliminate irrelevant features, a noise feature has been added to the dataset and each feature exhibiting a similar or lower SHAP feature importance has been removed. Lastly, a flag indicates if the current data point is in within an event.

To predict extreme frequency deviations we used gradient-boosted trees (GBT), specifically XGBoost [Chen and Guestrin 2016].. For hyperparameter optimization, we first performed a random search, followed by a grid search on the reduced search space. We evaluated the performance via 3-fold cross-validation on our training set. For more details refer to [Sharfeddine 2024].

An Algorithm for Modelling Rolling Intrinsic Battery Trading on the Continuous Intraday Market

LEO SEMMELMANN*, Karlsruher Institut für Technologie, Germany
JANNIK DRESSELHAUS, Karlsruher Institut für Technologie, Germany
KIM K. MISKIW, Karlsruher Institut für Technologie, Germany
JAN LUDWIG, Karlsruher Institut für Technologie, Germany
CHRISTOF WEINHARDT, Karlsruher Institut für Technologie, Germany

We introduce a novel rolling intrinsic algorithm to model battery-based trading on the continuous intraday market. Our approach leverages a discretization method with a subsequent step-wise optimization based on continuous price updates and complete uncertainty about future prices. We show the high potential of generating additional profits in the continuous intraday market in comparison to using an ID3-based representation of trading, which is frequently used in literature. Thereby, we realistically model actual trading behavior on the intraday market. We publish our code open-source to contribute to the further analysis of battery storage systems on the intraday market.

CCS Concepts: • **Computing methodologies** → *Model development and analysis.*

Additional Key Words and Phrases: battery trading, intraday market, continuous trading, energy markets

Availability of Data and Material:
The source code is available as an open-source Github repository.

1 INTRODUCTION

Battery storage systems (BESS) are playing an integral role in the clean energy transition to overcome the problem of intermittent solar and wind energy generation [Kittner et al. 2017]. To achieve a high proliferation of battery systems in modern power systems, developing viable business models to operate large-scale systems is integral. However, most studies find that the operation of BESS on wholesale markets is still unprofitable [Baumgarte et al. 2020]. One potential additional business model is "trading arbitrage" [Baumgarte et al. 2020], where the BESS is charged during low-price periods and discharged during high-price periods. Thereby, it also contributes to the smoothing of intermittent generation.

BESS can be operated on different markets, for instance, the auction-based day-ahead market or the continuous intraday market [Baumgarte et al. 2020; Braeuer et al. 2019]. The growing share of volatile generation from renewable energy requires market participants to balance their positions with shorter lead times and, hence, causes more traded volume on continuous intraday markets, as seen in Germany [Koch and Hirth 2019]. Internationally, many countries have implemented a continuous intraday market as well, making this a promising market for BESS participation [Scharff and Amelin 2016].

Studies that analyze BESS operations in the intraday market are typically based on a heavily simplified trading model of BESS, e.g. assuming that only one transaction every 15 minutes is executed [Braeuer et al. 2019], perfect knowledge about future prices is assumed [Cho and Kleit 2015] or potential purely financial shadow transactions, where the trades are later reversed, are neglected [Boukas et al. 2021]. These gaps in research lead to a potential underestimation of the generated profit of battery storage systems in the intraday market.

Hence, we provide an algorithm that extends the BESS trading options on the continuous intraday market. The proposed algorithm is easy to implement, corresponds to the continuous character of the market under uncertain future prices, and enables potential purely financial shadow trades, which lead to profits without requiring the battery to be discharged. The potential of purely financial, battery-backed transactions of the intraday market has not yet been discussed and analyzed in previous research. Overall, we thereby answer the following two research questions:

(1) How can continuous BESS trading on the intraday market be modeled with a rolling intrinsic model that considers uncertain future prices and possible financial countertrades?
(2) How high is the potential financial impact of the proposed rolling intrinsic intraday trading algorithm compared to an ID3 index-based benchmark model?

We answer these research questions with our study, which is structured as follows. First, we introduce related studies about intraday trading algorithms and develop the research gap at hand (Section 2). Then, in Section 3, we depict the German intraday market structure. Section 4 presents our methodology, consisting of the discretization of trades and a rolling optimization model. In addition, we introduce an ID3 benchmark model and relevant metrics. In Section 5, we present our results, which are discussed in Section 6. Finally, in Section 7, we conclude the study and provide an outlook on further research questions.

2 RELATED WORK

Trading in the electricity market is increasingly complex due to ongoing market evolutions. This complexity is evidenced by a spectrum of methods and varied problem scopes that are examined within the ambit of trading electricity studies. In contrast to day-ahead market modeling, literature that deals with the continuous intraday markets is scarce [Glas et al. 2020]. An extensive array of

*Corresponding author: leo.semmelmann@kit.edu

Authors' addresses: Leo Semmelmann, leo.semmelmann@kit.edu, Karlsruher Institut für Technologie, Kaiserstraße 12, Karlsruhe, , 76131, Germany; Jannik Dresselhaus, Karlsruher Institut für Technologie, Kaiserstraße 12, Karlsruhe, , 76131, Germany; Kim K. Miskiw, Karlsruher Institut für Technologie, Kaiserstraße 12, Karlsruhe, , 76131, Germany; Jan Ludwig, Karlsruher Institut für Technologie, Kaiserstraße 12, Karlsruhe, , 76131, Germany; Christof Weinhardt, Karlsruher Institut für Technologie, Kaiserstraße 12, Karlsruhe, , 76131, Germany.

literature does exist that analyzes the price movements and predictability aspects of intraday trading, as documented in sources such as [Hagemann 2013; Kiesel and Paraschiv 2017; Narajewski and Ziel 2020; Pape et al. 2016]. While these studies shed light on market determinants, they refrain from delving into the practical trading strategies in the continuous markets [Kath and Ziel 2020].

The methodologies employed to assess trading strategies in the intraday market vary widely, often necessitating an apriori quantification of underlying price movements. [Braun and Hoffmann 2016] present a case study based on a Mixed Integer Problem modeling approach for a pumped hydro storage system that considers both the continuous intraday and day-ahead markets in Germany. They focus mainly on technical constraints related to the storage system. [Kraft et al. 2023] formulate a Multi-Stage Stochastic Mixed Integer Linear Program for a portfolio consisting of biogas and photovoltaic assets. Their trading strategy considers day-ahead, balancing, and continuous intraday markets, as well as long and short positions exceeding the capacity of the underlying asset. [Ottesen et al. 2018] formulate an intraday trading strategy for an aggregator of demand flexibility. [Finnah et al. 2022] facilitate an Approximate Dynamic Program for the bidding problem of an energy storage in the day-ahead and intraday markets.

While the foregoing studies consider the electricity market interrelations, they simplify the continuity of the intraday market. This means that each intraday market product is tradeable with one auction where a volume-weighted price index is assumed to be realized, like the EPEX SPOT ID3. Such a simplification is suitable for a first approximation but neglects the trading character of the intraday market by reducing the dimension of different trading times [Kath and Ziel 2020]. In real-world scenarios, the continuous intraday market has frequent and sometimes rapid price changes due to various factors such as supply and demand imbalances, especially due to forecasting errors of intermittent RES [Kiesel and Paraschiv 2017]. These intraday price differentials allow market participants to re-optimize existing positions based on new prices, such as modeled by [Boukas et al. 2021], potentially leading to increased revenue for the BESS.

Some studies also explore the concept of continuity in trading strategies. We give an overview of exemplary studies in Table 1. Most notable, [Gönsch and Hassler 2016], [Bertrand and Papavasiliou 2019] and [Boukas et al. 2021] propose a trading algorithm that employs a Markov Decision Process MDP framework to identify optimal policies. [Bertrand and Papavasiliou 2019] and [Boukas et al. 2021] also provide a rolling intrinsic algorithm that simulates continuous trading of energy storage, which considers trading decisions based on the available information at each time step. Both studies show the advantages of an MDP consideration compared to a rolling intrinsic policy. For example, in [Bertrand and Papavasiliou 2019], the rolling intrinsic policy only considers buying power if it can sell it directly. This is because the method maximizes the profit of the current time step and ignores future trading. [Hassler 2017] also model the continuous intraday trading as an MDP, yet they derive applicable decision rules from the found solution for a real-world application. These rules performed well compared to a sophisticated Approximate Dynamic Programming algorithm adapted from literature. [Bertrand and Papavasiliou 2019] and [Hassler 2017] showcase the suitability of an accurate depiction of the continuity of the intraday auctions and bridge the MDP results to the real-world application. We build upon this narrative and propose an algorithm that also considers future trading possibilities in continuous markets while using application-ready optimization. Like this, we can omit the necessary steps to parameterize the price thresholds and optimize the resulting policy functions as in [Bertrand and Papavasiliou 2019].

We base the optimization on [Martin and Otterson 2018], which provides a model for the EPEX SPOT German continuous intraday market based on limited order book LOB data. They discuss possible trading decision rules and necessary preprocessing steps for EPEX SPOT market data. According to them, the proposed market model of the intraday market should be used to simulate intraday trading of flexible market participants to increase realism. The application of this model facilitates a detailed analysis of order execution within continuous intraday markets, an area that is in need of further investigation [Kath and Ziel 2020]. Given its growing importance in the electricity system, we selected a BESS as the underlying asset for our analysis. BESS are particularly compelling assets for intraday trading as they allow for the possibility to procure power from relatively cheap sell bids, store the power, and sell it back to subsequent buy bids that place a greater valuation on the procured power [Bertrand and Papavasiliou 2019].

With this approach, we fill the current research gap regarding application-oriented arbitrage trading using an underlying asset such as BESS in the continuous intraday market. We analyze the advantage of explicitly leveraging the continuity of the market and financial transactions by comparing an ID3 price simplification [Kraft et al. 2023; Ottesen et al. 2018] with our approach. To further deepen the application focus of the proposed algorithm, we consider the most important lithium-ion BESS specifics, such as cyclic warranty conditions. To assess the relevance of these constraints, we conduct a sensitivity analysis that examines the effect of adjusting these constraints on gross profits.

In addition, studies considering trading strategies for intraday auctions are based on data from 2011 to 2016 [Bertrand and Papavasiliou 2019; Boukas et al. 2021; Metz and Saraiva 2018]. Based on this data [Metz and Saraiva 2018] find that even arbitrage profits are not enough to justify the investment cost of BESS. To further accelerate these discussions and the proliferation of BESS we provide the proposed algorithm open-source. Moreover, we explore more recent market data (2022) for the evaluation of the trading strategies. As the market dynamics and prices have changed significantly in recent years, our analysis offers a timely and updated understanding of the current market landscape and a starting point for upcoming profitability discussions.

3 CONTINUOUS INTRADAY MARKET

Even though an intraday market is present in multiple electricity systems around the world [Cho and Kleit 2015] and in Europe mostly harmonized, some differences remain. In the following, we provide an overview of the German market structure and specific lead times. Yet, the proposed approach can be implemented in various continuous intraday markets.

Paper	Assets	Markets	Data	Model
[Bertrand and Papavasiliou 2019]	PHES	CI	LOB, DE/LU (2015, 2016)	MDP
[Boukas et al. 2021]	PHES	CI	LOB, DE/LU (2015)	MDP
[Braun and Hoffmann 2016]	PHES	DA, CI	LOB/Auctions, DE/LU (2015)	MILP
[Metz and Saraiva 2018]	Li-Ion BESS	IA	Auctions, DE/LU (2015)	MILP
[Braun 2016]	PHES	DA, IA	Auctions, DE/LU (2015)	MILP
[Kraft et al. 2023]	Biogas, PV	DA, FR, IA	ID3/Auctions, DE/LU (2019)	Stoch. MILP
[Ottesen et al. 2018]	Flexible Demand	DA, FR, IA	ID3/Auctions, NO (2019)	Stoch. MILP
[Gönsch and Hassler 2016]	Wind, Li-Ion BESS	CI	LOB, DE/LU (2013)	MDP/ADP
[Hassler 2017]	Wind, Li-Ion BESS	CI	LOB, DE/LU (2014)	MDP/ADP
[Löhndorf and Wozabal 2022]	PHES, BESS	DA, CI	LOB, DE/LU (2017)	Stoch. MILP

Abbreviations: PHES - Pumped Hydro Energy Storage, CI - Continuous Intraday, DA - Day-Ahead, IA - Intraday Auction, FR - Frequency Restoration, LOB - Limit Order Book, DE/LU - Germany/Luxembourg, NO - Norway, MDP - Markov Decision Process, MILP - Mixed-Integer Linear Programming, Stoch. MILP - Stochastic Mixed-Integer Linear Programming, ADP - Approximate Dynamic Programming

Table 1. Overview of related literature for trading in the intraday market

3.1 German case study

The main electricity intraday markets in Germany are operated by EPEX SPOT SE. They can be divided into two sub-markets: Intraday auctions of quarter-hourly products and the continuous intraday market [EPEX SPOT SE 2023].

In Germany, the intraday auctions occur at 15:00 CET for the quarter-hourly products of the following day. The higher resolution of 15 minutes allows market participants to adjust their positions more precisely and respond to smaller fluctuations in supply and demand [Baule and Naumann 2021].

In the continuous intraday market, market participants can continuously trade electricity throughout the day until shortly before delivery. This market operates in real-time, enabling participants to react to changing market conditions, unforeseen events, and new information that may impact their trading strategies. The continuous trading starts at 15:00 CET for Germany, with all 24 hours, 48 half hours, and 96 quarter hours of the following day as available products. [EPEX SPOT SE 2023]

Based on the area in which the matched trades are realized, the product lead times differ in the following way [EPEX SPOT SE 2023]:

(1) **Up to 1 hour before delivery:** EU cross-zonal intraday trading based on Single Intraday Coupling (*SIDC*)
(2) **Up to 30 minutes before delivery:** Cross-zonal trading within Germany
(3) **Up to 5 minutes before delivery:** Single delivery area trading (*SDAT*)

The continuous intraday market's structure differs vastly from the other auction-based electricity markets. Market participants have access to a LOB, as visualized in Figure 1), where they can submit their buy and sell orders for available products. The order book represents market participants' collective supply and demand, displaying the prices and volumes at which they are willing to trade. Market participants can adjust their orders quickly, modifying prices or volumes to reflect changing market conditions. This flexibility enables them to respond quickly to new information or events that may impact the supply and demand of the electricity market. By actively managing their orders, participants can re-optimize their positions and benefit from trading opportunities as prices deviate throughout the day [Martin and Otterson 2018].

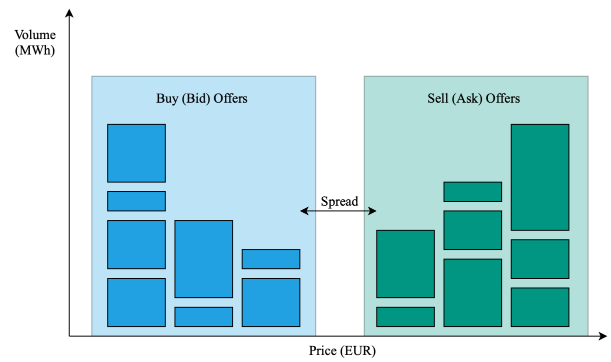


Fig. 1. Example of a limited order book for a specific product at a specific time

The LOB nature of the continuous intraday market means that market participants can observe the existing orders and their corresponding prices and volumes before deciding to execute their own trades. This is a feature not present in auction-based electricity markets.

Operators of electricity storage systems can leverage the continuous intraday market by capitalizing on price differentials and market fluctuations. Since all 24 hours (as well as all 48 half-hours and 96 quarter-hours) are traded simultaneously, there are risk-free arbitrage opportunities during the day, where an operator can buy a specific hourly product at a relatively low price and sell another, later hourly product at a higher price simultaneously [Boukas et al. 2021].

As outlined in this Section, the motivation for the detailed examination of trading models for the continuous intraday market in this study is based on the increasing relevance of the market [EPEX SPOT SE 2022b], as well as the additional revenue potential

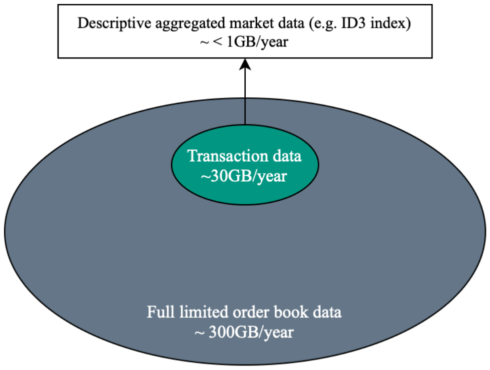


Fig. 2. Data sets for continuous intraday trading at EPEX SPOT (2022)

arising from the LOB format of the market, which results in price fluctuations during the day that flexibility providers can exploit.

3.2 Data structure

Data about past transactions on the continuous intraday market are available in various granularities. Understanding the underlying data structure is essential for building algorithms based on the LOB and reflecting daily market price fluctuations.

In Figure 2, we provide an overview of data sets of different granularity and detail available for the EPEX SPOT intraday market [Martin and Otterson 2018]. EPEX SPOT operates electricity markets in many European countries, running day-ahead and capacity auctions, as well as operating the continuous intraday market. In the following, we introduce the different data sets provided by EPEX SPOT post-trading for its intraday market:

(1) **Aggregated market data**: EPEX SPOT provides free aggregated market information for continuous markets. This data includes minimum and maximum prices for the available contracts, volume-weighted averages, and information about the traded volume for each contract. In addition, indices are defined to provide low-granularity information about the products traded: The ID1 index summarizes price information for the products by calculating a volume-weighted average price of overall transactions executed in the last hour before delivery. Respectively, the ID3 index calculates a volume-weighted average price over all transactions executed in the last three hours before delivery, and the IDFULL index provides a volume-weighted average over all transactions for a specific product. [Martin and Otterson 2018, 3]

(2) **M7 Order book data**: EPEX SPOT offers complete historical order book information for the continuous intraday market as a paid service. The order book offers an in-depth view of market dynamics and can be used to inform trading strategies and conduct detailed market analysis. The order book data provided by EPEX SPOT includes millions of orders processed daily. This includes both the orders that actually resulted in transactions and limit orders that never led to an execution. [Martin and Otterson 2018, 3]

(3) **Transaction data**: Besides full order book data, EPEX SPOT provides a subset of the order book that only includes orders that resulted in transactions (roughly one-tenth the size of full order book data). [Martin and Otterson 2018, 3]

While aggregated market data exhibit less intraday trading volatility due to the averaging over longer time-horizons, both full order book and transaction data could be used for trading simulation with shorter time steps. Full order book data is the most detailed and realistic market representation, capturing the depth and dynamics of trading activity. However, for the purpose of the presented trading simulation, we argue for employing the transaction data set for the following reasons [Rominger et al. 2019]. Transaction data provides a clear lens into the actual market outcomes — the prices and quantities where demand intersects with supply. The complete nature of order book data can also lead to noise in the data set [Martin and Otterson 2018]. Orders, especially those deep in the book and distant from realized prices, might be more speculative than genuine. By using only transaction data, the simulation focuses solely on confirmed, executed trades. While the full order book data is the most thorough representation of trades, it also introduces a significant computational overhead, as shown by the approximate data set sizes in Figure 2. Concentrating on transaction data simplifies the simulation process, eliminating the potential noise without compromising essential insights. While processing and storing the full order book data was not feasible with the resources available for this work, the transaction data subset provides a reasonable compromise between sufficient data granularity for estimating the results of the trading strategies discussed and the resources needed to process and store the data.

For these reasons, we use the transaction data set as the base data set for our rolling intrinsic trading simulation, which is described in the following chapter.

4 METHODOLOGY

In this Section, we provide an overview of our method to model BESS trading on the continuous intraday market based on the transaction data set. In the first step, we discretize the observations from the transaction data. Then, given past trades and new transaction data, we make trading decisions during every discretized time step. We describe our methodology in the following in detail.

4.1 Discretization

Successful trades on the continuous intraday market occur irregularly, as orders are matched continuously. For our approach, the transaction data is hence grouped into discrete intervals. The applied discretization methodology and the following notation are derived from [Kath and Ziel 2020]. Through discretization, the computational complexity of the intraday trading task is reduced, while a realistic representation of continuous orders is still maintained. In comparison, considering every available buy or sell order, at least one simulation step per second would be required. In the case of 5-minute discretization intervals, a second-wise optimization would lead to a 300-fold increase of time steps in the trading optimization. Hence, for the sake of computational efficiency, our study employs a bucket-wise discretization approach. We note that in future studies, also a consideration of all underlying orders could be an interesting benchmark scenario.

The transaction data can be formalized as a set Z. The execution times of the given transactions are defined as a set T such that $T \subseteq [t_{\text{start}}, t_{\text{end}}]$, where $[t_{\text{start}}, t_{\text{end}}]$ marks the trading period to be simulated:

$$Z = \{(y_t, v_t, p_t, d_t) | t \in T, T \subseteq [t_{\text{start}}, t_{\text{end}}]\} \tag{1}$$

The set Z is comprised of the trades at execution time t with the execution price of the trade denoted as $y_t \in \mathbb{R}$, the volume of the trade as $v_t \in \mathbb{R} | v_t > 0$, the product traded (e.g. delivery between 12:00-12:15) as $p_t \in P$, where P is the set of all tradable products, and the direction of the trade as $d_t \in buy, sell$. The set of all tradable products could, for instance, be all 15-minute or 1-hour delivery periods per day, depending on the market at hand.

With t_{start} being the start of the trading period, a bucket function $B(t)$ can be defined given a bucket size b:

$$B(t) = \left\lfloor \frac{t - t_{\text{start}}}{b} \right\rfloor \tag{2}$$

Based on the bucket size b and the trading period $[t_{\text{start}}, t_{\text{end}}]$ discrete time buckets can be defined as O_k:

$$O_k = \{(t, y_t, v_t, p_t, d_t) \mid t \in T, B(t) = k\} \tag{3}$$

This methodology splits the transaction data into equidistant time buckets indexed by k, where $k \in \left[0, \left\lfloor \frac{t_{\text{end}} - t_{\text{min}}}{b} \right\rfloor\right]$.

For every time bucket k (e.g., the first 5-minute time span of the simulation horizon), product p and trade direction d, a volume-weighted average price $\text{VWAP}_{k,p,d}$ can be defined as:

$$\text{VWAP}_{k,p,d} = \frac{\sum_{t \in T_k, p_t = p, d_t = d} y_t \cdot v_t}{\sum_{t \in T_k, p_t = p, d_t = d} v_t} \tag{4}$$

Due to the nature of the transaction data, where every buy trade matches a respective sell trade with the same volume and price, volume-weighted average prices are equal for both trading directions:

$$\text{VWAP}^{\text{buy}}_{k,p} = \text{VWAP}^{\text{sell}}_{k,p} \tag{5}$$

Based on Equation 5, Equation 4 can be simplified to:

$$\text{VWAP}_{k,p} = \frac{\sum_{t \in T_k, p_t = p} y_t \cdot v_t}{\sum_{t \in T_k, p_t = p} v_t} \tag{6}$$

We, therefore, simplify the trading reality in a continuous LOB market where spreads between bid and ask orders, as visualized earlier in Figure 1, must be considered.

For every time bucket k, a volume-weighted average can be defined. With the following equation, we only include time buckets with at least 10 transactions for a given product:

$$\text{VWAP}^{\text{valid}}_k = \left\{\text{VWAP}_{p,k} \mid W(p, O_k) > 10\right\} \tag{7}$$

where $W(p, O_k)$ is the number of transactions in the subset O_k for product p. The constraint $W(p, O_k) > 10$ is introduced for two reasons. First, a product that has been traded more frequently within a given time bucket generally signifies that there's sufficient market interest or liquidity for that product. This is important when calculating a metric like VWAP because, for products with very few trades, the VWAP can be highly sensitive to individual trades. By ensuring a minimum number of transactions for calculating the VWAP, we ensure that the VWAP represents a broad and realistic market consensus for a given product at a specific time bucket. Especially in the hours far away from the delivery time, fewer trades are made [Löhndorf and Wozabal 2022]. Occasional trades might occur at prices that deviate significantly from the norm for various reasons (e.g., input errors, strategic trades, etc.). Such trades can considerably skew metrics like the VWAP, especially if the volume associated with the outlier price is high. A recent extreme example of this can be seen in the transaction data for 18.09.2023 for the DE/LU market area, where the delivery period 12:00 to 12:15 was traded at approx. 190€/MWh before two distinct transactions were executed for a price of 6999.98 EUR/MWh. Prices returned to normal immediately afterward. By only considering products with a higher number of trades, the impact of such outliers on trading simulation results is reduced significantly. By setting $W(p, O_k) > 10$, we ensure a reasonable trade-off between realistic market consensus while still having enough tradable buckets. In addition, we provide a sensitivity analysis of minimum trades per bucket and resulting gross profit in the Appendix.

4.2 Rolling intrinsic trading model

In the following, we describe our rolling intrinsic trading algorithm, based on the discretized transaction data. The algorithm approximates continuous intraday trading by iterating over time buckets and making trading decisions while ensuring constraints are met, as described in Algorithm 1.

Algorithm 1 Battery Trading Simulation

1: **Initialization**: Battery and trading parameters are initialized. These parameter sets are provided as inputs to the simulation.
2: **for** $k = 0$ **to** K **do** ▷ Iterate through each time period, from $k = 0$ to $k = K$
3: Retrieve the set $VWAP^{\text{valid}}_k$ of tradable products in the given time period k.
4: Identify the most profitable combination of potential trades in the period k, given tradable products $VWAP^{\text{valid}}_k$, based on an optimization model
5: Apended these trades to the list of previous trades, serving as inputs for subsequent periods.
6: **end for**
7: **Output**: The set of trades is returned and can be used to calculate metrics such as full load cycles or gross revenues generated by trading.

The rolling intrinsic framework was first presented in the domain of gas storage valuation [Bjerksund et al. 2011; de Jong 2015; Gray and Khandelwal 2004]. The idea of the rolling intrinsic approach is to perform multiple optimization runs in a given frequency over the course of the trading period. This enables the storage operator to re-optimize the positions of previous optimization runs given new price information. The rolling intrinsic model implemented in this study is based on this basic idea and adopted for the use case of an electricity storage operator in the continuous intraday market [Boukas et al. 2021].

A mixed integer optimization model is formulated to find profit-maximizing trades at any point in time (in our case, within a time bucket).

Therefore, we introduce the following variables:

- **Buy Amount ($b_{k,p}$)**: This variable denotes the amount of energy bought in period p within the optimization run in time bucket k. It is measured in megawatt-hours (MWh).
- **Sell Amount ($s_{k,p}$)**: This indicates the quantity of energy sold from the battery in period p within the optimization run in time bucket k, also measured in MWh.
- **State of Charge (SOC_p)**: This variable reflects the battery's state of charge (SOC) at the start of period p, also in MWh.

Several parameters for the battery operations are implemented:

- **Round-trip Efficiency (η)**: This represents the combined efficiency of the charging and discharging processes, as typically used as a metric for life-cycle evaluations of BESS [Hiremath et al. 2015].
- **Battery Capacity (C)**: This is the maximum energy content the battery can store in MWh.
- **C-Rate (r)**: This represents the maximum rate at which the battery can charge or discharge as a percentage of the maximum battery capacity. The C-Rate can be formalized as $r = \frac{I}{C}$, with I being the charge–discharge current of the system [Collath et al. 2022, 5]. The choice of r is a critical decision, as high charging and discharging rates lead to increased cyclic aging [Naumann et al. 2020].
- **Total Cycles allowed per day (N)**: This indicates the total number of charging cycles the battery undergoes in a day. This parameter will simulate warranty conditions, which often include a maximum number of cycles per time period [Fares and Webber 2018; Kelly and Leahy 2020; Mongird et al. 2019; Xu 2022].

A trading algorithm can, therefore, be formulated as follows for a specific time bucket k. The optimization goal is maximizing the profits at each discretized bucket k. The profits are constituted by the sum over all tradable products p of sold quantities $s_{k,p}$, deducted by bought quantities $b_{k,p}$ at the respective prices $\mathrm{VWAP}_{k,p}$:

$$\max \sum_p \left(s_{k,p} \cdot \mathrm{VWAP}_{k,p} - b_{k,p} \cdot \mathrm{VWAP}_{k,p} \right) \tag{8}$$

The tradable product p can also be interpreted as the time step when the battery has to be charged or discharged. Various constraints have to be accommodated for realistically modeling continuous trading on the intraday market:

$$\mathrm{SOC}_1 = 0 \tag{9}$$

$$\mathrm{SOC}_p = \mathrm{SOC}_{k,p-1} + b_{k,p-1} \cdot \sqrt{\eta} - s_{k,p-1} \cdot \frac{1}{\sqrt{\eta}}, \quad \forall p > 1 \tag{10}$$

$$\mathrm{SOC}_p \leq C, \quad \forall p \tag{11}$$

$$b_{k,p} \leq C \cdot r, \quad \forall p \tag{12}$$

$$s_{k,p} \leq C \cdot r, \quad \forall p \tag{13}$$

$$s_{k,p} \cdot \frac{1}{\sqrt{\eta}} \leq \mathrm{SOC}_p, \quad \forall p \tag{14}$$

$$\sum_p b_{k,p} \cdot \sqrt{\eta} \leq N \times C \tag{15}$$

$$b_{k,p}, s_{k,p}, \mathrm{SOC}_p \geq 0, \quad \forall p \tag{16}$$

In Equation 9, the initial SOC of the battery is set at 0, Equation 10 controls the $SOC_{k,p}$ based on buy and sell decisions, Equations 11 - 13 enforce that the battery is operated within technical constraints. Equation 14 and 15 consider efficiency losses, Equation 16 enforces the non-negativity of $SOC_{k,p}$, the buy and sell quantities $b_{k,p}$ and $s_{k,p}$.

This algorithm returns trades for any given list of products and their respective prices ($VWAP_p$) while ensuring that technical restrictions are met. Because of the continuous trading format, this base algorithm has to be extended to account for trades made in previous time buckets.

This is achieved by introducing three more variables into the optimization problem:

- **Net Buy Amount ($\tilde{b}_p$)**: This is the quantity that is bought for one product after taking into account all previous trades as well as the trades within the current optimization run.
- **Net Sell amount ($\tilde{s}_p$)**: This is the quantity that is sold for one product after taking into account all previous trades as well as the trades within the current optimization run.
- **Charge Sign (λ_p)**: This is a binary variable, which is 1 if the battery is charging for a specific time period (or product) and 0 if it is discharging.

Following this modeling approach, the constraints affecting maximum net buy and sell amounts have to be adjusted based on the binary charge sign λ_p:

$$\tilde{b}_{k,p} \leq C \cdot r \times \lambda_p, \forall p \tag{17}$$

$$\tilde{s}_{k,p} \leq C \cdot r \times (1 - \lambda_p), \forall p \tag{18}$$

In addition, the following constraint is added:

$$\tilde{b}_{k,p} \times \lambda_p - \tilde{s}_{k,p} \times (1 - \lambda_p) = b_{k,p} + \sum_{i=1}^{k-1} b_{i,p} - s_{k,p} + \sum_{i=1}^{k-1} s_{i,p}, \forall p \tag{19}$$

While Equation 17 and Equation 18 are used to determine the current charge sign (whether the battery is charging or discharging), Equation 19 calculates the actual net charging or discharging amounts for each product (or delivery period) p. The differentiation between charging and discharging amounts is necessary to account for inefficiencies in the battery correctly: When calculating the battery SOC, net buy amounts have to be multiplied by the efficiency factor $\sqrt{\eta}$ (as not all energy bought makes it to the battery) and net sell amounts have to be divided by $\sqrt{\eta}$ (as more energy has to be extracted from the battery than was sold).

Overall, we replace in our optimization model the sell and buy amounts with the now introduced net amounts to consider previous trades:

$$\tilde{s}_{k,p} = s_{k,p}\tilde{b}_{k,p} = b_{k,p} \tag{20}$$

The introduced model now optimizes the positions without relying on forecast information by simultaneously capitalizing on the possibility of arbitrage between various contracts. It is important to highlight that the rolling intrinsic algorithm allows for repositioning. Unlike traditional strategies for auction-based markets, which can only hold a position based on an initial decision, this algorithm re-evaluates and might buy or sell back previous positions throughout the day. This dynamic approach ensures that the battery operates at its optimal state, taking advantage of the volatility in the intraday market. Furthermore, an important aspect of this approach is the occurrence of only financial trades. These are trades executed by the algorithm that do not necessarily lead to actual battery discharging or charging. Instead, they provide additional revenue potential by adjusting the battery's market position and capitalizing on price fluctuations throughout the trading period.

4.3 Benchmark: ID3

We benchmark our rolling intrinsic algorithm against a frequently used approach in research, that utilizes the ID3 price index to approximate continuous intraday trading [Kraft et al. 2023; Ottesen et al. 2018]. With this approach, studies, for example, attempt to depict the general option of repositioning on the intraday market within a multi-market setting. While aggregating prices on a higher level, the index gives a descriptive overview of the prices of the continuous markets and is therefore included as a benchmark for the evaluation. With this, we can assess the increased monetary benefit of the rolling intrinsic algorithm on the continuous intraday results in comparison to ID3 data.

In contrast to the previously described trading simulation with transaction data, the ID3 simulation is solved using a single optimization problem under perfect foresight, making trade, charge and discharge decisions based on the underlying ID3 price index. In contrast, the previously introduced rolling intrinsic algorithm has no perfect foresight, thereby making actual operations on the continuous market more realistic. The overall optimization goal in the benchmark approach is to maximize the difference between sold quantities s_p at price $\mathrm{VWAP}_p^{\mathrm{ID3}}$ and the bought quantities b_p at price $\mathrm{VWAP}_p^{\mathrm{ID3}}$ for each time step p:

$$\text{maximize} \quad \sum_{p} \left(s_p \cdot \mathrm{VWAP}_p^{\mathrm{ID3}} - b_p \cdot \mathrm{VWAP}_p^{\mathrm{ID3}} \right) \tag{21}$$

The optimization is subject to the initial technical constraints from Equation 10 to 16.

4.4 Performance metrics

We compare the rolling intrinsic and ID3-based simulation approach based on two metrics introduced in the following: gross trading profits and profits per cycle.

Gross Trading Profit (π)

Gross profit represents the total monetary gain from sales of energy while accounting for the direct costs of obtaining the energy but not accounting any other costs. Specifically, trading costs and wear on the battery are not considered.

The metric offers a standardized way of comparing various trading algorithms, focusing on their core performance without the complexities of individual cost structures. Given the adherence to a given number of full load cycles per year, battery wear concerns are primarily mitigated by warranty conditions [Fares and Webber 2018; Kelly and Leahy 2020].

The gross trading profit for the static priced ID3 benchmark model is given by the subtraction of all sold from the bought quantities multiplied by their prices:

$$\pi = \sum_{p \in P} \left(s_p \cdot \mathrm{VWAP}_p^{\mathrm{ID3}} - b_p \cdot \mathrm{VWAP}_p^{\mathrm{ID3}} \right) \tag{22}$$

The gross trading profit for the rolling intrinsic model is given by the sum of gross quantities bought and sold over all time buckets in the trading period weighted by their respective $VWAP$:

$$\pi = \sum_{k \in K} \sum_{p \in P_k} \left(s_p \cdot \mathrm{VWAP}_{k,p} - b_{k,p} \cdot \mathrm{VWAP}_{k,p} \right) \tag{23}$$

Profit Per Cycle (PPC)

While our study neglects a detailed degradation analysis of the simulated storage system, we assume that a certain number of load cycles should not be exceeded due to warranty conditions [Ahmadi et al. 2017; Viswanathan et al. 2022]. By introducing the Profit per Cycle (PPC) metric, a more nuanced understanding of a trading strategy's efficiency by relating the generated cash flow to the number of full load cycles is enabled.

As lithium-ion BESS are examined in this study, this metric is introduced to ensure that revenues are not simply examined on an absolute scale but are contextualized against the approximate wear they impose on the BESS.

The PPC can be formulated as:

$$PPC = \frac{\pi}{\mathrm{N}} \tag{24}$$

where N is the number of full load cycles used by the trading strategies. In our case, for the static priced ID3 benchmark model this can be calculated by:

$$N = \sum_{p \in P} b_p \times \sqrt{\eta} \tag{25}$$

In the case of the rolling intrinsic model the cycles used can be calculated by:

$$N = \sum_{k \in K} \sum_{p \in P_k} \tilde{b}_{k,p} \times \sqrt{\eta} \tag{26}$$

We apply these metrics in the following on a year-long case study.

5 RESULTS

This Section applies the proposed rolling intrinsic trading simulation algorithm to a year-long case study. We then compare the results with an ID3 index-based simulation frequently used in energy market trading studies [Kraft et al. 2023]. In the following, the rolling intrinsic models are abbreviated with $RI\ CONT$, while the ID3 models are denoted as $PF\ ID3$. For our case study, we use transaction data from the German continuous intraday market from 2022 [EPEX SPOT SE 2022a] for the quarterly hour products ($15m$) and hourly products ($60m$). Although our study utilizes German market data, the approach could easily be implemented on other international intraday electricity markets [Scharff and Amelin 2016].

We assume a battery capacity (C) of 1MWh, based on current large-scale projects [Petrova 2023], a round-trip efficiency (η) of 0.86 [Viswanathan et al. 2022] and one allowed equivalent full cycle per day, leading to 365 allowed warranty cycles (N) per year, based on [Komorowska et al. 2022]. The bucket size (b) is set at 15 minutes.

In Figure 3, we illustrate the rolling intrinsic simulation over one day for the hourly products based on the volume-weighted average prices ($\text{VWAP}_{k,p}$). Over the day, some hourly products are traded multiple times, thereby realizing financial revenues that are not necessarily associated with battery charging and discharging cycles. For example on the shown day the product for hour 13 p.m. to 14 p.m. is traded 32 times in total. This is a result of the underlying uncertainty in the price movement that we are considering. The reoptimisation in the rolling intrinsic optimisation facilitates new price movements as they occur. Ultimately, the battery is charged at lower prices mid-day and discharged during price peaks in the evening.

An overview over the behavior of the benchmark optimization on ID3 for an exemplary day is presented in Figure 4. Here, only one transaction per product is performed. Since no repositioning in the market is modeled here, all trades result in charging and discharging events of the battery. For the chosen exemplary day, this results in two full charging cycles. The $PFID3$ simulation facilitates the two largest price spreads and entirely charges/discharges the battery twice. Although we also restrict the yearly battery cycles in the $PFID3$ optimization, two cycles can be observed on the depicted day. This is due to the perfect foresight optimization over the whole year, thereby not operating the BESS on another, less profitable day. Contrary, the $RICONT$ trading results in only one full charging cycle. These differences can also be analysed with the introduced metrics.

Model	π (€)	PPC (Eur / Cycle)
RI CONT (15m)	**111631.86**	**305.84**
RI CONT (60m)	70749.62	193.83
PF ID3 (15m)	80244.57	219.85
PF ID3 (60m)	57731.58	158.17

Table 2. Performance metrics of trading model and full-foresight benchmarks

We compare the metrics of our battery-based continuous intraday trading simulation in Table 2. In the difference of $PFID3$ and $RICONT$, we can observe that active repositioning in the continuous intraday market increases the profit potential of the battery by 39.2%. For both simulation models, the $15min$ products yield remarkably higher profits than the $60min$ products. This might be because the 15-minute products are used to adjust the hourly power outputs in the day-ahead markets. The time interval leads during a ramping up of production to an underproduction of power in the first two-quarter hours and an overproduction in the second two-quarter hours. This systematic repositioning of quarter-hour products leads to price fluctuations that can be facilitated by storage. Furthermore, we can see that the higher profits of the rolling intrinsic simulation $RICONT$ are not caused by more cycles. Contrary even the profit per cycle (PPC) is higher for both product durations than for the ID3 counterpart. We see the ability for financial repositioning and the utilization of price fluctuations over the day as the main reason behind the higher profits exhibited by the $RICONT$ model.

In Figure 5, the resulting cumulative profits over the two investigated product types ($15min$ and $60min$) and two models are depicted. The illustration shows that the steepest increase in cumulative profits can be observed over all methods during summer months, utilizing heavy price fluctuations caused for example by high fluctuating photovoltaic infeed.

In the Appendix, we run a sensitivity analysis for the most important parameters of the rolling intrinsic simulation to assess their impact on the presented results. We find that the smaller the bucket size, the higher the profits from the simulation. One reason for this is that smaller bucket sizes lead to more daily trading steps, thereby more trading and repositioning opportunities. Especially for the $15min$ rolling intrinsic algorithm, a higher number of allowed equivalent full cycles per year leads to substantially increased profits.

6 DISCUSSION

The generated trading trajectory by the proposed algorithm generates notably higher profits than simplifying the continuity with the ID3 prices despite assuming full uncertainty. First, we show that a significant battery profit potential lies in purely financial trades that are later resolved through counter trades. Thereby, it underlines the drawback of a simplification of the time dimension ID3 index-based simulation since repositioning and buying positions back is not possible, as well as a consideration of frequent trades, which can be represented through bucket discretization. Another benefit of our approach is that no perfect foresight of prices is assumed. Instead, trading decisions are based on continuous price updates, which could also be observed in a real-world scenario. As shown in the results, the re-optimisation in every step of the rolling intrinsic algorithm leads to a high amount of trades in total. We note that for a detailed profitability cost of continuous BESS trading, transaction costs and spreads should be considered - an analysis not often made in literature [Kath and Ziel 2020]. [Narajewski and Ziel 2022] assume transactions costs in the day-ahead market of 0.05 €/MWh and in the intraday start auction of 0.10 €/MWh. Such values are easily implementable in our algorithm with a threshold of

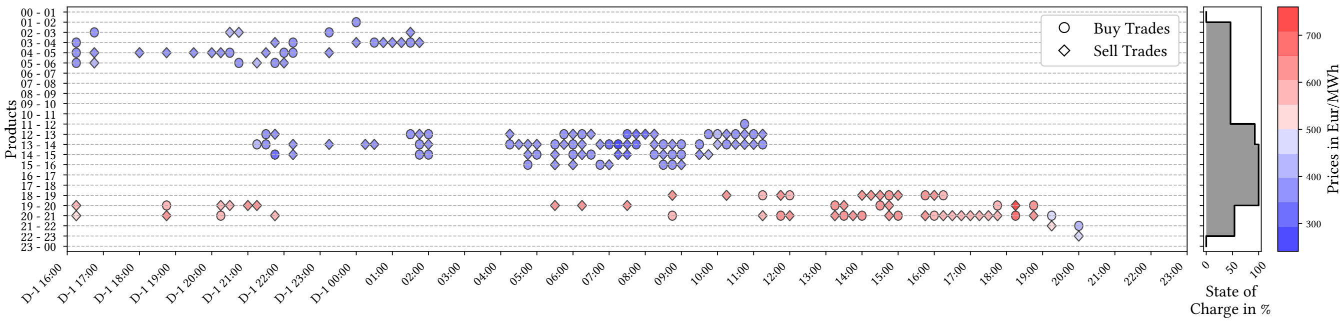


Fig. 3. Trade executions per product of the rolling intrinsic algorithm for 2022-09-07

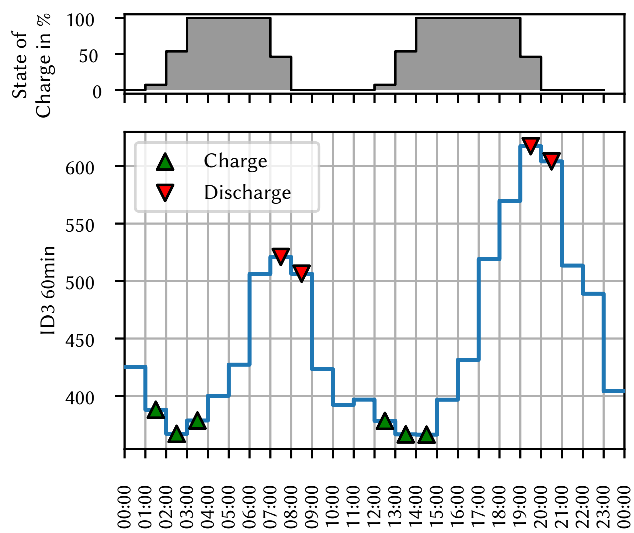


Fig. 4. Overview over the ID3 optimization results for 2022-09-07

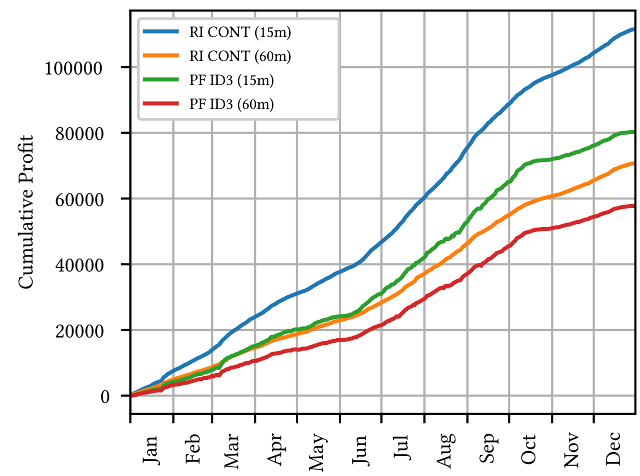


Fig. 5. Cumulative profits by trading model against full-foresight benchmarks for 2022

minimal profit made per trade, which consequently should be above the assumed transaction costs. We plan to implement a detailed profitability analysis in further studies.

Our rolling intrinsic algorithm depicts the market's reality more realistically since trades can be conducted continuously. Hence, an ID3 index-based simulation underestimated the real profit potential of battery storage systems active on the intraday market. For a profitability assessment, the continuity of the intraday markets should, therefore, be considered.

We note that our approach has a couple of limitations. First, the consideration of transaction data allows a realistic simulation based on continuously conducted trades, while keeping the computational complexity in an acceptable range. However, the whole order book data should be considered for a completely reliable analysis of profits. In addition, our algorithm assumes that a generated VWAP with 10+ trades ensures that our trade at the same price level would also be matched. Considering the low liquidity of products, even with 10+ trades and our own market impacts, this assumption might not hold in real-world settings. Especially for products with a high lead time as the price impact increases there significantly [Löhndorf and Wozabal 2022]. Second, our study only considers one single market. Previous studies have shown that considering sequential bidding on multiple markets, for instance, the reserve market, day-ahead, and intraday market, could further increase battery profitability [Miskiw et al. 2023]. We publish our source code open-source to enable a more in-depth analysis of battery systems active on the continuous intraday market in future studies, such as implementing a multi-market bidding process. Fourth, we have only benchmarked our approach against a simple ID3 simulation. Further studies should compare the algorithm against alternative methods, e.g. a Markov Decision Process. We also note that our study does not offer a detailed discussion of battery profitability on the intraday market, since we focus on the introduction of the underlying algorithm. However, future studies should employ the proposed rolling intrinsic intraday trading simulation to analyze battery-based trading profits on European intraday markets realistically.

7 CONCLUSION

Our study introduces a novel rolling intrinsic trading algorithm for battery-based trading on the continuous intraday market, leveraging a discretization method with subsequent step-wise optimization. The rolling intrinsic algorithm corresponds to the reality of the intraday market, with continuous price updates in the order book, and the thereby arising trading opportunities. In particular, the rolling intrinsic trading algorithm enables buying counterpositions, thereby realizing profits without the need for battery utilization. We show that the frequently used ID3 index-based simulation approach underestimates the profit potential of batteries active on the intraday market. We publish our algorithm open-source to enable a thorough analysis of battery trading on the intraday market in future studies. Future studies should consider the rolling intrinsic intraday market trading algorithm in combination with other markets in a multi-market bidding problem, to underline the profit potential of battery storage systems active on the power market, thereby contributing to the clean energy transition. Furthermore, a detailed analysis of BESS intraday trading profitability, considering BESS and trading costs, spreads and market access fees, has to be conducted in further studies.

REFERENCES

Leila Ahmadi, Steven B Young, Michael Fowler, Roydon A Fraser, and Mohammad Ahmadi Achachlouei. 2017. A cascaded life cycle: reuse of electric vehicle lithium-ion battery packs in energy storage systems. *The International Journal of Life Cycle Assessment* 22 (2017), 111–124.

Rainer Baule and Michael Naumann. 2021. Volatility and Dispersion of Hourly Electricity Contracts on the German Continuous Intraday Market. *Energies* 14, 22 (2021). https://doi.org/10.3390/en14227531

Felix Baumgarte, Gunther Glenk, and Alexander Rieger. 2020. Business models and profitability of energy storage. *IScience* 23, 10 (2020).

Gilles Bertrand and Anthony Papavasiliou. 2019. Adaptive trading in continuous intraday electricity markets for a storage unit. *IEEE Transactions on Power Systems* 35, 3 (2019), 2339–2350.

Petter Bjerksund, Gunnar Stensland, and Frank Vagstad. 2011. Gas storage valuation: Price modelling v. optimization methods. *The Energy Journal* 32, 1 (2011).

Ioannis Boukas, Damien Ernst, Thibaut Théate, Adrien Bolland, Alexandre Huynen, Martin Buchwald, Christelle Wynants, and Bertrand Cornélusse. 2021. A deep reinforcement learning framework for continuous intraday market bidding. *Machine Learning* 110 (2021), 2335–2387.

Fritz Braeuer, Julian Rominger, Russell McKenna, and Wolf Fichtner. 2019. Battery storage systems: An economic model-based analysis of parallel revenue streams and general implications for industry. *Applied Energy* 239 (2019), 1424–1440.

Sebastian Braun. 2016. Hydropower storage optimization considering spot and intraday auction market. *Energy Procedia* 87 (2016), 36–44.

Sebastian Braun and Rainer Hoffmann. 2016. Intraday optimization of pumped hydro power plants in the german electricity market. *Energy Procedia* 87 (2016), 45–52.

Joohyun Cho and Andrew N Kleit. 2015. Energy storage systems in energy and ancillary markets: A backwards induction approach. *Applied Energy* 147 (2015), 176–183.

Nils Collath, Benedikt Tepe, Stefan Englberger, Andreas Jossen, and Holger Hesse. 2022. Aging aware operation of lithium-ion battery energy storage systems: A review. *Journal of Energy Storage* 55 (2022), 105634. https://doi.org/10.1016/j.est.2022.105634

Cyriel de Jong. 2015. Gas storage valuation and optimization. *Journal of Natural Gas Science and Engineering* 24 (2015), 365–378. https://doi.org/10.1016/j.jngse.2015.03.029

EPEX SPOT SE. 2022a. EPEX SPOT Annual Market Review 2021. https://www.epexspot.com/sites/default/files/download_center_files/2022-01-19_EPEX%20SPOT_Annual%20Press%20Release-2021_final_clean.pdf Accessed 18.07.2023.

EPEX SPOT SE. 2022b. New record volume traded on EPEX SPOT in 2020. https://www.epexspot.com/en/news/new-record-volume-traded-epex-spot-2020

EPEX SPOT SE. 2023. Trading Products. https://www.epexspot.com/en/tradingproducts Accessed 18.07.2023.

Robert L Fares and Michael E Webber. 2018. What are the tradeoffs between battery energy storage cycle life and calendar life in the energy arbitrage application? *Journal of Energy Storage* 16 (2018), 37–45.

Benedikt Finnah, Jochen Gönsch, and Florian Ziel. 2022. Integrated day-ahead and intraday self-schedule bidding for energy storage systems using approximate dynamic programming. *European Journal of Operational Research* 301, 2 (2022), 726–746. https://doi.org/10.1016/j.ejor.2021.11.010

Silke Glas, Rüdiger Kiesel, Sven Kolkmann, Marcel Kremer, Nikolaus Graf Von Luckner, Lars Ostmeier, Karsten Urban, and Christoph Weber. 2020. Intraday renewable electricity trading: advanced modeling and numerical optimal control. *Journal of Mathematics in Industry* 10, 1 (Dec. 2020), 3. https://doi.org/10.1186/s13362-020-0071-x

Jochen Gönsch and Michael Hassler. 2016. Sell or store? An ADP approach to marketing renewable energy. *OR spectrum* 38, 3 (2016), 633–660.

Josh Gray and Pankaj Khandelwal. 2004. Towards a realistic gas storage model. *Commodities Now* 7, 2 (2004), 1–4.

Simon Hagemann. 2013. Price Determinants in the German Intraday Market for Electricity: An Empirical Analysis. *SSRN Electronic Journal* (2013). https://doi.org/10.2139/ssrn.2352854

Michael Hassler. 2017. Heuristic decision rules for short-term trading of renewable energy with co-located energy storage. *Computers & Operations Research* 83 (2017), 199–213.

Mitavachan Hiremath, Karen Derendorf, and Thomas Vogt. 2015. Comparative life cycle assessment of battery storage systems for stationary applications. *Environmental science & technology* 49, 8 (2015), 4825–4833.

Christopher Kath and Florian Ziel. 2020. Optimal order execution in intraday markets: Minimizing costs in trade trajectories. *arXiv preprint arXiv:2009.07892* (2020).

Joseph J Kelly and Paul G Leahy. 2020. Optimal investment timing and sizing for battery energy storage systems. *Journal of Energy Storage* 28 (2020), 101272.

Rüdiger Kiesel and Florentina Paraschiv. 2017. Econometric analysis of 15-minute intraday electricity prices. *Energy Economics* 64 (2017), 77–90.

Noah Kittner, Felix Lill, and Daniel M Kammen. 2017. Energy storage deployment and innovation for the clean energy transition. *Nature Energy* 2, 9 (2017), 1–6.

Christopher Koch and Lion Hirth. 2019. Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system. *Renewable and Sustainable Energy Reviews* 113 (2019), 109275. https://doi.org/10.1016/j.rser.2019.109275

Aleksandra Komorowska, Piotr Olczak, Emil Hanc, and Jacek Kamiński. 2022. An analysis of the competitiveness of hydrogen storage and Li-ion batteries based on price arbitrage in the day-ahead market. *International Journal of Hydrogen Energy* 47, 66 (2022), 28556–28572.

Emil Kraft, Marianna Russo, Dogan Keles, and Valentin Bertsch. 2023. Stochastic optimization of trading strategies in sequential electricity markets. *European Journal of Operational Research* 308, 1 (2023), 400–421.

Nils Löhndorf and David Wozabal. 2022. The value of coordination in multi-market bidding of grid energy storage. *Operations Research* 71, 1 (2022). https://doi.org/10.1287/opre.2021.2247

Henry Martin and Scott Otterson. 2018. German intraday electricity market analysis and modeling based on the limit order book. In *2018 15th international conference on the European energy market (EEM)*. IEEE, 1–6.

Dennis Metz and João Tomé Saraiva. 2018. Use of battery storage systems for price arbitrage operations in the 15-and 60-min German intraday markets. *Electric Power Systems Research* 160 (2018), 27–36.

Kim K. Miskiw, Emil Kraft, and Stein-Erik Fleten. 2023. Coordinated bidding in sequential electricity markets: Effects of price-making. *Under Review in Energy Economics* (2023). https://doi.org/10.2139/ssrn.4509211

Kendall Mongird, Vilayanur V Viswanathan, Patrick J Balducci, Md Jan E Alam, Vanshika Fotedar, V S Koritarov, and Boualem Hadjerioua. 2019. *Energy storage technology and cost characterization report.* Technical Report. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

Michał Narajewski and Florian Ziel. 2020. Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets* 19 (Sept. 2020), 100107. https://doi.org/10.1016/j.jcomm.2019.100107

Michał Narajewski and Florian Ziel. 2022. Optimal bidding in hourly and quarter-hourly electricity price auctions: trading large volumes of power with market impact and transaction costs. *Energy Economics* 110 (June 2022), 105974. https://doi.org/10.1016/j.eneco.2022.105974 arXiv:2104.14204 [q-fin, stat].

Maik Naumann, Franz B. Spingler, and Andreas Jossen. 2020. Analysis and modeling of cycle aging of a commercial LiFePO4/graphite cell. *Journal of Power Sources* 451 (2020), 227666. https://doi.org/10.1016/j.jpowsour.2019.227666

Stig Ottesen, Asgeir Tomasgard, and Stein-Erik Fleten. 2018. Multi market bidding strategies for demand side flexibility aggregators in electricity markets. *Energy* 149 (2018), 120–134. https://doi.org/10.1016/j.energy.2018.01.187

Christian Pape, Simon Hagemann, and Christoph Weber. 2016. Are fundamentals enough? Explaining price variations in the German day-ahead and intraday power market. *Energy Economics* 54 (2016), 376–387.

Veselina Petrova. 2023. Carlton Power Wins Permit for 1,040-MW/2,080 mwh battery in England. https://renewablesnow.com/news/carlton-power-wins-permit-for-1040-mw2080-mwh-battery-in-england-829157/ Accessed 17.06.2023.

Julian Rominger, Manuel Losch, Sebastian Steuer, Katrin Köper, and Hartmut Schmeck. 2019. Analysis of the German Continuous Intraday Market and the Revenue Potential for Flexibility Options. In *2019 16th International Conference on the European Energy Market (EEM)*. IEEE, 1–6.

Richard Scharff and Mikael Amelin. 2016. Trading behaviour on the continuous intraday market Elbas. *Energy Policy* 88 (2016), 544–557.

Vilayanur Viswanathan, Kendall Mongird, Ryan Franks, Xiaolin Li, Vincent Sprenkle, and R Baxter. 2022. 2022 grid energy storage technology cost and performance assessment. *Energy* 2022 (2022).

Bolun Xu. 2022. The role of modeling battery degradation in bulk power system optimizations. *MRS Energy & Sustainability* 9, 2 (2022), 198–211.

APPENDIX

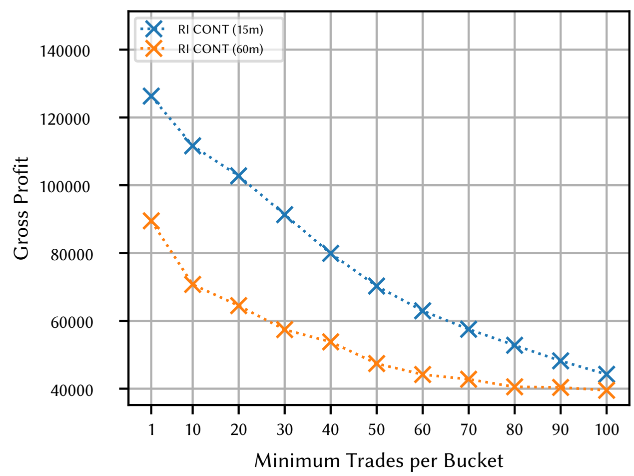


Fig. 6. Sensitivity analysis: Minimum Trades per Bucket (2022, DE/LU, 0.86% RTE, 365 yearly cycles, 0.5C C-Rate, 15m Bucket Size for RI)

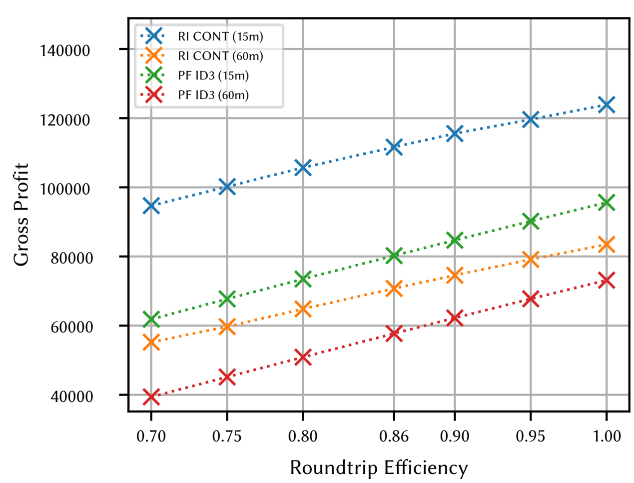


Fig. 7. Sensitivity analysis: Roundtrip efficiency (2022, DE/LU, 365 yearly cycles, 0.5C C-Rate, 15m Bucket Size for RI)

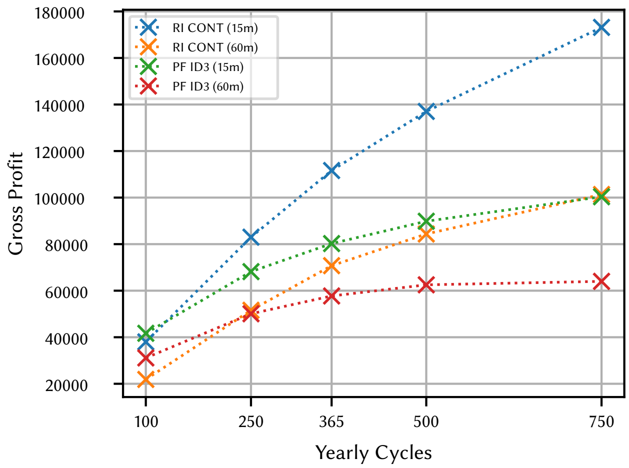


Fig. 8. Sensitivity analysis: Yearly cycles (2022, DE/LU, 86% RTE, 0.5C C-Rate, 15m Bucket Size for RI)

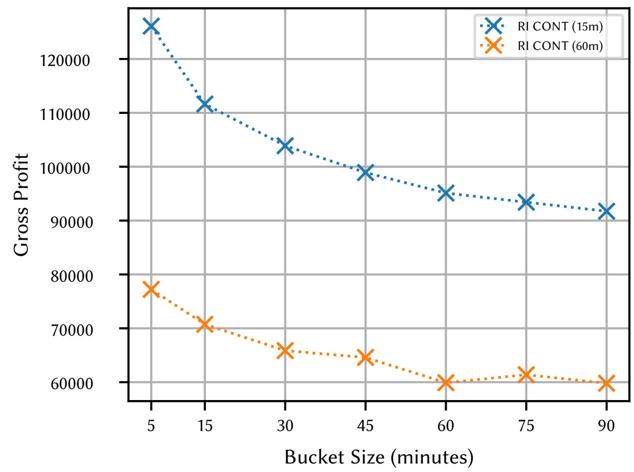


Fig. 10. Sensitivity analysis: Bucket size (2022, DE/LU, 365 yearly cycles, 86% RTE, 0.5C C-Rate)

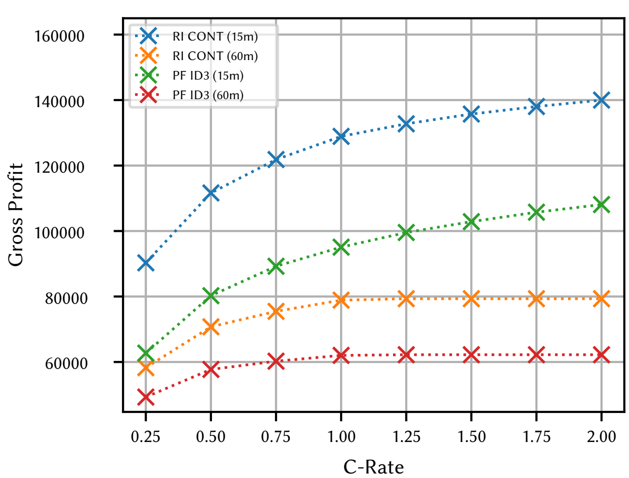


Fig. 9. Sensitivity analysis: C-Rate (2022, DE/LU, 365 yearly cycles, 86% RTE, 15m Bucket Size for RI)

Probabilistic energy forecasting through quantile regression in reproducing kernel Hilbert spaces

LUCA PERNIGO, Euler Institute, USI, Switzerland
ROHAN SEN, Euler Institute, USI, Switzerland
DAVIDE BAROLI, Euler Institute, USI, Switzerland

Accurate energy demand forecasting is crucial for sustainable and resilient energy development. To meet the Net Zero Representative Concentration Pathways (RCP) 4.5 scenario in the DACH countries, increased renewable energy production, energy storage, and reduced commercial building consumption are needed. This scenario's success depends on hydroelectric capacity and climatic factors. Informed decisions require quantifying uncertainty in forecasts. This study explores a nonparametric method based on *reproducing kernel Hilbert spaces (RKHS)*, known as kernel quantile regression, for energy prediction. Our experiments demonstrate its reliability and sharpness, and we benchmark it against state-of-the-art methods in load and price forecasting for the DACH region. We offer our implementation in conjunction with additional scripts to ensure the reproducibility of our research.

CCS Concepts: • **Mathematics of computing → Nonparametric statistics**; • **Computing methodologies → Feature selection**; • **Software and its engineering → Software libraries and repositories**.

Additional Key Words and Phrases: Kernel method, quantile regression, energy forecast, probabilistic forecast, climate, GEFCom, SP2050, DACH

1 INTRODUCTION

Climate shock and the penetration of renewable energy sources are pivotal issues in the modern energy system, particularly in the DACH region (comprising Germany, Austria, and Switzerland). Recently, the Swiss Federal Office of Energy has published a comprehensive analysis to assess how to secure and produce a cost-efficient energy supply in Swiss Energy Strategy 2050 and Energy Perspectives 2050+. Recently, Switzerland also adopted the long-term goal of climate neutrality, aiming to decrease its energy-building consumption and deploy more renewable energy technologies. In addition to being essential to mitigate climate change, the performance of renewable energy sources and the demand for building energy depend on weather data and energy storage.

Within the SURE SWEET energy initiative, supported by the SFOE, the development of future sustainable and robust systems is corroborated by techno-economic models that predict long-term scenarios and pathways, which are resilient to climate shocks [Panos et al. 2023]. These models require a large amount of technical and economic data and their quality influences the reliability of the results, for example, bottom-up techno-economics model [Kannan 2018], which provides hourly prediction, EXPANSE [Trutnevyte 2013], building stock model [Nägeli et al. 2020], macro-economics GEM-3M and life cycle assessment [Luh et al. 2023]. The projection of these models is affected by weather data. For example, the Swiss building stock model designs decarbonisation pathways for different buildings' archetypes, whose isolation and heating performance vary with external temperature. As a result, the prediction of hourly loads by transmission system operators is influenced by the variability and uncertainty of climate factors and is dependent on many parameters of the techno-economics models.

In particular, the pathways predicted by SURE models to achieve the Swiss net-zero scenario in conjunction with the RCP pathway 4.5 address the primary issue of reliable energy supply. This is essential because it requires an increase in renewable energy sources to meet annual net electricity demand of 80–100 TWh by 2050, compared to the current 60 TWh SFOE 2022b. One of the factors of such an increase in electricity due to sustainable mobility is up to 22 TWh [Kannan et al. 2022]. To guarantee uninterrupted energy supply, even under extreme weather conditions [Ho-Tran and Fiedler 2024], in SFOE 2022a various scenarios have been examined: dependence on importing electricity from the European market and expansion of technologies that can provide or save electricity in winter, for example wind, alpine photovoltaic, seasonal heat storage or nuclear power.

In such energy scenarios, forecasting models are needed to provide reliable energy management and probabilistic projections of socio-economical energy technologies [Zielonka et al. 2023]. In addition, these models serve as a decision support tool for the transmission system operator (TSO), such as SwisseGrid, to determine the balance of reserves [Abbaspourtorbati and Zima 2016] and for policy-makers to develop a transition to sustainable energy sources. Due to the challenges described above, this work aims to investigate probabilistic forecasting to assess the uncertainty of energy supply due to fluctuations in hydroelectric capacity at day frequency, meteorological, and the intermittency of renewable energy sources at high temporal frequency, i.e., at hour resolution.

In electricity forecasting state-of-the-art research, the focus has been mostly on point-forecast methods, that is, methods that output a single value for each target timestamp. Point forecasts are usually assessed using well-known criteria such as the *root mean squared error (RMSE)* and the *mean absolute error (MAE)*. Lately, the electricity forecasting community is shifting towards the probabilistic forecasting framework. The advantage of these methods is that they are more informative than a single-point prediction. [Gneiting et al. 2007] introduces how to assess the quality of probabilistic forecasts by maximizing the sharpness of prediction distributions under calibration constraints. Calibration refers to the statistical consistency between the predicted distributions and the observations, while sharpness refers to the spread of the forecast distributions. Scoring rules assign numerical scores to the forecasts based on the predicted distribution and the value that is materialized. In the assessment of the probabilistic model, the most widely used scores are pinball loss and the *continuous ranked probability score (CRPS)*. [Gneiting 2011]

Authors' addresses: Luca Pernigo, luca.pernigo@usi.ch, Euler Institute, USI, Via la Santa 1, 6962, Lugano, Switzerland; Rohan Sen, rohan.sen@usi.ch, Euler Institute, USI, Via la Santa 1, 6962, Lugano, Switzerland; Davide Baroli, davide.baroli@usi.ch, Euler Institute, USI, Via la Santa 1, 6962, Lugano, Switzerland.

studies the class of loss functions that lead to optimal predictors for the quantiles of a predictive distribution.

Probabilistic forecasting is gradually becoming an active research area for academia, where different researchers propose parametric and non-parametric models for forecasting specific outputs: wind, solar [Gneiting et al. 2023], prices [Nowotarski and Weron 2018a] or demand [Phipps et al. 2023]. A tutorial review on probabilistic load forecasting by [Hong et al. 2016] covers forecasting techniques, auxiliary methodologies, evaluation metrics, and good sources of reference. Another review paper by [Nowotarski and Weron 2018a] presents measures, tests, and guidelines for the rigorous use of probabilistic electricity forecasting methods. Another study, see [Van der Meer et al. 2018], provides a broad overview of probabilistic forecasting, specifically, the authors focus on solar power and load forecasting. In [Ziel and Steinert 2018] an extensive literature review has been carried out by classifying electricity price forecasting papers according to various attributes such as prediction horizon, data used, predictors, accuracy measures, and models proposed.

This article addresses probabilistic forecasting by adopting the *kernel quantile regression (KQR)* method within the RKHS framework. This method was introduced in [Takeuchi et al. 2006] and further investigated in [Li et al. 2007; Sangnier et al. 2016; Zhang et al. 2016; Zheng 2021]. The method offers a non-parametric and non-linear way to provide probabilistic forecasts. The main contribution of this article is to perform a probabilistic forecast with KQR for Swiss, Austrian, and German energy systems, where the data are extracted from ENTSO-E Transparency Platform, SECURES-Met [Formayer et al. 2023] and C3S Energy[Dubus et al. 2023], which is designed to assess the impacts of climate variability and climate change on the energy sector. The probabilistic forecast with KQR has also been validated in the GEFCom test case, where our Python-based open-source implementation compares favourably with the top teams in the probabilistic forecast of electricity load and price.

Kernel quantile regression has received little attention in the energy forecasting literature. [He and Li 2018] employs it to forecast wind power generation and compares it to a quantile regression neural network. [Moreira et al. 2016] uses it to forecast electricity prices for the Iberian electricity market. In that study, kernel quantile regression is compared with linear quantile regression, neural network quantile regression, random forest, and regression boosting. [He et al. 2017] studies the choice of kernel functions in the context of short-term load forecasting. However, no research has yet been done that thoroughly compares KQR to other state-of-the-art methods in the specific context of medium-term electricity load forecasting. Therefore, our second contribution is applying kernel quantile regression to the medium load forecasting setting, see section 4, sticking to best practices and guidelines of popular literature reviews in the field of probabilistic electric load forecasting (PLF) [Hong and Fan 2016; Lago et al. 2021; Nowotarski and Weron 2018b]. We implemented our version of KQR since there were no implemented Python packages available. We made our model class inherit from the scikit template classes `BaseEstimator` and `RegressorMixin`; by doing so, our KQR is compatible with useful scikit learn functionalities such as grid search, cross-validation, and metric scorers; that is, our method is compatible with the scikit-learn API, see [Pedregosa et al. 2011]. Sharing our implementation with the research community is the last contribution of this paper.

1.1 Outline

The rest of this article is structured as follows. In Section 2, we review quantile regression, in particular, kernel quantile regression. In Section 3, KQR is benchmarked against other popular quantile regressor models on data extracted from ENTSO-E Transparency Platform. Following, a kernel wise comparison is carried out on the SECURES-Met data [Formayer et al. 2023]. Furthermore, we evaluate the performance of KQR in the context of the challenge GEFCom2014 in Section 4. Finally, in Section 5, we conclude the article.

2 KERNEL QUANTILE REGRESSION

We first briefly review quantile regression here and then cover the details regarding kernel quantile regression. *Quantile regression (QR)* is a method used in various fields, such as econometrics, social sciences, and ecology, to analyse the empirical distribution. Estimates a target quantile of the response variable y based on a predictor vector, $\boldsymbol{x}$. QR is more robust to outliers in the data compared to the usual least-squares regression. It is also suitable for cases where there is heteroscedasticity in the errors. Additionally, using a series of quantile values provides a better description of the entire distribution than a single value, such as the mean. [Koenker and Bassett Jr 1978] showed that the *pinball loss* function

$$\rho_q(u) = \begin{cases} qu & \text{if } u \geq 0, \\ -(1-q)u & \text{if } u < 0, \end{cases} \tag{2.1}$$

can recover a target quantile of interest, q where $0 \leq q \leq 1$. We refer the reader to Appendix A for further details on the same.

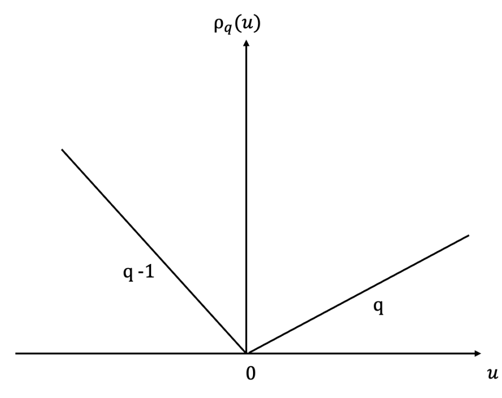


Fig. 1. Pinball loss function at the q quantile–the lower the pinball loss, the more accurate the quantile forecast is.

Given empirical samples $(\boldsymbol{x}_i, y_i)_{i=1}^n$, with $\boldsymbol{x}_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, quantile regression [Koenker 2005; Koenker and Hallock 2001] seeks to estimate the q-th conditional quantile of the response variable y as a linear function of the explanatory variables $\boldsymbol{x}_i$ by solving the following minimization problem

$$\underset{\boldsymbol{\beta}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_q(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}). \tag{2.2}$$

With the motivation to perform QR that can also capture additional complexity and non-linearity in the data, several extensions of Problem 2.2 were explored in [Hwang and Shim 2005; Li et al. 2007; Takeuchi et al. 2006; Xu et al. 2015; Zhang et al. 2016]. In the above works, the authors investigated non-parametric versions of the QR, called the kernel quantile regression, which is based on the framework of reproducing kernel Hilbert spaces. The key idea of KQR is to first transform the data samples non-linearly to a potentially higher-dimensional space of functions $\mathscr{H}$ via a feature map $\phi(\cdot)$, and then consider an affine function of the transformed data. Precisely, the conditional pinball loss of the response variable $\rho_q(y|\boldsymbol{x})$ given the predictor vector $\boldsymbol{x}$ is approximated by functions of the form

$$f(\boldsymbol{x}) = \langle w, \phi(\boldsymbol{x})\rangle_{\mathscr{H}} + b, \tag{2.3}$$

where $w \in \mathscr{H}$ is the regression coefficient and $b \in \mathbb{R}$ is the intercept term. In particular, the map $x \mapsto \phi(\boldsymbol{x})$ is implicitly defined by a reproducing kernel $\mathscr{K}$ that makes $\mathscr{H}$ an RKHS, which can contain a sufficiently rich class of functions. We first give herein the definition of RKHS and refer the reader to [Berlinet and Thomas-Agnan 2011; Schölkopf and Smola 2018] for more details of the same.

Definition 2.1. Let $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}})$ be a Hilbert space of real-valued functions on $\mathscr{X} \subset \mathbb{R}^d$. A function $\mathscr{K} : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is called a *reproducing kernel* of $\mathscr{H}$ if and only if

$$\mathscr{K}(\boldsymbol{x}, \cdot) \in \mathscr{H} \quad \text{for all } \boldsymbol{x} \in \mathscr{X}; \tag{2.4}$$

$$\langle h, \mathscr{K}(\boldsymbol{x}, \cdot)\rangle_{\mathscr{H}} = h(\boldsymbol{x}) \quad \text{for all } h \in \mathscr{H},\ \boldsymbol{x} \in \mathscr{X}. \tag{2.5}$$

If the above two properties hold, $\mathscr{H}$ is called a *reproducing kernel Hilbert space.* Associated to every RKHS $\mathscr{H}$, there exists the canonical *feature map* [Aronszajn 1950] $\phi : \mathscr{X} \to \mathscr{H}, \boldsymbol{x} \mapsto \mathscr{K}(\boldsymbol{x}, \cdot)$ such that

$$\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\rangle_{\mathscr{H}} = \mathscr{K}(\boldsymbol{x}, \boldsymbol{x}') \quad \text{for all } \boldsymbol{x}, \boldsymbol{x}' \in \mathscr{X}. \tag{2.6}$$

Furthermore, given a finite set of data samples $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathscr{X}$, the *kernel matrix* defined as $\boldsymbol{K} := [\mathscr{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1}^n \in \mathbb{R}^{n\times n}$ is a symmetric and positive semi-definite matrix.

In the context of KQR, [Li et al. 2007; Takeuchi et al. 2006; Zheng 2021] consider the following regularised objective function

$$R[f] := \frac{1}{m} \sum_{i=1}^{m} \rho_q(y_i - f(\boldsymbol{x}_i)) + \frac{\lambda}{2}\|w\|_{\mathscr{H}}^2, \tag{2.7}$$

for $f(\boldsymbol{x}) = \langle w, \phi(\boldsymbol{x})\rangle_{\mathscr{H}} + b$, cp. (2.3), where $w \in \mathscr{H}$ and $\lambda > 0$ is a hyper-parameter. The first term in (2.7) measures the empirical loss in terms of the pinball function, cp. Equation 2.1, and the second term measures the complexity of the model, see [Vapnik 1997]. In particular, [Li et al. 2007; Takeuchi et al. 2006; Zheng 2021] consider the following minimization problem

$$\underset{w\in\mathscr{H}, b\in\mathbb{R}}{\operatorname{argmin}}\ C \sum_{i=1}^{n} \rho_q\Big(y_i - \big(\langle w, \phi(\boldsymbol{x}_i)\rangle_{\mathscr{H}} + b\big)\Big) + \frac{1}{2}\|w\|_{\mathscr{H}}^2, \tag{2.8}$$

where $C = \frac{1}{\lambda m} > 0$ is the factor that balances the model complexity and the total empirical pinball loss on the sample data. Using the representer theorem, see [Schölkopf et al. 2001], we have that the optimal solution to Problem 2.8 can be written as a linear combination of kernel functions evaluated at the training examples, i.e. it has the following form

$$w^\star = \sum_{i=1}^{n} a_i^\star \phi(\boldsymbol{x}_i), \tag{2.9}$$

or equivalently,

$$f^\star(\boldsymbol{x}) = \sum_{i=1}^{n} a_i^\star \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}_i)\rangle_{\mathscr{H}} + b = \sum_{i=1}^{n} a_i^\star \mathscr{K}(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{2.10}$$

The optimal coefficients $a_i^\star$, $1 \le i \le n$ are obtained via

$$\begin{aligned} \boldsymbol{a}^\star = \underset{\boldsymbol{a}\in\mathbb{R}^n}{\operatorname{argmin}} \quad & \frac{1}{2}\boldsymbol{a}^\top \boldsymbol{K}\boldsymbol{a} - \boldsymbol{a}^\top \boldsymbol{y} \\ \text{s.t.} \quad & C(q-1)\mathbf{1} \le \boldsymbol{a} \le Cq\mathbf{1} \\ & \boldsymbol{a}^\top \mathbf{1} = 0, \end{aligned} \tag{2.11}$$

where $\boldsymbol{y} := [y_i]_{i=1}^n \in \mathbb{R}^n$, $\mathbf{1} := [1, \ldots 1]^\top \in \mathbb{R}^n$, and $\boldsymbol{K} \in \mathbb{R}^{n\times n}$ is the kernel matrix of the samples $(\boldsymbol{x}_i)_{i=1}^n$. We refer the reader to Appendix B for further details. Note that Problem 2.11 is a quadratic programming problem and, thus, can be solved by traditional solvers.

3 NUMERICAL EXAMPLES

All the numerical experiments have been performed in Python on a 3.2 GHz 16 GB Apple M1 Pro. Since KQR involves a quadratic programming problem, cp. (2.11), we used the interior-point method implemented in the `cvxopt` library to solve it. For a detailed description of `cvxopt` solvers and algorithms available, we refer the interested reader to the manual [Vandenberghe 2010]. All scripts are made publicly available at the Github repository along with the cleaned data.

3.1 Energy charts case study

In this case study, KQR is compared against popular quantile regressor models, the kernel of choice here is the Laplacian equipped with the Manhattan distance (Absolute Laplacian). The dataset for this case study comes from Energy charts which retrieves data from the ENTSO-E Transparency Platform. In predicting the national load we selected the following variables.

- `Weather temperature`;
- `Wind speed`;
- `Hour`;
- `Month`;
- `Is holiday`: a binary variable for holidays where holiday=1, working day=0;
- `Day of week`: an ordinal categorical variable corresponding to the day of the week, e.i. Monday=0, …Sunday=6;

We used the entire 2021 data as the training sample for fitting our models, and we tested them and computed their scores on the 2022 data. The results for Switzerland are reported in Table 1 and figure 2 while results for Germany can be found in Table 2 and figure 3. The error metric used is the pinball loss, scaled by the average load magnitude of each country. It is observed that the KQR model performs marginally better than other QR models and achieves results comparable to those of the GBQR. The superior performance is demonstrated in terms of CRPS to its competitors, quantile regressors, in predicting electricity load quantiles.

Table 1. Pinball loss for load in Switzerland (2022)

Quantile	LQR	GBMQR	QF	KQR
0.1	0.03595	0.01243	0.01479	**0.01210**
0.2	0.06161	0.01994	0.02200	**0.01962**
0.3	0.08064	0.02573	0.02743	**0.02495**
0.4	0.09396	0.02950	0.03073	**0.02853**
0.5	0.10224	0.03174	0.03218	**0.03048**
0.6	0.10393	0.03181	0.03115	**0.03109**
0.7	0.09892	0.03009	**0.02807**	0.02958
0.8	0.08528	0.02570	**0.02293**	0.02581
0.9	0.05892	0.01862	**0.01487**	0.01845
CRPS	0.08016	0.02506	0.02490	**0.02451**

Table 2. Pinball loss for load in Germany (2022)

Quantile	LQR	GBMQR	QF	KQR
0.1	0.04051	0.02678	**0.01585**	0.01774
0.2	0.06959	0.03020	**0.02431**	0.02517
0.3	0.09069	0.03110	0.02852	**0.02795**
0.4	0.10543	0.03011	0.03089	**0.02881**
0.5	0.11379	0.02782	0.03200	**0.02787**
0.6	0.11543	**0.02485**	0.02993	0.02558
0.7	0.10934	**0.02116**	0.02610	0.02208
0.8	0.09317	**0.01650**	0.02120	0.01737
0.9	0.06319	**0.01050**	0.01264	0.01128
CRPS	0.08901	0.02434	0.02460	**0.02265**

3.2 SECURES-Met case study

Combining the SECURES-Met data (predictors) [Formayer et al. 2023] and the load data from ENTSOE, we carried out a comparison between different classes of kernels. The kernel functions considered are: `Gaussian RBF`, `Laplacian`, `Matern` 0.5, `Matern` 1.5, `Matern` 2.5, `linear`, `periodic`, `polynomial`, `sigmoid`, and `cosine`. We used time series cross-validation to evaluate each model's performance, encompassing hyperparameter optimisation and feature selection. In Appendix C, Figure 6 shows the cross-validation process for the Absolute Laplace and Gaussian kernels. The analysis illustrates the criteria and scores to determine the optimal regularisation term and the hyperparameters of the kernel.

Details regarding the hyper-parameter selection for the different kernels can be found in the implementation here. The SECURES-Met data consist of historical data up to the end of 2020 while from 2021 onward, the data consist of forecasts modelled by the European Centre for Medium-Range Weather Forecasts (ECMWF). Therefore, we used the entire data of 2021 as the training set and then tested our kernels on the 2022 data. Note that there are two types of prediction for the data from 2021 onward, one for each of the emission scenarios RCP4.5 and RCP8.5. In this study, we restrict our attention to the RCP4.5 data, since it is part of SURE-SWEET's scenarios.

The predictors making up the dataset follow:

- `Direct irradiation: direct normal irradiation.`
- `Global radiation: mean global radiation.`
- `Hydro reservoir: daily mean power from reservoir plants in MW.`
- `Hydro river: daily mean power from run of river plants in MW.`
- `Temperature: air temperature.`
- `Wind potential: potential wind power production.`

In Appendix C, we have reported the table of quantiles and CRPS scores. Table 7 shows the results for Switzerland, Table 8 for Germany and Table 9 for Austria.

That study provides evidence of the superiority of the Gaussian kernel over the linear and polynomial kernels. In our research, we considered a larger set of kernels. From numerical experiments, the Absolute Laplacian kernel quantile has demonstrated superior performance to other Matern family kernels. In addition, a comprehensive cross-validation process for quantile 0.5 has been conducted to determine optimal hyperparameters and ridge regression parameters to prevent overfitting. The method has also been validated on an extended Secures Met dataset, which includes hydraulic capacity as a variable and technological-economic projections of the energy supply. By including high-resolution hours as a categorical variable, we have achieved accurate predictions and narrower confidence bounds for the Absolute Laplacian and Gaussian RBF Kernels as shown in Figure 8 and Figure 9. The superior performance of the Absolute Laplacian over the Gaussian RBF is largely attributable to the robustness of the Manhattan distance compared to the Euclidean distance[Aggarwal et al. 2001]. The empirical demonstration is provided by comparing the covariance kernels in Figure 7. To complete our analysis, Figures 2 and 3 show a satisfactory prediction and narrow confidence for Switzerland and Germany even when the method slightly fails to approximate the real value.

4 GEFCOM2014 CASE STUDY

We now apply KQR to the setting of probabilistic load and price forecasting. We use the GEFCom2014 [Hong et al. 2014] data to carry out our experiments. The GEFCom is a series of competitions that have been created with the intent of improving forecasting practices, addressing the gap between academia and industry, and fostering state-of-the-art research in the field of energy forecasting [Hong et al. 2016]. The GEFCom edition of 2014 consisted of four tracks, all involving probabilistic forecasting. The four tracks were load, price, wind power, and solar power forecasting. The reason for choosing the GEFCom2014 data is that it is an established benchmark in energy to compare against other valid methods. The data is freely available on Dr. Tao Hong's blog. Furthermore, a clear comparison can be performed due to the availability of the scores of each method for the load and price track. The score measure of the competition was the pinball loss, see section 2.1, averaged over the 99 quantiles,

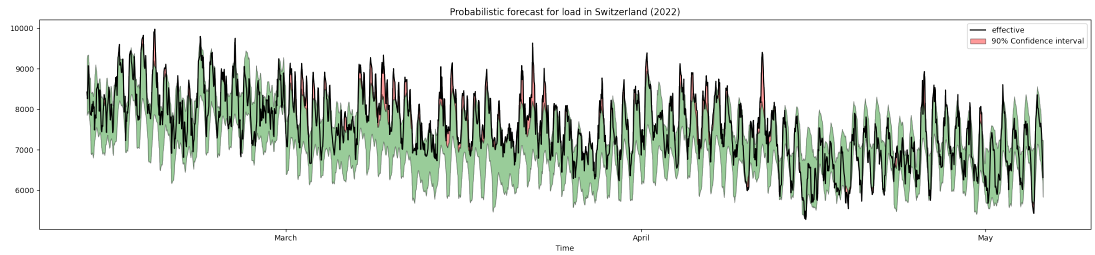


Fig. 2. Load 90% confidence interval for Switzerland Energy charts data using KQR Absolute Laplacian: Electric load probabilistic forecast for Switzerland 2022. The black line is the observed path for the load. The 90% confidence interval bands are plotted in green. Lower and upper red lines denote the 95% and 5% quantile forecast respectively.

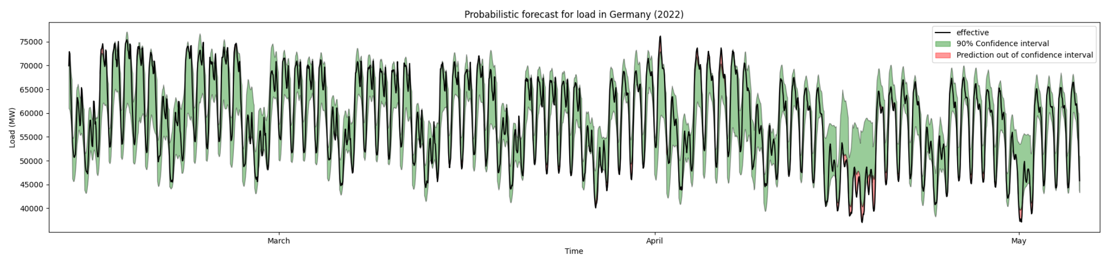


Fig. 3. Load 90% confidence interval for Germany Energy charts data using KQR Absolute Laplacian: Electric load probabilistic forecast Germany 2022. The black line is the observed path for the load. The 90% confidence interval bands are plotted in green. Lower and upper red lines denote the 95% and 5% quantile forecast respectively.

$q \in \{i/100\}_{i=1}^{n=99}$. Finally, to recreate the setting of GEFCom2014 and to provide a fair comparison, we adhere rigorously to the rules of the competition. Next, we study the performance of KQR in the load and price tracks. The kernel adopted throughout this study are the `Gaussian RBF and the Absolute Laplacian` kernel.

4.1 GEFCom2014 load track

This track was concerned with forecasting hourly loads of an anonymous US utility. The dataset provided at the start of the competition consisted of 69 months of load data and 117 months of weather data, both at hourly frequency. In this particular track, the challenge was to predict the load for the next month without the availability of weather temperature forecasts. Therefore, the primary task was to first accurately predict the weather and temperatures and then model the load accordingly. Since there were no attributes available for humidity or wind speed, we chose to predict weather temperatures by aggregating historical temperature data across different dimensions such as day, month, and hour. Then we proceeded with building KQR models for the load; we chose the following predictors.

- `Day: the number of the day`.
- `Hour`.
- `Day of week: an ordinal categorical variable for the day of the week`.
- `Is holiday: a dummy variable for holidays`.
- `w avg: average of weather temperatures across all the 25 stations`.

We built 12 models, one for each month, with each task model trained on the historical data of the month associated with it.

Table 3 reports our results for the load track. The top teams for the load forecasting track were Tololo, Adada, Jingrui(Rain) Xie, OxMAth, E.S. Managalova, Ziel Florian, and Bidong Liu; for a breakdown of the attributes of each method, see [Hong et al. 2016, Table 6].

We conclude this section with a visualisation of the 90% confidence interval forecast by our model for task number 9, that is the prediction for June, see Figure 4.

The results demonstrate that kernel quantile regression yields scores similar to those of the top five methods outlined in [He et al. 2017], specifically, the method has better performance with respect to non-quantile regression approaches, justifying the selection of this investigation.

4.2 GEFCom2014 price track

In this track, the objective was to forecast electricity prices for the next 24 hours on a rolling basis. The dataset provided consisted of 2.5 years of hourly prices and zonal and system load forecasts. The predictors fed to our KQR models are:

- `Day`;
- `Hour`;
- `Forecasted total load`;
- `Forecasted zonal load`.

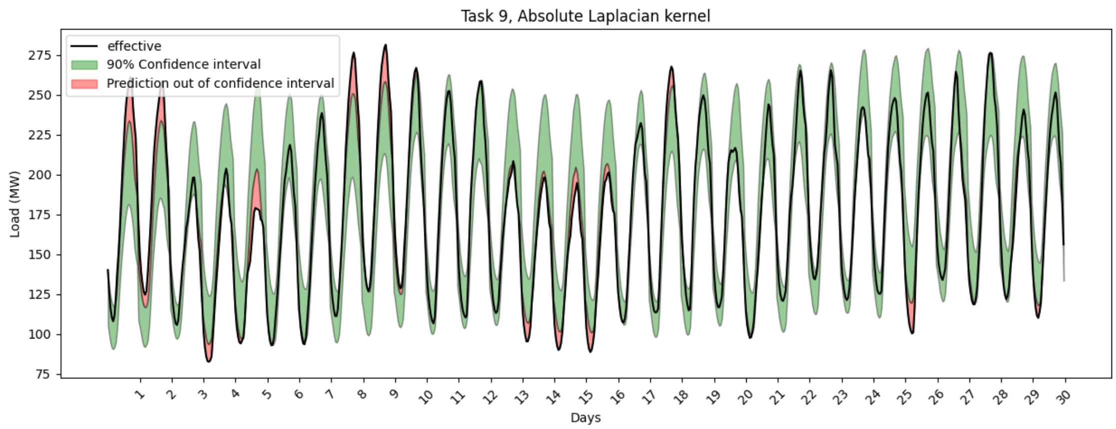


Fig. 4. Load 90% confidence interval task 9 using KQR Absolute Laplacian: Electric load probabilistic forecast for June 2011. The black line is the observed path for the load. The 90% confidence interval bands are plotted in green. Lower and upper red lines denote the 95% and 5% quantile forecast respectively. The prediction out-of confidence interval is denoted in red.

Like above, all models were trained on the historical data of the associated month.

Our results for this track are reported in table 5. In this track, the top entries came from the teams: Tololo, Team Poland, GMD, and C3 Green Team; for a breakdown of each method's attributes, see [Hong et al. 2016, Table 8]. Finally, our probabilistic prediction for the 13^{th} July 2013 zonal price at the 90% confidence interval is visualised in Figure 5.

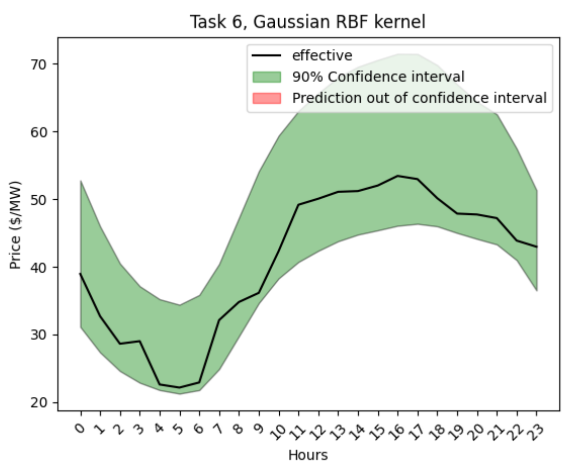


Fig. 5. Price 90% confidence interval task 6: Electricity price probabilistic forecast for the 13^{th} July 2013. The black line is the observed path for the price. The 90% confidence interval bands are plotted in green. Lower and upper red lines denote the 95% and 5% quantile forecast respectively.

5 CONCLUSION

In this article, we investigate a non-parametric probabilistic method, the kernel quantile regression, for estimating quantiles of load. To our knowledge, this method has not been explored before for load prediction. We show its effectiveness through several numerical tests on DACH data (Germany, Austria and Switzerland), illustrating that it performs competently compared to other well-known quantile regression techniques and exceeds the point regression results of GEFCom2014. In addition to these numerical experiments, we extend the test case of GEFCom2014 considering additional explanatory variables in hydrocapacity energy storage. We observe that KQR shows favourably and can forecast the medium-term horizon of the Secures-Met dataset. However, the forecasting for the short-term horizon has been demonstrated in the GEFCom2014 price track. The tuning of hyperparameters along with the ridge parameter was carried out using cross-validation, and we presented comprehensive analysis to support the results. In our investigation of kernel functions, we focus on evaluating the Absolute Laplace kernel in comparison to the RBF Gaussian Kernel during the kernel selection procedure. The results of this study confirm the durability of our selection in terms of precision of prediction. Finally, we provide an open-source implementation of KQR integrated with the most popular tools in the community. This investigation will serve as a case study for an uncertainty quantification model to predict the impact of climate shocks on pathways, which will be further refined in scenarios computed by the SURE SWEET energy models. Further investigation with different choices and combinations of kernels as well as a detailed comparison with gradient booster methods remains a future research area.

6 ACKNOWLEDGMENTS

We are grateful to three anonymous reviewers for insightful feedback. We are thankful for the fruitful discussions with Michael Multerer and Paul Schneider. Davide Baroli and Luca Pernigo carried out this research with the support of the Swiss Federal Office of Energy SFOE as part of the SWEET project SURE. The authors bear sole responsibility for the conclusions and the results presented

Table 3. Pinball loss GEFCom2014 load data

Team name\Task number	1	2	3	4	5	6	7	8	9	10	11	12
KQR Absolute Laplacian	12.5357	11.0209	9.4492	5.2240	6.6145	6.5418	11.2004	11.6325	5.9476	5.2219	7.5478	11.0587
KQR Gaussian RBF	12.4660	11.0894	9.4938	5.1826	6.9575	6.7947	10.8825	11.5542	5.9742	5.0779	7.3797	10.3110
Adada	10.5093	10.0801	**7.6238**	4.7289	**5.3936**	6.6242	8.0144	11.1366	5.7779	3.6379	7.0096	8.9109
Benchmark - Load	18.7384	22.7585	13.2163	8.3626	10.9162	16.9937	13.4038	17.3151	13.8374	6.4237	10.9380	34.0685
C3 Green Team	18.7384	19.2208	7.9637	4.6370	6.4543	8.3799	10.5546	10.6609	5.8867	4.4866	5.9396	10.3917
E.S. Mangalova	18.7384	13.3340	7.8025	**4.4096**	6.6330	6.2306	10.1511	10.9294	6.2224	4.2382	6.5464	8.8080
Jingrui (Rain) Xie	11.8700	10.9250	8.4938	4.9611	7.4442	6.9921	9.0523	11.2600	5.4864	**3.3602**	**5.9011**	9.7316
OxMath	14.4091	**8.9136**	7.6059	4.4548	7.2944	7.4551	7.9527	**10.2444**	5.4551	4.2111	6.4054	9.5520
Tololo	**10.4369**	12.5232	8.2695	4.4220	5.8976	**6.1878**	**7.3182**	10.8032	**5.4469**	3.9613	6.3173	**8.4787**

Table 4. GEFCom2014 load data ranking

Team name\Task number	1	2	3	4	5	6	7	8	9	10	11	12	Aggregate ranking
KQR Absolute Laplacian	6/362	7/362	14/362	10/362	4/362	3/362	20/362	15/362	9/362	16/362	18/362	12/362	11/362
KQR Gaussian RBF	5/362	8/362	15/362	9/362	8/362	7/362	18/362	12/362	10/362	15/362	17/362	8/362	10/362

Table 5. Pinball loss GEFCom2014 price data

Team name\Task number	1	2	3	4	5	6	7	8	9	10	11	12
KQR Absolute Laplacian	**1.02492**	3.35057	4.21374	7.33987	5.00981	6.96522	3.57168	1.77610	1.28765	2.73863	2.39831	23.30234
KQR Gaussian RBF	1.84673	2.81882	1.54608	8.31636	4.05988	6.60456	3.57818	2.02177	1.45779	2.20701	1.98184	21.41033
Benchmark - Price	4,02875	7,97208	5,70395	12,15104	38,33541	44,22979	18,22395	31,56729	42,94958	2,85583	3,20395	22,38333
C3 Green Team	1,85897	3,27786	1,2593	5,08886	6,87674	6,1505	4,42379	1,32639	1,25915	3,08224	1,55811	6,58123
GMD	3,7271	1,783	**0,92191**	5,08886	6,21331	**3,82599**	4,9342	1,47858	1,65933	**2,06134**	2,1235	6,84571
Team Poland	1,97477	1,81898	1,19162	2,82318	7,55914	4,20773	2,59715	1,04693	**1,24193**	4,06012	**1,08458**	**3,06512**
Tololo	1,70734	**1,45173**	1,10384	**2,01694**	9,15596	4,6821	**1,59517**	**0,75352**	2,45935	2,9614	1,34614	3,55819
pat1	2,36615	1,98567	1,07248	2,79465	**4,23269**	4,70614	8,40506	1,25376	2,23991	3,67952	1,06139	6,27517

Table 6. GEFCom2014 price data ranking

Team name\Task number	1	2	3	4	5	6	7	8	9	10	11	12	Aggregate ranking
KQR Absolute Laplacian	1/287	11/287	12/287	14/287	4/287	11/287	5/287	8/287	6/287	8/287	15/287	18/287	15/287
KQR Gaussian RBF	6/287	8/287	10/287	15/287	1/287	10/287	6/287	10/287	7/287	5/287	9/287	16/287	10/287

in this publication. Rohan Sen was supported by the SNF grant "Scenarios" (100018_189086).

REFERENCES

Farzaneh Abbaspourtorbati and Marek Zima. 2016. The Swiss Reserve Market: Stochastic Programming in Practice. *IEEE Transactions on Power Systems* 31, 2 (March 2016), 1188–1194.

Charu C. Aggarwal, Alexander Hinneburg, and Daniel A. Keim. 2001. On the Surprising Behavior of Distance Metrics in High Dimensional Space, In Proceedings of the 8th International Conference on Database Theory. *Lecture Notes in Computer Science* 1, 1, 420–434. https://doi.org/10.1007/3-540-44503-x_27

Nachman Aronszajn. 1950. Theory of reproducing kernels. *Transactions of the American mathematical society* 68, 3 (1950), 337–404.

Alain Berlinet and Christine Thomas-Agnan. 2011. *Reproducing kernel Hilbert spaces in probability and statistics.* Springer Science & Business Media, NY, USA.

Stephen P Boyd and Lieven Vandenberghe. 2004. *Convex optimization.* Cambridge University Press, Cambridge, UK.

Laurent Dubus, Yves-Marie Saint-Drenan, Alberto Troccoli, Matteo De Felice, Yohann Moreau, Linh Ho-Tran, Clare Goodess, Rodrigo Amaro e Silva, and Luke Sanger. 2023. C3SEnergy: A climate service for the provision of power supply and demand indicators for Europe based on the ERA5 reanalysis and ENTSO-E data.

Herbert Formayer, Imran Nadeem, David Leidinger, Philipp Maier, Franziska Schöniger, Demet Suna, Gustav Resch, Gerhard Totschnig, and Fabian Lehner. 2023. SECURES-Met: A European meteorological data set suitable for electricity modelling applications. *Scientific Data* 10, 1 (2023). https://doi.org/10.1038/s41597-023-02494-4

Tilmann Gneiting. 2011. Quantiles as optimal point forecasts. *International Journal of Forecasting* 27, 2 (2011), 197–207.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. 2007. Probabilistic Forecasts, Calibration and Sharpness. , 243–268 pages.

Tilmann Gneiting, Sebastian Lerch, and Benedikt Schulz. 2023. Probabilistic solar forecasting: Benchmarks, post-processing, verification. *Solar Energy* 252 (2023), 72–80.

Yaoyao He and Haiyan Li. 2018. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. *Energy conversion and management* 164 (2018), 374–384.

Yaoyao He, Rui Liu, Haiyan Li, Shuo Wang, and Xiaofen Lu. 2017. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory. *Applied energy* 185 (2017), 254–266.

Linh Ho-Tran and Stephanie Fiedler. 2024. A climatology of weather-driven anomalies in European photovoltaic and wind power production.

Tao Hong and Shu Fan. 2016. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting* 32, 3 (2016), 914–938.

Tao Hong, Pierre Pinson, and Shu Fan. 2014. Global energy forecasting competition 2012. , 357–363 pages.

Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J Hyndman. 2016. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. , 896–913 pages.

Changha Hwang and Jooyong Shim. 2005. A Simple Quantile Regression via Support Vector Machine. , 512–520 pages.

Ramachandran Kannan. 2018. Dynamics of long-term electricity demand profile: Insights from the analysis of Swiss energy systems. *Energy Strategy Reviews* 22 (Nov. 2018), 410–425. https://doi.org/10.1016/j.esr.2018.10.010

Ramachandran Kannan, Evangelos Panos, Stefan Hirschberg, and Tom Kober. 2022. A net-zero Swiss energy system by 2050: Technological and policy options for the transition of the transportation sector. *FUTURES; FORESIGHT SCIENCE* 4, 3–4 (March 2022), 126–148. https://doi.org/10.1002/ffo2.126

Roger Koenker. 2005. *Quantile Regression*. Vol. 38. Cambridge University Press, Cambridge, UK.

Roger Koenker and Gilbert Bassett Jr. 1978. Regression quantiles. *Econometrica: journal of the Econometric Society* 46, 1 (1978), 33–50.

Roger Koenker and Kevin F. Hallock. 2001. Quantile Regression. *The Journal of Economic Perspectives* 15, 4 (2001), 143–156. http://www.jstor.org/stable/2696522

Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy* 293 (2021), 116983.

Youjuan Li, Yufeng Liu, and Ji Zhu. 2007. Quantile Regression in Reproducing Kernel Hilbert Spaces. *J. Amer. Statist. Assoc.* 102, 477 (2007), 255–268.

Sandro Luh, Ramachandran Kannan, Russell McKenna, Thomas J Schmidt, and Tom Kober. 2023. How, where, and when to charge electric vehicles – net-zero energy system implications and policy recommendations. *Environmental Research Communications* 5, 9 (Sept. 2023), 095004. https://doi.org/10.1088/2515-7620/acf363

Rui Moreira, Ricardo Bessa, and Joao Gama. 2016. Probabilistic forecasting of day-ahead electricity prices for the Iberian electricity market. In *2016 13th International Conference on the European Energy Market (EEM)*. IEEE, NY, 1–5.

Claudio Nägeli, Martin Jakob, Giacomo Catenazzi, and York Ostermeyer. 2020. Towards agent-based building stock modeling: Bottom-up modeling of long-term stock dynamics affecting the energy and climate impact of building stocks. *Energy and Buildings* 211 (March 2020), 109763. https://doi.org/10.1016/j.enbuild.2020.109763

Jakub Nowotarski and Rafał Weron. 2018a. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* 81 (2018), 1548–1568.

Jakub Nowotarski and Rafał Weron. 2018b. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews* 81 (2018), 1548–1568.

Evangelos Panos, Ramachandran Kannan, Stefan Hirschberg, and Tom Kober. 2023. An assessment of energy system transformation pathways to achieve net-zero carbon dioxide emissions in Switzerland. *Communications Earth; Environment* 4, 1 (2023), 157–170. https://doi.org/10.1038/s43247-023-00813-6

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830.

Kaleb Phipps, Stefan Meisenbacher, Benedikt Heidrich, Marian Turowski, Ralf Mikut, and Veit Hagenmeyer. 2023. Loss-Customised Probabilistic Energy Time Series Forecasts Using Automated Hyperparameter Optimisation. https://doi.org/10.1145/3575813.3595204

Maxime Sangnier, Olivier Fercoq, and Florence d'Alché-Buc. 2016. Joint quantile regression in vector-valued RKHSs. , 3700-–3708 pages.

Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. 2001. A Generalized Representer Theorem. In *Computational Learning Theory*, David Helmbold and Bob Williamson (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 416–426.

Bernhard Schölkopf and Alexander J. Smola. 2018. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* The MIT Press, Cambridge, MA.

Ichiro Takeuchi, Quoc V. Le, Timothy D. Sears, and Alexander J. Smola. 2006. Nonparametric Quantile Estimation. *Journal of Machine Learning Research* 7, 45 (2006), 1231–1264.

Evelina Trutnevyte. 2013. EXPANSE methodology for evaluating the economic potential of renewable energy from an energy mix perspective. *Applied Energy* 111 (Nov. 2013), 593–601. https://doi.org/10.1016/j.apenergy.2013.04.083

Dennis W Van der Meer, Joakim Widén, and Joakim Munkhammar. 2018. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renewable and Sustainable Energy Reviews* 81 (2018), 1484–1512.

Lieven Vandenberghe. 2010. The CVXOPT linear and quadratic cone program solvers.

Vladimir N. Vapnik. 1997. The support vector method. In *Artificial Neural Networks — ICANN'97*. Springer Berlin Heidelberg, Berlin, Heidelberg, 261–271.

Qifa Xu, Jinxiu Zhang, Cuixia Jiang, Xue Huang, and Yaoyao He. 2015. Weighted quantile regression via support vector machine. *Expert Systems with Applications* 42, 13 (2015), 5441–5451.

Chong Zhang, Yufeng Liu, and Yichao Wu. 2016. On Quantile Regression in Reproducing Kernel Hilbert Spaces with the Data Sparsity Constraint. *Journal of Machine Learning Research* 17, 40 (2016), 1–45.

Songfeng Zheng. 2021. Fast quantile regression in reproducing kernel Hilbert space. *Journal of the Korean Statistical Society* 51, 2 (Oct. 2021), 568–588.

Florian Ziel and Rick Steinert. 2018. Probabilistic mid-and long-term electricity price forecasting. *Renewable and Sustainable Energy Reviews* 94 (2018), 251–266.

Nik Zielonka, Xin Wen, and Evelina Trutnevyte. 2023. Probabilistic projections of granular energy technology diffusion at subnational level.

A PINBALL LOSS FUNCTION FOR QUANTILE REGRESSION

In this section, we show why minimizing the pinball loss function leads to estimates of the quantile. This section has been sourced from [Koenker 2005] and hence, we would like to refer to the above reference for further details on the same. For a real-valued random variable X with distribution function $F(\cdot)$, the q-th quantile is defined as

$$Q(q) := \inf \{x \in \mathbb{R} : F(x) \geq q\} \quad \text{for } 0 \leq q \leq 1. \tag{A.1}$$

For a continuous distribution function $F(\cdot)$, the quantile becomes just the inverse, i.e. $Q = F^{-1}$. For example, $q = 0.5$ defines the median of the distribution of X. The quantiles arise from a simple optimization problem that is fundamental to all that follows. Consider a simple decision-theoretic problem: a point estimate is required for a random variable with (posterior) distribution function F. If the loss is described by the piecewise linear pinball loss, cp. (2.1), consider the problem of finding $\hat{x}$ to minimize the expected loss. We seek to minimize

$$\begin{aligned}\mathbb{E}\left[\rho_q(X - \hat{x})\right] = q \int_{\hat{x}}^{\infty} (x - \hat{x})\, dF(x) \\ - (1-q) \int_{-\infty}^{\hat{x}} (x - \hat{x})\, dF(x).\end{aligned} \tag{A.2}$$

To find the optimal $\hat{x}$, we consider the first-order conditions by differentiating problem A.2 with respect to $\hat{x}$ and setting it to zero:

$$0 = -q \int_{\hat{x}}^{\infty} dF(x) + (1-q) \int_{-\infty}^{\hat{x}} dF(x) = F(\hat{x}) - q$$

Since $F(\cdot)$ is a monotone function, any element of $\{x : F(x) = q\}$ minimizes expected loss. When the solution is unique, $\hat{x} = F^{-1}(q)$, otherwise, we have an "interval of q quantiles" from which the smallest element must be chosen - to adhere to the convention that the empirical quantile function be left-continuous. When $F(\cdot)$ is replaced by the empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}, \tag{A.3}$$

we may still choose $\hat{x}$ to minimize expected loss:

$$\int_{\mathbb{R}} \rho_q(x - \hat{x})\, dF_n(x) = \frac{1}{n} \sum_{i=1}^{n} \rho_q(x_i - \hat{x}) \tag{A.4}$$

where x_i's are now assumed to be generated from the independently and identically distributed random variables $X_i \sim F(\cdot)$; doing so now yields the q-th sample quantile. When qn is an integer there is again some ambiguity in the solution, because we really have an interval of solutions, $\{x : F_n(x) = q\}$, but this is of little practical consequence.

B ESTIMATOR DERIVATION FOR KERNEL QUANTILE REGRESSION

The goal is to solve the optimization problem, cp. (2.8)

$$\underset{w\in\mathcal{H},b\in\mathbb{R}}{\operatorname{argmin}}\ C\sum_{i=1}^{n}\rho_q\Big(y_i-\big(\langle w,\phi(\boldsymbol{x}_i)\rangle_{\mathcal{H}}+b\big)\Big)+\frac{1}{2}\|w\|_{\mathcal{H}}^2, \tag{B.1}$$

with the characterization $f(\boldsymbol{x})=\langle w,\phi(\boldsymbol{x})\rangle_{\mathcal{H}}+b$. Minimizing $\|w\|_{\mathcal{H}}^2$ is equivalent to minimizing the regularizer, see [Schölkopf and Smola 2018, Section 2.2.4]. Using the definition of the pinball loss function, cp. (2.1), we have the problem

$$\begin{aligned}\underset{w\in\mathcal{H},b\in\mathbb{R}}{\operatorname{argmin}}\quad & C\sum_{i=1}^{n}q(y_i-\langle w,\phi(\boldsymbol{x}_i)\rangle_{\mathcal{H}}-b)+\\ & +(1-q)(-y_i+\langle w,\phi(\boldsymbol{x}_i)\rangle_{\mathcal{H}}+b)+\frac{1}{2}\|w\|_{\mathcal{H}}^2,\end{aligned} \tag{B.2}$$

Introducing the slack variables ξ_i, ξ_i^*, $1\le i\le n$, we can rephrase Problem B.2 as the following constrained optimization problem, see [Hwang and Shim 2005; Takeuchi et al. 2006]

$$\begin{aligned}\underset{w\in\mathcal{H},\,b,\xi_i,\xi_i^*\in\mathbb{R}}{\operatorname{argmin}}\quad & C\sum_{i=1}^{m}q\xi_i+(1-q)\xi_i^*+\frac{1}{2}\|w\|_{\mathcal{H}}^2\\ \text{s.t.}\quad & y_i-\langle w,\phi(\boldsymbol{x}_i)\rangle_{\mathcal{H}}-b\le\xi_i,\\ & -y_i+\langle w,\phi(\boldsymbol{x}_i)\rangle_{\mathcal{H}}+b\le\xi_i^*,\\ & \xi_i\ge 0,\quad \xi_i^*\ge 0.\end{aligned} \tag{B.3}$$

Now, by the representer theorem, see [Schölkopf et al. 2001], any $w\in\mathcal{H}$ that minimizes Problem B.1 can be written as

$$w=\sum_{i=1}^{n}a_i\phi(\boldsymbol{x}_i)\implies w(\boldsymbol{x})=\sum_{i=1}^{n}a_i\mathcal{K}(\boldsymbol{x},\boldsymbol{x}_i).$$

Using (2.5), we have that $\langle w,\phi(\boldsymbol{x}_j)\rangle_{\mathcal{H}}=w(\boldsymbol{x}_j)=\sum_{i=1}^{n}a_i\mathcal{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)$ for $1\le j\le n$. Hence, denoting the coefficient vector as $\boldsymbol{a}:=[a_i]_{i=1}^n\in\mathbb{R}^n$, we have that

$$\big[\langle w,\phi(\boldsymbol{x}_j)\rangle_{\mathcal{H}}\big]_{i=1}^n=\boldsymbol{K}\boldsymbol{a},\qquad \|w\|_{\mathcal{H}}^2=\boldsymbol{a}^\top\boldsymbol{K}\boldsymbol{a}. \tag{B.4}$$

Using the notations $\boldsymbol{y}:=[y_i]_{i=1}^n\in\mathbb{R}^n$, $\boldsymbol{\xi}:=[\xi_i]_{i=1}^n\in\mathbb{R}^n$, $\boldsymbol{\xi}^*:=[\xi_i^*]_{i=1}^n\in\mathbb{R}^n$, we have the equivalent problem in matrix notation

$$\begin{aligned}\underset{\boldsymbol{a},\boldsymbol{\xi},\boldsymbol{\xi}^*\in\mathbb{R}^n,b\in\mathbb{R}}{\operatorname{argmin}}\quad & Cq\boldsymbol{\xi}^\top\boldsymbol{1}+C(1-q)(\boldsymbol{\xi}^*)^\top\boldsymbol{1}+\frac{1}{2}\boldsymbol{a}^\top\boldsymbol{K}\boldsymbol{a}\\ \text{s.t.}\quad & \boldsymbol{y}-\boldsymbol{K}\boldsymbol{a}-b\boldsymbol{1}\le\boldsymbol{\xi},\\ & -\boldsymbol{y}+\boldsymbol{K}\boldsymbol{a}+b\boldsymbol{1}\le\boldsymbol{\xi}^*,\\ & \boldsymbol{\xi}\ge\boldsymbol{0},\quad \boldsymbol{\xi}^*\ge\boldsymbol{0}.\end{aligned} \tag{B.5}$$

Through the Lagrange multipliers and the KKT conditions (see [Boyd and Vandenberghe 2004]), Problem B.5 can be solved to its equivalent dual formulation, see [Takeuchi et al. 2006; Xu et al. 2015] for more details. We can write the Lagrangian as

$$\begin{aligned}&L(\boldsymbol{a},b,\boldsymbol{\xi},\boldsymbol{\xi}^*,\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\nu},\boldsymbol{\nu}^*):=Cq\boldsymbol{\xi}^\top\boldsymbol{1}+C(1-q)(\boldsymbol{\xi}^*)^\top\boldsymbol{1}+\frac{1}{2}\boldsymbol{a}^\top\boldsymbol{K}\boldsymbol{a}\\ &-\boldsymbol{\alpha}^\top(\boldsymbol{\xi}-\boldsymbol{y}+\boldsymbol{K}\boldsymbol{a}+b\boldsymbol{1})-(\boldsymbol{\alpha}^*)^\top(\boldsymbol{\xi}^*+\boldsymbol{y}-\boldsymbol{K}\boldsymbol{a}-b\boldsymbol{1})\\ &-\boldsymbol{\nu}^\top\boldsymbol{\xi}-(\boldsymbol{\nu}^*)^\top\boldsymbol{\xi}^*,\end{aligned} \tag{B.6}$$

with the positivity constraints $\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\nu},\boldsymbol{\nu}^*\ge\boldsymbol{0}$. We proceed to derive the dual function by minimizing the Lagrangian

$$g(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\nu},\boldsymbol{\nu}^*)=\inf_{\boldsymbol{a},\boldsymbol{\xi},\boldsymbol{\xi}^*\in\mathbb{R}^n,b\in\mathbb{R}}L(\boldsymbol{a},\boldsymbol{\xi},\boldsymbol{\xi}^*,\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\nu},\boldsymbol{\nu}^*). \tag{B.7}$$

Setting the partial derivatives to zero, we have

$$\begin{cases}\frac{\partial L}{\partial \boldsymbol{a}}=\boldsymbol{0}\implies \boldsymbol{K}\boldsymbol{a}=\boldsymbol{K}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*),\\ \frac{\partial L}{\partial b}=0\implies(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^\top\boldsymbol{1}=0,\\ \frac{\partial L}{\partial \boldsymbol{\xi}}=\boldsymbol{0}\implies Cq\boldsymbol{1}=\boldsymbol{\alpha}+\boldsymbol{\nu},\\ \frac{\partial L}{\partial \boldsymbol{\xi}^*}=\boldsymbol{0}\implies C(1-q)\boldsymbol{1}=\boldsymbol{\alpha}^*+\boldsymbol{\nu}^*.\end{cases} \tag{B.8}$$

Substituting the conditions into B.6, we obtain the dual function as

$$\begin{aligned}g(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\nu},\boldsymbol{\nu}^*)=-\frac{1}{2}&(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^\top\boldsymbol{K}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^\top+(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^\top\boldsymbol{y},\\ \text{s.t.}\quad&(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)=\boldsymbol{a},\\ &(\boldsymbol{\alpha}-\boldsymbol{\alpha}^*)^\top\boldsymbol{1}=0,\\ &\boldsymbol{0}\le\boldsymbol{\alpha}\le\boldsymbol{C}q,\\ &\boldsymbol{0}\le\boldsymbol{\alpha}^*\le\boldsymbol{C}(1-q).\end{aligned} \tag{B.9}$$

In terms of the coefficient vector $\boldsymbol{a}$, we can consider the dual optimization problem which turns out to be (after switching signs)

$$\begin{aligned}\underset{\boldsymbol{a}\in\mathbb{R}^n}{\operatorname{argmin}}\quad & +\frac{1}{2}\boldsymbol{a}^\top\boldsymbol{K}\boldsymbol{a}-\boldsymbol{a}^\top\boldsymbol{y}\\ \text{s.t.}\quad & C(q-1)\boldsymbol{1}\le\boldsymbol{a}\le Cq\boldsymbol{1}\\ & \boldsymbol{a}^\top\boldsymbol{1}=0.\end{aligned} \tag{B.10}$$

From the constraint conditions of problem B.9 and the expression of w cp. (2.9), we have that the optimal w is given by the coefficients

$$w=\sum_{i=1}^{n}a_i^\star\phi(\boldsymbol{x}_i), \tag{B.11}$$

where $\boldsymbol{a}^\star=[a_i^\star]_{i=1}^n$ is the solution of Problem B.10. Finally, the data points for which $a_i^\star\notin\{C(q-1),Cq\}$ are called the support vectors. The intercept term b can be calculated using the fact that $f(\boldsymbol{x}_i)=y_i$ for the set of support vectors, see [Takeuchi et al. 2006; Xu et al. 2015] for more details. The latter holds due to KKT conditions on Problem B.10.

C ADDITIONAL NUMERICAL COMPARISONS

In this Appendix, we present additional investigations that demonstrate the reliability and robustness of the KQR method introduced in the GEFCom2014 and Secures MET case study.

Table 7. Pinball loss for load in Switzerland case study with SECURES-Met dataset

Quantile	Absolute Laplacian	Matern 0.5/ Laplacian	Matern 1.5	Matern 2.5	Matern ∞/ Gaussian RBF	Linear	Periodic	Polynomial	Sigmoid	Cosine
0.1	**0.018783**	0.018988	0.019108	0.019257	0.019547	0.019952	0.018891	0.019654	0.021456	0.021921
0.2	**0.030441**	0.030668	0.030809	0.031028	0.031546	0.032158	0.030683	0.031857	0.034032	0.034629
0.3	**0.038467**	0.038864	0.039050	0.039285	0.039907	0.040657	0.038968	0.040075	0.042944	0.043393
0.4	**0.043622**	0.044186	0.044516	0.044706	0.045286	0.046153	0.044354	0.045506	0.048701	0.048977
0.5	**0.046160**	0.046792	0.047205	0.047450	0.048116	0.048951	0.046956	0.047891	0.051423	0.051896
0.6	**0.045499**	0.046133	0.046824	0.047177	0.047840	0.048667	0.046446	0.047496	0.051292	0.051894
0.7	**0.041494**	0.042044	0.042928	0.043371	0.044104	0.044926	0.042458	0.043565	0.047656	0.048275
0.8	**0.033837**	0.034128	0.034797	0.035325	0.036166	0.037055	0.034290	0.035533	0.039501	0.040007
0.9	0.021883	**0.021871**	0.022431	0.022682	0.023169	0.023730	0.022000	0.022811	0.025561	0.026019
CRPS	**0.035576**	0.035964	0.036407	0.036698	0.037298	0.038028	0.036116	0.037154	0.040285	0.040779

Table 8. Pinball loss for load in German case study with SECURES-Met dataset

Quantile	Absolute Laplacian	Matern 0.5/ Laplacian	Matern 1.5	Matern 2.5	Matern ∞/ Gaussian RBF	Linear	Periodic	Polynomial	Sigmoid	Cosine
0.1	**0.025734**	0.026272	0.026642	0.026845	0.027090	0.027306	0.026139	0.026621	0.027954	0.028004
0.2	**0.043114**	0.044138	0.044733	0.045057	0.045469	0.045917	0.044088	0.044716	0.046925	0.047025
0.3	**0.054861**	0.056343	0.056930	0.057335	0.057920	0.058528	0.056229	0.057066	0.060092	0.060209
0.4	**0.061436**	0.063490	0.064291	0.064748	0.065433	0.066136	0.063324	0.064683	0.067979	0.068202
0.5	**0.064144**	0.066330	0.067229	0.067696	0.068275	0.068960	0.066325	0.071421	0.070957	0.071359
0.6	**0.062306**	0.064676	0.065836	0.066299	0.066881	0.067405	0.064660	0.069160	0.068494	0.068780
0.7	**0.055491**	0.057851	0.058848	0.059317	0.059878	0.060275	0.061442	0.061423	0.060988	0.060990
0.8	**0.044023**	0.045011	0.045441	0.045644	0.046008	0.046374	0.047348	0.047182	0.047071	0.047056
0.9	0.026178	**0.026126**	0.026160	0.026187	0.026207	0.026282	0.026772	0.026766	0.026581	0.026628
CRPS	**0.048587**	0.050026	0.050679	0.051014	0.051462	0.051909	0.050703	0.052115	0.053005	0.053139

Table 9. Pinball loss for load in Austrian case study with SECURES-Met dataset

Quantile	Absolute Laplacian	Matern 0.5/ Laplacian	Matern 1.5	Matern 2.5	Matern ∞/ Gaussian RBF	Linear	Periodic	Polynomial	Sigmoid	Cosine
0.1	**0.024365**	0.024592	0.024550	0.024611	0.024831	0.026853	0.025744	0.026004	0.027672	0.027816
0.2	**0.040971**	0.041406	0.041413	0.041487	0.041857	0.045007	0.043247	0.043720	0.045899	0.045879
0.3	**0.052316**	0.052969	0.052888	0.052973	0.053449	0.057880	0.055231	0.055446	0.058572	0.058454
0.4	**0.058286**	0.059018	0.058916	0.059078	0.059592	0.065770	0.061844	0.062495	0.066599	0.066335
0.5	**0.060344**	0.061100	0.061319	0.061541	0.062338	0.068197	0.063864	0.071731	0.069761	0.069574
0.6	**0.058724**	0.059507	0.059674	0.059930	0.060709	0.065828	0.062265	0.070269	0.067179	0.066970
0.7	**0.053148**	0.054189	0.054249	0.054372	0.054916	0.058548	0.055830	0.063015	0.059807	0.059419
0.8	**0.042740**	0.043368	0.043408	0.043510	0.043881	0.046017	0.044384	0.049813	0.047078	0.046662
0.9	**0.026430**	0.026583	0.026781	0.026817	0.026908	0.027050	0.026471	0.030279	0.028031	0.027986
CRPS	**0.046369**	0.046970	0.047022	0.047146	0.047609	0.051239	0.048764	0.052530	0.052289	0.052122

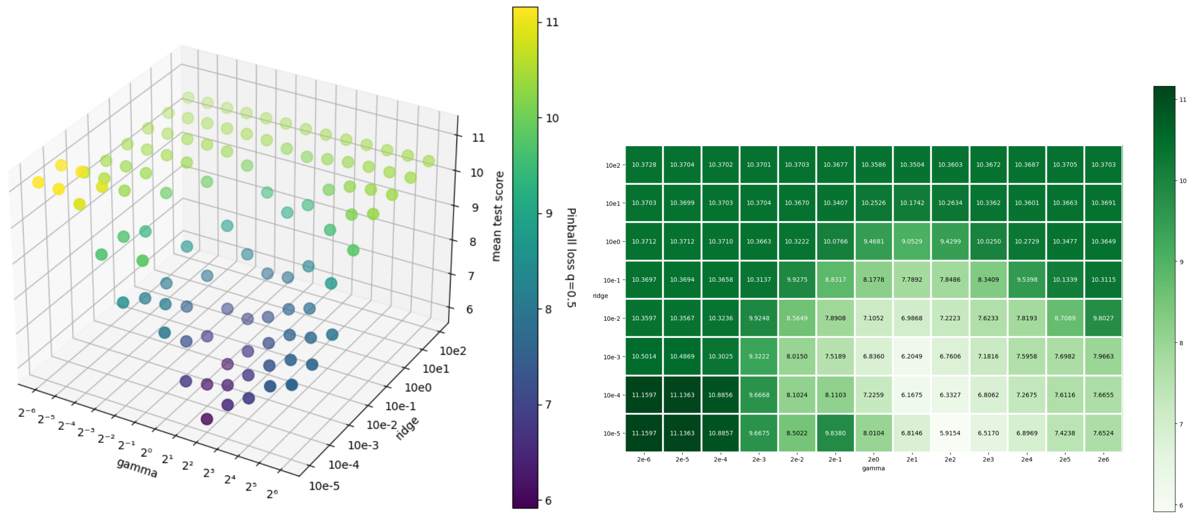


Fig. 6. Hyperparameter cross-validation for the RBF Kernel Quantile Method applied to the price task 4, specifically the cross-validation is obtained for the median quantile. On the left, the figure illustrates the optimal hyperparameters selection based on the mean quantile score. The same result is presented using a heatmap.

Fig. 7. Comparison between the kernel covariance evaluation for Absolute Laplacian and Gaussian RBF kernel for SECURE-Met study.

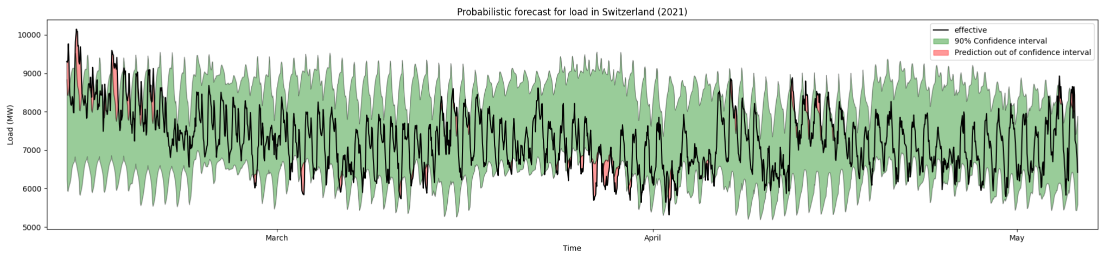


Fig. 8. Prediction and confidence bound of Absolute Laplacian kernel in Secures MET study. The black line is the observed path for the load. The 90% confidence interval bands are plotted in green. Lower and upper red lines denote the 95% and 5% quantile forecast respectively. The prediction out-of confidence interval is denoted in red.

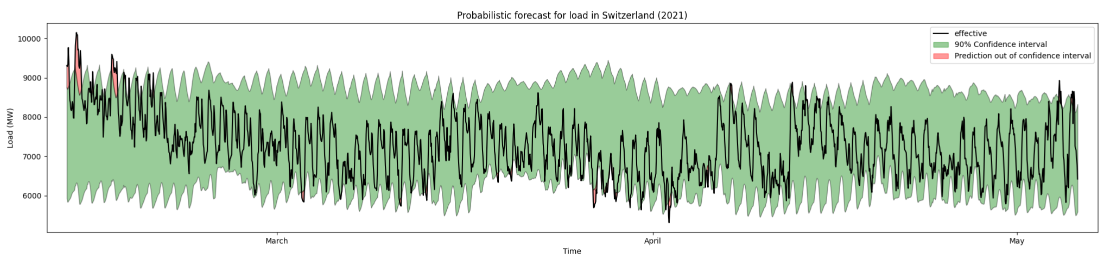


Fig. 9. Prediction and confidence bound of Gaussian RBF kernel in Secures MET study. The black line is the observed path for the load. The 90% confidence interval bands are plotted in green. Lower and upper red lines denote the 95% and 5% quantile forecast respectively. The prediction out-of confidence interval is denoted in red.

Poster Abstract: Forecasting Renewable Energy at European Markets

Hun Rim
rimh@usi.ch
Università della Svizzera italiana
Lugano, Ticino, Switzerland

Juraj Kardos
juraj.kardos@usi.ch
Università della Svizzera italiana
Lugano, Ticino, Switzerland

Olaf Schenk
olaf.schenk@usi.ch
Università della Svizzera italiana
Lugano, Ticino, Switzerland

ABSTRACT

The ambitious energy targets, accelerated by the recent energy crisis, are driving the increase of renewable energy share in gross energy consump- tion. However, the intermittent and seasonal nature of renewable energy sources presents challenges in predicting their production capacity. The abil- ity to accurately forecast the evolution of renewable energy's stake in the dynamic and ever-evolving energy market is a critical component in the de- cision making process of policy makers, and market participants alike. This project aims to explore and evaluate the performance of well-established forecasting methods in anticipating the trends of individual renewable en- ergy components, ultimately contributing to the fostering of a balanced, sustainable, and reliable energy market in the EU. The primary focus is to assess auto-regressive forecasting methods and advanced models incorporat- ing moving-average, exploit seasonality of time series data, or those utilising the correlation with exogenous variables. The results are presented for data considering recent history of the most significant energy component at the European energy markets.

KEYWORDS

Data analytics, Forecasting for energy systems, Energy Market, Renewable energy

1 INTRODUCTION

Policymakers and other energy market participants often stress that they need the grid operational statistics faster [3], particularly the European energy supply including renewable energy components, and consumption. The goal of the project is to develop a framework for forecasting the individual energy components contributing to the overall energy balance. The framework consists of well-established forecasting methods which can be easily applied to the energy market data, compare the forecast of various methods and assess the their reliability. Staple goods such as energy has certain predictability in relation to time, and ARIMA model is one of the most widely used statistical forecasting model which can exploit this trait [2]. ARIMA is a generalized version of AutoRegressive Moving-Average (ARMA) model and difference between the two models is their ability to handle non-stationary data. Autoregression is a technique used in time series analysis that assumes temporal trend between the values in the dataset, and uncover a function of order p. In the auto-regression model, the variable regresses against itself.ARIMA model [4] improves the accuracy by factoring in the non-stationarity of the data. All these factors are combined to formularize ARIMA model as the following:

$$\hat{L}_N^{(d)} = C + \sum_{i=1}^{p} \psi_i L_{N-i}^{(d)} + \sum_{j=1}^{q} \varphi_j \epsilon_{N-j} + \epsilon_N \tag{1}$$

EU energy market data often exhibit seasonal patterns. SARIMA is an extension that considers these seasonal patterns:

$$SARIMA = \hat{L}_N^{(d)} + \sum_{i=1}^{P} \Phi_i L_{N-im}^{D} + \sum_{j=1}^{Q} \Theta_j \epsilon_{N-jm} \tag{2}$$

The EU power sector, just like any other sector, experiences volatility due to exogenous factors such as the climate or the economy. The exogenous variable which has influence on the power sector is collected as additional indicators in the SARIMAX model.

Along with predictive statistics, deep learning (DL) is frequently used in forecasting. Especially, DL using Recurrent-Neural-Network (RNN) with Long Short-term Memory (LSTM) layers is very popular with time series forecasting [7]. RNN repeats the process of aggregation and activation unlike normal neural networks, and LSTM extends the memory of RNNs. This regression like property of RNN makes it very effective when building a forecasting model, and it would be interesting to evaluate it's performance in comparison to conventional statistical methods.

2 EXPERIMENTAL SETUP

This sections presents the setup of the forecasting methods considered in the study and presents the data that will be used to evaluate the accuracy of the approaches.

2.1 Parametrization of the Methods

The autoregressive models need to be parametrized by a set of parameters (p, d, q). Their values are determined by *auto_arima()*[8] function of Python, a built-in search algorithm from *pmdarima* library. The hyper-parameters are selected through step-wise approach using a mixture of conjugate gradient (gradient-based), and nelder-mead's (gradient-free) optimization method depending on the data. We split our data into two parts, training set and test set. We train the model using 80% of our data and assign the remaining 20% as test data set to compare it with the forecast. For training of the model, Python's functions from *statsmodels* library[6] are used. The derived order, and seasonal order are passed to the trainer functions along with the train data set. Each model has respective trainer function provided from the *statsmodels* library [5] and are configured accordingly with required training data, order of autoregression, differencing, and moving-average. Finally, the model is fitted to the passed training data. The trained model is used to make a forecast from the starting month to the ending month, in our case considering an interval of 3 months. To avoid the decrease of accuracy due to the large forecast window, forecasts are made at quarterly intervals. This approach is particularly necessary because the ARIMA model cannot incorporate new forecast errors into its formula, creating a need for iterative recalculation when the training set is updated. Therefore, we adopt the rolling horizon approach, where after each forecast, the training dataset is expanded by one month, and the model is re-trained before forecasting the future values of subsequent three months. By implementing the rolling horizon approach, the model becomes capable of incorporating real-time updates, mimics continuous learning, and avoids forecasting biases.

2.2 Analysis

The performance of forecasting method is measured in terms of the relative error between predicted and actual values using Root-Mean-Squared-Percentage-Error (RMSPE)[1]. Lower values in RMSPE indicate better performance. RMSPE measures the average percentage deviation between predicted and actual values, providing a standardized measure of prediction accuracy that is independent of the scale of the data. The data used for forecasting is not normalized and individual energy components have radically different scale. Hence, the RMSPE value provides better insight in evaluating forecasting models. Furthermore, their standard deviation values are observed to see the general precision of the models on the data.

2.3 Data

The dataset consists of monthly energy components contributing to the energy balance on the supply and demand side. The data are published by the national authorities of European countries (usually the transmission system operators or the national statistical offices) and are collected on the level of the high-voltage transmission grid. We use the time series on two different sampling frequency, one is the monthly value, while the other has the hourly resolution. The focus of this study is put on the key energy components or the components with the high level of variability, particularly electricity demand, and renewable energy generation such as wind, PV or hydro generation. Electricity demand is crucial for understanding how much power the system needs to meet consumer needs. Demand can vary significantly based on seasonal factors (heating/cooling), economic activity, and population growth. The increasing penetration of renewables significantly impacts supply dynamics. Understanding the variability of wind and solar generation, as well as the dispatchability (controllable generation) of hydro, is critical for managing grid integration challenges and optimizing renewable energy utilization. Data from the January of 2016 onwards is chosen as this period coincides with a significant ramp-up of renewable energy deployment across Europe.

3 RESULT & DISCUSSION

Forecasting model is applied to multitude of renewable energy sources of all EU countries. To ensure accurate evaluation of the forecasting model, the scope of assessment is focused on forecasting solar energy production in France. Table 1 presents the performance of the models in forecasting the solar energy production in France. Each model iteratively forecasts 3 months starting at each month within the historical data ranging from April-September 2023. The RMSPE between the test data and model's forecast is recorded. Figure 1 represents the box-plot of the data from table 1. The visualization of the data highlights accuracy, spread, and anomalies of model's performance. Forecast performance of the models showcased in figure 1 and table 1 highlights that, in general, SARIMAX model forecasting with collection of hourly data as an exogenous variable is the most accurate with all RMSPE being 3 to 5% with the median being approximately 4%. ARIMA model has second lowest median of RMSPE with the highest precision, while SARIMA has lower median but lower precision in comparison to the DL model.

Table 1: RMSPE of solar energy quarterly forecast in France

Date	Mean	ARIMA	SARIMA	DL	SARIMAX
2023-04-01	2501.90	11.52	10.35	5.31	5.17
2023-05-01	2668.51	9.18	8.78	11.96	5.04
2023-06-01	2667.33	5.12	7.97	4.93	4.20
2023-07-01	2523.00	5.18	8.15	7.79	3.20
2023-08-01	2161.00	8.89	8.11	14.82	3.13
2023-09-01	1574.33	19.80	8.74	11.45	4.46
Standard Deviation		4.95	0.98	3.64	0.80
mean		9.94	8.68	9.38	4.20

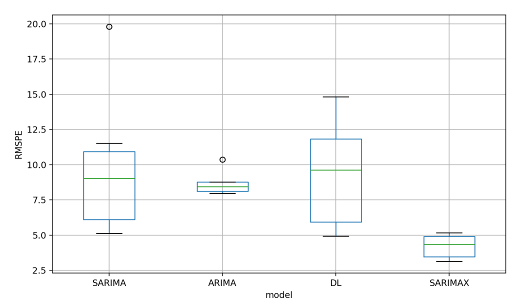


Figure 1: France solar energy production forecast performance visualization

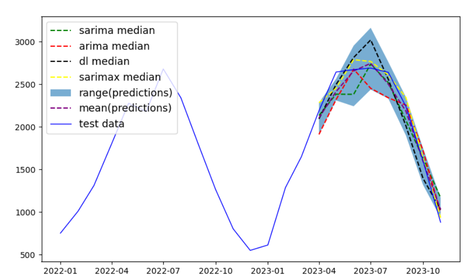
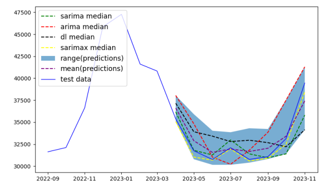
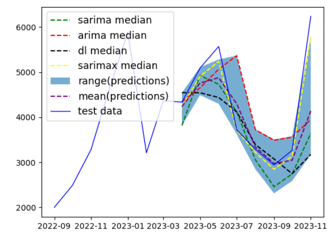


Figure 2: France solar energy production forecast

Figure 2 illustrates the median forecast values of each model against the test data. All models captured the seasonal trend of declining production from September. However, the solar energy production plateau from May to August was not forecasted correctly by any model. The models tended to predict a peak between June to July as it had occurred in the previous years. Especially the forecast made by the DL model predicted much higher peak and generally higher production of solar energy between May to August (Summer) compared to the true data. This is a reasonable expectation as the computed trend decomposition of solar energy production indicated an ascending trajectory, and in 2023, the production of solar energy surpassed that of previous years for every month until May. Although ARIMA and SARIMA models underestimated the production capacity, their accuracy was higher compared to the DL model. Possible reason for overfitting may be too many neurons in DL model, or in this case, may be due to sudden increase in residual factors which is visible from seasonal decomposition of the data. Due to seasonal component, forecasts made by SARIMA in June and July are much more accurate than ARIMA, which already started forecasting a decline in production from May. Forecast from the SARIMAX model is visibly the closest to the test data. Especially, it performs significantly better than SARIMA, its close variant. From this, we can deduce exogenous variables give significant edge to SARIMAX model, and notice the strength of well-chosen exogenous variable. However, while hourly data collection is a great exogenous variable, the time lag between availability of collection of hourly data and monthly data is very short. Hence, SARIMAX model's forecast range gets limited to the lag of the availability. As a result of combining all traits, the ensemble of forecasts accurately delineates the range within which the test data is expected to fall, as depicted by the blue highlighted region in the graphs. Furthermore, the purple dotted lines, representing a simple hybrid model averaging iterative forecasts from all models, demonstrate superior performance compared to individual models such as ARIMA, SARIMA, and DL models. In particular case such as depicted in Figure 2, the hybrid model even outperforms the SARIMAX model. Our analysis suggest that all models have unique characteristic which makes them perform better in certain scenarios. The well-established models are accurate in general, however, some models perform better than others depending on the variable. SARIMAX model promises consistently accurate and precise forecasts as it can be seen from the mean and standard deviation of SARIMAX model in the table, but it is constrained by the availability of reliable exogenous variables. Contrary to traditional models, the wisdom of artificial crowds [9] has demonstrated remarkable potential through a model that averages predictions from various models. This approach proved to be accurate, precise, and stable across all renewable energies, occasionally surpassing even the performance of the SARIMAX model. This discovery supported our belief that by tuning the wisdom of various forecasting models, it would be possible to create a hybrid model with high accuracy, precision, and consistent performance.



(a) France energy demand forecast



(b) France hydro energy production forecast

4 CONCLUSION AND FUTURE WORK

This paper has conducted an investigation into the forecasting performance of well-known statistical, and DL methods, providing insights into their strengths and limitations in forecasting demand and production capacity of various renewable energies of the EU countries. Our findings demonstrate that while conventional methods such as ARIMA, SARIMA, SARIMAX, and DL models have shown promise in modeling and forecasting renewable energy production, they have drawbacks or encounter limitations in accurately capturing rapidly shifting trend of evolving renewable energy. The conventional forecasting models showed variation in performance depending on the energy source. We realized that, for a balanced and reliable framework, there is a need for a tailor-made methodology that can better adapt to the complex dynamics of renewable energy production. During our research, we identified two promising approaches to enhance forecasting accuracy. Firstly, ensemble learning has demonstrated its effectiveness, and further development in this area could yield significant improvements. Secondly, incorporating data sources such as weather data, which have a causal relationship with renewable energy production, and utilizing their lagged correlation in the data for the SARIMAX model, could lead to better-performing frameworks. By exploiting diverse forecasting methodologies and utilising the advancements in data science, we can enhance the accuracy and reliability of renewable energy production forecasts, thus facilitating better informed decision making process of energy market participants and regulatory authorities.

REFERENCES

[1] Alexei Botchkarev. 2019. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. *Interdisciplinary Journal of Information, Knowledge, and Management* 14 (2019), 45–79. https://doi.org/10.48550/arXiv.1809.03006

[2] Walter Enders. 2014. An Introduction to ARIMA Models. *Journal of Time Series Analysis* 35, 1 (2014), 3–9.

[3] Eurostat. Accessed April 2024. Energy Nowcasting Challenge - Electricity. https://statistics-awards.eu/competitions/8.

[4] Tao Hong. 2019. *Core Concepts and Methods in Load Forecasting*. Springer.

[5] Samir Madhavan. 2015. *Mastering Python for Data Science: Explore the World of Data Science Through Python and Learn How to Make Sense of Data*. Packt Publishing Ltd.

[6] Sebastian Raschka and Vahid Mirjalili. 2019. *Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques*. O'Reilly Media.

[7] N. Raval et al. 2018. Forecasting Financial Time Series Using ARIMA vs. LSTM. *Journal of Finance and Economics* 42, 3 (2018).

[8] John Smith and Alice Johnson. 2020. Automated Time Series Forecasting with the pmdarima Python Library. *Journal of Machine Learning Research* (2020).

[9] Roman Yampolskiy and Ahmed El-Barkouky. 2011. Wisdom of artificial crowds algorithm for solving NP-hard problems. *IJBIC* 3 (01 2011), 358–369. https://doi.org/10.1504/IJBIC.2011.043624

Poster Abstract: A Flexible and Holistic Multi-Agent Framework for Local Energy Aggregation Approaches

MICHAEL BETTERMANN* and HERMANN DE MEER*, University of Passau, Germany

CCS Concepts: • **Computer systems organization** → *Grid computing*.

Additional Key Words and Phrases: Multi-Agent System, System Architecture, Energy Optimisation, Energy System

1 INTRODUCTION AND RELATED WORK

A growing number of communities have expressed a desire for self-sufficiency and the reduction of carbon emissions. These are two of the reasons why the interest of self-governed local energy community approaches has caught the attention of many. In [2], the notion of locality is kept rather vague. However, often the range of a low- or medium-voltage grid level is chosen for, e.g., local energy market designs [1], which coincides with the locality definition in this paper. A key difference between these concepts lies in the coordination approach. For instance, local energy markets focus on utilising market mechanisms, while other aggregation concepts, such as energy cells, rely on more technical approaches and may directly control assets to optimise energy consumption. This paper refers to these methodologies as local aggregation approaches. The goal of this paper is to provide a holistic and flexible model of local aggregation approaches, which provides structure during the design phase, but leaves enough flexibility to allow for adjustments, according to the local communities' needs and capabilities. Throughout this paper, a *holistic* model refers to a model, which operates in its whole local spectrum (e.g., a low-voltage grid), summarising all major entities and their concerns. The contributions of this paper are:

1. An approach for the development of a holistic model of local aggregation approaches, allowing a coherent realisation of variations, such as virtual power plants or energy islands
2. A holistic model which allows the integration of:
   - (a) Infrastructure provider signalling for intelligent grid operation, such as flexible grid tariffs
   - (b) Communication with external entities, such as the wholesale market
   - (c) Coordination algorithms adjusted to the local environment
   - (d) Strategies for preference maximisation on the consumer level

The realisation of a local aggregation approach is not a trivial task, since different stakeholders, e.g., consumer and infrastructure providers, with different goals and capabilities have to interact and coordinate properly. Several approaches exist trying to realise such a system, but those mostly focus on specific sub-parts of the entire system. For instance, in [7], a electricity market is modelled, where prosumers can trade with each other or with the network operator. The model provided by [5] separates the prosumer into an energy controller and a market controller, where the latter utilises the information from the energy controller to form the optimal bid. However, none of these examples propose a holistic approach and only focus on a subset of the whole system. To be able to form a complete concept, individual stakeholders have to be identified and their dependencies need to be analysed thoroughly to get a deeper understanding of the underlying system. The information gathered from this analysis then needs to be formed into a general model, which provides a clear structure for local aggregation approaches, while still being flexible enough to allow space for the integration of new concepts.

*Both authors contributed equally to this research.

Authors' address: Michael Bettermann, Michael.Bettermann@uni-passau.de; Hermann de Meer, Hermann.DeMeer@uni-passau.de, University of Passau, Innstraße 41, Passau, Bayern, Germany, 94032.

2 SYSTEM ARCHITECTURE MODELLING

This section addresses the process of defining the Business Use Case (BUC) of a general model for the energy procurement of local aggregation approaches. By identifying the stakeholders involved, and analysing their dependencies both individual and shared goals are reached. For this, the Smart Grid Architecture Model (SGAM) is applied, which is a well-established architectural framework to model the interactions between entities inside the smart grid. The authors of [3] define multiple business models, usually created by citizen groups trying to benefit from cheaper energy supply. Each of these business models often involves the proactive matching of supply and demand. Therefore, the general BUC for local aggregation approaches can be summarised as follows:

The feasible and efficient coordination of local energy-prosuming entities.

The notion of locality is defined as one low-voltage or medium-voltage grid. This conforms with the locality definition of energy islands [6]. This general use case can be split into two High-Level Use Cases (HUCs):

HUC1: Optimal coordination of consuming and producing entities
HUC2: Optimal operation of the power grid

HUC1 focuses on the optimal coordination between local entities to establish an agreement on consumption and production matching. This typically consists of multiple energy providers or consumers and a centralised entity, which monitors this process. This coordination optimises in accordance with the preferences of providers and consumers, but does not necessarily consider the limitations of the underlying infrastructure. HUC2 addresses the technical perspective of the coordination process and involves business models such as community prosumerism or community flexibility aggregation. This involves the infrastructure provider, which verifies the feasibility of the coordination process by checking for network constraints such as congestion, and an external entity, which consumes or provides energy in case of over- or underproduction, respectively.

Utilising the descriptions of the BUC and the corresponding HUCs, the involved stakeholders can be identified. In the following, a short description of each stakeholder is provided:

Prosumer: The goal of a prosumer is to consume or provide energy in a preferable way. These preferences can range from minimal operation cost to reduced carbon emissions.

Management unit: The goal of the management unit is to optimally coordinate the consumption and production inside the locally delimited area. The objective of the coordination process depends on the environment of the underlying system.

Infrastructure provider: The primary goal of the infrastructure provider is the economic, efficient, safe, and secure operation of the underlying network, which can be realised by a bilateral communication between management unit and infrastructure provider.

External energy connection: The goal of the external energy connection is to provide and consume energy in an economic way. Additionally, the external connection guarantees the energy balance inside the local network by providing or consuming energy from external sources (e.g., acquired from the wholesale market).

The identified HUCs are further decomposed into multiple Primary Use Cases (PUCs) and the interrelation between each PUC is depicted in an activity diagram for each HUC.

Table 1. Primary use case sketching for high-level use case 1: optimal coordination of Consuming and Producing entities.

Optimal Coordination of Consuming and Producing Entities	
PUC1	Information Collection
PUC2	Coordination Process
PUC3	Information Publication

HUC 1 depicted in Table 1 consists of three PUCs. First, the information necessary for the coordination process needs to be collected (PUC1). The content of this information depends on the information that participants can or are willing to provide. PUC2 can be achieved by, e.g., aggregating prosumers and then buying or selling the residual energy from external sources. At last, other entities are informed about the results of the coordination process (PUC3).

Table 2. Primary use case sketching for the high-level use case 2: optimal operation of the power grid.

Optimal Operation of the Power Grid	
PUC4	Identification of Network Constraints
PUC5	Utilisation of System Mechanisms
PUC6	Information Publication

Table 2 shows the three PUCs derived from HUC2. First, the identification of network constraints need to be considered (PUC4), which, e.g., involves line limitations. Second, the mechanisms provided by the aggregation approach is utilised (PUC5), such as dynamic grid tariffs. At last, the respective entities are informed about the measures taken (PUC6). Following the identification of individual PUCs, primary use cases can be analysed and described in detail by utilising, e.g., sequence diagrams. A summarising sequence diagram (cf. Figure 1) of the entire system is presented, to provide an outline of interactions between stakeholders.

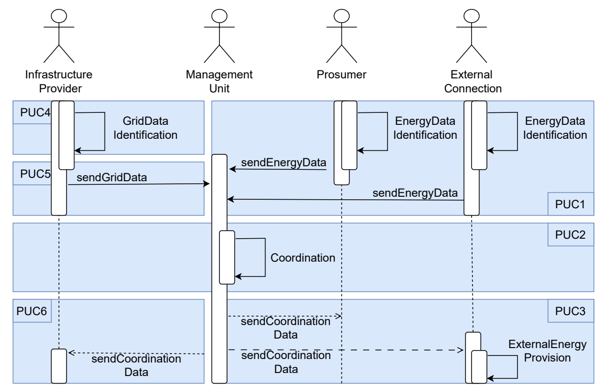


Fig. 1. Sequence diagram sketching the interactions between stakeholders, when realising high-level use cases 1 and 2

Figure 1 provides an overview of the interactions between stakeholders for each PUC. Note that this figure only sketches the message exchanges between stakeholders. The actual syntax and semantics of the message exchange are highly dependent on the design of the local aggregation approach.

3 MULTI-AGENT SYSTEM DEFINITION

To exploit the advantages of the SGAM, such as the provision of a structural approach to define an architecture for involved actors, and the benefits of agent-based systems, such as the design and implementation of software systems, both methodologies need to be connected. In addition to aligning the stakeholders to logical actors as defined in the SGAM, an agent mapping is carried out (Figure 2). To provide more in-depth description of each stakeholder, the coordination of major logical functionalities, so-called modules, of agents is defined as concretely as necessary, but as generically as possible to cover a wide range of implementation options.

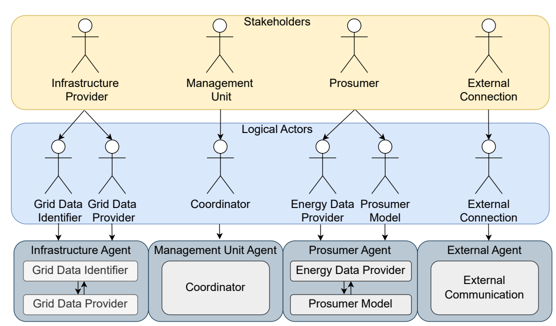


Fig. 2. Relationship between agents, stakeholders and logical actors.

Figure 2 shows the derived agents of the identified logical actors of each stakeholder. An in-depth explanation of each module in the internal structure of the derived agents is provided in the following:

Infrastructure agent.

(a) **Grid data identifier:** The grid data identifier maintains known infrastructure information (e.g., line capacities) and establishes a model of the underlying network to analyse the impact of scheduled energy prosumption. Since knowledge about the power grid can be obtained by different methods (e.g., power flow analysis), the grid data identifier may be implemented differently depending on the available capabilities or equipment.

(b) **Grid data provider:** The grid data provider communicates with the grid data identifier to obtain grid-related data, which is then utilised to determine the necessary information for the coordination process (e.g., dynamic grid tariffs). This information is then forwarded to the management unit agent.

Management unit agent.

Coordinator: The design of the coordination method is vital to the other agents, since it defines the information that the agents have to provide or will receive.

Prosumer agent.

(a) **Prosumer model:** The prosumer model is an abstraction of the underlying local energy system. The minimal requirements the prosumer has to fulfil is the provision of an estimation of the scheduled energy prosumption, since it is necessary for the management unit to properly coordinate the entities in the system. Optionally, the prosumer model reacts to information sent by the energy data provider (e.g., dynamic prices) to adjust its scheduled prosumption accordingly.

(b) **Energy data provider:** The energy data provider serves two major functions in the model. First, the energy data provider is responsible for the communication with the management unit, by translating the information gathered from the prosumer model. Second, by processing the information obtained from the coordinator, the module relays information to the prosumer model (e.g., reduced consumption prices).

External agent.

External communication: The external communication handles the communication with entities outside of the local environment. First, it provides information about expected energy cost from external sources to the management unit. Second, once the coordination process ends, the external agent uses the information of the aggregated consumption of local participants and ensures the energy provision from external sources (e.g., at the wholesale market).

In the proposed agent model, the intrinsic structure of these entities is encapsulated into modules, which increases the flexibility by allowing a wide range of approaches to be implemented. For instance, the prosumer model may be a sophisticated energy management system, such as in [4], or a collection of heuristics estimating the behaviour of a simple household.

4 CONCLUSION

The paper presents a flexible and holistic MAS framework for local energy aggregation approaches such as local energy markets. A systematic approach is shown, following the principles of the business and function layer of the SGAM, and derives agents from the logical actors of the defined stakeholders, establishing a MAS and allowing the incorporation of, e.g., energy management systems or coordination schemes. Even though the model provides a holistic view on the low- or medium-voltage level, it does not provide a holistic view on the entire power grid. While the external agent defines a gateway to the external grid and potentially may function as a prosumer agent for an aggregation approach on a higher-voltage grid level, the coordination and communication between lower- and higher-level local aggregation approaches is still unanswered.

ACKNOWLEDGMENTS

This paper has received funding by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957845: "Community-empowered Sustainable Multi-Vector Energy Islands - RENergetic"

REFERENCES

[1] Timothy Capper, Anna Gorbatcheva, Mustafa A. Mustafa, Mohamed Bahloul, Jan Marc Schwidtal, Ruzanna Chitchyan, Merlinda Andoni, Valentin Robu, Mehdi Montakhabi, Ian J. Scott, Christina Francis, Tanaka Mbavarira, Juan Manuel Espana, and Lynne Kiesling. 2022. Peer-to-peer, community self-consumption, and transactive energy: A systematic literature review of local energy market models. *Renewable and Sustainable Energy Reviews* 162 (2022), 112403. https://doi.org/10.1016/j.rser.2022.112403

[2] Aura Caramizaru, Andreas Uihlein, et al. 2020. *Energy communities: an overview of energy and social innovation*. Vol. 30083. Publications Office of the European Union Luxembourg.

[3] Inês F.G. Reis, Ivo Gonçalves, Marta A.R. Lopes, and Carlos Henggeler Antunes. 2021. Business models for energy communities: A review of key issues and trends. *Renewable and Sustainable Energy Reviews* 144 (2021), 111013. https://doi.org/10.1016/j.rser.2021.111013

[4] Stepan Gagin, Michael Bettermann, and Hermann de Meer. 2023. Multi-vector optimization scheme for distributed components in energy islands. *e & i Elektrotechnik und Informationstechnik* 140, 5 (2023), 460–470.

[5] Rahim Ghorani, Mahmud Fotuhi-Firuzabad, and Moein Moeini-Aghtaie. 2019. Optimal Bidding Strategy of Transactive Agents in Local Energy Markets. *IEEE Transactions on Smart Grid* 10, 5 (2019), 5152–5162. https://doi.org/10.1109/TSG.2018.2878024

[6] Sonja Klingert, Michael Niederkofler, Hermann de Meer, Mona Bielig, Stepan Gagin, Celina Kacperski, and Matthias Strobbe. 2023. The Best of both Worlds : Social and Technical Challenges of Creating Energy Islands. In *Proceedings of the 12th International Conference on Smart Cities and Green ICT Systems*, Cornel Klein and Matthias Jarke (Eds.). SCITEPRESS - Science and Technology Publications, Setúbal, Portugal, 129–136. https://doi.org/10.5220/0011974600003491

[7] M. Yebiyo, R.A. Mercado, A. Gillich, I. Chaer, A.R. Day, and A. Paurine. 2020. Novel economic modelling of a peer-to-peer electricity market with the inclusion of distributed energy storage—The possible case of a more robust and better electricity grid. *The Electricity Journal* 33, 2 (2020), 106709. https://doi.org/10.1016/j.tej.2020.106709

Field Survey of Wireless M-Bus Encryption for Energy Metering Applications in Residential Buildings

FRIEDRICH HILLER V. GAERTRINGEN, JOHANNES GALENZOWSKI, KAIBIN BAO, SIMON WACZOWICZ, and VEIT HAGENMEYER, Karlsruhe Institute of Technology (KIT), Germany

The wireless Metering-Bus (M-Bus) is widely used in Germany to transmit meter data for heat cost allocation as well as cold and warm water consumption in multi-family apartment buildings. This metering data poses significant privacy risks as it can reveal inhabitants' behaviors. Consequently, the German Heating Cost Ordinance demands this transmission to be both interoperable and secure. However, the wireless M-Bus standard EN 13757 specifies security features as optional. In our work, we conducted a field study by recording sensor telegrams in four cities to assess the implementation of these security features. We analyzed the presence of encryption and the types of metering applications in use. Our findings reveal that about 48.5 % of the recorded sensor devices did not have encryption enabled. Additionally, the use of encryption was found to correlate with specific manufacturers, indicating a systematic acceptance of privacy risks. To demonstrate the impact of unencrypted wireless M-Bus radio telegrams on the privacy, we recorded a wireless M-Bus based warm water meter over a period of several weeks and show that inhabitants' presence and sleep cycles can be inferred from the recordings. These findings underscore the need for mandatory security features in the operation of wireless M-Bus based metering applications to protect consumer privacy.

CCS Concepts: • **Security and privacy** → **Mobile and wireless security**; **Social aspects of security and privacy**; • **Hardware** → *Energy metering*; • **Networks** → Transport protocols.

Additional Key Words and Phrases: Wireless M-Bus, Encryption, Privacy, Field Survey, Cybersecurity, Submetering

Availability of data:
The pseudonymized and processed data of this field study is available online at https://doi.org/10.5281/zenodo.13234545.

1 INTRODUCTION

The German Heating Cost Ordinance (HeizkostenV), amended in 2021, demands the mandatory usage of remotely readable heat cost allocators and warm water meters in buildings in Germany (§ 5 Section 2 of HeizkostenV [7]). The implementation of this ordinance is associated with the obligation to use an interoperable communication standard. In Germany, wireless M-Bus (wM-Bus) is used for this application (Section 3.3.5.2.1 of the BSI-TR-03109-1 [6]). The wM-Bus protocol utilizes frequencies in the ISM band (Industrial, Scientific and Medical Band), as described in EN 13757-4.

For our explorative study, we analyzed broadcasting communication modes. In these modes, each sensor data concentrator records every received radio telegram, and then backend processing later determines which records are used. Following the standard EN 13757, the metering data in the payload of the telegrams can be encrypted to ensure confidentiality. This should ensure that only the metering operator can decrypt the metering data.

Authors' address: Friedrich Hiller v. Gaertringen, friedrich.gaertringen9@kit.edu; Johannes Galenzowski, johannes.galenzowski@kit.edu; Kaibin Bao, kaibin.bao@kit.edu; Simon Waczowicz, simon.waczowicz@kit.edu; Veit Hagenmeyer, veit.hagenmeyer@kit.edu, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131, Karlsruhe, Germany.

As mentioned, the widespread application of wM-Bus is to be expected due to legal requirements. Yet, we identified that statistics about the usage of wM-Bus in the field are not reflected in scientific literature. We saw a need to conduct a field study on how wM-Bus is used in practice. This study also investigates whether there are any obvious security risks to the privacy of residents due to the operation of these continuously transmitting devices.

Consequently, in our work, we target the following research questions:

- **RQ1** Which impact on privacy do meters using wireless M-Bus have in practice?
- **RQ2** Is it possible to capture, interpret and spatially allocate wireless M-Bus telegrams in the field?
- **RQ3** Are the captured wireless M-Bus telegrams protected by encryption or other security measures?
- **RQ4** Do the unprotected telegrams or unprotected telegram parts expose sensitive data?

The remainder of the paper is structured in the following way: We investigate at the related literature in Section 2. In Section 3, we present the most relevant parts of the wM-Bus standard which we used to conduct the field study and data analysis. The design of the device to capture and spatially allocate wM-Bus telegrams is described in Section 4. Subsequently, we present the results of our field study in Section 5, including statistics of used encryption, device types and a demonstration of the impact on privacy. In Section 6, we discuss the limitations of our field study. Finally, we describe the responsible disclosure process of our findings in Section 7 and conclude our study in Section 8.

2 RELATED WORK ON WIRELESS M-BUS SECURITY

Most related literature focuses on the advancement of the functional aspect of wM-Bus. This includes the work of Squartini et al., who explore wM-Bus in the context of smart water grid applications, focusing on energy consumption and efficiency [17]. The work of Spinsante et al. [15, 16] investigates the suitability and efficiency of the wM-Bus protocol for smart water grids. Spinsante et al. highlight the wM-Bus protocol's potential in automatic monitoring and smart metering of water consumption with an energy-efficient network architecture. Pavel Masek et al. [9, 10] demonstrate the potential of wM-Bus protocol for smart electricity grids and 5G-grade home automation.

Another line of research focuses on the privacy issues associated with smart metering. Lisovich et al. [8] demonstrate the possibility to infer which movies inhabitants watch based on a time series of the household's electricity consumption. Chen et al. [5] propose a statistical framework for disaggregating utility consumption from smart meters with low sample rates into specific appliance usage associated with human activities. Asghar et al. [2] review the uses of metering data in the smart grid and related privacy legislation. They provide a structured overview of security solutions for privacy-preserving meter data delivery and management. Moreover, they survey recent work on privacy-preserving technologies for billing, operations, and value-added services including demand response. As a key result, they identify the need for metering solutions that balance privacy concerns with the functionality of smart metering systems.

Researchers also propose extensions of the wM-Bus protocol for enhanced security. Anani and Ouda [1] propose a security framework for wM-Bus within the smart grid communication network, introducing a new lightweight security profile with less impact on battery life.

A limited number of publications perform a security analysis and expose shortcomings of the protocol or issues of the real-world implementations of wM-Bus. Polčák and Matoušek [12] highlight a concern with wM-Bus meters that monitor energy consumption at short intervals, such as every minute, increasing privacy and security risks. In their case-study, they investigated devices from two wM-Bus vendors and found four CVE[1]-indexed security vulnerabilities:

Broken Key Management The authors demonstrated that metering data acquisition devices of a manufacturer could receive meter data without specifying the encryption key. Either a shared key is used across all devices, or the key can be derived using the serial number.

Zero Consumption Detection Even though the payload of the wM-Bus telegrams are encrypted, it is still distinguishable whether the metering values change between transmissions.

Vulnerability to Replay Attacks The readout software does not adequately detect the replay of previously recorded telegrams.

Misleading Event Detection The readout software showed misleading events related to tampering detection.

Additionally, a white paper by Brunschwiler [4] presents a detailed security analysis of the wM-Bus standard. Several vulnerabilities were found in this paper, including:

Inadequate Key Length The standard suggests that only half of the key length shall be different for each meter, reducing the key strength significantly.

Zero Consumption Detection Inappropriate key and initialization vector use allows inferring unchanged meter counter values, indicating zero consumption detection.

Consumption Values Inappropriate key and initialization vector derivation may disclose plain texts including consumption values.

Manipulation of encrypted telegrams Missing integrity protection allows for manipulation of consumption values in transit.

Clock synchronization Lack of authentication with clock updates may lead to key stream repetition.

Network Management Lack of authentication for network management could allow adversaries to become a rogue relay.

Unencrypted tamper detection Plain text error and alarm notifications allow an adversary to recognize if tamper switches have been triggered.

Unencrypted identification Unencrypted disclosure of device manufacturer, meter type and version ID in the device's telegrams simplify identification of vulnerable targets.

Key Management Loosely specified key update mechanism may lead to key disclosure. Some transceiver chips support the transparent decoding of encrypted telegrams. The encryption key is not protected and can be read out.

Finally, there are open-source implementations to transceive or decode wM-Bus radio telegrams. For example, wmbusmeters[2] is a decoder for wM-Bus telegrams, including some manufacture-specific encodings. Also worth mentioning is the rtl-wmbus[3] library, a Software Defined Radio library for decoding the physical layer of wM-Bus.

All existing works do not include extensive field studies investigating the usage and security of a large number of metering devices in real-world operation. Existing work does not always describe the exact number of investigated metering devices in total. But where numbers are given, either the number of metering device models or the number of individual metering devices were below 20. By recording wM-Bus signals in the field, we expect to capture at least several hundred individual metering devices.

Our field study targets the research gap regarding real-world usage and draws inspiration from Peter Shiple's concept of War-Driving presented at the DEF CON 9 conference [14]. War-Driving describes the process of mapping Wi-Fi networks by logging Wi-Fi Access Point beacons and correlating it with a location gathered using a Global Navigation Satellite System. In a non-representative initial study dating back to the year 2000, Shiple located around 1500 Access Points from which 85% were not using encryption [14]. Ten years later, Said et al. scanned 1228 Access Points in Dubai from which 35% were not encrypted and 26% used weak encryption [13].

3 WIRELESS M-BUS TELEGRAM PROCESSING BASED ON THE EN 13757 STANDARD

This chapter aims to explain all the technical principles necessary for analyzing the wM-Bus data collected in our field study. As the EN 13757 wM-Bus standard comprises eight parts and over 750 pages, this section cannot provide a comprehensive breakdown. Rather, it specifically isolates and describes the knowledge that is essential for our use case. It for the proposed research questions, it is essential, to differentiate whether an individual sensor uses encryption, of what device type it is, and from which manufacturer it originates. Additionally, this chapter describes how sensor data of unencrypted wM-Bus telegrams are decoded to access the impact on privacy.

[1]Common Vulnerability Enumeration, https://cve.mitre.org/

[2]https://github.com/wmbusmeters/wmbusmeters

[3]https://github.com/xaelsouth/rtl-wmbus

3.1 History and overview of EN 13757

The Metering-Bus (M-Bus) was originally developed by Horst Ziegler in 1992 [3]. Originally, DIN EN 1434 proposed the M-Bus protocol for the metering of heat energy consumption. Later, M-Bus is described as a separate standard set DIN EN 13757. Following the European Union Directive 2006/32/EG, which set the goal to save final energy consumption through optimizing energy-related services and other energy efficiency measures by 9%, a new means for measuring energy consumption was required. The Open Metering System (OMS) was developed to address the metering of energy consumption and proposed a technical solution based on the M-Bus protocol.

At the time of publication of the present paper, the European standard EN 13757 specifies all aspects of the (wireless) M-Bus protocol, from the physical layers up to the application layer. The EN 13757 standard is divided into eight parts, as shown in Table 1.

Table 1. Overview of the eight parts of the M-Bus standard EN 13757.

Part Nr.	Brief Description
EN 13757-1	General overview of the data exchange and communication, as well as the protocol, used on different layers for remote reading of metering values.
EN 13757-2	Physical and data link layer of the *wired* M-Bus protocol.
EN 13757-3	Application layer of the meter bus protocol, which is independent of the physical transmission medium.
EN 13757-4	Physical and data link layer of the *wireless* M-Bus protocol.
EN 13757-5	Retransmission, relaying and routing for the wireless protocol.
EN 13757-6	Description of a wired local bus for meter readings over short distance as an alternative to the M-Bus specified in part two.
EN 13757-7	Definition of the session and transport layers, including transport layer encryption.
EN 13757-8	Extension to usability of M-Bus in wireless communication outside the standard of part four.

In the following subsections, we will briefly outline which of these parts are particularly relevant and structure their content to enable the extraction of the relevant information from the wM-Bus telegrams.

3.2 OSI-layer-wise description of M-Bus telegrams

We obtain a structured overview by mapping each part of EN 13757 to the layers of the OSI reference model in Table 2. This allows a dedicated analysis of the different sections of a M-Bus telegrams.

Starting from the physical layer at the bottom of Table 2, the first distinction is between wired and wireless M-Bus. Parts two and six of the standard cover the wired part and can therefore be discarded for a field study focussing on wireless drive-by readout. Since in the case of meters, there is only the sending device and no forwarding via gateways, routed transmission as described in EN 13757-5 is also disregarded. Therefore, only EN 13757-4 is the decisive standard for analyzing the physical layer in the context of our study (see Section 3.2.1).

The parts of EN 13757 relevant for layer 2 to 7 are also depicted in Table 2. Our particular interest in encryption requires a more

Table 2. OSI layers of M-Bus and related standard parts (based on EN 13757-1 Table 3 and EN 13757-7 Table 1).

OSI layer		Wired M-Bus	Wireless M-Bus
7	Application Layer	EN 13757-3	
6 5 4	Presentation Layer Session Layer Transport Layer	optional, depending on CI-Field (EN 13757-7)	
3	Network Layer	optional, depending on CI-Field (EN 13757-5)	
2	Data Link Layer	EN 13757-2 / extended by EN 13757-4 / based on IEC 60870-5	
1	Physical Layer	wired M-Bus (EN 13757-2) wired local bus (EN 13757-6)	direct transmission (EN 13757-4) routed transmission (EN 13757-5)

in-depth analysis of the layers 2 (13757-2 and 13757-4) and 4 (13757-7). Furthermore, to analyze whether privacy-relevant data can be extracted, it is necessary to look at the application layer (EN 13757-3). To summarize, we need an analysis of the standard parts 2, 3, 4, 5, and 7.

3.2.1 Physical Layer (EN 13757-4). An understanding of the physical layer is primarily relevant for designing a suitable data acquisition device. The physical layer of wM-Bus can be operated in different modes that are optimized towards specific use cases. Mode S is used for communication among stationary devices. Mode F and N are used for longer distances. Mode C and T are optimized towards drive-by readouts.

The radio frequencies and symbol rates of each mode are different, such that a transceiver usually needs to be set to a specific mode. However, the frequency and symbol rate of mode C and T used for the channel from the sensor to the collector is equivalent. That is, the receiver can receive both C and T telegrams if tuned to the same frequency and symbol rate. Despite this, the encoding schemes are different. The byte rate also differs due to the encoding scheme. However, using a separate decoding scheme for each mode, C and T mode can be received in parallel.

The modes are further divided into sub-modes to distinguish between unidirectional communication (designated as S1, C1, etc.) and bidirectional communication (designated as S2, C2, etc.).

In our work, we only analyze sensor telegrams received by passively listening. We do not request the sensors for readouts and rely on unsolicited frequent transmissions. This field study therefore only considers mode C1 and T1.

3.2.2 Data Link Layer (EN 13757-2 and EN 13757-4). The data link layer consists of a header containing the length of the telegram, manufacturer and serial number, as well as the device type and telegram type contained in the telegram (see Figure 1). Another part of the data link layer is the trailer field, which contains a checksum for cyclic redundancy check (CRC). Since integrity checking is commonly integrated in the receiving device, the CRC is relevant when selecting a suitable data acquisition device, but is not required

e.g.	Data Link Layer (2)	
39_h	L-Field	
44_h	C-Field	
43_h	**M-Field**	
04_h		
76_h	Serial Number	A-Field
35_h		
25_h		
12_h		
18_h	Device Version	
06_h	**Device Type**	
$7A_h$	CI-Field	
	optional: if CI = $8C_h$ or **$8D_h$** Extended Link Layer (see EN 13757-4)	

Fig. 1. Example of a data link layer of one of our captured telegrams. In **bold** the fields included in our statistical evaluation. For the overview of a full telegram, see Figure 12.

to be analyzed for our research questions. It is important to note that the header fields of the data link layer is always unencrypted, which always allows the identification of the sending device. All the header fields required for our analysis are briefly explained below.

L- and C-Field. The one byte long L-Field states of how many bytes the telegram consists of (for example, 39_h indicates a telegram with 57 bytes). The subsequent one byte long C-Field describes the sending mode. As described in Section 3.2.1 we captured communication mode C1/T1. With only listening to one-directional openly broadcasted telegrams, we only captured C-Fields with 44_h corresponding to "SEND/NO REPLY, Meter initiative" according to EN 13757-4. These two fields are neither statistically evaluated nor technically required for further evaluation.

M-Field. The first field of the header that we evaluated statistically is the M-Field. The M-Field contains the manufacturer ID, which is assigned by the DLMS User Association and encoded according to EN 13757-7 7.5.2 in the form of three letters coded in two bytes. We evaluate the M-Field to be able to analyze differences in encryption usage between manufacturers.

A-Field. The A-Field serves two purposes. Firstly, it allows the device to be unambiguously identified. This is possible through the unique combination of the four byte serial number and the one byte device version. This part of the A-Field not statistically evaluated, but is technically required to identify and group several telegrams from the same device. Secondly, the A-Field contains one byte specifying the device type. The device type field was evaluated according to EN 13757-7 Table 13.

CI-Field. The most important field for subsequent decoding is the CI-Field. For all layers above the data link layer, CI-Fields are used to specify what the next telegram segment looks like. For our evaluation, it is relevant to be able to understand the following multiple numbers of telegram sequences so far, until the segment concerning encryption or interpretable user data is encountered and can be evaluated. The segment containing the relevant information may already be defined in the CI-Field found on the data link layer, but may also be encoded in the subsequent telegram segment on the transport layer or the application layer. In those cases, further CI-Fields are present on the data link layer. In any case, the CI-Fields must be evaluated segment by segment to be able to analyze the subsequent data.

The lowest level at which an insight can be gained is in the data link layer in the case of a CI-Field of $8D_h$. A CI-Field of $8D_h$ indicates that the telegram is encrypted at the data link layer. We can therefore count these telegrams as "encrypted at the data link layer" meaning no further evaluation of their content is done.

In the case of all other CI-Fields we encountered, it is necessary to decode also the next higher (transport) layer, according to the CI-Field defined in the header of the data link layer. In the simplest case, those CI-Fields are CI = 78_h (no header), CI = 72_h (short header) and CI = $7A_h$ or $7B_h$ (long header).

In a more complex case, we find some further sequences between the initial CI-Field of the header and the CI-Field of interest (also 78_h, 72_h, $7A_h$or $7B_h$). This is the case when an extended link layer exists, that is not encrypted (CI = $8C_h$) and we find a further sequence on the transport layer (CI = 90_h), the so-called Authentication and Fragmentation Layer (AFL). Only after those two sequences, we then find the CI-Field of interest for further analysis.

Data Link Layer Encryption (EN 13757-4). Taking a further look at the encryption, we observed that the majority of encryption is applied at the transport layer (see Section 3.2.3). In our data set, only a total of 18 telegrams were encrypted at the data link layer, indicated by a CI-Field of $8D_h$. The used encryption scheme for the data link layer encryption (DLL encryption) is AES-128 in Counter Mode (CTR). The Initial Vector consists of a relative timestamp, a session counter and a block counter. The telegram payload does not need to be padded, as the CTR mode produces a cipher stream that can be truncated at any position.

3.2.3 Transport Layer (EN 13757-7). After deriving possible CI values of 78_h, 72_h, $7A_h$ or $7B_h$, in the following, we discuss, how subsequent telegram sequences can be analyzed on the transport layer (also see Figure 2). For this, we have to differentiate the case for each individual CI value. Telegrams with no header on the transport layer (CI = 78_h) can be classified as not encrypted according to the EN 13757 standard. For those, the next analysis step can continue on the application layer. For the other cases with a short or long header, a Configuration Field exists in the telegram, which is explained in more detail later. In addition to the Configuration Field, other header fields like ACC and STS are present (as shown in Figure 2).

ACC- and STS-Field. The one byte access number in the ACC-Field is counted up when new values are transmitted in a new telegram to indicate that it is not simply a retransmission of the same telegram content. The one byte status field (STS-Field) describes the error state of the device. The STS-Field's bits zero and one (underlined) define the error type:

- $0000\ 00\underline{00}_b$= no error
- $0000\ 00\underline{01}_b$= application busy
- $0000\ 00\underline{10}_b$= arbitrary application error
- $\cdots t\ pe\underline{11}_b$= unusual state/alarm

e.g.	Session / Transport layer (4) → different depending on CI-Field		
	optional: if CI = 90_h Authentication and Fragmentation Sublayer (see EN 13757-7)		
	if CI = $7A_h$ or $7B_h$ Short Header:	if CI = 72_h Long Header:	if CI = **78_h** No Header:
96_h	ACC-Field	Structure according to EN 13757-7	-
00_h	STS-Field		
00_h	**Configuration Field**		
20_h			

Fig. 2. Example of the transport layer of one of our captured telegrams. In **bold** the fields included in our statistical evaluation.

Bits two to seven further specify the error. Bit two (e) is set to one if energy supply is low, bit three (p) is set if a permanent error is present, and bit four (t) is set if a temporary error is present. Bit five's to seven's functions are not standardized, and reserved for manufacturer-specific error codes. The ACC- and STS-Fields are not used in our evaluation.

Configuration Field. The Configuration Field consists of two bytes. It specifies the protection mode (type of encryption), other required bits, further required transport layer fields (header or trailer) and, if present, the Configuration Field extension. It contains information relevant for a security analysis of the wM-Bus telegrams. The information in the Configuration Field is encoded in binary. It needs to be decoded according to Table 3. In particular, we derive information about the encryption from the Configuration Field. The five protection mode bits (bits 8 to 12 as shown in Table 3) can store values from 0 to 31. The security mode is indicated directly by the number calculated from the bits, as described in the following paragraph.

Transport Layer Encryption Modes (EN 13757-7). For security and privacy, wireless M-Bus supports optional encryption modes at the transport layer for the application data. A list of all 16 encryption modes is described in EN 13757-7 (p. 33 Table 19). We observed encryption modes 0, 5, 7 and 10 being used in our field study and therefore describe them more detailed in the following.

- **In mode 0,** the application data is unencrypted.
- **In mode 5,** the application data is encrypted using the symmetric block cypher AES-128 in cipher block chaining (CBC) mode with a static key. The encryption key is static for each metering device. The Initialization vector (IV) is constructed using the manufacturer identifier, the meter identifier, version, device type and access number. All information, from which the IV is calculated, is present in the headers of the telegram.
- **In mode 7,** the used encryption is also the symmetric block cypher AES-128 in cipher block chaining (CBC). But the encryption key is dynamically derived based on a static master key, a use-case-dependent constant, a telegram counter and the sensor id. The freshness of the ciphertext is thus provided by this key derivation instead of using different IVs (in mode 5). Consequently, the IV is not needed to provide security features and is therefore set to zero and not utilized in mode 7. For mode 7, a cypher-based message authentication code (CMAC) is recommended using the authentication and fragmentation sub-layer (AFL).
- **In mode 10,** the 128-bit AES block cypher is used in CCM mode, which is counter mode with CBC-MAC. A 32-bit counter is encrypted using the AES block cypher. The resulting key stream is xored with the plaintext to get the ciphertext. A CBC-MAC is calculated in parallel where the most significant 4, 8, 12, or 16 bytes are appended to the telegram as an authentication tag. The authentication tag size is defined in the extended configuration field of the transport header. The counter is either defined as an optional field of the configuration frame or as a field in the authentication and fragmentation layer (AFL).

Mode 0 does not provide any security guarantees. Mode 5 provides confidentiality only and mode 7 optionally provides integrity and authenticity, if a CMAC is used in conjunction with the AFL. Mode 10 always provides confidentiality, integrity, and authenticity.

3.2.4 Application Layer (EN 13757-3). In cases where either no header is defined on the transport layer (CI = 78_h) or encryption mode 0 is defined in the Configuration Field, it must be assumed that the user data is not encrypted in accordance with the EN 13757 standard. In this case, it can be further investigated whether the actual application data can be extracted.

e.g.	Application Layer (7) → different depending on CI-Field	
	if CI = $7A_h$ Full Frame:	if CI = **$7B_h$** Compact Frame:
$0C_h$	**DIF**	Structure according to EN 13757-3
13_h	**VIF**	
15_h	Data	
08_h		
01_h		
00_h		
…	…	

Fig. 3. Example of the application layer of one of our captured telegrams. In **bold** the fields included in our statistical evaluation.

Payload structure. Application data is formatted either as a compact frame or a full frame (compare Figure 3). The frame type is determined by the preceding CI-Field. In the compact frame, the metadata and the actual metering data are transmitted in separate telegrams that must be combined for interpretation. This allows a data-saving (compact) transmission of the metering data. However, it also means that pure data telegrams containing a compact frame cannot be interpreted. Further knowledge is required for their evaluation. Full frames, on the other hand, contain the metering and metadata in the same telegram. We therefore use telegrams with a full frame as an example to analyze the received data regarding privacy problems. The structure of a full frame is therefore briefly described below.

Table 3. Decoding of the Configuration Field (CF) according to EN 13757-7 Table 27 (with the least significant bit (LSB) corresponding to bit zero and the most significant bit (MSB) to 15) for an unencrypted telegram example: $00_h 20_h$ resp. 00100000_b,00000000_b in binary (for details see references in table).

Bit number	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Content	Bidirectional communication	Accessibility	Synchronized	Security mode	Security mode	Security mode	Security mode	Security mode	Reserved	Reserved	Reserved	Reserved	Telegram content	Telegram content	Repeater access	Hop counter
Short	B	A	S	M	M	M	M	M	0	0	0	0	C	C	R	H
Reference	EN 13757-7 7.7.1		EN 13757-7 7.7.1	EN 13757-7 Tab 19					0b	0b	0b	0b	EN 13757-7 7.7.1. Tab 21 & 22		EN 13757-5	EN 13757-5
Example	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Description example	no access window (unidirectional meter)		synchronized transmission	no security mode									standard telegram		meters are always zero here	

Full frame. Application data, i.e., sensor values, are arranged in consecutive data records. Each data record starts with a data record header (DRH) describing the data type, encoding, length and metadata of the data record.

The DRH consists of a data information field (DIF) and a value information field (VIF). There are optional data information field extensions (DIFE), value information field extensions (VIFE) and a variable length field (LVAR).

Both standard data types (in the DIF) and standard units (in the VIF) can be used, as well as proprietary codes. The DIF values for standard data types can, for example, be found in Table 4 of EN 13757-3. The VIF values for standard units can be found in Table 10 of EN 13757-3.

In any case, knowledge of the meaning of the DIF and VIF, either from the standard or the manufacturer's documentation, is required for the interpretation of the values contained in the data block. In each of these cases, user data can be accessed and analyzed. More specific, the data information field (DIF) contains two function bits, that indicates if the corresponding value is maximum, minimum, error state or instant value (both bits of the function field are zero).

In our work, if we identify successfully that the telegram contains a full frame (CI = $7A_h$) and the first data block contains an instant value (function bits both zero), we can conclude that this telegram contains real instant metering values. In this case, we consider this telegram as interpreted. Otherwise, the telegrams would be considered as still unencrypted but also as not interpreted by us.

For our example, with a data point with DIF = $0C_h$ we know that it is an instant value, Binary Coded Decimal number. With VIF = 13_h we know, it is a volume in the range 10^{3-6} with the unit m^3. And with the data being $15_h 08_h 01_h 00_h$ we know, the value is 10815. We can therefore conclude that the given data represents the current meter value with 10 815 m^3.

With this understanding of EN 13757, the following section describes how the data of our field study were collected and analyzed.

4 DATA ACQUISITION

The goal of the field study was to collect wM-Bus communication samples from as many unique sensor devices as possible. As the sensors are stationary, the data acquisition device must be portable so that it can be transported to multiple locations to capture telegrams of different sensors. The acquisition device therefore is built to be battery-powered and also includes a GPS module to geo-reference each captured telegram. This setup allows a data collection process similar to war-driving to map wireless networks (see Section 2).

We limited our field study to listening to telegrams from continuously sending devices. In this way, our acquisition system does not interfere with the metering system. The setup consists of low-cost commercial off-the-shelf modules and enables undetectable acquisition of metering data from third-party devices.

In preliminary experiments, we noticed that many wM-Bus devices regularly send metering data with an interval between 1 and 15 minutes, which is in line with the observations by Polčák and Matoušek [12].

4.1 Portable Wireless M-Bus Data Acquisition Device

The portable data acquisition device is depicted in Figure 4. The components are referenced in the following using the numbers in the brackets.

The wM-Bus telegrams were captured with a iM871A-USB dongle from IMST GmbH (Figure 4, I). After a firmware update, the iM871A-USB is configured to capture wM-Bus telegrams in combined C1/T1 mode. The CRC-Field is not passed on by the iM871A-USB and only the header and data blocks were recorded. The geolocation was acquired using a GY-GPS6MV2 GPS Module (Figure 4, II) and its antenna (Figure 4, III). A Raspberry Pi Zero W was used as the

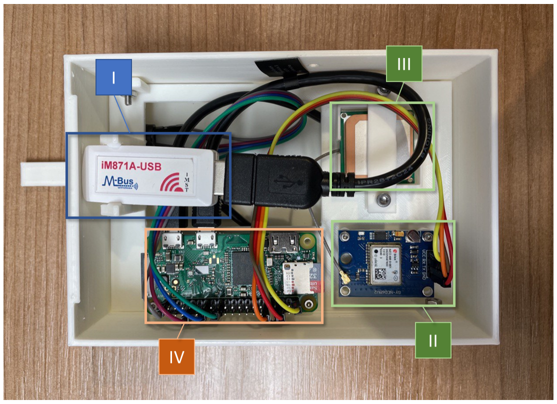


Fig. 4. The main components of our portable wM-Bus data acquisition device, referenced in Section 4.1.

processing and control unit (Figure 4, IV). Additionally, a basic buzzer was used for audible feedback each time a telegram was received. We used a 5 V USB external battery pack with 10 400 mA h to power the device. This battery allowed to operate the acquisition device for more than 24 hours. All components were mounted inside a self-designed 3D-printed enclosure.

4.2 Data Collection Process

The data acquisition device records all received telegrams together with a time stamp and the current geolocation. The recording device itself does not carry out the decoding of the received telegrams. Each of the telegrams is stored in hexadecimal representation with a sequence number, timestamp, and a numerical location identifier in CSV format. The locations are stored separately in JSON format.

The data collection was undertaken in the years 2022 and 2023 by walking through the German cities Munich, Trier, and Karlsruhe, as well as Luxembourg. The majority of data was captured in Munich. We walked along major streets of these cities, starting from multiple randomly selected starting locations. Sometimes, tramlines were used.

Decoding of the telegrams was done later in an iterative process with a proprietary library written by us in Python. The library does not implement the comprehensive wM-Bus standard but was extended incrementally to handle the observed telegrams.

5 RESULTS

In this section, we present the data collected using the data acquisition device presented in Section 4 and the knowledge about decoding derived in Section 3. In Section 5.1, we confirm RQ2 that it is possible to capture, interpret and spatially allocate wM-Bus telegrams from arbitrary sensors in the field. The metering application for which the wM-Bus devices are used and whether manufacturers specialize on specific metering applications is described in Section 5.2. In Section 5.3, we analyze which encryption mode is used by the devices which addresses RQ3, inquiring if wM-Bus telegrams are protected. Finally, Section 5.4 demonstrates how metering data recorded from wM-Bus devices impacts privacy, answering RQ4 that wM-Bus telegrams expose sensitive data. The pseudonymized and processed data of this field study is available online[4].

5.1 Statistics of the Field Study

To answer RQ2, whether it is possible to receive data from arbitrary wM-Bus sensors, we listened to wM-Bus telegrams in the German cities Munich, Trier, and Karlsruhe as well as in Luxembourg using our data acquisition device.

The telegrams were captured during several measuring rounds between 2022-07-22 and 2023-05-21. In total, 13.537 wM-Bus telegrams were recorded. As described in Section 3.2, the unencrypted header allows the unambiguous identification of the device that is sending a given telegram.

Identifying Unique Devices. The 13.537 wM-Bus telegrams could be assigned to 4.986 unique wireless M-Bus devices based on manufacturer ID and serial number.

The distribution of how many telegrams were recorded per device during the entire field study is shown in Figure 5. For about 65 % of the devices, only a single telegram was recorded. For the other 35 % of devices, two or more telegrams were received.

There was no structural difference in the telegrams that we received from the 35 % of devices from which we captured multiple telegrams. In particular, static metadata from the header has remained the same for all telegrams from the same device. Only the user data such as power consumption has changed. For this reason, in the following, we do not analyze the metadata of the header for each telegram, but for each unique device.

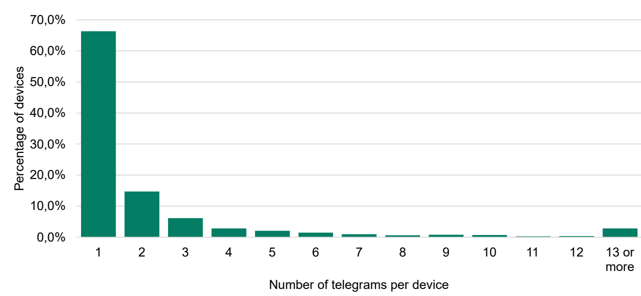


Fig. 5. The percentage of devices over the number of telegrams received from the same device. Only from one third of the devices, more than one telegram was captured during the entire field study.

Distribution of Recorded Devices by Year. The field study was conducted in the years 2022 and 2023, partially repeating the measurement in the same cities. Figure 6 shows how many unique devices were registered in the respective years. Twelve percent of the devices were recorded both in 2022 and 2023. No specific method was used to ensure that the same receiving conditions applied in both years. For example, no effort was made to be in the exact location at the exact time of day or to walk at the exact speed on the same shoulder of the road. If one wanted to compare the years not only

[4]https://doi.org/10.5281/zenodo.13234545

qualitatively, but also quantitatively, a methodically defined walking procedure would be necessary.

If the telegrams had been recorded exactly in the same area, an overlap rate of 84% to 90% would have been expected due to the calibration validity and therefore lifetime of a single meter of six years (according to annex 7 (to § 34 paragraph 1 number 1) of [11]). However, the overlap between the years is lesser, as the data was only partially recorded in the same areas. From the 4.986 devices,

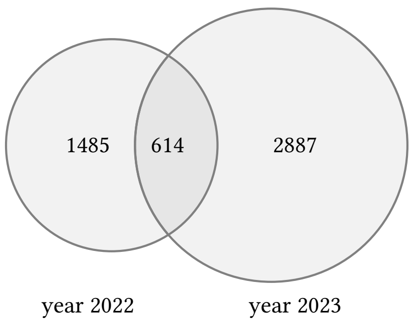


Fig. 6. Observed number of unique devices per year.

1.485 devices were only captured in 2022, 2.887 only in 2023, and 614 devices in both years (see Figure 6). This corresponds to twelve percent of devices being captured in both years.

Spacial Allocation. Consequently, it can be concluded that it is possible to record a considerable number of third-party devices with moderate effort (compare RQ2). The spatial allocation was not carried out at the apartment level. Rather, the meter's location was roughly determined via an average of the GPS positions, at which the meter's telegrams were received. This results in an accuracy of only several meters. However, other works such as [12] show that a more precise spatial allocation is technically feasible with further effort. Therefore, we can answer RQ2's aspect of spatial allocation, with: yes, it is possible to specially allocate the received telegrams.

Regarding the first general analysis before answering the question about privacy, it remains to be answered which data can be exploited from the telegrams.

5.2 Device Types and Device Manufacturers

As described in Section 3.2, all header information can be decoded in plain text according to the EN 13757 standard. Analyzing this information is used to reveal correlations between encryption modes, device types and device manufacturers. The device type is crucial information to determine the severity and duration of privacy violations. Regarding the severity, cold or warm water meters generate the most sensitive data, heating cost allocator data is medium severe and smoke detector data has the lowest impact on privacy. Regarding the duration, the exchange rate depends on the device type due to calibration period or battery lifespan.

As shown in Fig. 7, the majority of the recorded devices are heat cost allocators. These are followed by hot water meters and water meters (which do not define whether they are used for hot or cold water).

If heat cost allocators are used, they must be installed on every radiator, while heat meters and water meters (cold & hot) are usually only installed at the pipe near the connection point of the accommodation unit or at a maximum on every tap in the apartment. Based on these facts, it is plausible that most of the devices recorded are heat cost allocators. Since consumption-based billing is usually legally obligatory for centralized hot water supply, it is also understandable that hot water meters represent the second-largest group of devices.

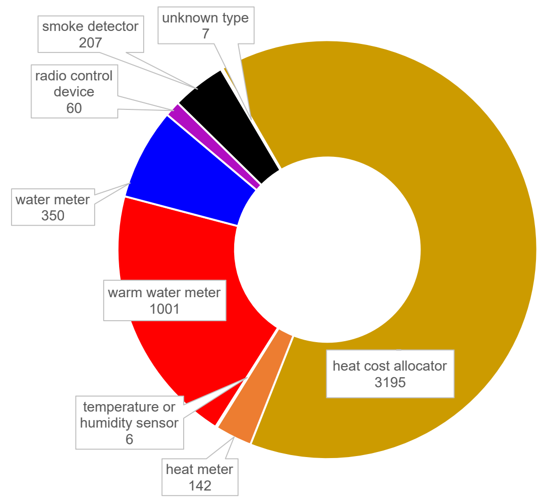


Fig. 7. Device Types. For details, see Section 3 and EN 13757-7 Table 13.

Furthermore, the examined telegrams could be assigned to 15 different manufacturers based on the manufacturer IDs. Thereby, we only consider the manufacturer IDs of the wireless transmitting device. In the case of an attached transmitter module whose manufacturer ID differs from that of the meter, we only consider the manufacturer ID of the transmitter module, as our investigation relates to the radio protocol. The most frequent observed manufacturers are shown in Figure 8. For a detailed description, of how this information can be derived from the M-Field and Device Type field, see Section 3.2.2. In our sample, MAN1 and MAN2 appear to be predominant. MAN3 and MAN4 are still partially widespread, and the other eleven manufacturers only to a minuscule extent. In general, according to our field study, it appears that there are only a few manufacturers who dominate the market.

At this point, all data from the header were evaluated. This header data is required for all telegrams. Although the header data allows conclusions to be drawn about the manufacturer and the devices used, this data does not initially reveal sensitive information. However, the number and the type of appliance recorded in a certain area certainly allow conclusions to be drawn about the structural condition or the heating system used. At first glance, however, no personal data can be generated from this.

For the telegram header, RQ4 can be answered to the extent that the non-encrypted header does not contain any critical information. For a complete response, however, the data block of the telegrams

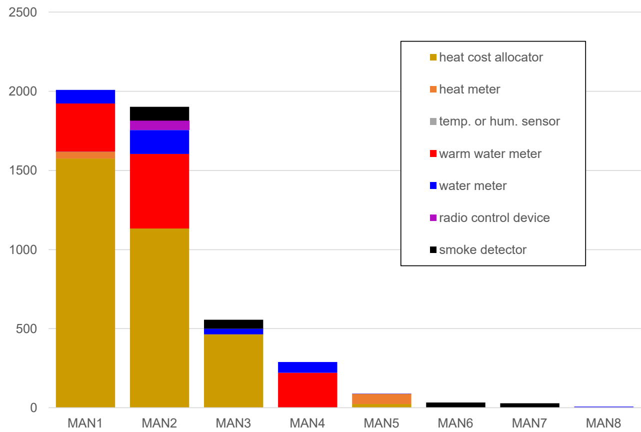


Fig. 8. Number of captured devices and their types ordered by manufacturer identifier (pseudonymized). Manufacturers observed less than five times are not shown (20 devices in total omitted).

must also be considered. The first step is to check whether the data blocks are encrypted to answer RQ3.

5.3 Usage of Encryption

The wM-Bus telegrams are encrypted at the transport layer (see Section 3.2.3) or optionally use the data link layer (DLL) encryption (see Section 3.2.2). In our field study, telegrams with the modes 0, 5, 7, 10 (see Section 3.2.4) as well as DLL (see Section 3.2.2) encryption were observed. The overview of the identified devices separated by the encryption methods is shown in Figure 9. In 2022, 56.6 % of the devices did not use encryption. In 2023, this figure is reduced to 46.7 %. Of the total number of devices recorded over both years, 48.5 % were not encrypted. Among the encrypted devices, mode 5 dominates with around a third of the encrypted devices.

The decoder used for this work has only a basic implementation of the wM-Bus protocol, and therefore cannot decode or interpret the full range of manufacturer-specific modifications to the wM-Bus telegram. For this reason, an additional distinction was made between *interpretable* and *non-interpretable* in the evaluation for mode 0. *Interpretable* are all telegrams for which the first data record could be evaluated according to Section 3.2.4.

The share of encrypted devices seems to have slightly increased in 2023. However, this is not confidentially deductible from the data, due to the capturing areas of 2022 and 2023 not having maximum achievable spatial overlap (see Section 4.2). When looking at Figure 10 it seems more likely, that simply more devices of MAN1, who uses proper encryption, were captured in 2023, explaining the difference.

As it can be seen in Figure 10, the use of encryption differs greatly from manufacturer to manufacturer. While we almost exclusively recorded encrypted devices from MAN1, MAN4, and MAN5, the devices of MAN2 and MAN3 mostly do not use encryption.

Within this context, it is important to note that the choice in which encryption mode the devices are operated in is made by the operator (metering provider). The depicted manufacturers here cannot always influence in which modes their devices are used. The responsibility for resolving the problem of missing encryption therefore lies with the metering point operators.

Evolution of encryption from 2022 to 2023. When comparing the results of the two years, the difference is almost negligible. In relation to the other encryption states, the number of not-encrypted and not-interpreted transmitting devices has decreased. At the same time, more devices with Mode 7 were recorded. However, when looking at Figure 10 the general tendency of the implementation of encryption or lack thereof per manufacturer seems to have remained unchanged. Regarding RQ3, it can be concluded that at least the devices of some manufacturers are operated without using encryption.

5.4 Impact on Privacy

Finally, the impact of unencrypted telegrams on privacy needs to be examined (RQ4). As mentioned in Section 5.3 some unencrypted telegrams can be decoded using data types defined in the standard, while others are proprietary encoded by the manufacturer.

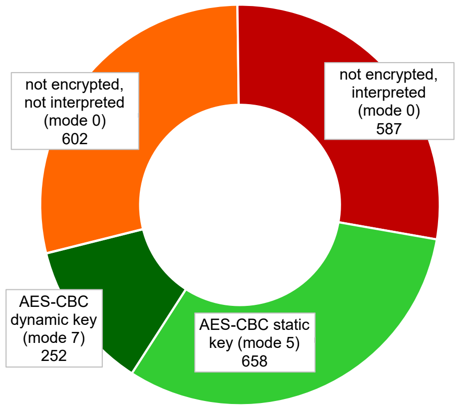
Manufacturer-specific encoding could be a limited protection by security through obscurity. However, we found it to be easy to interpret how a manufacturer-specific data record is structured based on the data physically displayed on the sensor and through visually inspecting the received telegrams. Particularly on newly installed devices, many values are zero in manufacturer-specific data records, which considerably simplifies reverse engineering. To demonstrate the possibility to also decode unencrypted proprietary telegrams, we implemented custom decoders for a limited number of sensor types.

Unencrypted and unsolicited transmission of metering data over the air poses a serious risk to privacy because they open the path to behavioral tracking. Hence, for the present paper, an apartment was sniffed (with the approval of the apartment inhabitants). By the data of one meter, the presence, and the behavior of the inhabitants could be silently tracked without significant effort.

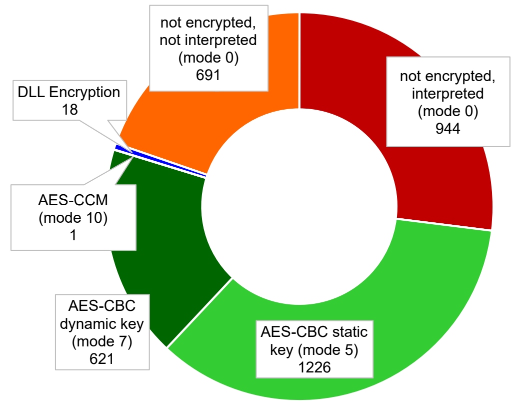
The severity of the sensitivity of the data is highlighted in Figure 11. The heatmap shows the data from one warm water meter that has been tracked for several weeks. Plotted on a heat map, it is easy to identify the periods of high and low consumption. Consequently, it can be concluded that the apartment was probably empty between the 2023-06-05 and the 2023-06-12. Furthermore, it can be stated that the inhabitant usually sleeps between midnight and 7 a.m. to 10 a.m. in the morning.

In this long-time experiment, only the data from one device were analyzed in detail over a long period. We used the actual metering device installed by the metering point operator in the home of one of the authors. However, based on the results so far, it can be assumed that if a larger number of devices (e.g., from an entire building complex or a street) are recorded, complex and precise behavioral profiles of hundreds of people can easily be created.

These could then allow statements to be made about the number of people, their presence, and absence, their sleeping and rising times, and their heating and showering behavior. It is even possible to track whether inhabitants stay out at night and use the bathroom. Detailed monitoring of residents is therefore possible.

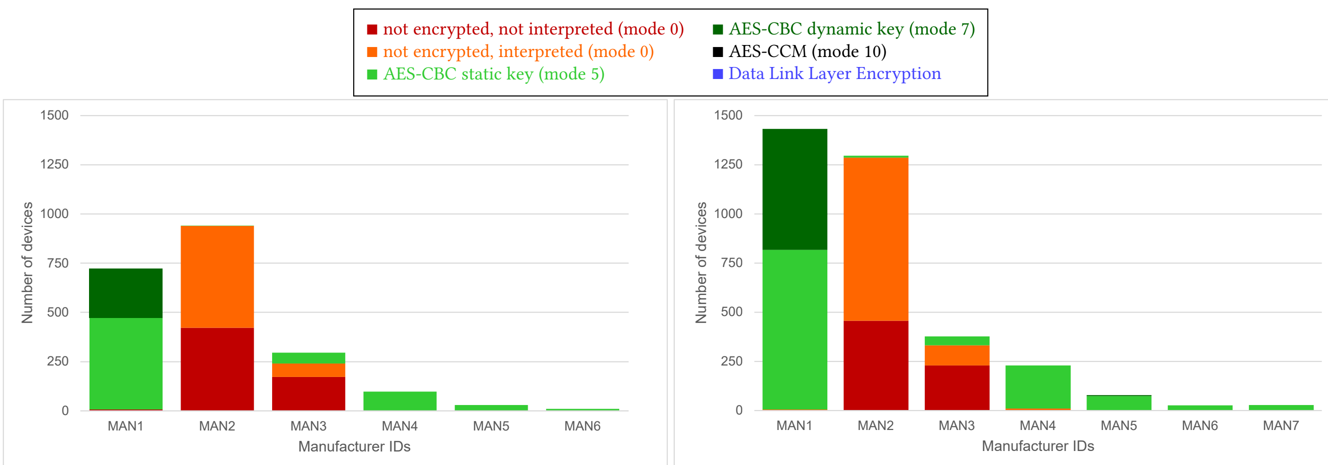


(a) Encryption mode usage in 2022



(b) Encryption mode usage in 2023

Fig. 9. Number of Devices ordered according to the Encryption Method. About half of all devices are not encrypted, and of those about half were directly interpretable without additional effort of implementing manufacturer-specific data formats. No significant change can be detected between 2022 and 2023.



(a) Encryption modes used by manufacturers in 2022

(b) Encryption modes used by manufacturers in 2023

Fig. 10. Number of sensors with encryption modes grouped by manufacturer. Manufacturers observed less than five times in a given year are not shown.

If one goes further, one can probably also determine social demographic factors such as age, employment status, number of people in the household, etc. with a more in-depth data analysis.

To summarize, we can answer RQ4 (Do the unprotected telegrams or unprotected telegram parts expose sensitive data?) as follows: Based on the results of the field study regarding the encryption status and the results of the sniffing experiment, it must be stated that even one single unencrypted water meter makes extensive personal data publicly accessible. Without effective encryption, anyone with the simplest hardware can spy on people and households anonymously and discreetly. Privacy is therefore by no means guaranteed. For this reason, all meters must transmit in encrypted form, even though this is only one prerequisite for protecting the personal rights of inhabitants.

6 DISCUSSION AND OUTLOOK

This field study is a first investigation on the usage of wM-Bus in the real world, focusing on the privacy aspect. Although we do not claim the study to be either representative or comprehensive due to the following reasons, the results still show that noteworthy shortcomings are threatening the privacy of tenants using wM-Bus metering.

Spacial Scope of the Field Study. The first limitation of this paper lies in the scope of the field study. Data were only collected in three southern German federal states and Luxembourg. Market participants that only operate in other federal states are therefore excluded from this study.

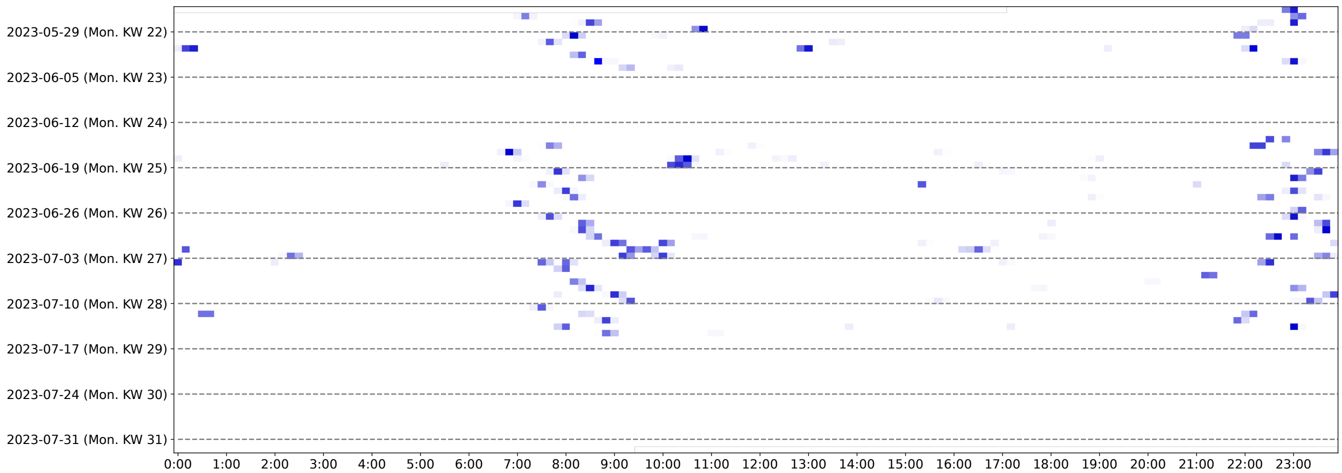


Fig. 11. Heatmap of daily warm water usage. Lines marking the weeks. The absence of the inhabitant in calendar week (KW) 23 as well as the weekly rising and sleeping pattern is clearly visible.

Deviation of Observed Market Share and Expected Market Share. Using our more comprehensive non-pseudonimized data set, we compared the manufacturers' names[5] with the existing market environment. Some well-known manufacturers of wireless metering devices were not present. This suggests that different manufacturers may use different operating modes or operate in different areas not covered by our study.

Distribution of Metering Devices. 4.986 metering devices were identified, most of which were heat cost allocators. The majority of data was recorded in the Munich city center/west area, where a travel distance of around 65 km was covered. In addition, routes were traveled in southern Munich, Karlsruhe, Trier, and Luxembourg. In total, a distance of more than 100 km was covered, which led to the desired spread of the measuring points and thus also to an extrapolation of the data.

Duration of Stay. Frequent transmission may not be often enough that a walk along the street was able to capture all sensors of the buildings. It also has to be mentioned that not the same walking path was traversed in 2022 and 2023. And no systematic methodology was employed to ensure the comprehensive capture of all existing devices (such as duration of stay, consistent route, etc.).

Telegram loss. In a stationary experiment with more than 100 devices, we noticed that some devices were only received for the first time after more than an hour, although it was ensured that these devices were all transmitting at the same rate of 3 Minutes. We assume that these devices were lost due to telegram collisions or other physical limitations in the wireless transmission. This suggests that a certain number of devices were also missed in the field study due to telegram collision or physical limitations.

Restrictions on Operating Modes. Due to the configuration of the hardware, only telegrams with transmission mode C1 or T1 were recorded. C1 and T1 are unidirectional transmission modes (as described in 3.2.1). Devices for which the transmission has to be requested using bidirectional communication are therefore not recorded (mode C2 / T2). Communication modes apart from C and T (e.g., N, S, R and F) were not considered in this study. The limitations mean that the results presented in this work relate exclusively to modes C1 / T1 and conclusions about devices with other transmission modes are inadmissible.

Decryption of encrypted telegrams. Decryption of telegrams was not considered in this work. However, it is theoretically possible that telegrams can be easily decrypted if the operator uses a weak key or a single key for all devices. This would mean that not only the unencrypted half of the devices examined in this study would represent a security risk for privacy, but also the supposedly securely encrypted devices would reveal private data.

Outlook

The implementation of this field study only provides a first investigation of how wM-Bus is used in real-world applications. We propose the following ideas for improvement in future work.

- **Increase sample size** The field study was only conducted in three urban areas in Germany and one in Luxembourg. It is known to the authors that metering operators and deployed sensor devices vary between cities and federal states. For a representative field study, more cities from different states and countries should be considered.
- **Wireless M-Bus Modes** This work focused on the wM-Bus modes C1 and T1. Other modes of operation should be considered in future campaigns to get a comprehensive view of the wM-Bus usage.
- **Software Defined Radio** Software Defined Radio (SDR) facilitates the simultaneous recording of multiple wM-Bus

[5]https://www.dlms.com/flag-id-directory/

modes at once. This alleviates some limitations of the presented study and enables future studies using inexpensive hardware.

- **Secret Key Management** As highlighted by [14], the wM-Bus standard proposes disadvantageous methods for the management of the wM-Bus encryption keys. It would be of interest to investigate weaknesses in how secret keys are handled in the field by the operator and manufacturers.

7 DISCLOSURE PROCESS

In the process of this field study, we informed all wireless M-Bus manufacturers from which we have captured telegrams, and operators known to us prior to the publication of this paper and requested comments. Only one manufacturer replied, commenting their devices strictly uses encryption. We could confirm that in our dataset.

In this study, we did not attempt to break any protection mechanisms like the decryption of the encrypted telegrams, and we collected all telegrams by listening to openly non-directionally broadcasted telegrams only. Only the non-directional, public telegrams from sending mode 44_h were recorded (see also Section 3.2.2). These are telegrams that can be read by anyone.

It is infeasible to coordinate a remedy for the issues demonstrated in this study, as it would require the physical exchange of many sensors. With this publication, we hope to promote the usage of adequate encryption and privacy protection in the future.

8 CONCLUSION

Our field study analyzed an untargeted sample of the real-world usage of wireless M-Bus (wM-Bus) in Germany and Luxembourg in 2022 and 2023 with the goal to expose the existence of security or privacy risks. We created a portable device to capture wM-Bus sensor telegrams which are broadcasted frequently. Using the device, we walked through four urban areas and recorded broadcasted wM-Bus telegrams as well as the position where the telegram was received. Almost 5000 unique sensor devices were identified throughout all campaigns. The application of the sensors was identified by decoding the transmitted device type of each sensor. Demonstrating, we were able to capture, interpret and spatially allocate wM-Bus telegrams (RQ2).

We displayed a statistic of how often encryption was used to protect the metering information. We described the capturing process, the used hardware, and the relevant parts of the wM-Bus standard which we used to derive our statistics. To our surprise, we found that about half of the metering devices do not use encryption (RQ3) to protect the privacy of the apartment inhabitants. We demonstrated that non-encrypted metering data transmitted by a commercially available wM-Bus sensor poses a privacy threat (RQ4). To conclude, we found no inherent flaws in the design of the wM-Bus protocol that poses a practical risk to privacy. However, our sample of wM-Bus usage in practice demonstrated that misconfiguration puts the privacy of inhabitants of apartment buildings at risk (RQ1).

The major contributions of this paper are:

- We showed that about half of the wireless sensors in real-world use do not use encryption (Section 5.3).
- The non-encrypted telegrams can be captured in a short interval (5 - 15 min) and most telegrams can be decoded easily.
- The decoded data enables precise tracking of inhabitant behavior.

We want to raise awareness of existing privacy problems in the real-world usage of wM-Bus and point out that standards exist that protect privacy. Especially BSI-TR-03109-1[6] requires wM-Bus using encryption mode 7 if these devices are to be integrated into the German Smart Metering System. This study highlights the necessity of binding requirements, as manufacturers and operators do not always implement privacy-friendly solution on their own initiative. As discussed in Section 6, further investigation needs to be conducted to create a comprehensive data set of wM-Bus usage to representatively quantify the real dimension of the identified privacy concern.

ACKNOWLEDGMENTS

This work was supported by the project 'Smart East'[6] (funding reference L75 21113 by Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg) as well as by the Helmholtz Association under the programs 'Energy System Design (ESD)' (topic number 37.12.01) and 'Engineering Digital Futures (EDF)' as part of the KASTEL Security Research Labs, Karlsruhe (topic number 46.23.02).

AUTHOR CONTRIBUTIONS

Author contributions according to Contributor Roles Taxonomy (CRediT, https://credit.niso.org/): *Conceptualization:* F.H.; *Methodology:* F.H.; *Software:* F.H.; *Formal analysis:* F.H.; *Investigation:* F.H.; *Data curation:* F.H.; *Writing - original draft:* F.H., K.B., J.G.; *Writing - review and editing:* F.H., K.B., J.G., S.W., V.H.; *Visualization:* F.H., K.B.; *Supervision:* J.G., S.W., V.H.; *Funding acquisition:* V.H.

REFERENCES

[1] Wafaa Anani and Abdelkader Ouda. 2022. Wireless Meter Bus: Secure Remote Metering within the IoT Smart Grid. en. In *2022 International Symposium on Networks, Computers and Communications (ISNCC)*. IEEE, Shenzhen, China, (July 2022), 1–6. ISBN: 978-1-66548-544-9. DOI: 10.1109/ISNCC55209.2022.9851807.

[2] Muhammad Rizwan Asghar, Gyorgy Dan, Daniele Miorandi, and Imrich Chlamtac. 2017. Smart Meter Data Privacy: A Survey. *IEEE Communications Surveys & Tutorials*, 19, 4, 2820–2835. DOI: 10.1109/COMST.2017.2720195.

[3] Carsten Bories. 1995. *Einrichtung Einer Intelligenten Ausleseeinheit Für Verbrauchsmeßzähler.* Diploma Thesis. (Mar. 1995).

[4] Cyrill Brunschwiler. 2013. Wireless M-Bus Security Whitepaper Black Hat USA 2013 June 30th, 2013. https://www.compass-security.com/fileadmin/Datein/Research/Praesentationen/blackhat_2013_wmbus_security_whitepaper.pdf.

[5] Feng Chen, Jing Dai, Bingsheng Wang, Sambit Sahu, Milind Naphade, and Chang-Tien Lu. 2011. Activity Analysis Based on Low Sample Rate Smart Meters. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '11: The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Diego California USA, (Aug. 21, 2011), 240–248. ISBN: 978-1-4503-0813-7. DOI: 10.1145/2020408.2020450.

[6] Bundesamt für Sicherheit in der Informationstechnik. 2021. Technische Richtlinie BSI TR-03109-1. Anforderungen an die Interoperabilität der Kommunikationseinheit eines intelligenten Messsystems. Version 1.1 (6b75fb88). (2021). https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR03109/TR03109-1.pdf?__blob=publicationFile&v=4.

[6] https://smart-east-ka.de/

[7] HeizkostenV. 2023. Verordnung über Heizkostenabrechnung in der Fassung der Bekanntmachung vom 5. Oktober 2009 (BGBl. I S.3250), die zuletzt durch Artikel 3 des Gesetzes vom 16. Oktober 2023 (BGBl. 2023 I Nr. 280) geändert worden ist. (2023). Retrieved Apr. 18, 2024 from https://www.gesetze-im-internet.de/heizkostenv/HeizkostenV.pdf.
[8] Mikhail A. Lisovich, Deirdre K. Mulligan, and Stephen B. Wicker. 2010. Inferring Personal Information from Demand-Response Systems. *IEEE Security & Privacy*, 8, 1, 11–20. DOI: 10.1109/MSP.2010.40.
[9] Pavel Masek, Martin Stusek, Krystof Zeman, Radek Mozny, Aleksandr Ometov, and Jiri Hosek. 2019. A Perspective on Wireless M-Bus for Smart Electricity Grids. In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. 2019 42nd International Conference on Telecommunications and Signal Processing (TSP). IEEE, Budapest, Hungary, (July 2019), 730–735. ISBN: 978-1-72811-864-2. DOI: 10.1109/TSP.2019.8768840.
[10] Pavel Masek, Krystof Zeman, Zenon Kuder, Jiri Hosek, Sergey Andreev, Radek Fujdiak, and Franz Kropfl. 2016. Wireless M-BUS: An Attractive M2M Technology for 5G-Grade Home Automation. In *Internet of Things. IoT Infrastructures*. Vol. 169. Benny Mandler et al., (Eds.) Springer International Publishing, Cham, 144–156. ISBN: 978-3-319-47062-7 978-3-319-47063-4. DOI: 10.1007/978-3-319-47063-4_13.
[11] MessEV. 2024. Mess- und Eichverordnung vom 11. Dezember 2014 (BGBl. I S. 2010, 2011), die zuletzt durch Artikel 1 der Verordnung vom 29. Januar 2024 (BGBl. 2024 I Nr. 27) geändert worden ist. (2024). Retrieved Apr. 18, 2024 from https://www.gesetze-im-internet.de/messev/MessEV.pdf.
[12] Libor Polčák and Petr Matoušek. 2022. Metering Homes: Do Energy Efficiency and Privacy Need to Be in Conflict?: en. In *Proceedings of the 19th International Conference on Security and Cryptography*. SCITEPRESS - Science and Technology Publications, Lisbon, Portugal, 47–58. ISBN: 978-989-758-590-6. DOI: 10.5220/0011139000003283.
[13] Huwida Said, Mario Guimaraes, Noora Al Mutawa, and Ibtesam Al Awadhi. 2011. Forensics and War-Driving on Unsecured Wireless Network. In *2011 International Conference for Internet Technology and Secured Transactions*. 2011 International Conference for Internet Technology and Secured Transactions. (Dec. 2011), 19–24.
[14] Peter Shipley. 802.11b War Driving and Lan Jacking. (2001). Retrieved Apr. 3, 2023 from https://www.youtube.com/watch?v=bWH-3OZJ0vo.
[15] S. Spinsante, S. Squartini, L. Gabrielli, M. Pizzichini, E. Gambi, and F. Piazza. 2014. Wireless M-Bus Sensor Networks for Smart Water Grids: Analysis and Results. *International Journal of Distributed Sensor Networks*, 10, 6, (June 1, 2014), 579271. DOI: 10.1155/2014/579271.
[16] Susanna Spinsante, Mirco Pizzichini, Matteo Mencarelli, Stefano Squartini, and Ennio Gambi. 2013. Evaluation of the Wireless M-Bus Standard for Future Smart Water Grids. In *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)*. 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC 2013). IEEE, Sardinia, Italy, (July 2013), 1382–1387. ISBN: 978-1-4673-2480-9 978-1-4673-2479-3 978-1-4673-2478-6. DOI: 10.1109/IWCMC.2013.6583758.
[17] S. Squartini, L. Gabrielli, M. Mencarelli, Mirco Pizzichini, S. Spinsante, and F. Piazza. 2013. Wireless M-Bus sensor nodes in smart water grids: The energy issue. *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, 614–619. DOI: 10.1109/ICICIP.2013.6568148.

Received 19 April 2024; revised 2 August 2024; accepted 17 June 2024

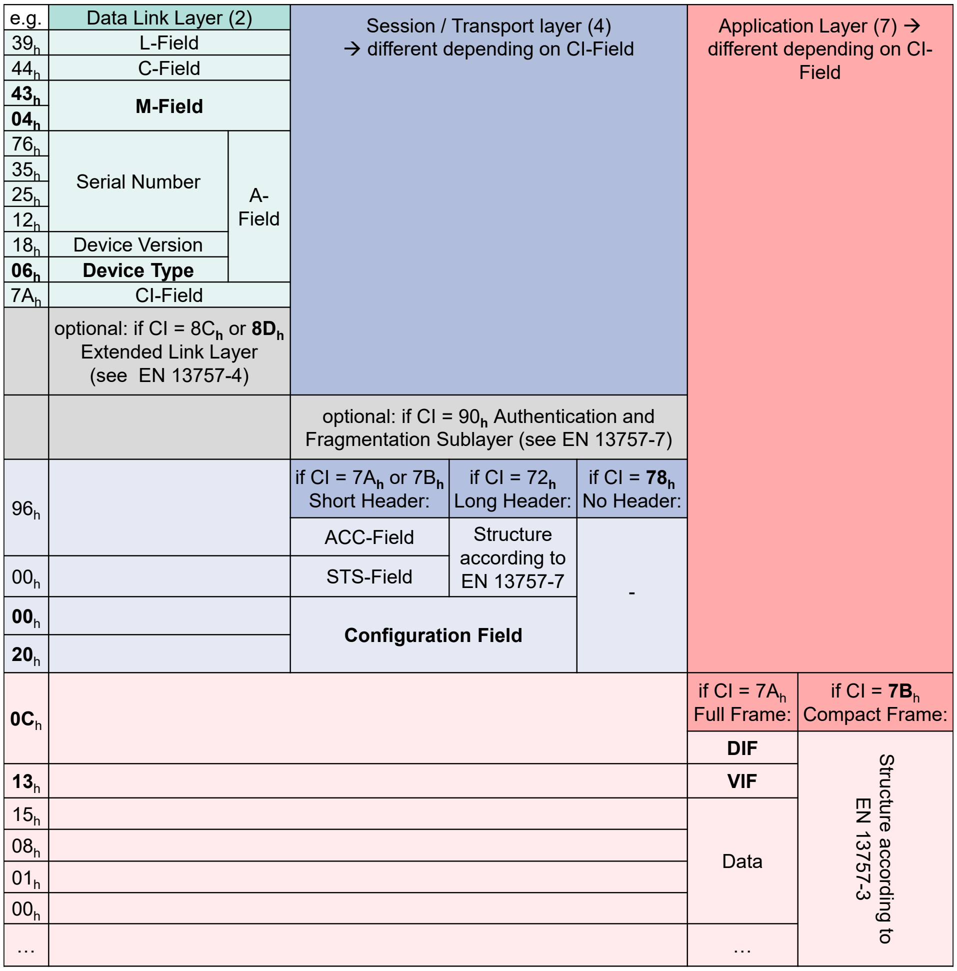


Fig. 12. M-Bus telegrams and variations depending on the value of the CI fields. **Bold** font marks data that is used to derive the results shown in Section 5.

A Distributed Game Theoretic Approach for Optimal Battery Use in an Energy Community

NICOLAS KIRSCH, GIULIO SALIZZONI, and MARYAM KAMGARPOUR, EPFL, Switzerland
TOMASZ GORECKI, Sustainable Energy Center, CSEM, Switzerland

With the recent rise of decentralized energy resources like solar panels and batteries, energy communities have grown in popularity. Without collaborative load management, some of the locally produced electricity in the energy community could be unused and sold back to the grid, hindering its economic and environmental benefits. In this paper, we introduce a game to model an energy community with a shared energy source and to determine how to distribute the generated electricity. The goal is to compute a Nash equilibrium so that no participant has an incentive to leave the community. To do so with minimal information sharing between members, a distributed approach is considered. The game structure should incentivize self-consumption in the energy community. We then integrate the game in a receding horizon control framework. This approach is tested on real photovoltaic data and reduces costs for the energy community. Its robustness to imperfect solar forecasting is also evaluated, showing that uncertainty induces minimal increases in cost.

1 INTRODUCTION

Energy communities are a promising way to restructure energy systems to accelerate the energy transition [1]. They have been growing in popularity, especially in Europe, where nearly 4000 citizen-led energy initiatives were recorded in 2023. This has been fostered by the transposition of the EU directives [[2],[3]] into national laws between 2020 and 2022 [4]. An energy community is a localized group of individuals, businesses, or organizations that collaboratively generate, manage, and share distributed energy resources. By combining energy communities with photovoltaic panels and batteries, the amount of energy exchanged with the grid can be significantly reduced, potentially smoothing peaks in the demand during the day [5].

In current energy community, each member often does not have an energy management software; and if they do it does not account for the action of others. Lack of load coordination can lead to sub-optimal community behavior. Having a fully centralized energy community planning also presents issues, especially regarding privacy and fairness within the energy community.

Game theory is a powerful framework that can overcome the shortcomings described above. Game theory can indeed successfully capture the interactions of self optimizing agents with coupled objectives or constraints [6]. Each player has individual objective and a game framework allows one to find a common solution, namely, a Nash equilibrium (NE), from which no player has incentive to deviate. Furthermore, for some specific game, the Nash equilibrium can be computed in a distributed way, reducing privacy concerns.

Some previous works in the literature have applied game theory to energy systems [6–16]. A popular framework to do so is the mean field game. In such a game, each player do not react to the exact strategies of everyone, but to a statistical aggregation of it. The game can be solved distributively because each player only need to know this shared parameter in addition to their own information. In [7–9] the shared parameter represents the price of electricity to be purchased from the grid and the Nash equilibrium is found using an iterative algorithm. For it to be successful, the studied energy community are assumed to be very large, so that the mean parameter is statistically significant and the community consumption decision could influence the grid tariff. In [6] and [10], the shared parameter is the mean consumption load of all the players over the horizon. This is used to compute the expected price at each time step. This method is applied to the control of plug-in electric vehicles. It is again assumed that the number of players is large enough to influence the cost of electricity. The example used in both papers controls a fleet of 10^7 electric vehicles.

Another studied game theoretic approach is the Stackelberg game, for which the community manager is decision-maker along with the community members. For example, [11] presents a version of a Stackelberg game where the aggregator has a centralized battery to manage flows between the grid and the consumers. In [12], the leader is a central storage facility aiming at maximising its profits and the followers are the consumers willing to minimize the cost of external energy supply.

In energy systems, the physical constraints can depend on the actions of all the players. In this case, the game becomes a generalized Nash equilibrium problem [17]. An example is [13], which proposes an approach for prosumers community day-ahead planning where the coupling constraint is the energy balance between the different photovoltaic installations. The generalized Nash equilibrium is computed with each player needing the exact strategies of the others.

Several papers combine generalized Nash equilibrium problems with aggregative games when the coupled constraint represents the limited availability of a resource, as it is the case in this work. In [14], the total demand is constrained by the limited capacity of an electrical feeder and a gas pipe. In [15] the total electrical demand is constrained by the grid power limitation and in [16] the coupled constraint is for PV energy consumption. Even if considering setups close to ours, these papers present some difference from our work. For the first two papers, it is assumed that the average consumption level in the community influences the resources per unit cost. However, we consider a community which is too small to influence the price, so this assumption does not apply to our scenario. The approach in [16] consider some global objective to share between the agent as part of each player objective function, which is not present in our work. Furthermore, they apply their approach to a case where each agent has an individual PV installation, which differs from our scenario where a common central resource has to be shared between the agents.

Authors' Contact Information: Nicolas Kirsch, nicolas.kirsch@epfl.ch; Giulio Salizzoni, giulio.salizzoni@epfl.ch; Maryam Kamgarpour, maryam.kamgarpour@epfl.ch, EPFL, Lausanne, Vaud, Switzerland; Tomasz Gorecki, Sustainable Energy Center, CSEM, Neuchâtel, Switzerland, tomasz.gorecki@csem.ch.

Overall, these approaches are interesting, but they either are not applicable to small scale energy communities or do not find a solution in a distributed manner. Furthermore, most of them consider day ahead planning rather than real-time. Our objective is to provide a tool for real-time, privacy preserving, consumption load management in energy communities with a shared energy source. This tool should reduce costs for community members by encouraging local energy consumption. To this end, the contributions of this paper are as follows:

- We formulate a game theoretic framework for optimizing battery charge and discharge plans for each members in a small scale energy community. This framework has a clear physical and economical interpretation and allows the community to maximize self-consumption.
- We design an algorithm that can compute a generalized Nash equilibrium for the energy community under consideration in a distributed way, mitigating privacy concerns. This algorithm is also applicable to real-time planning.
- Our extensive simulations show that the proposed approach can significantly reduce costs for energy community members by increasing self-consumption, and that it is robust to imperfect solar forecasts.

2 PROBLEM FORMULATION

Table 1. Nomenclature

		Variables and Cost Functions
N	-	Number of agents
T	[h]	Time horizon
e_i	[kW]	Electricity consumed by agent i
s	[kW]	Electricity produced by the photovoltaic panels
$p_{lc,i}$	[kW]	Electricity from the panels used by agent i
$p_{buy,i}$	[kW]	Electricity imported from the grid by agent i
p_{sld}	[kW]	Electricity sold to the grid by the community
$p_{ch,i}$	[kW]	Electricity charged in the battery by agent i
$p_{ds,i}$	[kW]	Electricity discharged from the battery by agent i
c_{lc}	$\left[\frac{\text{CHF}}{\text{kWh}}\right]$	Local price for the electricity
c_{buy}	$\left[\frac{\text{CHF}}{\text{kWh}}\right]$	Price for the electricity bought from the grid
c_{sld}	$\left[\frac{\text{CHF}}{\text{kWh}}\right]$	Price for the electricity sold to the grid
SoC_i	[kW]	State of charge of agent i's battery
η_{ch}	-	Charging efficiency of agent i's battery
η_{ds}	-	Discharging efficiency of agent i's battery
β_i	[kWh]	Battery storage capacity
C_{cyc}	$\left[\frac{\text{CHF}}{\text{kWh}^2}\right]$	Degradation cost of agent i's battery
y_i	-	Decision variables of agent i
y	-	Vector of joint actions

We consider an energy community with a set $\mathcal{N} = 1, \ldots, N$ members and a horizon $T \in \mathbb{R}$. Each member i is characterized by an initial demand $e_i \in \mathbb{R}^T$ to be met. To modify their loads, each member possesses a battery energy storage system with a storage capacity $\beta_i \in \mathbb{R}$. Energy communities are often constructed around a community manager whose role is to organize and overview the different physical and economic flows to and from the members of the energy community. The community manager receives the expected solar production $s \in \mathbb{R}^T$ from the central solar installation and dispatches it to the community members. Each member i receives a share of this local resource $p_{lc,i} \in \mathbb{R}^T$, buying it at a cost of $c_{lc} \in \mathbb{R}$. The value of c_{lc} depends on the annualised investment cost and the maintenance costs of the photovoltaic installation. If this energy is not sufficient to meet the members' demand at each time step, they will also consume electricity from the grid, $p_{buy,i} \in \mathbb{R}^T$, at a price $c_{buy} \in \mathbb{R}$. The sum of all these grid imports, $\sum_{i=1}^{N} p_{buy,i}$, corresponds to the imports of the community $p_{buy} \in \mathbb{R}^T$. If at some point the solar production is greater than the demand, this excess load $p_{sld} \in \mathbb{R}^T$ will be sold back to the grid at a price $c_{sld} \in \mathbb{R}$. The local cost c_{lc} is often designed in a way that is greater than c_{sld} and less than c_{buy}. In this way, consumption within the community is encouraged.

The mutual influence among group members is modeled by diminishing the solar energy available to each individual. At every time steps over the horizon, each agent in the energy community can only consume locally what has not been consumed by the other members. Given the local power consumed by all the other players, $p_{lc,-i} = \sum_{j \neq i} p_{lc,j}$, the local power consumption of player i must satisfy:

$$p_{lc,i} \leq s - p_{lc,-i}, \tag{1}$$

over the entire horizon.

2.1 Game theoretic formulation of the objectives

The game we introduce is a non-cooperative game with N players. Each player aims at minimizing a cost function depending on their joint decision. At every time step, each player can choose how much to consume locally, and how much to charge and discharge their batteries. For a player i, let the vector $y_i = (p_{ch,i}, p_{ds,i}, p_{lc,i}) \in \mathbb{R}^{3T}$ represents the strategy choice. To satisfy the power balance over the horizon, the following constraint is added:

$$p_{lc,i} + p_{buy,i} = e_i + p_{ch,i} + p_{ds,i}. \tag{2}$$

Each player wants to minimize their billing cost, while accounting for the degradation cost of using the battery. The billing cost contains the cost of using local electricity and the cost of using grid electricity, and it is equal to:

$$J_i^{bill}(y_i) = {p_{lc,i}}^\top c_{lc} + p_{buy,i}^\top c_{buy} \tag{3}$$

Regarding electricity costs, it is here assumed that the energy community is too small to have an impact on the cost of grid electricity c_{buy}. The cost of local electricity c_{lc} is also assumed to be fixed. Thus, both c_{buy} and c_{lc} are set price vectors which are independent of the behavior of energy community members behaviors.

2.2 Battery model

For each agent i, the following constraints are added to model battery behavior over the entire horizon [7]. Although more complex battery models can be developed, we consider the following constraints as sufficiently precise for the purpose of planning.

- Battery dynamics:

$$SoC_i^{t+1} = SoC_i^t + \frac{\eta_{ch}}{\beta_i} p_{ch,i}^t + \frac{1}{\eta_{ds}\beta_i} p_{ds,i}^t.$$

- State of charge constraints:

$$SoC_{min} \leq SoC_i^t \leq SoC_{max}.$$

- Charge and discharge constraints:

$$p_{ds,max} \leq p_{ds,i}^t \leq 0 \text{ and } 0 \leq p_{ch,i}^t \leq p_{ch,max},$$

where $SoC_i \in [0,1]^T$ is the state of charge of the battery, $p_{ds,i} \in \mathbb{R}_{\leq 0}^T$ is the discharge power of the battery and $p_{ch,i} \in \mathbb{R}_{\geq 0}^T$ the charging power. The charging and discharging efficiencies are, respectively, $\eta_{ch} \in \mathbb{R}$ and $\eta_{ds} \in \mathbb{R}$.

In practice, battery degradation occurs because of several factors, such as the gradual wear and tear of the electrodes due to repeated charging and discharging cycles. Our aim is to minimize this and we thus introduce a cost associated to it. This cost can be defined as a quadratic function of the net power exchanges with the battery [18]. With the diagonal matrix $C_{cyc} \in \mathbb{R}^{T\times T}$ representing the degradation cost per kW used, this cost is:

$$J_i^{batt}(y_i) = \left(\eta_{ch} p_{ch,i} + \frac{p_{ds,i}}{\eta_{ds}}\right)^\top C_{cyc} \left(\eta_{ch} p_{ch,i} + \frac{p_{ds,i}}{\eta_{ds}}\right). \quad (4)$$

2.3 Individual Players Optimization problem

While the objective function of player i is independent of the actions of the other players, these actions affect the constraints of player i because of (1). This means that the game at hand is a generalized Nash equilibrium problem [17]. This constraint will need to be satisfied for all players when computing their optimal strategies.

Each player i aims to find the strategy y_i^* that minimizes the following optimization problem:

$$\begin{aligned} \arg\min_{y_i} \quad & J_i^{bill}(p_{lc,i}, p_{buy,i}) + J_i^{batt}(p_{ch,i}, p_{ds,i}) \\ \text{s.t.} \quad & p_{lc,i} + p_{buy,i} = e_i + p_{ch,i} + p_{ds,i} \\ & 0 \leq p_{buy,i}, \quad 0 \leq p_{lc,i} \\ & SoC_i^{t+1} = SoC_i^t + \eta_{ch} p_{ch,i}^t + \frac{1}{\eta_{ds}} p_{ds,i}^t \\ & SoC_{min} \leq SoC_i^t \leq SoC_{max} \\ & p_{ds,max} \leq p_{ds,i}^t \leq 0, \ 0 \leq p_{ch,i}^t \leq p_{ch,max} \\ & \underbrace{p_{lc,i} \leq s - p_{lc,-i}}_{\text{Dependence on other players}} \quad . \end{aligned} \quad (5)$$

Let the set $K_i \subset \mathbb{R}^{3T}$ represent the vectors y_i satisfying all the individual constraints, i.e. all but the last constraints in (5). To model the influence of agents on each other, the vector of joint actions $y = [y_1 \dots y_N] \in K = K_1 \times \dots \times K_N$ has to belong to the global coupling constraint set C, namely:

$$C = \{y \in K : g(y) \leq 0\}, \quad (6)$$

with $g : \mathbb{R}^{NT} \to \mathbb{R}^T = (\sum_{i=1}^N p_{lc,i} - s)$ corresponding to the last constraint in (5). The optimization problem can be rewritten as:

$$\begin{aligned} \arg\min_{y_i} \quad & J_i(y_i) \\ \text{s.t.} \quad & y_i \in K_i \\ & g(y) \leq 0. \end{aligned} \quad (7)$$

The solution set $S_i(y_{-i})$ of the problem (7) depends on the joint actions y_{-i} of the other agents, where $y_{-i} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_N)$. A generalized Nash equilibrium of the game is vector y such that

$$y_i \in S_i(y_{-i}), \quad \forall i \in \{1, \dots, N\}.$$

Due to pratical limits in the computation of a Nash equilibrium, we aim for a solution that approximates it within a small margin of error. This type of solution is called ϵ-Nash equilibrium. An ϵ-Nash equilibrium is a vector y such that

$$J_i(y_i) \leq J_i(\tilde{y}_i) + \epsilon, \ \forall \tilde{y}_i \in \{g(\tilde{y}_i, y_{-i}) \leq 0\} \cup K_i, \forall i \in \{1, \dots, N\}. \quad (8)$$

In the next section, we prove that there exists at least one generalized Nash equilibrium, and we propose an algorithm to compute an ϵ-Nash equilibrium.

3 ANALYSIS AND ALGORITHM

Our aim in this section is first to show that the game has at least one generalized Nash equilibrium, and then to design an algorithm that can compute an equilibrium in a distributed way.

3.1 Existence of a Nash equilibrium

Proposition 3.1. *A game with N agents, each one with a cost function as the one described in (7), has at least a generalized Nash equilibrium.*

Proof. Notice that the set C, which is defined as the intersection of hyperplanes and half-spaces, is closed and convex. Then, we introduce the map $F : \mathbb{R}^{3NT} \to \mathbb{R}^{3NT}$ which is the pseudo-gradient of the game, defined as $F(y) = [(\nabla_{y_i} J_i(y_i))_{i=1}^N]$. Given that the objective function J_i is quadratic and convex, F is continuous and convex. Theorem 5 in [17] states that, for a so-called *variational inequality* problem, $VI(C, F)$, the set of solutions is a subset of the generalized Nash equilibrium set of the game. Also, corollary 2.2.5 in [19] states that if C is compact and convex and F is continuous, then the set of solutions is non-empty. As this is the case with our definitions, there exists a generalized Nash equilibrium for the game. □

3.2 Algorithm for computing a Nash equilibrium

We now look at how to design an algorithm that converges to a Nash equilibrium in a distributed way. Doing so would mitigate the data sharing and privacy concerns linked to centralized resolution. The algorithm should converge without each agent knowing the exact strategies of the other agents in the game. The algorithm in [8] also does this to find the Nash equilibrium of a mean field game.

Our approach is inspired by this algorithm, but differs in the shared parameter. Let the parameter $z \in \mathbb{R}^T$ be an estimation of the community load of local energy consumed, namely $\sum_{i=1}^N p_{lc,i}$, and the only information shared with players. Each player would require an individual version of z, z_{-i} which does not account for their own local consumption. The coupling constraint for player i defined in (1) becomes:

$$p_{lc,i} \leq s - z_{-i}. \tag{9}$$

Each player would still solve problem (7), but with this updated coupling constraint.

The optimal strategies could be computed for all players through a two-step iterative algorithm. Given a current iteration step k, each player would first compute their optimal action by solving (7) given $z_{(k)}$, the value of z at iteration k. By collecting each agents optimal $p^*_{lc,i(k)}$, the central coordinator would then compute the resulting estimation of the community local load $\Lambda(z_k)$ as:

$$\Lambda(z_{(k)}) = \sum_{i=1}^{N} p^*_{lc,i(k)}. \tag{10}$$

Notice that the central coordinator does not perform any optimization and does not control any parameters. Rather, it sums the loads of all agents and communicates this aggregate value to everyone, without disclosing the private consumption information. If privacy is not an issue, the agents could directly share among themselves their consumption without the need of the central coordinator. This would then be used to update the value of z for the next iteration:

$$z_{(k+1)} = (1-\eta) z_{(k)} + \eta \Lambda(z_{(k)}), \tag{11}$$

where $\eta \in (0,1)$ is the learning rate. These iterations would continue until the updates of $z_{(k)}$ become smaller than a convergence criterion ϵ_{stop}.

Algorithm 1 summarizes the steps of the approach described above. It is similar to the Jacobi-type algorithm presented in [17], but applied in a distributed way. Some other approaches linked with variational inequality theory also exist, such as projected gradient descent. However, to the best of our knowledge, no algorithm has been proved to converge to a Nash equilibrium in this setting [17]. We chose the following algorithm due to its simplicity in distributed implementation, its potential for ensuring privacy guarantees, and its ability to provide a feasible solution at each iteration.

Algorithm 1 Distributed Jacobi-type algorithm

```
initialize:
    z_(1) ← min(∑_{i=1}^N e_i, s), z_(1) ∈ ℝ^T
    p_{lc,i,(0)} ← min(e_i, s/N) for i = 1...N , p_{lc,i,(0)} ∈ ℝ^T
    k ← 1
repeat
    for i ∈ N do
        z_{-i,(k)} ← z_(k) − p*_{lc,i,(k−1)}
        Solve local optimization problem to compute (y*_{i,(k)})
    end for
    z_(k+1) ← (1 − η) z_(k) + η(∑_{i=1}^N p*_{lc,i,(k)})
    k ← k + 1
until ∑_{i=1}^N ||p*_{lc,i(k+1)} − p*_{lc,i(k)}|| ≤ ε_stop
```

Although we did not prove global convergence of the algorithm, it is clear that if the algorithm converges, it converges to a generalized Nash equilibrium. Looking at the convergence criteria, it can indeed be seen that the algorithm converges if no players change their strategies anymore, which corresponds to a Nash equilibrium.

4 SIMULATION STUDIES

This section presents several simulation results evaluating the applicability of the proposed approach given real photovoltaic data. The objective is to gain qualitative and quantitative insights on the benefits and robustness of our approach for energy communities.

We first analyzed the outcome of the algorithm over a single planning horizon of 48h. We found that all the agents were able to coordinate to maximize local energy consumption, without having to share their exact consumption plan.

The above behavior corresponds to a day-ahead planning approach, where the forecasts are fixed over the entire horizon. However, PV forecasts become increasingly uncertain with time, so the predictions at the end of the horizon can be significantly off. To mitigate this impact, our approach is implemented in a model predictive control framework with receding horizon. To do so, Algorithm 1 is run at each time step. Every time, it generates an entire optimized battery use sequence, but only the value at the first timestep is used to update the state of charge of the batteries. Before advancing to the next step, the forecast is updated. This is simulated using the python library NRGMaestro™ designed by the Centre Suisse d'Electronique et de Microtechnique [20].

Every time Algorithm 1 was used, convergence was achieved, and so we can conclude that players always reached an ϵ-Nash equilibrium (8).

This section analyzes the performance of receding horizon controllers using the load game approach for energy community management. The energy community studied is composed of a population of size $N = 10$, each member possessing its own battery. Each member's baseload is constructed from a common baseload $e_{base} \in \mathbb{R}^T$ representing a standard residential load with morning and evening consumption peaks. The differences between members are then induced by adding some normal noise to e_{base}. The noise is sampled from a normal distribution with a mean of 0 and a standard deviation of 0.2.

The following parameters were selected:

Battery parameters:

- Storage capacity: $\beta_i = 15$ kWh
- Charging and discharging efficiency: $\eta_{ch} = \eta_{ds} = 95\%$
- Maximum charge and discharge rate: $p_{ch,max} = p_{ds,max} = 7.5$ kW
- Starting *SoC*: 7.5 kWh (50%)

Economical parameters:

- Grid electricity: 0.2 CHF/kWh
- Local electricity: 0.1 CHF/kWh
- Feed-in tariff: 0.05 CHF/kWh

Community parameters:

- Number of members: $N = 10$

The economical parameters have been selected based on the study in [21]. The battery capacity has been arbitrarily chosen to match the scale of each agent's consumption. All the other battery parameters have been kept at the default values of the NRGMaestro™ tool [20].

4.1 Real photovoltaic data

The load game approach has been applied to real photovoltaic data from an installation close to Zurich. The dataset used contains forecasted and real solar power time series for several years and locations, with quarter hourly granularity. One installation is selected from the dataset, and two days of data are sampled from February 2023.

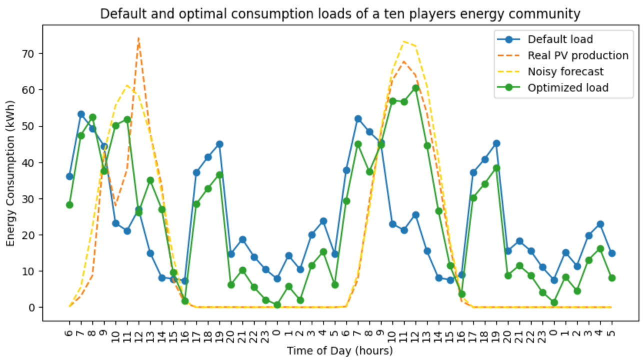


Fig. 1. Default and optimal community loads

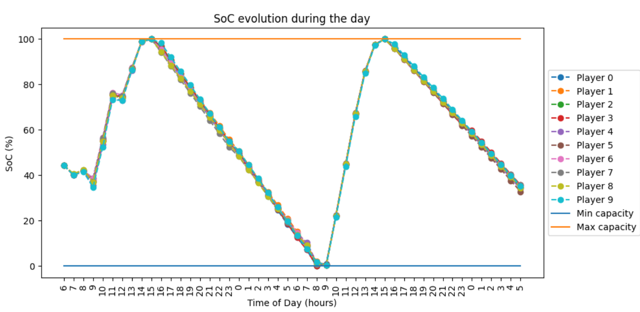


Fig. 2. Optimal SoC profiles

Figure 1 shows the initial community load ($\sum_{i=1}^{N} e_i$) and the community load at the generalized Nash equilibrium ($\sum_{i=1}^{N} p_{lc,i} + p_{buy,i}$) and Figure 2 shows the resulting *SoC* profiles for all the energy community members. Even if the optimizations are only done at the individual level, the load game control approach seems to improve the self-consumption within the entire community. Indeed, the morning and evening peaks relying on grid electricity have been shaven, and have been replaced by another peak during the day, aligned with photovoltaic production.

Looking at the *SoC* profile, the load game approach successfully pushed energy community members to charge their battery when photovoltaic production is high and discharge it when it is low. The coordination enabled by the load game also ensures that the total power consumed for battery charging by the community does not exceed the photovoltaic power available.

As seen in figure 1, even if using an imperfect photovoltaic forecast, the load game still manages to improve self-consumption in the energy community.

4.2 Sensitivity to photovoltaic power available

Our goal in this section is to evaluate the sensitivity of the load game results to the quantity of solar energy available. This is important to ensure that the load game approach remains useful for a wide range of implementation conditions, for which the solar production could vary significantly. To quantify this, the parameter $\gamma \in \mathbb{R}$ is introduced. It represents the self production rate, which is the share of total demand over a horizon which can be produced locally over the same period. It is defined as:

$$\gamma = \frac{s^\top 1}{x^\top 1}.$$

The higher the γ, the more solar energy is available. Note that γ represents aggregate values of the parameters over the horizon, but does not assess their time granularity. Having $\gamma = 1$ does not necessarily mean that no electricity is imported from the grid. Imports depends on when the consumption takes place and on the maximum capacity of the batteries. This analysis is done on a one day control horizon with a simulated typical solar load, with peak production during the day and no production during the night.

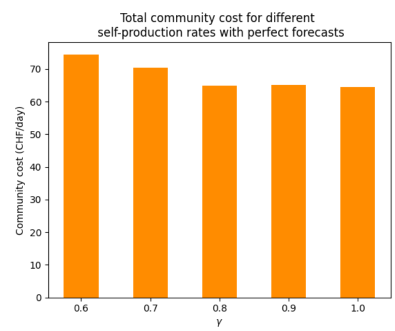


Fig. 3. Community costs for different values of self-production rates

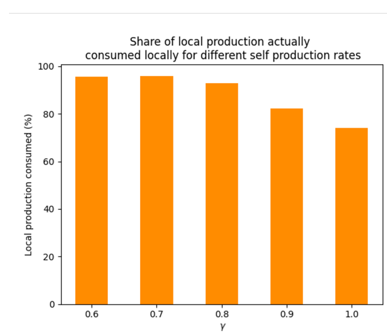


Fig. 4. Local production consumed locally for different values of self-production rates

Figure 3 shows the community costs for different values of γ. As expected, the more solar energy is available the lower is the cost for the community. This is because a greater share of the demand can be met using the cheaper local resources. However, for γ greater than 0.8, the cost reaches a plateau, after which increasing the amount of local energy does not reduce the cost.

To understand why increasing the local energy available does not decrease cost after a certain point, it is interesting to look at whether all the available solar energy is consumed. Figure 4 shows the share of available local energy consumed locally ($p_{lc,i}{}^{\top}1/s^{\top}1$) for different values of γ. For values of γ where the cost is decreasing, the entirety of the available solar energy is consumed. The amount of solar energy consumed increases, leading to cost savings. However, for $\gamma \geq 0.8$, this percentage gradually decreases as γ increases. A lower percentage of a higher amount of energy is consumed, which makes the absolute amount of solar energy consumed nearly constant. This explains why no additional savings are induced by increasing γ.

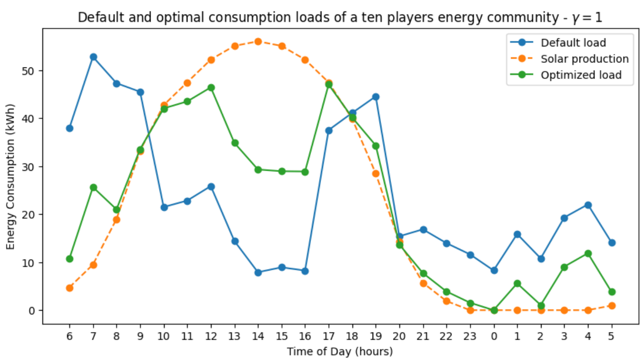


Fig. 5. Default and optimal community loads with $\gamma = 1$

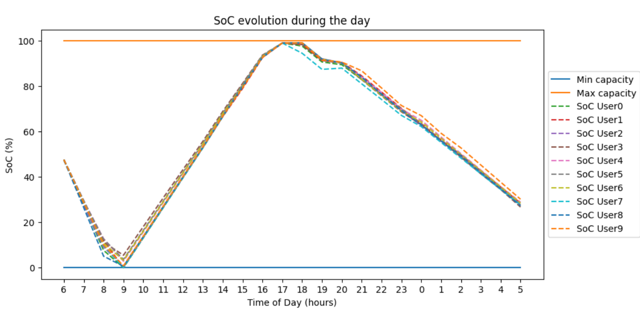


Fig. 6. Optimal SoC profiles with $\gamma = 1$

The optimal loads and battery use with $\gamma = 1$ are presented in Figures 5 and 6. Not all the photovoltaic energy available is consumed by the community. There is also still some dependence on the grid at the beginning of the horizon and at the end. Looking at the *SoC* profiles, without using the entire solar energy available, the battery of each player already reaches its maximum capacity. If more photovoltaic energy were consumed earlier on to match the available resources at that time, the maximum capacity would have been reached earlier on, causing the same amount of local resources not to be consumed.

4.3 Effect of uncertainty

The effect of uncertainty is assessed using the same simulated solar load as in section 4.2. The total community cost is compared for different scenarios ranging from 30% underestimations to 30% overestimation of the real photovoltaic production. This analysis also includes a perfect forecast scenario to use as benchmark. This is done for $\gamma = 0.6$ and $\gamma = 0.8$.

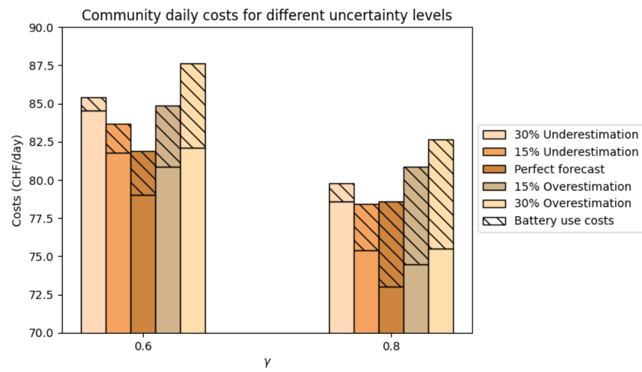


Fig. 7. Impact of imperfect forecasts on community costs for different self-production rates

Figure 7 shows the community costs for the different scenarios studied. The cost of importing energy (without accounting for battery use) is higher for all uncertainty scenarios than for perfect forecast.

For underestimation, the battery use cost described in equation (4) gets gradually smaller, but the energy cost described in (3) increases. Because the expected solar production is smaller, the incentive to use the battery is low, resulting in optimized loads being closer to the initial load.

For overestimations, the battery use cost and the energy costs increases. As more photovoltaic production is expected the incentive is now to use the battery more. However, the real solar production is smaller, resulting in the use of imported energy to charge the battery, causing the cost increase.

The above analysis provide interesting insights on the qualitative impact of uncertainty on performance, but when looking quantitatively at the cost increases caused by uncertainty, they seem relatively low. In the worst cases, when uncertainty reaches 30%, the cost does not increase by more than 10%.

Using a load game distributed optimizer seems to ensure that forecasting errors only cause small and acceptable losses in the performance of the controller.

5 CONCLUSIONS

In this paper, we have designed a new game theoretic framework to optimally manage small scale energy communities. The interactions between participants are modeled as a reduction in the available local energy for each player. Such a game can be solved distributively and can be implemented as the optimizer of a receding horizon controller. We have showed that doing so enables the community to maximize self-consumption given the battery capacity at hand. It also allows to generate battery use sequences which are resilient to noisy forecasts.

To build upon the work of this paper, the load game can be applied to more complex energy systems. For instance, it could be applied to systems that integrate both electrical and heating components, or those featuring dynamic energy tariffs, or scenarios where each agent possesses photovoltaic panels instead of relying on a central shared installation. Another possible direction involves stochastic events, such as fluctuations in solar production and agents' energy consumption patterns. Although we investigated the impact of solar production variability in our setting, our algorithm was not optimized for handling such stochastic elements. These represent promising areas for future exploration within the context of the load game framework.

REFERENCES

[1] IEA. Empowering people – the role of local energy communities in clean energy transitions, 2023.

[2] Council of the European Union European Parliament. Directive (eu) 2018/2001 of the european parliament and of the council on the promotion of the use of energy from renewable sources. *Off. J. Eur. Union*, pages 82–209, 2018.

[3] IEM. Directive (eu) 2019/944 on common rules for the internal market for electricity and amending directive 2012/27/eu. *Off. J. Eur. Union*, pages 125–199, 2019.

[4] Maksym Koltunov, Simon Pezzutto, Adriano Bisello, Georg Lettner, Albert Hiesl, Wilfried van Sark, Atse Louwen, and Eric Wilczynski. Mapping of energy communities in europe: Status quo and review of existing classifications. *Sustainability*, 15(10):8201, 2023. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[5] Anna-Riikka Kojonsaari and Jenny Palm. Distributed energy systems and energy communities under negotiation. *Technology and Economics of Smart Grids and Sustainable Energy*, 6(1):17, 2021.

[6] Dario Paccagnan, Maryam Kamgarpour, and John Lygeros. On aggregative and mean field games with applications to electricity markets. *2016 European Control Conference (ECC)*, 2016. ISBN: 9781509025916 Meeting Name: 2016 European Control Conference (ECC) Place: Aalborg, Denmark Publisher: IEEE.

[7] Hesam Farzaneh, Hamed Kebriaei, and Farokh Aminifar. Deterministic Mean Field Game for Energy Management in a Utility with Many Users. In *2018 Smart Grid Conference (SGC)*, pages 1–6, Sanandaj, Iran, November 2018. IEEE.

[8] Hesam Farzaneh, Mohammad Shokri, Hamed Kebriaei, and Farrokh Aminifar. Robust Energy Management of Residential Nanogrids via Decentralized Mean Field Control. *IEEE Transactions on Sustainable Energy*, 11(3):1995–2002, July 2020.

[9] Mohamad Aziz, Hanane Dagdougui, and Issmail Elhallaoui. A decentralized game theoretic approach for virtual storage system aggregation in a residential community. *IEEE Access*, 10:34846–34857, 2022. Conference Name: IEEE Access.

[10] Zhongjing Ma, Duncan S. Callaway, and Ian A. Hiskens. Decentralized charging control of large populations of plug-in electric vehicles. *IEEE Transactions on Control Systems Technology*, 21(1):67–78, 2013.

[11] Batchu Rajasekhar, Naran Pindoriya, Wayes Tushar, and Chau Yuen. Collaborative energy management for a residential community: A non-cooperative and evolutionary approach. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3(3):177–192, 2019.

[12] Edstan Fernandez, M. J. Hossain, Khizir Mahmud, Mohammad Sohrab Hasan Nizami, and Muhammad Kashif. A Bi-level optimization-based community energy management system for optimal energy sharing and trading among peers. *Journal of Cleaner Production*, 279:123254, January 2021.

[13] Louise Sadoine, Zacharie De Grève, and Thomas Brihaye. Valuing the electricity produced locally in renewable energy communities through noncooperative resources scheduling games, 2023.

[14] Yile Liang, Wei Wei, and Cheng Wang. A generalized nash equilibrium approach for autonomous energy management of residential energy hubs. *IEEE Transactions on Industrial Informatics*, 15(11):5892–5905, 2019.

[15] Dario Paccagnan, Basilio Gentile, Francesca Parise, Maryam Kamgarpour, and John Lygeros. Distributed computation of generalized nash equilibria in quadratic aggregative games with affine coupling constraints. In *2016 IEEE 55th conference on decision and control (CDC)*, pages 6123–6128. IEEE, 2016.

[16] Lorenzo Nespoli, Matteo Salani, and Vasco Medici. A rational decentralized generalized nash equilibrium seeking for energy markets. pages 1–6, 09 2018.

[17] Francisco Facchinei and Christian Kanzow. Generalized nash equilibrium problems. *4or*, 5(3):173–210, 2007.

[18] Mohammad Amin Tajeddini and Hamed Kebriaei. A mean-field game method for decentralized charging coordination of a large population of plug-in electric vehicles. *IEEE Systems Journal*, 13(1):854–863, 2019.

[19] Francisco Facchinei and Jong-Shi Pang, editors. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research and Financial Engineering. Springer, 2004.

[20] Tomasz T. Gorecki and William Martin. Maestro: A python library for multi-carrier energy district optimal control design. *IFAC-PapersOnLine*, 53(2):13293–13298, 2020. 21st IFAC World Congress.

[21] Sébastien Faivre. Rcp - regroupement dans le cadre de la consommation propre (rcp). Technical report, CECB, 2022.

Data-Driven Identification and Operational Optimization of Energy-Flexible Thermal Supply Systems

JAN ZANGENBERG*, Institute of Production Management, Technology and Machine Tools, Germany
JONAS WENDT, Institute of Production Management, Technology and Machine Tools, Germany
TOBIAS KOCH, Institute of Production Management, Technology and Machine Tools, Germany
TOBIAS KAPSER, Technical University of Darmstadt, Germany
MATTHIAS WEIGOLD, Institute of Production Management, Technology and Machine Tools, Germany

The accelerating expansion of renewable energies in Europe and the rest of the world leads to the challenge of adapting energy demand to the increasingly volatile renewable electricity generation. To contribute to this goal, thermal supply systems can be used for demand response by utilizing their dedicated and inherent thermal energy storages and optimizing the loading and unloading operation. In this paper we present a method for the data driven system identification and operational optimization of industrial thermal supply systems. This method aims at developing optimization problems for the control of thermal supply system and identifying system dynamics specifically for this purpose. The method includes parameter identification and validation of individual components, such as compression chillers and thermal storages, using the system's measurement data. From the identified component dynamics an overall optimization model is formulated and tested, taking into account the interactions between single components. Finally, the optimization model is deployed, and the optimized control signals are passed onto the real systems components. For permanent application of the optimization, monitoring und maintenance of the model are integrated as a last and recurring step of the method. This includes updating model parameters in case of increasing mismatch between model and reality over time. These steps are successfully implemented in the real-world example of the cooling supply system of a climatized room at the ETA Research Factory at the Technical University of Darmstadt. In this example the room's thermal capacity is used to vary its cooling supply depending on time-variable electricity prices. The model is deployed as a model predictive control loop, regularly updating its optimal trajectory using electricity prices of a representative period from 2023. Just by optimizing the room's temperature control a reduction in energy cost of 10.06 % compared the conventional operation was achieved. This results in a reduction of electricity related carbon emissions of 4.93 %.

Additional Key Words and Phrases: Demand response, energy-flexibility, system identification, mathematical optimization, optimal control, model predictive control

Authors' Contact Information: Jan Zangenberg, j.zangenberg@ptw.tu-darmstadt.de, Institute of Production Management, Technology and Machine Tools, Darmstadt, Hessen, Germany; Jonas Wendt, Institute of Production Management, Technology and Machine Tools, Darmstadt, Hessen, Germany, j.wendt@ptw.tu-darmstadt.de; Tobias Koch, Institute of Production Management, Technology and Machine Tools, Darmstadt, Hessen, Germany, t.koch@ptw.tu-darmstadt.de; Tobias Kapser, Technical University of Darmstadt, Darmstadt, Hessen, Germany, tobias.kapser@outlook.com; Matthias Weigold, Institute of Production Management, Technology and Machine Tools, Darmstadt, Hessen, Germany, M.Weigold@PTW.TU-Darmstadt.de.

Abbreviations	
AHU	Air handling unit
API	Application programming interface
APOPT	Advanced process optimizer
CC	Compression chiller
CR	Climate room
EER	Energy efficiency ratio
ENTSO-E	European Network of Transmission System Operators for Electricity
IBC	Intermediate bulk container
MAE	Mean absolute error
MPC	Model predictive control
OPC UA	Open Platform Communications Unified Architecture
PI	Proportional–integral
RC	Resistance capacitance
REST	Representational state transfer
TCP	Transmission control protocol
VDI	The Association of German Engineers
Variables, parameters and indices	
$\bar{A}$	Average peak-to-peak amplitudes
b	Index representing base data
β	Estimation parameter
C	Electricity price
c	Specific heat capacity
D	Multivariate distance function
G	Thermal conductivity
h	Number of hours in time horizon
i	Index representing time segment
j	Index representing peak to peak element
M	Number of extrema
m	Mass
n	Number of time segments
P^{el}	Electrical power
$\dot{Q}$	Cooling output of compression chiller
σ	Standard deviation
T	Temperature
y	Measured values
Indices	
amb	Ambient air
m2	"virtual mass" representing masses beside air in the room
rep	Represents
train	Training set
val	Validation set

1 Introduction

In its Renewable Energy Directive the European Union raised its binding target for the share of renewable energies in the European energy mix to 42.5 % by 2030 [36]. The progressive expansion of renewable energies and the associated fluctuating energy generation result in increasingly volatile energy prices. The development of electricity price frequency from 2019 to 2023 can be seen in Figure 1. Taking into account the energy crisis in 2022, increasing average and variance in electricity prices over the last years can be observed in the bidding zone of Germany and Luxemburg. To counteract the imbalance of renewable energy generation and consumption, demand side energy-flexibility can be utilized. This enables adapting the energy demand to the volatile renewable energy supply and thus, stabilizing the energy system. [31] Given the right energy tariffs a market participant can economically benefit from price-driven demand response [33] . Previous studies have shown a significant positive correlation between Day-Ahead electricity prices and the associated greenhouse gas emissions [14]. Thus, a price-driven economical optimization of the flexible operation can also reduce emissions caused by electricity generation. Thermal supply systems are utilized in various processes, but they are especially significant in the building sector (both industrial and residential), which accounts for 33 % of the worldwide final energy demand. The share of thermal energy in the building sector is 77 %. [30] Given the thermal inertia of the building itself and integrated storage tanks, thermal supply systems of buildings have high potential for demand response measures [7]. In addition to providing heat in industrial and residential areas, the cooling supply is also becoming increasingly important, particularly in the building sector for air conditioning. The number of air conditioning systems is rapidly rising due to global warming and growing prosperity [8].

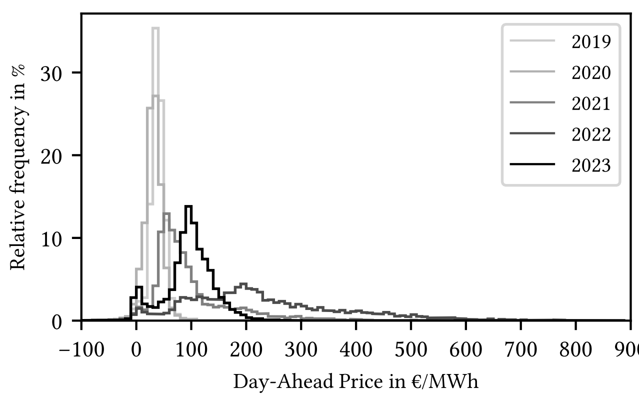


Fig. 1. Relative frequency of Day-Ahead electricity prices for bidding zone Germany-Luxembourg between 2019 and 2023. Data from ENTSO-E [12]

To maximize the economic benefit of energy-flexibility in thermal supply systems, mathematical optimization can be utilized. This enables optimal decision-making for energy-flexibility measures. Thermal supply systems can vary in their structure and complexity, but often contain recurring components, like pumps, valves, heat pumps, compression chillers and storages tanks. This motivates the development of standardized models for these single components which can then be assembled in an optimization problem for the overall system, an approach that is already existing in simulation of thermal systems [5]. Research efforts in optimal control of thermal supply systems for demand response often demonstrate improvements, by developing individual, use case specific models (see [25] and [21]), that come with little generalization capability. There are multiple general software frameworks for formulating optimization problems like Pyomo [6], GEKKO [4] or CasADi [2], and even simulation tools like Dymola for Modelica, where extensive model libraries for thermal systems exist, e. g. the Modelica Buildings Library [42]. Model collections specifically for creating dynamic optimization problems of thermal systems remain rare. In this paper we present a method that aims at structured component-based development of optimization problems to enable transfer of the developed models to further use cases. To address the problem of a mismatch between optimization model and the real system, in our approach, model parameters are estimated, using measurement data of the real system, as Weber proposes [40].

In Section 2 of this paper, we summarize the basics of energy-flexibility and mathematical optimization. Since the use case for application of the method is a cooling system, we provide the relevant fundamentals of cooling supply systems. In Section 3 we present the method of data-driven identification and optimization for thermal supply systems and explain each step of the method and its advantages. The method is applied to a use case of a climatized room and its cooling supply system. This system as well as the process of selecting representative electricity prices for an energy-flexible experiment is described in Section 4. In Section 5 we apply the parameter estimation, optimization and deployment to the mentioned system, and evaluate the results of this experiment. We draw a conclusion and give an outlook to further topics in this research area.

2 State of the Art

In the following Section we describe the basics of energy-flexibility, cooling supply systems, as a subcategory of thermal systems, and mathematical optimization.

Energy-flexibility is defined by Eurelectric as the "modification of generation injection and/or consumption patterns in reaction to an external signal (price signal or activation) in order to provide a service within the energy system" [13]. The definition of Demand Response provided by the U.S. Department of Energy, as cited in [26] closely aligns with the interpretation of Eurelectric and our understanding of Energy-Flexibility. Given the minor variances between these definitions in our specific context, we consider them synonymous and use the term Energy-Flexibility in the following chapters.

The potential for energy-flexibility in factories is harnessed through the implementation of strategies known as energy-flexibility measures, which are standardized in the VDI 5207 guideline [37]. These measures can span various timeframes, encompass different levels of enterprise hierarchy or automation, and range in complexity. Examples of energy-flexibility measures include organizational decisions made on a management level, such as "adjust shift times" to meet with periods of high solar generation as well as technical solutions like energy price oriented tank temperature controlling ("store energy inherently") in aqueous parts cleaning machines via a digital twin, as Fuhrländer-Völker et al. demonstrate in [15]. From an industrial standpoint, the primary goal of energy-flexibility is to optimize (energy) costs. This can be achieved by objectives such as reacting to volatile energy prices, maximizing self-consumption of on-site generation, but also saving infrastructural costs by increasing a factory's energy resilience or even gaining extra revenue streams like external marketing of energy-flexibility (e.g. to grid operators). While the focus might be either a single or a multi-objective optimization, it is important to recognize that these objectives are interconnected [37]. Hence, analyzing and determining these objectives with respect to their (technical) requirements is essential prior to implementing energy-flexibility measures.

Cooling supply systems are utilized not only in air conditioning but also in various industrial processes, such as machine cooling [29]. They are therefore a crucial component of modern industrial infrastructure and maintain specific temperatures for several processes. Cooling supply systems consist of a heat source (e.g. a production machine), a cooling medium (e.g. a water-glycol mixture) and a heat sink (e.g. a compression chiller). In this context, chillers transfer heat against a temperature gradient and consume energy in the process [22]. Evaluating these systems for energy efficiency measures is important, given their widespread necessity across numerous processes and their consistent requirement for energy to facilitate heat transport. One exception is free cooling by natural convection [19]. The main energy consumers are chillers. However, peripheral devices such as pumps also significantly contribute to the final energy demand of these systems [20]. The efficiency of a chiller is usually quantified by the energy efficiency ratio (EER). This is the ratio of the total cooling capacity to the effective power consumption of the chiller [9]. Ideally, the waste heat from the compression chiller can be reused as useful energy for other processes as shown in [27], but in most cases, it is transferred directly to the environment. The efficiency of heat dissipation to the environment increases with a larger temperature difference between the cooling medium and the ambient temperature. Since the ambient temperature varies greatly throughout the day and season, the EER of most chillers also varies. Another significant factor in EER variation is cooling demand. Such systems should generally be designed for a specific application. If they operate outside of an optimal operating point, the EER will decrease [1]. Thermal buffer storage tanks are frequently utilized to maintain the cooling system operating at or near the optimal operating point, even when ambient conditions or cooling demand fluctuate. These storage tanks, as well as the entire cooling supply systems, can therefore be used for energy-flexible operation. Dedicated thermal energy storages with bypass and mixing valves are even better suited for such measures than pure permanent flow buffer tanks [41]. As a result, wider temperature ranges can be operated in the thermal energy storage and thus greater energy-flexibility potential can be realized. For the combination of a cooling supply systems with a ventilation system (see Section 3), it is recommended to use so-called inherent thermal energy storages [37] since previous works have concluded that inherent thermal energy storages in production systems have sufficient energy-flexibility potential to be utilized for demand response [35]. These storage units are already existing thermal inertia masses, consisting of the room air, stored content, and the structure of the room itself. They can be specifically temperature-controlled to bridge expensive price periods or periods with high CO2 intensity. The wider the permitted temperature limits are set, the greater the potential for energy-flexibility [34].

To maximize the economic benefits of energy-flexible control, and automate the decision-making process, mathematical optimization models are suitable for finding the optimal control signals of thermal supply systems and their components. Compared other data-driven methods to achieve optimal control, e. g. Reinforcement Learning, mathematical optimization models are interpretable and have better generalization capability, when operating conditions differ from the training data. This paper addresses the challenge of developing optimization programs and estimating model parameters for mathematical optimization of thermal supply systems. Mathematical optimization problems consist of decision variables that can be adjusted, an objective function that is to be minimized or maximized and constraints that must be met when finding an optimal solution. [6] Given the problem's variables are indexed over multiple discrete time steps and are dependent on one another, we speak of dynamic optimization problems [17]. To build accurate optimization models the model parameters can be estimated based on measurement data [18]. We present an adequate method for the data-driven identification, structured model development and deployment of optimization problems for the energy-flexible operation of thermal supply systems.

3 Method

In this section, we introduce a method aimed at enhancing the data-driven identification and operational optimization of energy-flexible thermal supply systems. Initially, we contextualize our approach in the overall method for energetically flexibilizing factories, as presented in VDI Case Study [3]: First, a comprehensive energy-flexibility potential analysis is conducted for all systems within a factory, assessing both economic and technical potential. This analysis prioritizes systems and energy-flexibility measures. In the subsequent step of conception and planning, the previously prioritized energy-flexibility measures are examined in detail, measurements are taken, and implementation plans are devised. The realization and implementation phase follows, which includes the enablement of the information and communication technology infrastructure (such as sensor implementation and adjustment of the control architecture), as well as the implementation of energy-flexibility measures with automation and/or organizational approaches. The subsequent steps are marketing of energy-flexibility, controlling and monitoring and the overall process optimization. The method, presented in this paper, is located in the realization and implementation phase, when the information and communication technology infrastructure is already fully enabled. Figure 2 presents the method for a data-driven identification and operational optimization of energy-flexible thermal supply systems. It is designed to develop, implement and deploy optimization problems according to the "offline identification and online control" loop by Huang and Gao [18]. We integrate a data-driven model estimation on component level and structured assembly of the optimization problem to ensure transferring the identified models to further use cases.

The method consists of the identification, the optimization, and the deployment phase. The identification is based on the system

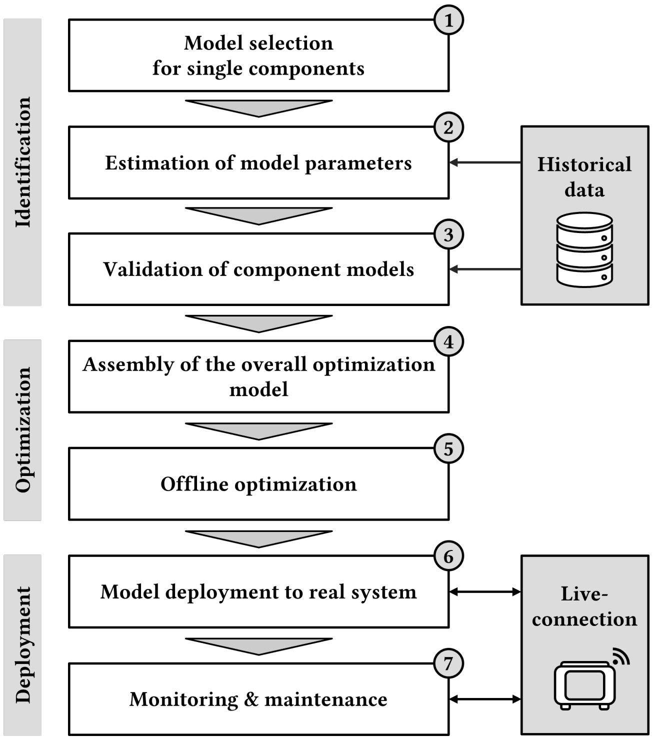


Fig. 2. Method for data-driven identification and operational optimization of energy-flexible thermal supply systems.

identification loop by Ljung [24]]. To create an optimization problem for complex systems, such as thermal supply systems, the single components must be mathematically modelled individually. Firstly, these models need to be selected depending on their complexity, performance, and overall suitability for optimization (1). Component models can reach from physics based white-box models to purely statistically identified black box models as well as physics-informed based grey-box models, as a combination of both [10]. Model type and complexity of the component models depend on design of the overall optimization problem. For example, in Section 5 we create a linear optimization problem, therefore only linear component models can be used. Model parameters can sometimes be determined from the component's datasheet but often, e.g. in statistical models, parameters have no physical based value. Also, purely physics-based models can be very complex. This can result in too long solving durations of the final optimization problem for optimal control. In that case, model simplification is necessary. For those reasons we suggest a parameter estimation (2) based on grey box models using measurement data recorded on the real system. Once parameterized, the single component models are validated using a validation dataset from the real system (3). For the operational optimization of the system a mathematically formulated optimization problem is required. The overall optimization model is assembled from the identified single components, the system consists of (4). The mathematical equations describing the behavior of components serve as constraints for the optimization problem. E.g. a model equation for a compression chiller of the form

$$P_t^{\mathrm{el}} = f(\dot{Q}_t) \tag{1}$$

will result in an equality constraint between the electrical power P^{el} and the function of the heat flow $\dot{Q}$ for each time step t. Since the system's parameters are estimated on the components level, no parameter estimation on the system level is needed. This approach requires the individual components to be equipped with the necessary sensors for data acquisition. The next step (5) involves one or more offline optimization runs to test solvability, solving duration and plausibility of the optimization result. If the results are valid the optimization can be deployed to the real system and validated (6) – reading live data, optimizing for minimal cost, and writing optimal control signals to the corresponding system components. Like machine learning models, the performance of an optimization model in production can degrade. Therefore, model performance should be monitored (7). To ensure permanent operation of the model, an online system identification is required to minimize the long-term model mismatch [28]. If a model's performance degrades the parameters should be updated with more recent measurement data. The parameter estimation can also be conducted within the control loop based on recent measurement data, so that the model's parameters are regularly updated as described by Weber [40]. This process fits in with the approach of post-identification control described by Huang and Gao [18] since the model is identified before the optimization takes place.

This method allows for building standardized component models and applying them to different systems by estimating the individual model parameters. This transferability and adaptability enable rapid and structured development of optimization models, provided that the required historical data is available and data connection for the required measurements and control signals is implemented.

4 Use case

To showcase the presented method, we conduct the central steps for a real-world example of a climatized room with a dedicated cooling supply system. The goal is to store energy inherently by using the room itself as a thermal capacity. Keeping the room within specified temperature limits, its cooling supply can be shifted to times, when electricity prices are low. When prices rise, the cooling system can then be relieved to reduce overall cost of energy.

4.1 Description Climate Room

The optimizer's validation tests are conducted in the climate room of the ETA-Factory [38] at the Institute of Production Management, Technology and Machine Tools (PTW). This controlled environment minimizes external disturbances and allows for repeated testing under almost identical conditions. The temperature, electrical, and thermal performance of the climate room and its components are precisely measured, enabling detailed energy balancing. For specified descriptions of the climate room and its possible applications, please see [39] and [32]. These sources describe the overall system as shown in Figure 3. However, only the components highlighted in the figure are relevant for the present work. These are a compression chiller (CC), an air handling unit (AHU) and the climate room (CR) with an intermediate bulk container (IBC), including the associated measurement instruments.

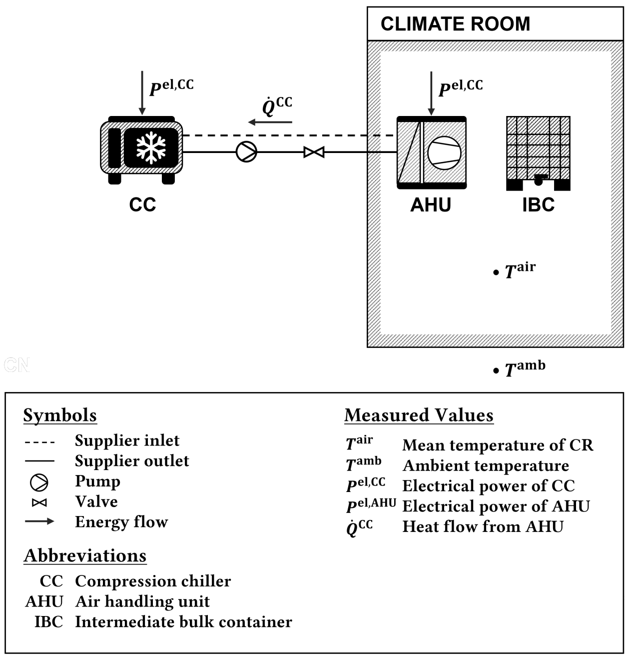


Fig. 3. Schematic representation of the climate room and its components (chiller, air handling unit, and IBC tank). Highlighting of the components relevant to this work. Based on [41] and [37].

The compression chiller can operate within a cooling output range of 3.5 to 10 kW [39] and release the waste heat either to a central water network or, as in this case, to the environment via an air cooler. Deionized water is used as the cooling medium for the components of the climate room. The cooled water is pumped to the AHU via the outlet of the chiller. There, it flows through a heat exchanger and absorbs heat from the room, in which the fans of the AHU blow the warm air to the heat exchanger. The now heated deionized water flows back to the inlet of the chiller, where it is cooled down again. The compression chiller's electrical connection allows for a power consumption of up to 4.99 kW. To generate an artificial thermal inertia mass in the room, an IBC containing 1000 liters of water is added. As a result, the room does not react directly to temperature changes, but with a time delay, as required for the use case in Section 4.

The ETA Research Factory utilizes a comprehensive energy management system to manage all measured and control variables. Communication primarily occurs via OPC UA and Modbus TCP communication protocols. This also applies to measured variables in Section 4. The average room temperature T_{CR} is calculated from eight temperatures measured in the individual corners of room. The ambient temperature T^{amb} refers to the temperature near the outside of the room's walls and is approximately equal to the hall temperature of the ETA-Factory. Power meters are used to measure the electrical power of the compression chiller P^{CC} and of the AHU P^{AHU}. These measurements are particularly important for calculating operating costs. The heat flow Q^{CC} from the AHU to the cooling medium is measured using a heat meter and is necessary for thermal balancing and calculating the COP. The climate room and its components also have additional measurement and control technology. However, this is not shown in Figure 3 for the sake of simplicity.

4.2 Selecting Representative Electricity Prices

Given that electricity prices constitute a significant component of the cost function in the optimization problem, they should be selected to be as representative as possible for this use case. A temporal horizon h of 168 hours (7 days) is suitable for the experiment duration, as it includes both weekdays and weekends. For this purpose, a one-year electricity price dataset $C(x)$ is divided into n=52 seven-day segments. The goal is to identify the segment i_{rep} that best represents the entire dataset based on various metrics. Given the objective of optimized energy-flexible operation, the following metrics have been chosen:

- Mean price $\bar{C}$, to account for the average price.
- Standard deviation σ, for the fluctuation around $\bar{C}$.
- Number of extrema M, as an indication of the number of inexpensive and expensive cycles.
- Average peak-to-peak amplitudes $\bar{A}$, to determine how large the spreads between cheap and expensive cycles are on average.

For each segment $i \in \{1, 2, \ldots, n\}$ a multivariate distance function D_{i} is calculated, which indicates how far the segment deviates from the base dataset. The function for D_{i} is an extension of the Euclidean distance [23]. The individual components are normalized to standardize the different dimensions of the metrics. The required segment i_{rep} is represented by the smallest D_{i} and can be determined as follows:

$$i_{\mathrm{rep}} = \arg\min_i (D_i) \tag{2}$$

In the following, the equations for D_{i} and their components are derived. Here, the indices i refer to the segments and b to the entire base dataset, with x representing the hours since the start x_o. The extreme values were filtered such that if the j-th peak-to-peak amplitude $A_{(\mathrm{i,j})}$ between two local extremes in segment i is less than 10 €/MWh, this pair of extreme values is not considered further and is not included in the count of extreme values N_i. Smaller amplitudes are not significant for the system under consideration.

$$D_i = \sqrt{\left(\frac{\bar{C}_i - \bar{C}_b}{\bar{C}_b}\right)^2 + \left(\frac{\sigma_i - \sigma_b}{\sigma_b}\right)^2 + \left(\frac{M_i - M_b}{M_b}\right)^2 + \left(\frac{\bar{A}_i - \bar{A}_b}{\bar{A}_b}\right)^2} \tag{3}$$

with

$$\bar{C}_i = \frac{1}{h} \sum_{j=h\cdot(i-1)}^{h\cdot i} C(x_0 + j) \tag{4}$$

$$\bar{C}_b = \frac{1}{h \cdot n} \sum_{j=0}^{h\cdot n} C(x_0 + j) \tag{5}$$

Table 1. Overview of the metrics for the base data set b and the identified representative time segment i_{rep}

Metric	b	i_{rep}
Mean price $\bar{C}$	95.45 €/MWh	91.31 €/MWh
Standard deviation σ	47.50 €/MWh	39.91 €/MWh
Number of extrema M	25.06	24
Average amplitudes $\bar{A}$	70.33 €/MWh	74.35 €/MWh

$$\sigma_i = \sqrt{\frac{1}{h-1} \sum_{j=h(i-1)}^{h \cdot i} \left(C(x_0 + j) - \bar{C}_i\right)^2} \tag{6}$$

$$\sigma_b = \sqrt{\frac{1}{h \cdot n - 1} \sum_{j=0}^{h \cdot n} \left(C(x_0 + j) - \bar{C}_b\right)^2} \tag{7}$$

$$M_b = \sum_{i=1}^{n} M_i \tag{8}$$

$$\bar{A}_i = \frac{2}{M_i} \sum_{j=1}^{M_i/2} A_{i,j} \tag{9}$$

$$\bar{A}_b = \frac{1}{n} \sum_{i=1}^{n} \bar{A}_i \tag{10}$$

The base data for this use case contains the day-ahead electricity prices for the entire year 2023. The prices are shown in Figure 4a. The data can be loaded via the ENTSO-E Transparency Platform RESTful API [12]. After dividing the base data into n segments and calculating the respective D_{rep}, the result for the representative segment is $i_{\text{rep}} = 11$ (which corresponds to the 12th until 19th of March 2023). The detailed view of the representative segment is shown in Figure 4b. The calculated metrics for the base dataset and for i_{rep} are shown in Table 1.

5 Application and Validation

In the following, the method presented in Section 2 is applied to the cooling supply use case described in Section 3. Therefore, we conduct a data-driven parameter estimation for a stationary linear model of the compression chiller and for two different dynamic models of the climatized room using a training data set. Models for both components are then validated on a validation data set. The selected component models are used to formulate an optimization problem, which is then deployed to the real system.

5.1 Parameter Identification

To estimate the model parameters of both chiller and room, we use the corresponding mathematical models (sections 5.1.1 and 5.1.2), calculate the loss function for the variable to be explained and minimize this loss function by applying an optimization algorithm. As a loss function for the parameter estimation problem, the Mean Absolute Error (MAE) is used to calculate the error between predicted values $\hat{y}_i$ and measured values y_i:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \tag{11}$$

5.1.1 Compression Chiller. For the compression chiller, a simple linear model is used for the relation between the electrical power $P_t^{\text{el,CC}}$ and the cooling output $\dot{Q}_t^{\text{CC}}$ at each time step t:

$$P_t^{\text{el,CC}} = \beta_0 + \beta_1 \dot{Q}_t^{\text{CC}} \tag{12}$$

Of course, this simple linear model does not fully represent the behavior of a compression chiller and every single influencing factor. But since certain variables, e.g., ambient temperature, remain constant in this specific use case, they are left out of the equation.

In Table 2, the results of the parameter estimation for the compression chiller are displayed. The model behavior can be analyzed in Figure 5a for the training dataset and Figure 5b for the validation dataset. The identification was conducted with data from an experiment with hysteresis control of the room temperature over a duration of 37 hours.

Table 2. Overview of the identified parameters β_0 and β_1 for the compression chiller model and error metrics on training and validation data sets.

Type	**Symbol**	**Value**
Input variable	$\dot{Q}^{\text{CC}}$	-
Output variables	$P^{\text{el,CC}}$	-
Estimated model parameters	β_0	407.47 W
	β_1	0.311
Error	$\text{MAE}^{\text{train}}$	55.77 W
	MAE^{val}	60.20 W
Algorithm	scipy.optimize.minimize (L-BFGS-B) [38]	

In this model, the parameter β_0 can be seen as the chiller's base power consumption, while β_1 describes the variable part that is related to the produced cooling output. With a MAE of 55.77 W on the training dataset and 60.20 W on the validation set, model mismatch is very small. From those metrics, we conclude that sufficiently accurate model parameters have been found and the model is not overfitted to training data. In Figure 5c and Figure 5d, we observe some dynamic behavior of the measured electrical power in the form of power spikes that the model is not able to represent. This seems to be especially the case in periods of no provided cooling when the chiller is on standby. All in all, the error metrics and plots for the validation dataset suggest good overall model performance, given the simplicity of the model. It will be used as a component model for the compression chiller in the overall optimization problem in Chapter 4.2.

5.1.2 Climatized Room. For the room, two models are tested. Firstly, a simple model that depicts the room as a homogeneous body. The model describes the change of the room temperature $\dot{T}^{\text{air}}$ as a function of the current room temperature T^{air}, the ambient temperature T^{amb}, and the cooling input $\dot{Q}^{\text{CC}}$ applied to the room via AHU.

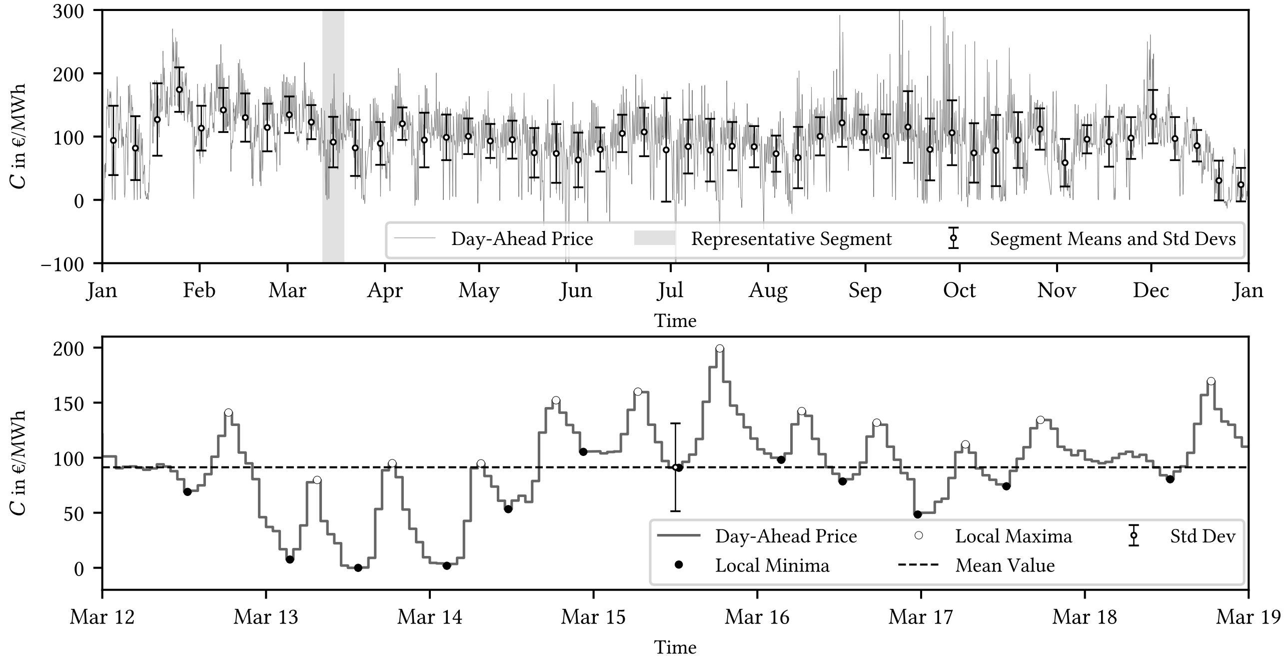


Fig. 4. a) Segmented day-ahead electricity prices C of 2023 for bidding zone Germany-Luxembourg [12] with the segment mean values $\bar{C}_b$ and standard deviations σ_b, as well as the highlighting of the representative segment i_{rep}. b) Detailed view of electricity prices, segment mean $\bar{C}_{i_{\text{rep}}}$, standard deviations $\sigma_{i_{\text{rep}}}$, and local extrema for i_{rep}.

$$m^{\text{air}} c^{\text{air}} \dot{T}^{\text{air}} = G^{\text{air}} (T^{\text{amb}} - T^{\text{air}}) - \dot{Q}^{\text{CC}} \tag{13}$$

Figure 6 shows the corresponding RC-thermal circuit model. It consists of the thermal capacitance $m^{\text{air}} c^{\text{air}}$ describing the potential of the room to store thermal energy inherently. The thermal conductivity G^{air} connected to the room capacitance allows for heat transfer between the ambient air and the room, proportional to their temperature difference. Heat is additionally removed from the room capacitance by the cooling input $\dot{Q}^{\text{CC}}$.

Table 3 shows the results and error metrics of the parameter estimation. With an MAE of 1.1 °C on the training dataset and 0.931 °C on the validation dataset, the 1R1C model does not sufficiently represent the real dynamic behavior of the room temperature given the model's input variables. This model mismatch can also be observed in Figure 8, where the model behavior on the training dataset is depicted. Both plots, training and validation, show the parameterized model's dynamical behavior given the initial conditions for the room temperature T_0^{air} as well as the input variables T^{amb} and $\dot{Q}^{\text{CC}}$. We can see that the model cannot reproduce the quick cooling and heating phases of the real system and therefore is too simple for this application.

To get a better representation of the real system, we apply a more complex 2R2C model and conduct the parameter identification. This second model extends the first by introducing one more mass. This "virtual mass" allows for heat transfer with the room air and the ambient air. It can represent the walls of the room as well as storage mass m_2 in the room, like the IBC tank.

Table 3. Overview of the identified thermal capacitance $m^{\text{air}} c^{\text{air}}$ and thermal conductivity G^{air} for the 1R1C room model and error metrics on training and validation data sets.

Type	Symbol	Value
Input variable	$T^{\text{amb}}, T^{\text{air}}, \dot{Q}^{\text{CC}}$	-
Output variables	$\dot{T}^{\text{air}}$	-
Estimated	$m^{\text{air}} c^{\text{air}}$	3,494,087.1 J/K
model parameters	G^{air}	134.3 W/K
Error	$\text{MAE}^{\text{train}}$	1.100 °C
	MAE^{val}	0.931 °C
Algorithm	GEKKO Optimization Suite (APOPT) [4]	

$$m^{\text{air}} c^{\text{air}} \dot{T}^{\text{air}} = G^{\text{air,m2}} (T^{\text{m2}} - T^{\text{air}}) - \dot{Q}^{\text{CC}} \tag{14}$$

$$m^{\text{m2}} c^{\text{m2}} \dot{T}^{\text{m2}} = G^{\text{amb,m2}} (T^{\text{amb}} - T^{\text{m2}}) - G^{\text{air,m2}} (T^{\text{m2}} - T^{\text{air}}) \tag{15}$$

In comparison to the 1R1C model, a second thermal capacitance $m^{\text{m2}} c^{\text{m2}}$ describing the virtual mass is added to the second room model (Figure 7). It is connected to the thermal room capacitance by the thermal conductivity $G^{\text{air,m2}}$ and to the ambient temperature by the thermal conductivity $G^{\text{amb,m2}}$.

The results for the 2R2C room model in Table 4 show that by introducing a second thermal mass to the model and estimating its

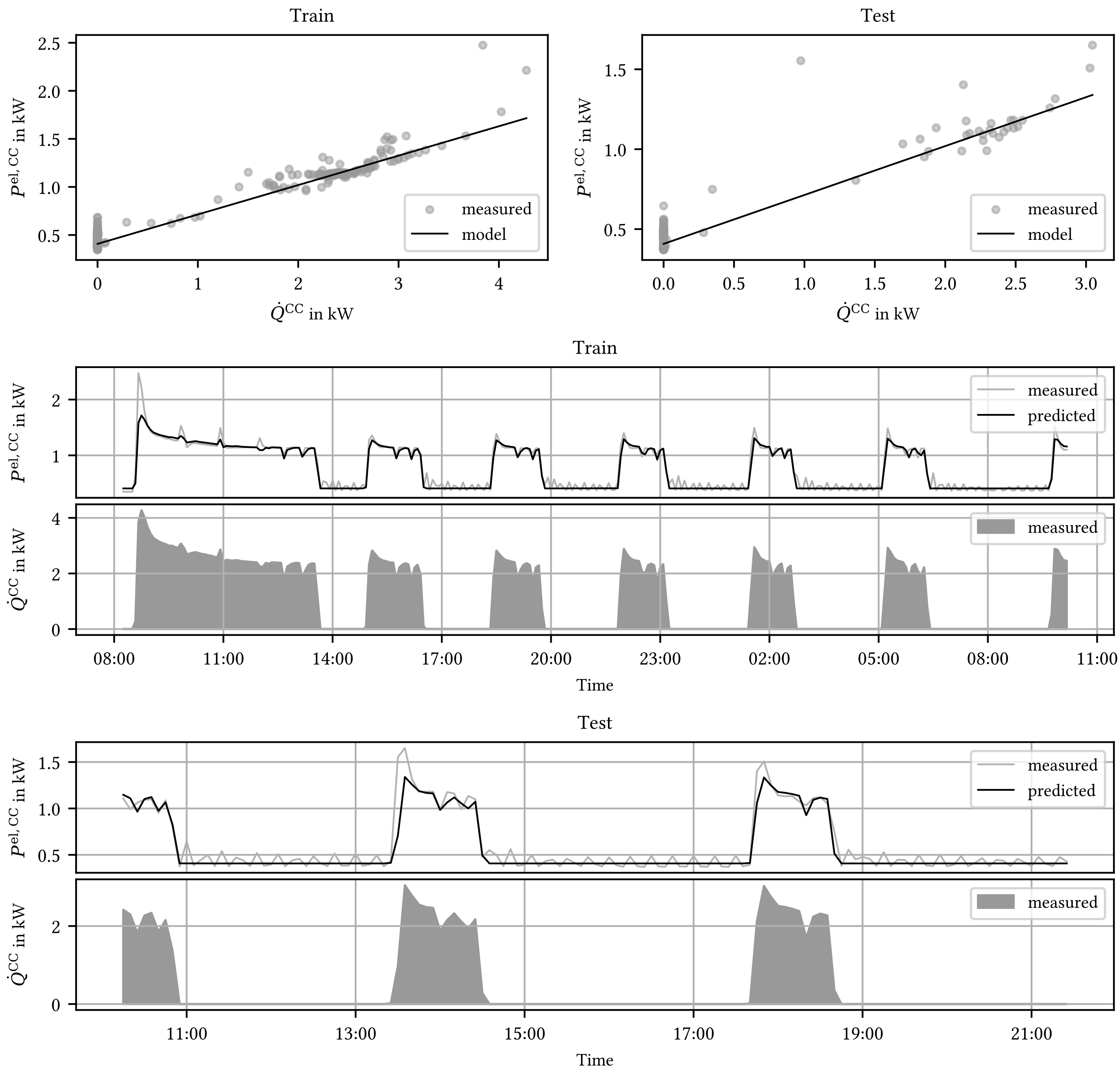


Fig. 5. a) Measured and predicted electrical power consumption P^{el} for the linear compression chiller model against the measured cooling output $\dot{Q}^{CC}$ for training dataset and b) validation dataset. c) Time series plot of measured and predicted electrical power consumption for the linear compression chiller model and measured cooling output for training dataset and d) validation dataset.

parameters, we obtain a much better model fit to the training data with an MAE of 0.201 °C. The model represents the measurement data very well. The validation error has also improved, although it is larger than the training error by 0.603 °C. This could suggest that the 2R2C model is slightly overfitted to the training data. This error could be improved by model adjustment and by using multiple and/or longer time periods for parameter estimation.

Although the validation error of the 2R2C model has only improved a little compared to the simpler 1R1C model, we observe that the model behavior of the 2R2C model is also a qualitatively

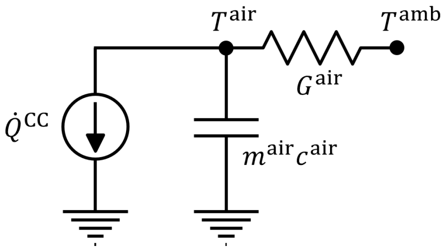


Fig. 6. RC-thermal circuit model for the 1R1C room model.

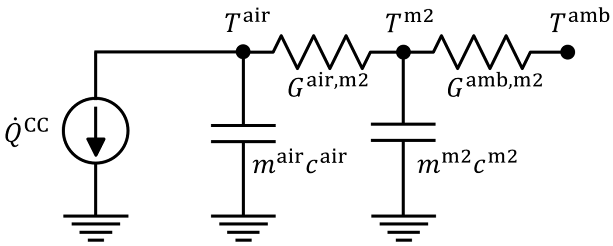


Fig. 7. RC-thermal circuit model for the 2R2C room model.

Table 4. Overview of the identified thermal capacitances $m^{\mathrm{air}}c^{\mathrm{air}}$ and $m^{\mathrm{m2}}c^{\mathrm{m2}}$, thermal conductivities $G^{\mathrm{air,m2}}$ and $G^{\mathrm{amb,m2}}$ for the 2R2C room model and error metrics on training and validation data sets.

Type	**Symbol**	**Value**
Input variable	$T^{\mathrm{amb}}, T^{\mathrm{air}}, \dot{Q}^{\mathrm{CC}}$	-
Output variables	$\dot{T}^{\mathrm{air}}$	-
Estimated model parameters	$m^{\mathrm{air}}c^{\mathrm{air}}$	755,383.3 J/K
	$m^{\mathrm{m2}}c^{\mathrm{m2}}$	6,816,201.5 J/K
	$G^{\mathrm{air,m2}}$	466.5 W/K
	$G^{\mathrm{amb,m2}}$	144.3 W/K
Error	$\mathrm{MAE}^{\mathrm{train}}$	0.201 °C
	$\mathrm{MAE}^{\mathrm{val}}$	0.804 °C
Algorithm	GEKKO Optimization Suite (APOPT) [4]	

better representation of the real system dynamic. Another characteristic is the model's capability to represent the real system over time until it diverges. On training and validation data, the model mismatch of the 1R1C model seems to be changing over time, while the 2R2C model's error remains mostly constant. Since both models are still abstract representations of the real system, consisting of multiple thermal masses with different thermal characteristics, a mismatch between model and measurement of the real system cannot be avoided. The 2R2C has better performance on both training and validation dataset and still a low model complexity to be used in optimization. Therefore, in the next section this model is selected for assembling the optimization problem of the overall system - the fourth step of the method (Section 5.2).

5.2 Optimization and Deployment

5.2.1 Optimization Problem. To build the overall optimization problem, the single component models for the chiller and room are coupled together by their shared variable $\dot{Q}^{\mathrm{CC}}$. The mathematical formulas described in Section 5.1 are used as model constraints that cannot be violated while finding an optimal solution.

For the objective function to be minimized by the optimizer, the electricity price C_t^{el} and the electrical power consumption $P_t^{\mathrm{el,CC}}$ of the compression chiller are multiplied and summed up over the prediction horizon of the optimization model:

$$\min \sum_{t=1}^{N} C_t^{\mathrm{el}} \cdot P_t^{\mathrm{el,CC}} \tag{16}$$

In the optimization model, only the electrical power of the compression chiller is included in the cost function. This is because the chiller is by far the largest consumer, and the components pump and AHU are not explicitly modeled as components.

This optimization model is deployed within a model predictive control (MPC) loop. Every 15 minutes, the current system variables and prediction time series of the system are read, and the optimization problem is solved for the horizon of 24 hours. From the calculated optimal room temperature trajectory, the value of the next timestep T_{t+1}^{air} is passed as a setpoint to the PLC of the room's cooling system using the OPC UA connector from the eta-utility Python library [16]. The setpoint temperature is then realized by the room's own PI-controller. For the electricity price, the identified representative period of 12th until 19th of March 2023 is used. In this experiment, the room temperature variable is allowed between a lower bound $T_{\min}^{\mathrm{air}}$ of 10 °C and an upper bound $T_{\max}^{\mathrm{air}}$ of 15 °C.

$$T_{\min}^{\mathrm{air}} \leq T_t^{\mathrm{air}} \leq T_{\max}^{\mathrm{air}} \quad \forall t \in \{1, \ldots, N\} \tag{17}$$

The optimization model was formulated using the GEKKO Optimization Suite (APOPT) [4]. To quantify the result of the energy-flexible optimized control, a reference experiment for a conventional scenario was conducted. For the conventional control, the room's PI controller was set to maintain a constant room temperature of 15 °C.

5.2.2 Result. Figure 10 shows the optimized control signals as well as the resulting behavior of the climate room and its cooling supply system in comparison to the electricity prices. In times of low electricity prices followed by high prices, it can be observed that the room is cooled below its upper limit of 15 °C, resulting in an increased electrical power consumption of the compression chiller. Once prices increase again, the chiller's cooling utilization is reduced, and the room temperature rises.

This temporal load shifting measure is only utilized when the described price fluctuation is high enough. Cooling the room down below 15 °C results in a higher temperature gradient and thus more heat input from the environment into the room, leading to a higher required cooling input. Since the 2R2C room model includes this dynamic, the optimizer avoids this measure at times when this trade-off is not economically beneficial.

Comparing the setpoint of T^{air} and the realized room temperature, we observe some deviation. While the real system follows

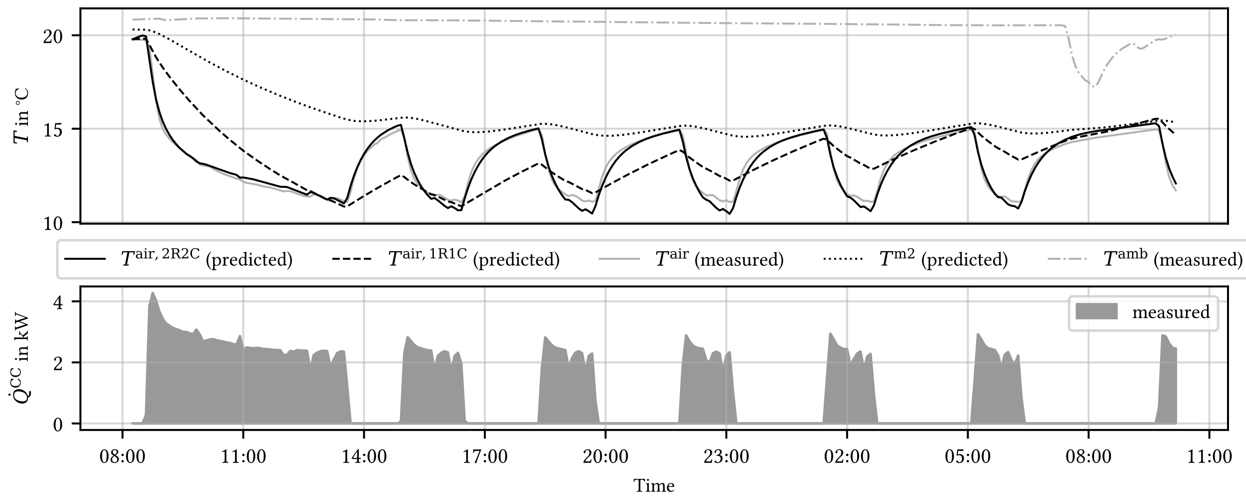


Fig. 8. a) Time series plot of measured and predicted room temperature T^{air} for 1R1C room model and 2R2C model, predicted temperature T^{m2} of the 2R2C model and cooling load $\dot{Q}^{\text{CC}}$ for the training dataset.

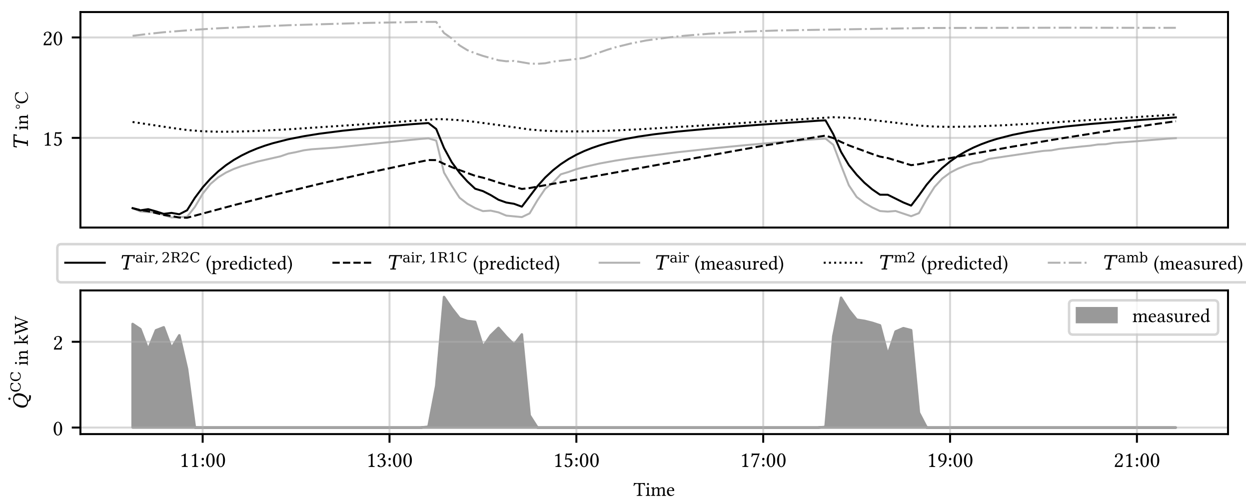


Fig. 9. a) Time series plot of measured and predicted room temperature T^{air} for 1R1C room model and 2R2C model, predicted temperature T^{m2} of the 2R2C model and cooling load $\dot{Q}^{\text{CC}}$ for the validation dataset.

its setpoint temperature very well in higher temperature areas, at lower temperatures the real system does not cool down as quickly as the found optimization solution predicts. This can be explained by the limited heat transfer over the AHU and the fluid circuit between the chiller and room, which is not directly represented by the optimization model.

The results of the optimized control strategy are presented in Table 5. In this experiment, an overall reduction of the energy costs of 10.06 % was achieved by controlling the room temperature. This is achieved by exploiting the volatile electricity prices but also partially due to a more efficient utilization of the compression chiller.

We calculate the associated CO_2 emissions using data from the Electricity Maps platform [11]. By implementing the price-driven control strategy, the electricity-related CO_2 emissions were reduced

Table 5. Results of the optimized energy-flexible MPC control compared to the conventional PI control strategy.

Metric	**PI Control**	**MPC**
Electrical energy	98.32 kWh	94.85 kWh
Thermal energy	65.35 kWh	78.37 kWh
Average EER of the CC	0.665	0.826
Electricity costs	8.69 €	7.81 €
Savings in cost of energy	-	10.06%
CO_2 emissions	34.08 kgCO_2eq	32.4 kgCO_2eq
Savings in electricity related CO_2 emissions	-	4.93%
Experiment duration	24 h (extrapolated)	144 h

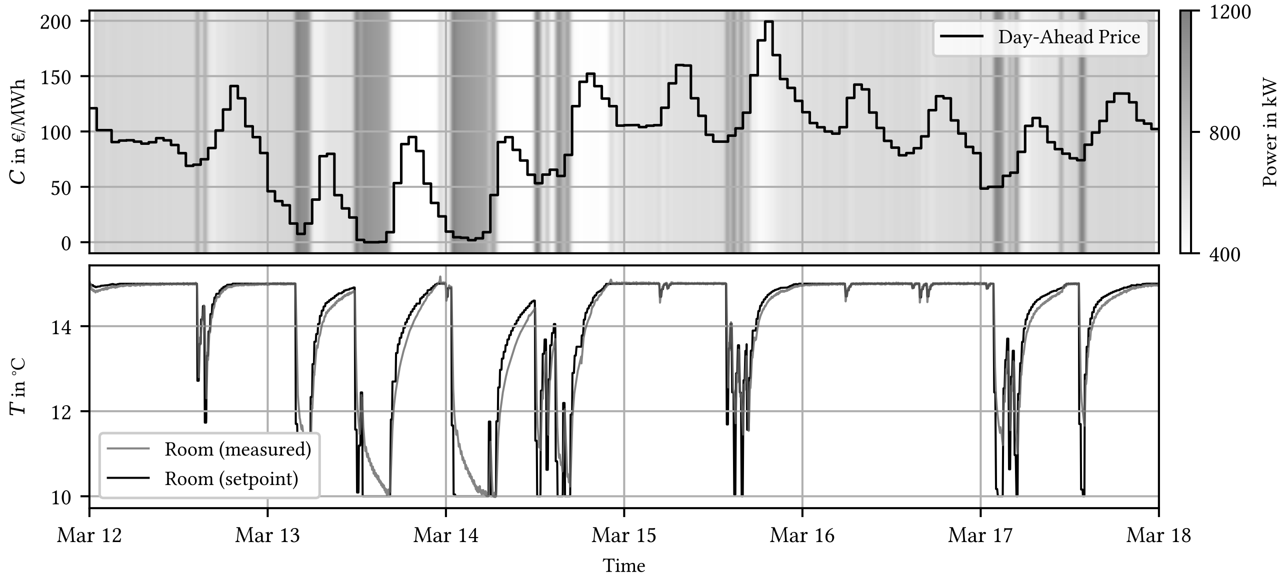


Fig. 10. Time series plot of the optimized and implemented setpoint and measured room temperature T^{air} as well as Day-Ahead electricity prices C and measured electrical power consumption $\dot{P}^{\text{el, CC}}$ of the compression chiller.

by 4.93 %, due to the high correlation between electricity price and greenhouse gas emissions mentioned in Chapter 1.

6 Conclusion and Outlook

This paper presents a method for optimizing the energy-flexible operation of thermal supply systems. The method includes the data-driven identification on a single component basis, structured assembly of identified models into an overall optimization problem for the system, and deployment of the optimization algorithm to the real system.

The method was implemented for a demonstrator in the ETA Research Factory at TU Darmstadt, including a climatized room and a dedicated cooling supply system. For this system, a simple statistical linear model for the compression chiller and a dynamic 2R2C model for the climate room were selected and parameterized on measurement data. From the identified component models, we derived an optimization model and deployed it to the real system in a 144-hour experiment. By using the climate room as an inherent thermal energy storage, with electricity prices representative of 2023, the overall cost of energy was reduced by 10.06 % in comparison to a reference scenario with a constant room temperature. As a consequence of this electricity price-driven optimization, the energy-related carbon emissions of the system were also reduced by 4.93 %.

Thermal supply systems are employed across multiple industries and sectors and mostly exhibit similar technologies for energy converters and storages. A component identification based on measurement data ensures low-effort transferability to further and more complex systems, once a broad model component library for optimization is in place.

Besides increasing the variety of considered systems, the method should also be tested on far more complex thermal supply systems, with multiple storage and converter components as well as multiple utilized energy sources.

Since weather is a dominant factor in the characteristics of thermal supply systems, and weather forecasts are becoming increasingly available, weather conditions can be included in the optimization problem as well if the considered components are influenced by it (e.g., efficiency of an air-cooled heat pump).

In the presented example, we used already existing or manually built mathematical models for the single components that were then parameterized on measurement data. In future research, model identification purely based on measurement data for the specific case of energetic optimization should be explored to find simple yet accurate models and close the gap between purely physics-based and statistical models.

Acknowledgments

The authors gratefully acknowledge the financial support of the Kopernikus-Project "SynErgie" (Grant Number 03SFK3A0-3) by the Federal Ministry of Education and Research of Germany (BMBF) and thank the Projektträger Jülich (PtJ) for the project supervision.

References

[1] Sayyed Faridoddin Afzali and Vladimir Mahalec. 2017. Optimal design, operation and analytical criteria for determining optimal operating modes of a CCHP with fired HRSG, boiler, electric chiller and absorption chiller. *Energy* 139 (2017), 1052–1065. https://doi.org/10.1016/j.energy.2017.08.029

[2] Joel A. E. Andersson, Joris Gillis, Greg Horn, James B. Rawlings, and Moritz Diehl. 2019. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* 11, 1 (2019), 1–36. https://doi.org/10.1007/s12532-018-0139-4

[3] Andreas Bachmann, Lukas Bank, Carlo Bark, Dennis Bauer, Bruno Blöchl, Martin Brugger, Hans Ulrich Buhl, Benjamin Dietz, Julia Donnelly, Thomas Friedl, Stephanie Halbrügge, Heribert Hauck, Joachim Heil, Aljoscha Hieronymus, Torben Hinck, Svetlina Ilieva-König, Carl Johnzén, Carsten Koch, Jana Köberlein,

Ekrem Köse, Stefan Lochner, Martin Lindner, Tim Mayer, Alexander Mitsos, Stefan Roth, Alexander Sauer, Claudia Scheil, Johannes Schilp, Jens Schimmelpfennig, Julia Schulz, Jan Schulze, Johannes Sossenheimer, Nina Strobel, Alejandro Tristan, Susanne Vernim, Jonathan Wagner, Felix Wagon, Martin Weibelzahl, Matthias Weigold, Jan Weissflog, Simon Wenninger, Moritz Wöhl, Jan Zacharias, and Michael F. Zäh. 2021. Energieflexibel in die Zukunft – Wie Fabriken zum Gelingen der Energiewende beitragen können: VDI-Handlungsempfehlungen, Oktober 2021. https://doi.org/10.24406/FIT-N-638765

[4] Logan Beal, Daniel Hill, R. Martin, and John Hedengren. 2018. GEKKO Optimization Suite. *Processes* 6, 8 (2018), 106. https://doi.org/10.3390/pr6080106

[5] Fabian Borst, Michael Georg Frank, Lukas Theisinger, and Matthias Weigold. 2023. ThermalSystemsControlLibrary: A Modelica Library for Developing Control Strategies of Industrial Energy Systems. *Modelica Conferences* (2023), 209–215. https://doi.org/10.3384/ecp204209

[6] Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. 2021. *Pyomo – Optimization Modeling in Python* (3rd ed. 2021 ed.). Springer Optimization and Its Applications, Vol. 67. Springer International Publishing and Springer, Cham. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6531628

[7] Yongbao Chen, Zhe Chen, Peng Xu, Weilin Li, Huajing Sha, Zhiwei Yang, Guowen Li, and Chonghe Hu. 2019. Quantification of electricity flexibility in demand response: Office building case study. *Energy* 188 (2019), 116054. https://doi.org/10.1016/j.energy.2019.116054

[8] Lucas W. Davis and Paul J. Gertler. 2015. Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences of the United States of America* 112, 19 (2015), 5962–5967. https://doi.org/10.1073/pnas.1423558112

[9] DIN Deutsches Institut für Normung e. V. July 2019. Air conditioners, liquid chilling packages and heat pumps for space heating and cooling and process chillers, with electrically driven compressors: Part 1: Terms and definitions.

[10] Inga Döbel, M. Leis, Manuel Molina Vogelsang, D. Neustroev, Henning Petzka, A. Riemer, Stefan Rüping, Angelika Voß, M. Wegele, and Juliane Welz. 2018. *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung.* Fraunhofer-Gesellschaft, München. https://publica.fraunhofer.de/entities/publication/47f259bb-c191-4b17-84f4-2b3824cc3a0a/details

[11] Electricity Maps ApS. [n. d.]. Climate influence of the regions. https://app.electricitymaps.com/map

[12] ENTSO-E. 2024. Transparency Platform RESTful API. https://transparency.entsoe.eu/content/static_content/Static%20content/web%20api/Guide.html

[13] Eurelectric. 2023. Power System of the Future: Keys to delivering capacity on the distribution grid. https://www.eurelectric.org/publications/power-system-of-the-future-keys-to-delivering-capacity-on-the-distribution-grid

[14] Robert Förster, Sebastian Harding, and Hans Ulrich Buhl. 2024. Unleashing the economic and ecological potential of energy flexibility: Attractiveness of real-time electricity tariffs in energy crises. *Energy Policy* 185 (2024), 113975. https://doi.org/10.1016/j.enpol.2023.113975

[15] Daniel Fuhrländer-Völker, Benedikt Grosch, and Matthias Weigold. 2023. Modelling and Control of Aqueous Parts Cleaning Machines for Demand Response. In *th Conference on Production Systems and Logistics CPSL 2023*. 790–800. https://doi.org/10.15488/13498

[16] Benedikt Grosch, Heiko Ranzau, Bastian Dietrich, Thomas Kohne, Daniel Fuhrländer-Völker, Johannes Sossenheimer, Martin Lindner, and Matthias Weigold. 2022. A framework for researching energy optimization of factory operations. *Energy Informatics* 5, S1 (2022). https://doi.org/10.1186/s42162-022-00207-6

[17] Karl Hinderer, Ulrich Rieder, and Michael Stieglitz. 2017. *Dynamic Optimization: Deterministic and Stochastic Models.* Springer International Publishing AG, Cham. https://doi.org/10.1007/978-3-319-48814-1

[18] Jing-Wen Huang and Jia-Wen Gao. 2020. How could data integrate with control? A review on data-based control strategy. *International Journal of Dynamics and Control* 8, 4 (2020), 1189–1199. https://doi.org/10.1007/s40435-020-00688-x

[19] Werner Kast, Herbert Klan, and (Revised by André Thess). 2010. *F2 Heat Transfer by Free Convection: External Flows* (2. ed. ed.). Springer, Berlin and Heidelberg. https://doi.org/10.1007/978-3-540-77877-6{_}120

[20] Durmuş Kaya, Fatma Çanka Kılıç, and Hasan Hüseyin Öztürk. 2021. *Energy Management and Energy Efficiency in Industry.* Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-25995-2

[21] Kevin J. Kircher and K. Max Zhang. 2015. Model predictive control of thermal storage for demand response. In *American Control Conference (ACC), 2015*. IEEE, Piscataway, NJ, 956–961. https://doi.org/10.1109/ACC.2015.7170857

[22] Klaus Langeheinecke, André Kaufmann, Kay Langeheinecke, and Gerd Thieleke. 2020. *Thermodynamik für Ingenieure.* Springer Fachmedien Wiesbaden, Wiesbaden. https://doi.org/10.1007/978-3-658-30644-1

[23] Leo Liberti and Carlile Lavor. 2017. *Euclidean distance geometry: An introduction.* Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-60792-4

[24] Lennart Ljung. 2009. *System identification: Theory for the user* (2. aufl. 11. print ed.). Prentice-Hall, Upper Saddle River, NJ.

[25] Yudong Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves. 2012. Model Predictive Control for the Operation of Building Cooling Systems. *IEEE Transactions on Control Systems Technology* 20, 3 (2012), 796–803. https://doi.org/10.1109/TCST.2011.2124461

[26] U.S. Department of Energy. 2006. *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving them: A Report to the United States Congress* (rev. 2008 ed.). U.S. Department of Energy, Washington, D.C.

[27] Gbemi Oluleye, Megan Jobson, Robin Smith, and Simon J. Perry. 2016. Evaluating the potential of process sites for waste heat recovery. *Applied Energy* (2016), 627–646. https://doi.org/10.1016/j.apenergy.2015.07.011

[28] Peter Radecki and Brandon Hencey. 2017. Online Model Estimation for Predictive Thermal Control of Buildings. *IEEE Transactions on Control Systems Technology* 25, 4 (2017), 1414–1422. https://doi.org/10.1109/TCST.2016.2587737

[29] Matthias Rehfeldt, Tobias Fleiter, and Felipe Toro. 2018. A bottom-up estimation of the heating and cooling demand in European industry. *Energy Efficiency* 11, 5 (2018), 1057–1082. https://doi.org/10.1007/s12053-017-9571-y

[30] REN21. 2022. Renewables 2022: Global status report. https://www.ren21.net/wp-content/uploads/2019/05/GSR2022_Full_Report.pdf

[31] Alexander Sauer, Eberhard Abele, and Hans Ulrich Buhl (Eds.). 2019. *Energieflexibilität in der deutschen Industrie: Ergebnisse aus dem Kopernikus-Projekt - Synchronisierte und energieadaptive Produktionstechnik zur flexiblen Ausrichtung von Industrieprozessen auf eine fluktuierende Energieversorgung (SynErgie).* Fraunhofer Verlag, Stuttgart. https://opus.hs-offenburg.de/frontdoor/index/index/docId/4061

[32] Alexander Sauer, Hans Ulrich Buhl, Alexander Mitsos, and Matthias Weigold (Eds.). 2022. *Energieflexibilität in der deutschen Industrie: Band 2: Markt- und Stromsystem, Managementsysteme und Technologien energieflexibler Fabriken.* Fraunhofer Verlag. https://doi.org/10.24406/publica-258

[33] Michael Schreiber, Martin E. Wainstein, Patrick Hochloff, and Roger Dargaville. 2015. Flexible electricity tariffs: Power and energy price signals designed for a smarter grid. *Energy* 93 (2015), 2568–2581. https://doi.org/10.1016/j.energy.2015.10.067

[34] Nina Strobel. 2021. *Einsatz inhärenter Energiespeicher in Produktionssystemen zum elektrischen Lastmanagement.* Darmstadt. https://doi.org/10.26083/tuprints-00017581

[35] Nina Strobel, Daniel Fuhrländer-Völker, Matthias Weigold, and Eberhard Abele. 2020. Quantifying the Demand Response Potential of Inherent Energy Storages in Production Systems. *Energies* 13, 16 (2020), 4161. https://doi.org/10.3390/en13164161

[36] The European Parliament and the Council of the European Union. 2023. Directive (EU) 2023/2413 of the European Parliament and of the Council of 18 October 2023 amending Directive (EU) 2018/2001, Regulation (EU) 2018/1999 and Directive 98/70/EC as regards the promotion of energy from renewable sources, and repealing Council Directive (EU) 2015/652. http://data.europa.eu/eli/dir/2023/2413/oj

[37] Verein Deutscher Ingenieure e. V. July 2020. Energy-flexible factory: Fundamentals.

[38] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* 17, 3 (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[39] Adrian von Hayn, Jonas Wendt, Nicolas Lieberth, Nicole Weisel, and Matthias Weigold. 2023. Development of Energy Flexible and Sustainable Operation Strategies of Air Conditioning Systems for Industrial Production Environments. *Procedia CIRP* 116 (2023), 155–160. https://doi.org/10.1016/j.procir.2023.02.027

[40] Thomas Weber. 2023. *Betriebsoptimierung industrieller Kälte- und Wärmeversorgungssysteme über mathematische Programmierung.* Dissertation. Technische Universität Darmstadt and Shaker Verlag, Düren.

[41] Jonas Wendt, Jan Zangenberg, and Matthias Weigold. 2023. Optimized Energy-Flexible Operation Strategies of Air Conditioning Systems in Production Environments. *Procedia CIRP* 120 (2023), 255–260. https://doi.org/10.1016/j.procir.2023.08.046

[42] Michael Wetter, Wangda Zuo, Thierry S. Nouidui, and Xiufeng Pang. 2014. Modelica Buildings library. *Journal of Building Performance Simulation* 7, 4 (2014), 253–270. https://doi.org/10.1080/19401493.2013.765506

A Platform Ecosystem Providing New Data For The Energy Transition

MARKUS DUCHON, JESSY MATAR, MAHSA FARAJI SHOYARI, ALEXANDER PERZYLO, and INGMAR KESSLER, fortiss GmbH, Germany
PATRICK BUCHENBERG, PHILIPP KUHN, and THOMAS HAMACHER, Technische Universität München, Germany
THORSTEN SCHLACHTER and WOLFGANG SÜSS, Karlsruhe Institute of Technology (KIT), Germany
NGUYEN XUAN THINH, HANIYEH EBRAHIMI SALARI, and JASMIN LATKO, Technische Universität Dortmund, Germany
MINSHENG XU, MAXIM SHAMOVICH, DOMINIK SCHLÜTTER, and JÉRÔME FRISCH, RWTH Aachen University, Germany
KUSHAGAR RUSTAGI and MARKUS KRAFT, Computational Modelling Pirmasens GmbH, Germany
CAROLIN AYASSE and FLORIAN STEINKE, Technical University of Darmstadt, Germany
MICHAEL METZGER and LAURA KUPER, Siemens AG, Technology, Germany

There is a great need for high-quality and comprehensive data in the energy sector. This data is collected and preprocessed at considerable expense and is not only required for research, but also by planning offices and other industries in connection with planning activities, such as the creation of municipal heat planning. The NEED ecosystem will accelerate these processes establishing an efficient, robust, and scalable energy data ecosystem. Heterogeneous energy-related data sources will be brought together and automatically linked consistently across different sectors as well as temporal and spatial levels. In this context, existing data sources will not be replaced but rather integrated into the NEED ecosystem as dedicated sources including a semantic description on how to utilize them. In addition to conventional data sources from the various planning levels, we envision a quality assessment scheme based on the FAIR criteria. In reality, we are often faced with missing data, too. To close this gap we explore data-driven, model-driven, AI-based, and tool-driven generation of synthetic data. These heterogeneous data sources will be interlinked using ontology modules which will be represented in a knowledge graph. Via a semantic API, queries will be generated to identify the required data sources, which will be orchestrated to provide the data needed. This will enable researchers, planners, and others including their tools to interact with the NEED ecosystem, while a tool proxy will be able to translate the resulting data into proprietary formats, required by some tools to operate. The NEED ecosystem is planned to be a robust, easy-to-maintain, and flexible infrastructure to enhance planning energy measures at different spatial levels and with different time horizons. We envision to evaluate our NEED approach for the transparent provision of data by integrating relevant data sources as microservices, definition and analysis of application scenarios in the planning domain, as well as the integration of various tools for different planning purposes. With these elements, we will be able to quantify the efficiency of data procurement and demonstrate the functionality of the approach using practical use cases.

Authors' addresses: Markus Duchon, duchon@fortiss.org; Jessy Matar, matar@fortiss.org; Mahsa Faraji Shoyari, farajishoyari@fortiss.org; Alexander Perzylo, perzylo@fortiss.org; Ingmar Kessler, ikessler@fortiss.org, fortiss GmbH, Guerickestr. 25, München, Germany, 80805; Patrick Buchenberg, patrick.buchenberg@tum.de; Philipp Kuhn, pkuhn@tum.de; Thomas Hamacher, thomas.hamacher@tum.de, Technische Universität München, Lichtenbergstraße 4a, Garching b. München, Germany, 85748; Thorsten Schlachter, thorsten.schlachter@kit.edu; Wolfgang Süß, wolfgang.suess@kit.edu, Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, Karlsruhe, Germany, 76131; Nguyen Xuan Thinh, nguyen.thinh@tu-dortmund.de; Haniyeh Ebrahimi Salari, haniyeh.ebrahimi@tu-dortmund.de; Jasmin Latko, jasmin.latko@tu-dortmund.de, Technische Universität Dortmund, August-Schmidt-Straße 10, Dortmund, Germany, 44227; Minsheng Xu, xu@e3d.rwth-aachen.de; Maxim Shamovich, shamovich@e3d.rwth-aachen.de; Dominik Schlütter, schluetter@e3d.rwth-aachen.de; Jérôme Frisch, frisch@e3d.rwth-aachen.de, RWTH Aachen University, Mathieustraße 30, Aachen, Germany, 52074; Kushagar Rustagi, krustagi@cmpg.io; Markus Kraft, mkraft@cmpg.io, Computational Modelling Pirmasens GmbH, Delaware Avenue 1-3, Pirmasens, Germany, 66953; Carolin Ayasse, carolin.ayasse@eins.tu-darmstadt.de; Florian Steinke, florian.steinke@eins.tu-darmstadt.de, Technical University of Darmstadt, Landgraf-Georg-Str. 4, Darmstadt, Germany, 64283; Michael Metzger, michael.metzger@siemens.com; Laura Kuper, laura.kuper@siemens.com, Siemens AG, Technology, Otto-Hahn-Ring 6, München, Germany, 81739.

1 INTRODUCTION

The procurement and provision of data is still a very time-consuming and cost-intensive part of planning energy technology systems and is estimated by the project partners to account for 30 % of the total costs. Depending on requirements and use cases, different data—such as LoD2 data[1], information on existing infrastructures, geodata and cadastral data, consumption curves, primary energy requirements, plant data, information on mobility or geothermal potential—have to be procured and preprocessed accordingly for dedicated applications or tools. This process of obtaining and preparing data and information of a sufficient quality and up-to-dateness for the application must be repeated for each and every planning case. Additionally, in order to advance the energy transition, cross-connections must also be taken into account. Consequently, both, the heat and electricity side must be considered in all planning perspectives, e.g., from the perspective of a building or with regard to the infrastructure. This also applies to other energy vectors such as gas, mobility, or water. In many areas, however, hardly any data of sufficient quality is available in machine-readable form and the origin and transparency of data generation and quality are not reliable. With this in mind, the publicly funded NEED research project[2] was launched in September 2023 to develop an energy data ecosystem for future energy planning with numerous research institutions and industrial partners.

With the help of the NEED ecosystem, the planning basis will be made digitally available in the form of data from different levels and domains and linked with each other by the application of ontologies

[1] 3D building models for Germany from Federal Agency for Cartography and Geodesy
[2] https://enargus.de/pub/bscw.cgi/?op=enargus.eps2&q=%2201256602/1%22&v=10&s=13, accessed 10.05.2024

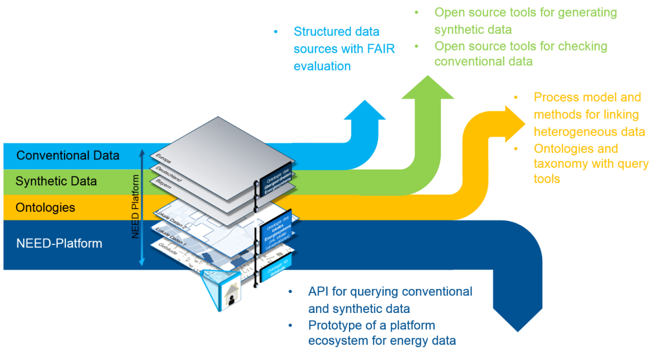


Fig. 1. Multilayered NEED ecosystem: The NEED-platform aims to provide a communication interface for energy data. To provide the interface, underlying ontologies link heterogeneous data sources. The ecosystem will increase accessibility and coherence of data of different aggregation levels, using conventional data sources complemented by synthetic data.

resulting in a knowledge graph. This will create a modular system for end-to-end planning tasks from the building to the infrastructure, which enables automated, model-based analyses across system boundaries. In addition to an improvement in quality through transparent, digital, and verified data, the costs and time required for data acquisition and preparation may be significantly reduced.

In this context, the aim is not to replace existing data sources, but rather to integrate them into the ecosystem as dedicated sources. In addition to conventional data sources (such as energy atlases, state offices, grid operators, building data, geothermal energy deposits, weather data, census data) at the various planning levels (e.g., buildings, districts, regions), we aim to explore ways of closing existing data gaps with synthetic data. By integrating conventional data and deriving synthetic data, the NEED ecosystem may provide a robust, easy-to-maintain, and flexible tool for deriving energy measures at different spatial levels without losing sight of the overall picture.

Finally, the partners' tools and models should access the required data via semantic queries and suitable interfaces in order to fulfill the respective (planning) tasks. This illustrates the NEED approach to the transparent provision of current data, particularly in the examples of heat management planning and the creation of a dynamic energy usage plan. The following subgoals are pursued:

- Subgoal 1: Supplement and integrate existing data sources and data formats, such as the energy atlas of Bavaria[3] or the energy atlas of Thuringia[4], LoD2 data of Bavaria[5], etc., to integrate different conventional data sources.
- Subgoal 2: Provide and validate synthetic data created through derivation and aggregation.
- Subgoal 3: Bringing together different sources and levels on the topic of energy in a knowledge graph to realize semantic queries and, based on this, to set up monitoring, if possible in real time, and to run knowledge-augmented 'what-if scenarios'.
- Subgoal 4: Validation and verification by means of practical applications for energy planning based on conventional and synthetic data using real examples in order to demonstrate the benefits and applicability with the aim of stabilization.

The paper is structured as follows. Section 2 will introduce the stakeholders addressed in our work including the main application scenarios. Section 3 deals with conventional data sources and the corresponding investigations. In Section 4, synthetic data for closing existing gaps in the conventional data or to validate the developed data-driven approaches is introduced. To interconnect the available and existing data sources with actual planning tools, in Section 5, we present the concept of ontology models resulting in a knowledge graph to query relevant information. To combine the so far presented building blocks we propose a distributed service-based architecture in Section 6 before we summarize the work conducted and give a brief outlook.

2 USE CASES AND CASE STUDIES

In order to develop the NEED ecosystem addressing actual challenges and problems of practice, use cases and case studies are defined. The use cases describe what methods, tools, and data are used, as well as the respective customers addressed. One or two specific questions for each use case are elaborated, which are then used as guidance. Based on the questions, the necessary analysis methods, tools, and data are identified.

The description of used methods and tools provides information about the interfaces to the NEED ecosystem. That also includes the data requirements for providing information (e.g., population development, spatial development, time series for forecasts, geographical allocation, defined units). As a result, the overall requirements regarding data and interfaces to the NEED ecosystem can be specified. Within the research project, the partners agreed on focusing on the following customers:

- Homeowners, housing associations, energy consultants
- SMEs (small and medium-sized enterprises)
- Network operators
- Municipalities, communities

The first two types of users reflect the point of view from an individual perspective either at residential or industrial level. The third and fourth types of users represent the supply side and are responsible for operating energy networks and ensuring energy supply.

The scientific partners and especially the industry partners have a variety of tools available for this purpose. With regard to the different tools, the resulting use cases are summarized in application scenarios, which represent a thematic cluster. The following four application scenarios were developed by the project partners to take energy planning requirements into account:

- Energy utilization and municipal heat planning
- Network planning (electricity and heat)
- District planning
- Energy consulting (renovation/transformation/electrification)

[3] https://energieatlas.bayern.de/, accessed on 10.05.2024
[4] https://karte.energieatlas-thueringen.de/, accessed on 10.05.2024
[5] https://geodaten.bayern.de/opengeodata/OpenDataDetail.html?pn=lod2, accessed on 10.05.2024

The creation of these four application scenarios covers both the expertise of the consortium partners and the demand of the above-mentioned customers. In addition, we have ensured that a set of suitable case studies exist for each of the application scenarios to validate the benefits of the proposed NEED ecosystem over the course of the project.

2.1 Case Studies

A case study is defined as a real project carried out by the consortium partners in collaboration with local stakeholders. The case studies are intended to demonstrate and validate the benefits of the NEED ecosystem at a later stage of the project. Each case study is assigned to a previously defined use case with an associated customer. There is at least one case study for each use case.

Various industrial partners in the NEED consortium are already working on projects that can be assigned to a specific application scenario. Some of these projects will be converted into case studies and then reworked with the help of the NEED ecosystem. The aim is to show how the planning process for the various use cases changes through the use of the NEED ecosystem. The aim is to significantly reduce the time and effort required to record and preprocess the input data, as the relevant information may be efficiently retrieved in the desired spatial and temporal resolution and in the desired format via suitable NEED interfaces and proxies.

In addition to relying on projects already being worked on by industry partners, new methods for processing the planning tasks of the various use cases are also being investigated. This also includes investigations into new efficiency potentials made possible by the NEED ecosystem. The possibility of querying data in different temporal and spatial resolutions should facilitate the inclusion of synergy effects of neighboring units in a variety of planning tasks. In addition, research is being carried out into the automated provision of data, e.g., for energy consulting for residential buildings or for municipal heat planning.

A selection of case studies is presented below:

- Municipal heat planning for a city in southern Hesse
- Detailed heat grid planning in a city in southern Hesse and in a city in Bavaria
- Generation of synthetic electricity distribution networks in a city in northern Bavaria
- Automated data provision for energy efficiency consulting of (residential) buildings

Within these case studies, we aim to increase the level of automation in the planning processes. For example, for heat grid planning, we intend to develop a method to handle the (computational) complexity of topology design to make automated heat grid planning applicable in real planning processes. However, the executions of the case studies and the integration of automated processes rely on a wide range of different data sources at the various planning levels. Some data are publicly available, others are provided by government institutions, others are not available at all. Nevertheless, it is no trivial task to collect and preprocess the right data in the right quality and up-to-dateness. In this context, the project aims to analyze and process existing, so-called conventional data sources to integrate them into the NEED ecosystem.

3 CONVENTIONAL DATA SOURCES

In the context of the NEED project, conventional data refers to data that is already accessible and typically published by the authors of the data in different data sources, e.g., web-based map services such as the *Energieatlanten* (energy atlases) or online registries such as the *Marktstammdatenregister* (master data on the electricity and gas market)[6]. Usually, these data sources serve specific purposes, resulting in limited connections between sources. Even though in some instances the data of different sources overlap (for example some of the energy atlases include data about global sun radiation which is a dataset released by the *Deutscher Wetterdienst*, Germany's institution for meteorological services[7]), sources like energy atlases only include data for a specific time frame or a specific spatial extent instead of holistic datasets.

Given the broad spectrum of data sources with ready-to-use data, it is necessary to gain an overview of the data in order to integrate them into the NEED ecosystem. We are therefore working on finding a uniform characterization for all data in order to assess data quality and characterize further criteria in the variety of data sources and authors of the data. Testing our characterization approach requires the use of a range of data sources with comparable datasets, which led to the decision to use the Energy Atlases.

3.1 Evaluation of Energy Atlases

An energy atlas typically refers to a web-based map service hosted by a federal state of Germany. Generally, the hosted data relates to the field of energy planning in the broad sense and ranges from simple geometric data used to describe the visualized area to the global sun radiation over the course of one year. As of now, not every state has an energy atlas and those that do, all have their own instead of one energy atlas for the whole country (see Figure 2). The atlases differ regarding the hosted content and data quality. Since all these atlases were probably created with similar intentions, working with them would be much more efficient if they all followed the same standards. In order to realize a common standard, it is necessary to gain an understanding of what data is contained in the individual energy atlases and in what quality. To this end, the energy atlases are analysed and two matrices are introduced to characterize and evaluate the individual data sets and their metadata. The first step is to develop the columns of the two matrices. The data to be analyzed in the matrices consists mainly of geodata. While most of the data comes from the field of energy planning, there is a focus on the inclusion of basic geodata, as this represents an essential basis in any planning process. The matrices must therefore be designed in such a way that they allow data from different areas to be analyzed. This led to the columns being strongly oriented towards the way in which data is analyzed in the geoinformation sciences. To ensure the reusability of the data and to allow users to easily assess the quality of the data, a strong focus is placed on the spatial and temporal aspects of the datasets and their metadata.

[6]https://www.marktstammdatenregister.de/MaStR, accessed on 10.05.2024
[7]https://www.dwd.de/DE/leistungen/solarenergie/solarenergie.html;jsessionid=2A01EAA2CFC9627F02AC72EF3D5DDDD9.live21074?nn=16102, accessed on 29.04.2024

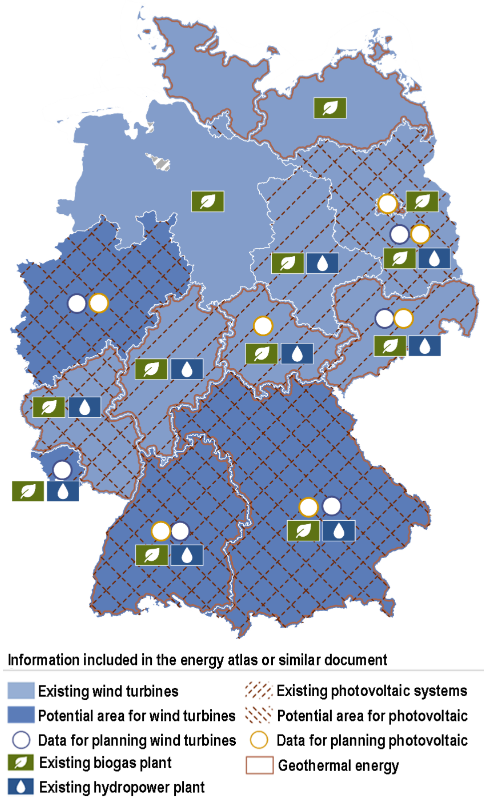


Fig. 2. Information included in the energy atlases of Germany's federal states.

The first matrix focuses on the metadata to gain a basic understanding of the corresponding energy atlas and its quality. It includes the following aspects:

- The name of the dataset as it is shown in the energy atlas,
- Contact information of the dataset's author,
- Information about when the metadata was updated last,
- The data's usage license,
- The hyperlink to the metadata, and
- The option to evaluate the data regarding the FAIR principles with certain indicators for each principle.

The above-mentioned 'FAIR principles' evaluate the findability, accessibility, interoperability and reusability of data and serve as a benchmark for data quality and applicability [Wilkinson et al. 2016]. For the NEED ecosystem, we pre-select FAIR principles that are relevant in the context of energy planning processes and evaluate them for the datasets under consideration according to [Bahim et al. 2020]. We plan to make the evaluation of these principles together with the datasets available to the users of the platform in order to enable an individual evaluation of the relevance of the individual principles in specific planning processes.
The second matrix focuses on the specific datasets and their contents, and the following aspects are included in the matrix:

- The name of the dataset as it is shown in the energy atlas,
- The topic the dataset that it can be associated with, such as electricity, mobility, or heat energy,
- Information about the dataset's source, e.g., if it is commercial data or open data,
- The classification in spatial data with or without references to a specific subject or research field,
- The geometric dimension and whether the dataset includes temporal information,
- The specific content of the data such as buildings, different kinds of infrastructure, etc.,
- The data's spatial extent, ranging from single georeferenced locations to a whole country,
- The data's resolution or scale,
- Information about how often the data is updated,
- Information about when the dataset was initially released,
- Information about when the data was updated last,
- Information about the data format,
- Information whether the data can be downloaded,
- Contact information of the dataset's author,
- A hyperlink to the dataset in the energy atlas, and
- The option to evaluate the data regarding the FAIR principles with certain indicators for each principle.

Each of these aspects serves as one of the columns of the matrix and all information must be added to the matrix for each relevant dataset. Once all information on the datasets of all energy atlases has been collected in the matrix, it is possible to compare the different atlases. The comparison can be used to make a proposal to the institutions responsible for the atlases in order to identify possible weaknesses and deficiencies and to establish a nationwide standard. In addition, the matrix can also be used to characterize data from other data sources. Since all data is characterized in the same way, it can be more easily implemented into the NEED ecosystem for future use without the need to implement different parsers for each data source. The data itself is not necessarily included in the same way as it is usually retrieved, e.g. on an energy atlas website. Instead, the data can be mapped to specific areas based on the geographical information about the datasets. When information about these areas is retrieved, the datasets and their respective entries from the matrices can be displayed and accessed via hyperlinks or other references to where they were originally published. In this way, additional datasets can be added to the ecosystem, even if not all data can be downloaded or accessed in the form of a Web Map Service. In the following, we describe the *Marktstammdatenregister (MaStR)* as another example of conventional data.

3.2 Open-MaStR

Germany's publicly available *Marktstammdatenregister* is a register that keeps track of energy units (including power and gas). It is provided by the German Federal Network Agency *(Bundesnetzagentur / BNetzA)* and is updated on a daily basis. The *MaStR* open dataset can be browsed online on the website of the *BNetzA*. To facilitate access to the database, a Python package called *open-mastr* has been developed by the *RLI* and *fortiss* to provide an interface to improve the usability of the register. The package includes methods to clean and write the data into a local database[8].

The package provides a Python interface for accessing data via the bulk download and the web service API, and methods to clean the data. Through the bulk method, one can download all units (e.g., all solar farms in Germany) and through the API method, one can retrieve specific information regarding single units (requires registration for an API token). The cleaned data is then written into a database. The *open-mastr* package hence provides easy access to the dataset, which is especially useful in the community of energy system researchers.

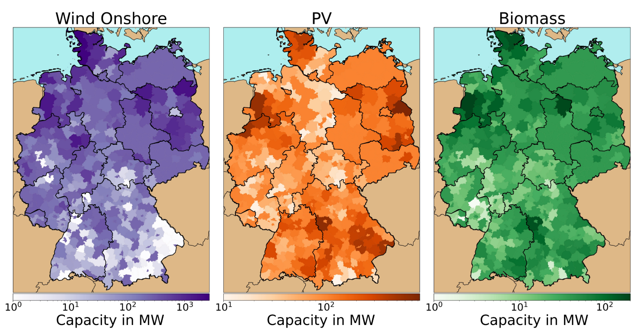


Fig. 3. Installed capacity in Germany per district in 2024 extracted from the *Marktstammdatenregister* using *open-mastr*.

Easy and direct analysis of various energy sources installed nationwide, in every state, district, and municipality can be tracked through this package. Figure 3 illustrates the capacity provided from wind, PV, and biomass in 2024 in Germany per district. The data found in the *Marktstammdatenregister* are manually entered, resulting in a large potential of errors and data gaps. The latter issue will be tackled in the next section, where we introduce synthetic data generation and its importance in filling the gaps of conventional data sources.

4 SYNTHETIC DATA

Despite the considerable amount of conventional data presented in Section 3, there are also significant data gaps. In some cases, needed datasets are not available at all, in other cases the necessary temporal or spatial resolution is lacking, or existing datasets cannot be used or have to be modified for data protection reasons, for example.

In addition, datasets from different sources often have different technical and semantic representations, which is why it is necessary to map, link, and harmonize datasets. It therefore makes sense not only to think about generating *data*, but also to create, provide, and use *metadata* for all (both conventional and synthetic) datasets to ensure data can be (automatically) converted, mapped, linked, etc.

Metadata for both synthetic datasets and processed conventional datasets should contain traceability of their origin and all processing steps applied (so-called *provenance data*). This is essential for the reproducibility and transparency of the models, simulations, and plans created along the data.

4.1 Strategy for the Creation and Use of Synthetic Data

The lack of reliable data and the need for preserving the privacy of the data are major concerns, e.g., when working with low-voltage networks or residential data, particularly when such data is unavailable to the general public or cannot be measured, recorded, or documented. However, reliable outputs can be achieved and gaps can be filled with synthetic data created using the characteristics of conventional data.

Our primary goal is to create an ecosystem that provides and integrates conventional and synthetic data necessary for energy planning. In this section, we focus on exploring the current ideas related to synthetic data generation, with a particular focus on exploring various methodologies. We introduce methods such as image processing, data fusion, and AI, ensuring validation and verification of the generated data [Dankar and Ibrahim 2021].

The project partners will focus on three strategies in particular:

- Disaggregation of data that is available at a lower temporal or spatial resolution
- Reconstruction of data using established planning principles
- The utilization of unconventional data sources, such as aerial and satellite photos

First and foremost, we examine methodologies used in diverse fields for synthetic data creation. We explore methods from geoinformatics to be implemented for energy planning purposes. We identify gaps and challenges in current conventional data sources in the energy sector, and use synthetic data generation approaches to fill these gaps.

Our proposed methodologies for generating synthetic data encompass a range of approaches. Starting from the use of image data to extract information and characteristics relevant to energy systems. Additionally, we tackle methodologies that harness data sources to derive information, particularly in the context of building sectors and regional analysis, through aggregation and fusion techniques.

4.2 Methods for the Generation of Synthetic Data

We will also explore both the standard methods, such as Random Oversampling (ROS), Cluster-Based Oversampling, and Gaussian Mixture Models, as well as deep learning methods such as Generative Adversarial Networks (GAN) used for generating synthetic data (see Figure 4). Deep learning has led to the emergence of several promising techniques for creating synthetic data, most notably the GAN. Furthermore, GAN can produce new synthetic samples that closely resemble the original dataset's underlying data distribution. Since the time it was proposed, the architecture of GAN has been modified based on the use case or field of application. Hence, we will delve into finding the most suitable architecture for generating synthetic data.

4.2.1 Random Oversampling (ROS).

Oversampling is a method in which the minority classes are duplicated. ROS is one of the classical approaches to oversampling,

[8] https://github.com/OpenEnergyPlatform/open-MaStR/, accessed on 10.05.2024

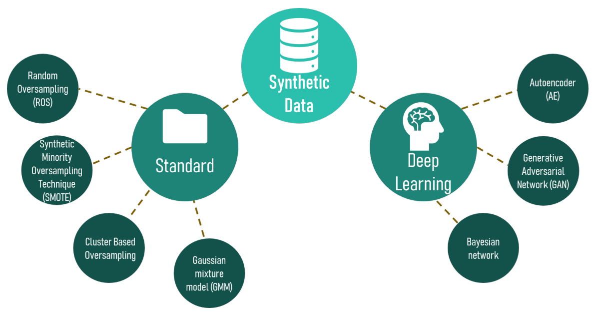


Fig. 4. Different methods for generating synthetic data [Figueira and Vaz 2022].

which expands the dataset with new observations by randomly selecting a replacement sample from the minority class. Although this is the most direct approach to growing a dataset, this method merely replicates the existing samples rather than generating new ones [Batuwita and Palade 2010].

4.2.2 *Synthetic Minority Oversampling Technique (SMOTE).*

SMOTE is an oversampling technique in which new instances are generated for each training observation by selecting points at random. There are two components to the SMOTE algorithm: the selection mechanism and data generation mechanism [Chawla et al. 2002]. In the selection mechanism, a minority class observation and its nearest neighbors are selected at random. The data generation mechanism is responsible for generating synthetic data. Despite its advantages, SMOTE has certain limitations, such as generating noisy data. Hence, there are many variants to the classical SMOTE method that address this issue, such as the Borderline-SMOTE, Safe-level SMOTE, and Adaptive Synthetic Sampling approach (ADASYN).

4.2.3 *Cluster-Based Oversampling.*

Cluster-Based Oversampling, first proposed by [Jo and Japkowicz 2004], involves clustering the training data in the minority and majority classes separately and then applying ROS to each cluster. This method aims to improve the between-class and within-class imbalances.

4.2.4 *Gaussian Mixture Model (GMM).*

In cases where the dataset displays multiple areas of elevated density, a single Gaussian model may struggle to adequately fit the data. This is where GMM proves to be particularly useful. GMM is a probabilistic model that assumes that the data is a mixture of many Gaussian distributions, representing different subpopulations within a dataset, each of which contributes a certain weight to the whole distribution. The model learns these subpopulations through training using the Expectation Maximization algorithm.

4.2.5 *Bayesian Networks.*

A Bayesian Network, also known as a *belief network*, is a type of graphical model that represents the joint probability distribution for a group of variables [Young et al. 2009]. The two main components of the Bayesian model are a set of graphical structures and a set of conditional probability distributions. The graphical structure is a set of nodes. Each node in a Bayesian Network corresponds to a probability distribution that quantifies the likelihood of a variable taking on different values given the values of its parent variables [Heckerman et al. 1995].

4.2.6 *Autoencoders (AE).*

An autoencoder is a type of artificial neural network used in unsupervised learning. It consists of an encoder network and a decoder network that work together to learn the diverse representation of the input dataset [Ackley et al. 1985]. The encoder compresses the input data into low-dimensional data, called 'latent space', and the decoder then reconstructs the original input data from the low-dimensional representation. Despite their advantages, autoencoders also suffer from some disadvantages. They may struggle to capture the full diversity of the input data and as a result the generated samples might lack diversity or fail to represent all possible instances in the dataset. When trained on a limited dataset, autoencoders could also suffer from overfitting. This can lead to samples being generated that are very similar to the training data, but do not accurately capture the underlying distribution [Figueira and Vaz 2022].

4.2.7 *Generative Adversarial Networks (GAN).*

The idea behind GAN is to train two neural networks, a generator (G) and a discriminator (D), simultaneously (see Figure 5). The generator model tries to imitate the underlying original distribution of the data while the discriminator model tries to classify a given observation as real (coming from the original dataset) or fake (generated by the generator model) [Goodfellow et al. 2020]. During the training, the two models compete with each other. Based on the discriminator's feedback, also known as the *loss function*, the generator attempts to improve its performance by generating more realistic samples (see Figure 5). Through this process, both the generator and discriminator models improve their performance, leading to the generation of high-quality synthetic data.

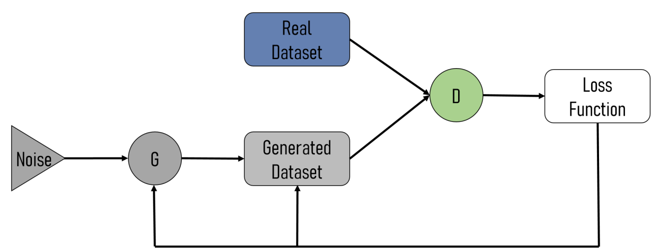


Fig. 5. Overview of GAN. Given a random noise, G generates a set of datapoints, the generated dataset. The generated dataset and the real dataset are fed to D, which labels the output as a loss function. This is then fed back to G and D to improve their performances [Figueira and Vaz 2022].

4.3 Validation Metrics

To ensure the proper validation of synthetic data , we will consider evaluation metrics tailored to our specific use case. Existing metrics often fall short in capturing complex relationships and dependencies within data, such as temporal correlations in time-series data. We will address these limitations essential to ensure a more accurate and rigorous validation process, ultimately leading to higher quality and more reliable synthetic data.

Relying on the existing validation techniques, we will identify key characteristics in conventional data, and design and test new metrics through simulations and empirical analysis. We will employ a combination of statistical tests, visual inspection, and domain-specific criteria for comparison, such as Kolmogorov-Smirnov tests for distributional comparison and Pearson correlation coefficients for relationship assessment [Whitnall et al. 2011]. By providing a comprehensive assessment, these new metrics will enhance the reliability and applicability of synthetic data, ensuring it meets the rigorous standards required for its intended use.

4.4 Ontology-Based Descriptions for Spatio-Temporal Compliance

Ontology descriptions can guide synthetic data generation to produce data that complies with specific scales. For instance, if the ontology specifies that the data should have a specific resolution, the synthetic data generation algorithm can be configured to produce images at this resolution. This ensures consistency with real satellite imagery, facilitating seamless integration into existing datasets. In our initial work, we investigate high resolution aerial images and lower resolution satellite images from Sentinel-2 [Spoto et al. 2012] and use these immense monitoring tools to segment solar installations in Germany. Super-resolution images can be synthesized from the widely available Sentinel-2 data, allowing a regular automated update of the MaStR registry inputs.

By leveraging semantically rich annotations, we can generate high-quality synthetic satellite imagery that aligns with specific requirements. The generated data will enhance the robustness and accuracy of machine learning models, fill the data gaps, support comprehensive analysis, and provide valuable insights across various applications, from environmental monitoring to urban planning enriching further the NEED platform.

5 KNOWLEDGE GRAPH

5.1 Challenges of Data Integration in Energy System Planning

Energy system planning is a multifaceted approach that requires a holistic understanding of the systems and processes involved. Numerous researchers have contributed to this field by proposing energy system models and tools aimed at optimizing energy system planning [Eveloy and Ahmed 2022; Henke et al. 2022; Metzger et al. 2021; Prina et al. 2018]. The accuracy and effectiveness of energy system modeling and planning depends on the quantity and quality of input data [Keirstead et al. 2012].

In the context of energy system planning, data availability stands out as one of the fundamental challenges. These input data originate from diverse databases, encompassing both conventional data discussed in Section 3 and synthetic data explored in Section 4. These data span multiple energy domains, including electricity, heating, and cooling, and cover a wide range of technological, economic, social, and political aspects. Cross-regional cooperation, renewable energy source integration, and environmental impact considerations become more prevalent. Thus, the complexity of energy system planning escalates, accompanied by a dramatic increase in data requirements.

Moreover, data quality assumes critical significance in energy system planning. Errors or inaccuracies within data sources can lead to additional efforts, necessitating the use of proxy data or calculation adjustments [Keirstead et al. 2012]. Such discrepancies may ultimately result in flawed models and suboptimal planning outcomes. In addition, ensuring that data aligns precisely with the intended purpose and context is a foundational principle. Valid data faithfully represent the phenomena under study and significantly contribute to effective planning efforts.

Consequently, there is an urgent need for data integration to enhance interoperability and ensure the availability of reliable data.

5.2 Ontology-Based Data Integration

In recent years, Semantic Web Technologies (SWT) and ontology-based data integration (OBDI) have garnered increasing attention as innovative and effective approaches for addressing challenges related to data management in the energy domain [Sicilia and Costa 2017]. OBDI, as an information management system, comprises three key components: an ontology, a set of data sources, and the mapping between them [Liu and Özsu 2018]. The ontology serves as a set of formal, explicit specifications of a common conceptualization that captures a shared understanding of the matters involved [Lork et al. 2019]. It maps the required terminological aspects within a knowledge representation.

In the context of the OBDI, ontologies provide a high-level, conceptual view of the set of data sources [Calvanese et al. 2011]. By constructing a Knowledge Graph (KG) based on ontologies, this framework enables the interlinking of information across diverse domains at data level.

In the NEED project, most data sources initially provide only tabular rows of information. Classification is necessary to integrate these datasets into a larger framework. This will be achieved using a taxonomy constructed from both existing and synthetic data examples. This taxonomy will delineate key concepts within each domain of the NEED ecosystem, facilitating the clustering and extension of these concepts. Additionally, it will enable data mapping into our knowledge network without necessitating direct replication into semantic triples.

The energy and buildings sectors already boast a substantial collection of existing ontologies [Pritoni et al. 2021]. Therefore, the goal of the NEED project is not to create a new standalone ontology, but rather to utilize these existing resources by aligning with and incorporating new concepts as needed.

Thus, the design of the NEED ontology will be modular. The core module will amalgamate shared concepts and terms that are fundamental across all domains considered, as well as recurring attributes and relations among entities. These shared elements will be identified during the initial data acquisition phase, which will be followed by the dissemination and definition of key entities and attributes.

Each domain within the project, such as buildings, electricity, or heating, will have its own module enriched with specialized classes and categories. These will be linked to the broader concepts established in the core module. The configuration of these domain-specific modules will be shaped by the systems they represent and the available data types, such as georeferenced data, potentially requiring the integration of a common ontology module across different system categories. Furthermore, these modules will be designed to align with other domain-specific ontologies, enabling module-to-module alignments without necessitating full-scale alignment across the entire NEED ontology, which is known in the literature as a hybrid approach.

The final design of the ontology will encompass all necessary data sources and metrics, accurately reflecting the systems under consideration. The modular setup will ensure that the ontology stack can be flexibly instantiated based on the specific needs of each use case and application, thus ensuring adaptability and relevance across a variety of scenarios.

5.3 Implementation of OBDI via Knowledge Graph

After the ontology design phase, our focus will shift to ontology-based data integration. This will involve mapping data sources to the corresponding modules within the domain ontology. The culmination of this process will be the creation of a comprehensive knowledge graph. This graph will play a vital role in supporting tools for planning and analysis within the application layer, enabling them to perform queries and interact efficiently with the data. Such functionality is crucial as it acts as a semantic bridge, seamlessly linking diverse data sources and enabling the application's features without requiring users to know details of the underlying database structures. A diagram of the envisioned framework is shown in Figure 6.

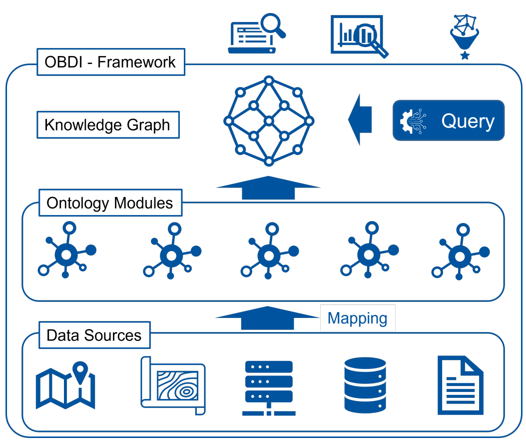


Fig. 6. Possible structure of OBDI in the NEED project.

Implementation of knowledge graph in NEED can be done using The World Avatar (TWA) Project. The World Avatar (TWA) project uses knowledge models and technologies from the Semantic Web to seamlessly integrate data and computational agents [Akroyd et al. 2021]. It offers a general and scalable way to connect heterogeneous data sources to provide an aligned world view, provide up-to-date insights, analyse complex what-if-scenarios, and provide robust control functionalities as a bridge between the physical and the digital worlds, supporting more effective and coordinated decision-making processes.

What sets TWA apart from other approaches is its dynamic behaviour. Computational agents act as autonomous knowledge components to manage and update data. Inputs and outputs from the agents can be semantically annotated to form chains of dependent information. A provenance framework ensures that the consequences of any changes to data are cascaded throughout TWA [Bai et al. 2024a]. This automation, together with input agents that continuously assimilate data feeds into the system, allows TWA to remain current and responsive to new information and scenarios. TWA is scalable by design. The computational agents can wrap around existing software, and new ontologies can be incorporated continuously while maintaining connections to everything existing in real world. The underlying knowledge models (i.e., ontologies) facilitate interoperability and knowledge retention by explicitly codifying domain expertise.

TWA has been applied in a wide variety of contexts including representing molecular scale information to automate calculations [Farazi et al. 2020] [Pascazio et al. 2023], chemical experiments [Bai et al. 2024b] and materials discovery [Kondinski et al. 2022], the development of natural language methods to query data and trigger calculations [Zhou et al. 2021] [Zhou et al. 2022] [Tran et al. 2024], the integration of data from Geographical Information Systems (GIS), Building Information Models (BIM), Building Management System (BMS) [Quek et al. 2024], digitalisation of city planning processes in Singapore [Chadzynski et al. 2021] [Silvennoinen et al. 2023], optimisation of district heating operations in Germany [Hofmeister et al. 2024c], and evaluation of cascading risk to improve the climate resilience of connected infrastructure networks in the UK [Hofmeister et al. 2024a] [Hofmeister et al. 2024b], including the Climate Resilience Demonstrator.[9] TWA separates data and knowledge representation from technical implementation, allowing it to exceed the capabilities of connected digital twins. It is platform-agnostic and open-source, eliminating the risk of vendor lock-in and fostering collaborative and transparent development. It has a distributed architecture that supports safe data access, including local hosting and access control, ensuring data security and privacy. TWA is designed for agents to wrap around existing software, facilitating seamless integration of both current and future technologies. The ideas developed in TWA project will support the development of the underlying knowledge models and technological aspects of the NEED project.

5.4 Semantic Georegistration of Data

As most of the data that are relevant to NEED cover aspects of specific geographic regions or are obtained at specific locations, we target to semantically annotate the data with their geographic references. Ontologies can facilitate the creation of synthetic geographical data by providing a structured and semantically rich framework that defines the entities, relationships, and constraints specific to geographical domains. For this, a hierarchical approach will be established that combines the various relevant scopes for our planning domains from building layouts, to city districts, city limits, administrative districts, states, and the overall country. A

[9]https://digitaltwinhub.co.uk/climate-resilience-demonstrator-credo/, accessed on 10.05.2024

related approach using an OntoCityGML ontology that describes a CityGML-based conceptual schema was used to create a semantic model of the city of Berlin [Chadzynski et al. 2022].

A suitable approach is to rely on the Open Geospatial Consortium's GeoSPARQL 1.1 standard[10], which on the one hand provides a standardized vocabulary for describing geospatial linked data and their corresponding geometric properties and on the other hand a set of extensions to the SPARQL query language to properly interpret the represented models. By employing this technique, represented data in the NEED platform can be easily segmented along the defined geospatial layers, as geospatial properties can be used in the formulation of semantic queries for filtering the data. This will enable the extraction of required data and the structured comparison of different regions.

6 SOFTWARE ARCHITECTURE

In order to supply the application scenarios with the data sources mentioned in the previous sections and link them to a knowledge graph using ontologies, a suitable infrastructure must be created. This is intended to ensure flexible and efficient processing within the framework of a platform ecosystem and requires a modular, scalable, and robust software architecture. Since no existing sources are to be replaced and we are therefore dealing with a highly distributed system, we do not rely on a monolithic architecture but rather on a service-based architecture (SBA) using cloud computing technologies. In this way, we eliminate vendor lock-in and interoperability issues. First, we will highlight the basic concept, and then explain each component of the architecture, before we conclude with a possible instantiation.

Service-Based Architecture (SBA): . SBA is an approach for developing an application as small services, so-called microservices. The characteristics of SBA can be summarized as follows [Lyu et al. 2020; Söylemez et al. 2024]:

- Modularity: The application is broken down into multiple microservices. Each microservices can be developed, deployed, and managed independently. Each microservice is associated with a specific business capability within a certain context boundary, resulting in low complexity and small size.
- Lightweight communication mechanism: The microservices communicate with lightweight mechanisms, often an HTTP resource API.
- Decentralized governance: The scalability of a microservice-based application can be significantly improved through decentralized service governance and data management.
- Agility: SBA makes it easier to adapt to new requirements and change management. In a monolithic architecture, any changes require the whole system to be rebuilt and completely redeployed. However, with SBA, only the affected services need to be rebuilt and deployed independently. Microservices are highly maintainable and testable, making them a great choice for modern software development.

This will enable conventional and synthetic data sources to instantiate their own microservices. These microservices will enable both integration with the ontologies through corresponding API descriptions and the actual, automated retrieval of the requested information. The latter can be understood as wrappers that will offer a NEED-compliant API and translate it into the often proprietary interfaces or implement access to the raw data.

Container Technology: Containerization is a popular virtualization approach where isolation of applications and services occurs at the host level through containers. This provides several advantages for users [Behravesh et al. 2019; Lyu et al. 2020]:

- Lightweight and efficient deployment: Deploying applications using lightweight container images offers practical advantages since these images contain all the necessary dependencies for the seamless operation of an application.
- Portability: A container provides a self-sufficient and executable computational space, encompassing the code, runtimes, system tools, libraries, and configurations necessary for an application.
- Scalability: Containers facilitate agile scaling of applications in response to demand, ensuring optimal resource utilization. This scalability empowers industrial organizations to replicate and distribute cloud-native application instances across various edge devices or servers. This distributed approach enables workload balancing and efficient device utilization, preventing specific infrastructure components from becoming overloaded.
- Secure and reliable: Container platforms improve fault tolerance through robust mechanisms that isolate and manage application failures effectively. In the event of a container failure, it can be quickly restarted or substituted without disrupting the functionality of other containers or the entirety of the infrastructure.

Having a robust and flexible infrastructure is a result of the above characteristics of SBA and container technology. The robustness of a system is an important life-cycle attribute, since it is a strategic attribute that supports business, development, and operation needs. As a result, robust infrastructure uses some techniques in order to detect faults (such as monitoring, heartbeat, condition monitoring, voting, etc.) and recover from faults (such as redundant spares, rollbacks, restarts, etc.), as well as to prevent faults (such as predictive models) [Kazman et al. 2022; Paparistodimou et al. 2020].

A containerized infrastructure also means that NEED components may be deployed relatively easily on other systems and therefore will not necessarily have to be run on a central platform. Individual process chains may be set up through suitable orchestration of the data sources and the integration of preprocessing steps. A simple preprocessing step could be, for example, changing the resolution of time series data. The individual process chain, which will be made up of containers of the data sources in combination with basic functions, may then run locally on the respective client side and thus enable the scaling of the overall system. Figure 7 highlights the proposed NEED architecture, designed to address the shortcomings present in a monolithic architecture. The figure highlights several innovative features as outlined below.

[10] https://docs.ogc.org/is/22-047r1/22-047r1.html, accessed on 10.05.2024

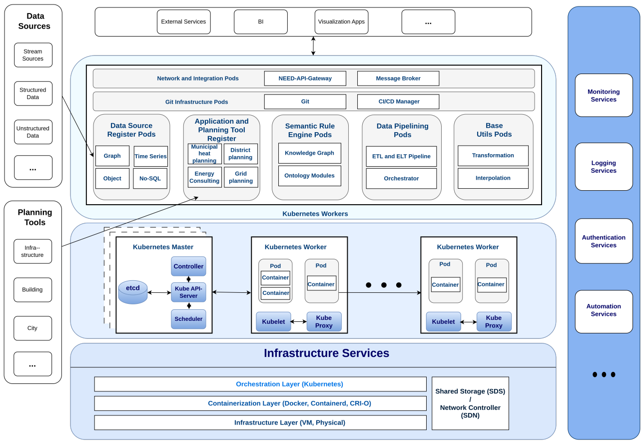


Fig. 7. NEED architecture.

Containerized Deployment Mode: In practice, the proposed NEED architecture will be implemented within a cluster administered by Kubernetes (K8s). This setup will manage various aspects like deployment, maintenance, and scalability of containers, thereby facilitating the management of microservice-oriented applications. Several methods exist for establishing a Kubernetes cluster, with one cost-effective approach being the design and deployment of Kubernetes nodes on predefined cloud infrastructure rather than physical hardware resources. This deployment model typically utilizes Infrastructure as a Service (IaaS). Moreover, to ensure continuous availability of applications, strategies will be employed to enhance the resilience and reliability of the Kubernetes cluster, as depicted in Figure 7. These strategies may include employing multiple nodes for data replication, implementing shared storage solutions, and so forth.

Containerized Microservices as Essential Components: Every NEED application will be housed within its own container, ensuring that its operating environment is isolated from the host system. Consequently, any errors within a container instance do not disrupt the normal functioning of other services. Additionally, container instances can be generated, updated, and terminated with ease and speed, facilitating independent scaling of each microservice as required.

- Data Source Register: Here, the data source microservices will register themselves, so that the platform services will be aware of their data availability and what kind of data (see Sections 3 and 4) will be provided.
- Application and Planning Tool Register: Similar to the Data Source Register, the planning tools or their wrappers that will enable interaction with the Semantic Rule Engine Pods or the knowledge graph will register here.
- Semantic Rule Engine: These pods will hold the required services and functionality for the ontology-based data integration as described in Section 5.2.
- Data Pipelining: The functionality of these pods will enable the orchestration and merging of different data sources, whereby transformations and preprocessing steps may also be inserted in substeps depending on the requirements of the requesting entity.
- Base Utils: They will provide the necessary preprocessing steps, transformations, and manipulations of the data in order to realize seamless orchestration.
- Git and CI/CD Infrastructure Pods: These pods will enable the continuous updating of client and server code throughout the development process. By leveraging these tools, development teams may maintain a streamlined workflow for code integration, testing, and deployment, promoting agility and efficiency in software development cycles.
- Network and Integration Pods: In a microservices architecture, client applications typically interact with various microservices to access different functionalities. However, direct consumption of these microservices necessitates managing multiple calls to their endpoints, which can pose challenges as an application evolves by introducing new microservices or updating existing ones. Moreover, developers may employ different technology stacks and communication protocols for each microservice within their application. Additionally, the elastic nature of the cloud enables services to horizontally scale in response to fluctuating demand, enhancing application resilience. Nonetheless, this scalability necessitates load balancing, simplified service discovery, and features like timeouts and retries for effective recovery. To address these challenges, networking pods such as the API Gateway container will be employed.

Shared Services (Monitoring, Logging, Authentication, and Automation): Logging and monitoring are indispensable components of any containerized cluster, especially in a microservices architecture where numerous services are distributed across multiple machines or containers. The complexity of such deployments necessitates robust logging and monitoring solutions to swiftly diagnose issues. Additionally, each microservice requires its own security measures to manage access for team members, making authentication and authorization crucial considerations. Moreover, the implementation and maintenance of large systems like these would be challenging without automation tools. Automation mechanisms enable administrators to effectively implement and maintain the infrastructure, mitigating the risk of operational disasters.

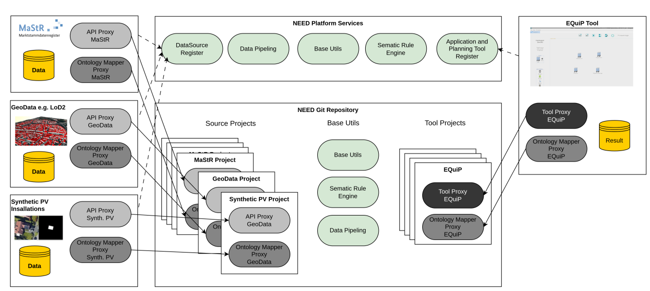


Fig. 8. Possible instantiation of the NEED ecosystem with API Proxy (light gray), Ontology Mapper (dark gray), Tool Proxy (black), Core Services and Base Utils (light green).

Figure 8 illustrates a possible instantiation of three data source providers (*Marktstammdatenregister*, LoD2 data, synthetic PV installations), which would register their capabilities at the Data Source Register (dashed arrow). This way, the Semantic Rule Engine may utilize the Ontology Mapper to integrate the new data sources. In parallel, the resources (API Proxy and Ontology Mapper) would be provided as Git projects. These may be deployed locally on the client side to enable the data access. A similar process would be conducted for tools. The requirements and data needs of the tools would be extracted via the Ontology Mapper and processed with the help of the knowledge graph. The result would provide the blueprint for the orchestration and preprocessing steps to retrieve the required data from the available data sources. The combined data records may then be forwarded to the Tool Proxy and finally processed by the tool. In this example, we would collect the already installed PV systems, update and crosscheck the result with synthetic PV installations, and calculate the still available roof area out of the LoD2 data to estimate the PV potential for a certain area.

7 SUMMARY AND OUTLOOK

In this work, we have emphasized the importance and need for energy-related and up-to-date data in order to perform reliable and repeatable planning tasks. In order to reduce the considerable effort required to collect and preprocess the necessary data, we will create an efficient, robust, and scalable infrastructure as part of the NEED research project. The project analyzed typical planning tasks and identified the requirements, stakeholders, and processes. The use cases developed for this were summarized in application scenarios and suitable case studies, which have already been implemented by the industry partners involved in the project, were documented for evaluation.

The idea of the project is to make numerous data sources available without replacing them, but rather integrating them. On the one hand, various conventional data sources were examined and their data and metadata evaluated with regard to the FAIR criteria. On the other hand, we want to close existing data gaps with the help of synthetic data generation approaches. In this context, we will also validate the quality of the developed methods, for example with available conventional data, in order to assess the transferability of the data. This will make it possible to transfer such data collection and provision to other countries that do not have corresponding data sources.

A major difficulty lies in combining different data for the different planning levels. Here the NEED project relies on ontologies or a knowledge graph. We believe that creating interoperability by using these technologies makes more sense and is more efficient than establishing a standard for data provision. Ultimately, the modules presented must be combined so that existing planning tools can access the required datasets. In some cases, further processing steps are required before the data can be merged and orchestrated. For this purpose, the project is developing a distributed service-based architecture, the essential components of which have been presented.

In the further course of the research project, the application scenarios presented will be implemented using conventional and synthetic data sources, which will be linked to the requirements of the planning tools using a knowledge graph, and the added value of the NEED ecosystem will be presented using the case studies. In this context, we will also evaluate the availability, scalability, and performance of the NEED platform. Overall, we are trying to make planning tasks more transparent, efficient, and uniform in order to further advance the energy transition.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support through the project executing agency Jülich (PTJ) with funds provided by the Federal Ministry for Economic Affairs and Climate Action (BMWK) due to an enactment of the German Bundestag under Grant No. 03EN3077A.

REFERENCES

David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. 1985. A learning algorithm for Boltzmann machines. *Cognitive science* 9, 1 (1985), 147–169.

Jethro Akroyd, Sebastian Mosbach, Amit Bhave, and Markus Kraft. 2021. Universal digital twin-a dynamic knowledge graph. *Data-Centric Engineering* 2 (2021), e14.

Christophe Bahim, Carlos Casorrán-Amilburu, Makx Dekkers, Edit Herczog, Nicolas Loozen, Konstantinos Repanas, Keith Russell, and Shelley Stall. 2020. The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal* 19 (Oct 2020). https://doi.org/10.5334/dsj-2020-041

Jiaru Bai, Kok Foong Lee, Markus Hofmeister, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. 2024a. A derived information framework for a dynamic knowledge graph and its application to smart cities. *Future Generation Computer Systems* 152 (2024), 112–126.

Jiaru Bai, Sebastian Mosbach, Connor J Taylor, Dogancan Karan, Kok Foong Lee, Simon D Rihm, Jethro Akroyd, Alexei A Lapkin, and Markus Kraft. 2024b. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nature Communications* 15, 1 (2024), 462.

Rukshan Batuwita and Vasile Palade. 2010. Efficient resampling methods for training support vector machines with imbalanced datasets. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

Rasoul Behravesh, Estefanía Coronado, and Roberto Riggio. 2019. Performance Evaluation on Virtualization Technologies for NFV Deployment in 5G Networks. In *2019 IEEE Conference on Network Softwarization (NetSoft)*. 24–29. https://doi.org/10.1109/NETSOFT.2019.8806664

Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. 2011. The MASTRO System for Ontology-Based Data Access. *Semantic Web* 2, 1 (2011), 43–53. https://doi.org/10.3233/SW-2011-0029

Arkadiusz Chadzynski, Nenad Krdzavac, Feroz Farazi, Mei Qi Lim, Shiying Li, Ayda Grisiute, Pieter Herthogs, Aurel von Richthofen, Stephen Cairns, and Markus Kraft. 2021. Semantic 3D City Database—An enabler for a dynamic geospatial knowledge graph. *Energy and AI* 6 (2021), 100106.

Arkadiusz Chadzynski, Shiying Li, Ayda Grisiute, Feroz Farazi, Casper Lindberg, Sebastian Mosbach, Pieter Herthogs, and Markus Kraft. 2022. Semantic 3D City Agents—An intelligent automation for dynamic geospatial knowledge graphs. *Energy and AI* 8 (2022), 100137. https://doi.org/10.1016/j.egyai.2022.100137

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

Fida Dankar and Mahmoud Ibrahim. 2021. Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation. *Applied Sciences* 11 (02 2021), 2158. https://doi.org/10.3390/app11052158

Valerie Eveloy and Wasiq Ahmed. 2022. Evaluation of Low-Carbon Multi-Energy Options for the Future UAE Energy System. *Sustainable Energy Technologies and Assessments* 53 (Oct. 2022), 102584. https://doi.org/10.1016/j.seta.2022.102584

Feroz Farazi, Nenad B Krdzavac, Jethro Akroyd, Sebastian Mosbach, Angiras Menon, Daniel Nurkowski, and Markus Kraft. 2020. Linking reaction mechanisms and quantum chemistry: An ontological approach. *Computers & Chemical Engineering* 137 (2020), 106813.

Alvaro Figueira and Bruno Vaz. 2022. Survey on synthetic data generation, evaluation methods and GANs. *Mathematics* 10, 15 (2022), 2733.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

David Heckerman, Dan Geiger, and David M Chickering. 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning* 20 (1995), 197–243.

Hauke T.J. Henke, Francesco Gardumi, and Mark Howells. 2022. The Open Source Electricity Model Base for Europe - An Engagement Framework for Open and Transparent European Energy Modelling. *Energy* 239 (Jan. 2022), 121973. https://doi.org/10.1016/j.energy.2021.121973

Markus Hofmeister, Jiaru Bai, George Brownbridge, Sebastian Mosbach, Kok F Lee, Feroz Farazi, Michael Hillman, Mehal Agarwal, Srishti Ganguly, Jethro Akroyd, et al. 2024a. Semantic agent framework for automated flood assessment using dynamic knowledge graphs. *Data-Centric Engineering* 5 (2024), e14.

Markus Hofmeister, George Brownbridge, Michael Hillman, Sebastian Mosbach, Jethro Akroyd, Kok Foong Lee, and Markus Kraft. 2024b. Cross-domain flood risk assessment for smart cities using dynamic knowledge graphs. *Sustainable Cities and Society* 101 (2024), 105113.

Markus Hofmeister, Kok Foong Lee, Yi-Kai Tsai, Magnus Müller, Karthik Nagarajan, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. 2024c. Dynamic control of district heating networks with integrated emission modelling: A dynamic knowledge graph approach. *Energy and AI* 17 (2024), 100376.

Taeho Jo and Nathalie Japkowicz. 2004. Class imbalances versus small disjuncts. *ACM Sigkdd Explorations Newsletter* 6, 1 (2004), 40–49.

Rick Kazman, Phil Bianco, Sebastián Echeverría, and James Ivers. 2022. *Robustness*. Technical Report TR-004. Carnegie Mellon University. DOI: 10.1184/R1/16455660.

James Keirstead, Mark Jennings, and Aruna Sivakumar. 2012. A Review of Urban Energy System Models: Approaches, Challenges and Opportunities. *Renewable and Sustainable Energy Reviews* 16, 6 (Aug. 2012), 3847–3866. https://doi.org/10.1016/j.rser.2012.02.047

Aleksandar Kondinski, Angiras Menon, Daniel Nurkowski, Feroz Farazi, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. 2022. Automated rational design of metal–organic polyhedra. *Journal of the American Chemical Society* 144, 26 (2022), 11713–11728.

Ling Liu and M. Tamer Özsu (Eds.). 2018. *Encyclopedia of Database Systems*. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-8265-9

Clement Lork, Vishal Choudhary, Naveed Ul Hassan, Wayes Tushar, Chau Yuen, Benny Kai Kiat Ng, Xinyu Wang, and Xiang Liu. 2019. An Ontology-Based Framework for Building Energy Management with IoT. *Electronics* 8, 5 (April 2019), 485. https://doi.org/10.3390/electronics8050485

Zhongliang Lyu, Hua Wei, Xiaoqing Bai, and Chunjie Lian. 2020. Microservice-Based Architecture for an Energy Management System. *IEEE Systems Journal* PP (04 2020), 1–12. https://doi.org/10.1109/JSYST.2020.2981095

Michael Metzger, Mathias Duckheim, Marco Franken, Hans Joerg Heger, Matthias Huber, Markus Knittel, Till Kolster, Martin Kueppers, Carola Meier, Dieter Most, Simon Paulus, Lothar Wyrwoll, Albert Moser, and Stefan Niessen. 2021. Pathways toward a Decarbonized Future—Impact on Security of Supply and System Stability in a Sustainable German Energy System. *Energies* 14, 3 (2021). https://doi.org/10.3390/en14030560

Giota Paparistodimou, Alex Duffy, Robert Ian Whitfield, Philip Knight, and Malcolm Robb. 2020. A network science-based assessment methodology for robust modular system architectures during early conceptual design. *Journal of Engineering Design* 31, 4 (2020), 179–218.

Laura Pascazio, Simon Rihm, Ali Naseri, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. 2023. Chemical species ontology for data integration and knowledge discovery. *Journal of Chemical Information and Modeling* 63, 21 (2023), 6569–6586.

Matteo Giacomo Prina, Lorenzo Fanali, Giampaolo Manzolini, David Moser, and Wolfram Sparber. 2018. Incorporating Combined Cycle Gas Turbine Flexibility Constraints and Additional Costs into the EPLANopt Model: The Italian Case Study. *Energy* 160 (Oct. 2018), 33–43. https://doi.org/10.1016/j.energy.2018.07.007

Marco Pritoni, Drew Paine, Gabriel Fierro, Cory Mosiman, Michael Poplawski, Avijit Saha, Joel Bender, and Jessica Granderson. 2021. Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis. *Energies* 14, 7 (April 2021), 2024. https://doi.org/10.3390/en14072024

Hou Yee Quek, Markus Hofmeister, Simon D Rihm, Jingya Yan, Jiawei Lai, George Brownbridge, Michael Hillman, Sebastian Mosbach, Wilson Ang, Yi-Kai Tsai, et al. 2024. Dynamic knowledge graph applications for augmented built environments through "The World Avatar". *Journal of Building Engineering* 91 (2024), 109507.

Álvaro Sicilia and Gonçal Costa. 2017. Energy-Related Data Integration Using Semantic Data Models for Energy Efficient Retrofitting Projects. In *The Sustainable Places 2017 (SP2017) Conference*. MDPI, 1099. https://doi.org/10.3390/proceedings1071099

Heidi Silvennoinen, Arkadiusz Chadzynski, Feroz Farazi, Ayda Grišiūtė, Zhongming Shi, Aurel von Richthofen, Stephen Cairns, Markus Kraft, Martin Raubal, and Pieter Herthogs. 2023. A semantic web approach to land use regulations in urban planning: The OntoZoning ontology of zones, land uses and programmes for Singapore. *Journal of Urban Management* 12, 2 (2023), 151–167.

Mehmet Söylemez, Bedir Tekinerdogan, and Ayça Kolukısa Tarhan. 2024. Microservice reference architecture design: A multi-case study. *Software - Practice and Experience* 54, 1 (Jan. 2024), 58–84. https://doi.org/10.1002/spe.3241

Francois Spoto, Omar Sy, Paolo Laberinti, Philippe Martimort, Valerie Fernandez, Olivier Colin, Bianca Hoersch, and Aime Meygret. 2012. Overview Of Sentinel-2. In *2012 IEEE International Geoscience and Remote Sensing Symposium*. 1707–1710. https://doi.org/10.1109/IGARSS.2012.6351195

Dan Tran, Laura Pascazio, Jethro Akroyd, Sebastian Mosbach, and Markus Kraft. 2024. Leveraging text-to-text pretrained language models for question answering in chemistry. *ACS omega* 9, 12 (2024), 13883–13896.

Carolyn Whitnall, Elisabeth Oswald, and Luke Mather. 2011. An Exploration of the Kolmogorov-Smirnov Test as Competitor to Mutual Information Analysis. Cryptology ePrint Archive, Paper 2011/380. https://eprint.iacr.org/2011/380

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (15 Mar 2016), 160018. https://doi.org/10.1038/sdata.2016.18

Jim Young, Patrick Graham, and Richard Penny. 2009. Using Bayesian networks to create synthetic data. *Journal of Official Statistics* 25, 4 (2009), 549–567.

Xiaochi Zhou, Daniel Nurkowski, Angiras Menon, Jethro Akroyd, Sebastian Mosbach, and Markus Kraft. 2022. Question answering system for chemistry—A semantic agent extension. *Digital Chemical Engineering* 3 (2022), 100032.

Xiaochi Zhou, Daniel Nurkowski, Sebastian Mosbach, Jethro Akroyd, and Markus Kraft. 2021. Question answering system for chemistry. *Journal of Chemical Information and Modeling* 61, 8 (2021), 3868–3880.

Empowering Energy Communities and P2P Energy Sharing: A Novel End-to-End Ecosystem for Planning, Deployment, and Operation

SHIEVAM KASHYAP, CHRISTOPH SCHAFFER, TOBIAS FISCHER, MICHAEL ZAUNER, FRANZ FISCHER, MARC KURZ, GEORG HARTNER, STEFAN GRÜNBERGER, and OLIVER HÖDL, University of Applied Sciences Upper Austria, Austria

The widespread adoption of renewable energy sources has necessitated innovative solutions to address their inherent intermittency and facilitate energy democracy through Energy Communities (ECs). However, the successful planning, deployment, and operation of ECs face multifaceted socio-technical challenges. This publication presents a novel end-to-end ecosystem designed to foster the planning, deployment, and operation of ECs and peer-to-peer (P2P) energy relations. The ecosystem empowers stakeholders by addressing various challenges associated with the formation and management of ECs. A user-centric mobile application facilitates the identification and engagement of prospective EC members, evaluation of feasibility, and negotiation of operational, business, and financial models. The application also enables users to monitor the performance of their ECs or P2P energy-sharing relationships. An Energy Community Marketplace serves as a platform for interactions between feasible clusters and EC operators, acting as a broker to finalize the EC operator and establish connections with users' near real-time data access. The ecosystem inherently includes historical and real-time data integration based on user consent and GDPR best practices. The proposed ecosystem streamlines entire EC lifecycle, fostering active user participation, data-driven decision-making, and seamless integration of diverse stakeholders, ultimately supporting the energy transition and democratization goals.

CCS Concepts: • **Information systems → Data streaming**; **Process control systems**; *Online analytical processing*; • **Human-centered computing → Ubiquitous and mobile computing**; **Visualization**.

Additional Key Words and Phrases: Renewable Energy Communities, Citizen Energy Communities, Mobile Application, Energy data access, Energy community control, Energy Community Marketplace

1 INTRODUCTION

With the increased adoption of Renewable Energy Sources (RES) globally, the energy landscape is witnessing a paradigm shift to manage the intermittent nature of RES. A key development in this regard is the emergence of Energy Communities (ECs) and the accompanying concept of energy democracy. ECs are viewed as a promising avenue for enhancing self-sufficiency [Iazzolino et al. 2022], bolstering local economic circulation [Ruggieri et al. 2023], fostering social acceptance of renewable energy [Bauwens and Devine-Wright 2018; Biresselioglu et al. 2021], and optimising individual investments of RES [Manuel De Villena et al. 2022]. Furthermore, these communities can contribute to grid stability by providing grid flexibility services [Rocha et al. 2023] and facilitating demand response [Iazzolino et al. 2022].

The term *Energy Community* encompasses a wide range of meanings and configurations, given the varied types of actors, technologies, and complex network infrastructures involved. In a broad sense, an EC comprises a diverse array of stakeholders, including consumers, prosumers, social entrepreneurs, public authorities, and community organizations, as mentioned by [Iazzolino et al. 2022]. This multifaceted composition reflects the intricate nature of energy community structures and their adaptability to various contexts.

In response to the evolving energy landscape, various legislative bodies are incorporating the concept of Energy Communities into their respective jurisdictions. For example, European energy policies are shifting away from incentive-based programs, with the goal of attracting private financing. In this context, energy communities and collective self-consumption initiatives are gaining recognition as crucial elements in the transition. Their significance has been further underscored by directives issued such as the Renewable Energy Directive (RED II) (2018/2001/EU) [Union 2018] and the Internal Electricity Market Directive (IEMD) (2019/944/EU) [Union 2019]. The 2018 RED II marked a turning point by introducing the official definition of *renewable energy communities* (RECs) under Article 2(16) and proposing a regulatory framework specifically for them. Subsequently, the IEMD, which entered into force in 2019, built upon RED II by introducing a broader definition encompassing *citizens' energy communities* (CECs). This broader definition acknowledges the potential for communities to engage not only with renewable energy sources but also with other aspects of energy management. RED II furthermore defines peer-to-peer renewable energy trading as *"the sale of renewable energy between market participants by means of a contract with pre-determined conditions governing the automated execution and settlement of the transaction, either directly between market participants or indirectly through a certified third-party market participant, such as an aggregator."*

Despite the legal endorsement for energy communities, technical obstacles persist [Roversi et al. 2022] that impede user adoption of such solutions. While ECs are fundamentally conceived as people-driven initiatives, a socio-technical gap emerges [Ryghaug et al. 2018] in identifying users who would be well-suited participants for an EC or P2P energy trading. In the planning phase of ECs, users currently face limited avenues to explore the potential of a prospective EC by discovering compatible energy consumers, prosumers, or producers. Access to energy data is of paramount importance for such investigations [Kashyap et al. 2021], and it must be obtained in compliance with the General Data Protection Regulation (GDPR). As a result, multiple challenges arise in finding a cluster of users well-suited for an EC. In the subsequent deployment phase, after potential users have been identified, bridging the gap between an EC operator and the participants presents another significant hurdle.

Authors' Contact Information: Shievam Kashyap, shievam.kashyap@fh-hagenberg.at; Christoph Schaffer, christoph.schaffer@fh-hagenberg.at; Tobias Fischer, tobias.fischer@fh-hagenberg.at; Michael Zauner, michael.zauner@fh-hagenberg.at; Franz Fischer, franz.fischer@fh-hagenberg.at; Marc Kurz, marc.kurz@fh-hagenberg.at; Georg Hartner, georg.hartner@fh-hagenberg.at; Stefan Grünberger, stefan.gruenberger@fh-hagenberg.at; Oliver Hödl, oliver.hoedl@fh-hagenberg.at, University of Applied Sciences Upper Austria, Austria.

The ensuing operational phase heavily relies on near real-time data access from the users, contingent upon their data-sharing consent.

To address these challenges, we propose a novel IT-based infrastructure ecosystem for energy communities that encompasses the planning, deployment, and operational phases of an EC and P2P energy-sharing relationship. This infrastructure adopts GDPR-compliant best practices to facilitate active user participation.

2 ENERGY COMMUNITIES LANDSCAPE

The establishment and operation of an EC is a multi-faceted process that requires a systematic approach spanning from initial planning to long-term maintenance. To ensure successful implementation and sustainable functioning, a comprehensive set of steps must be undertaken as delineated below:

(1) *Identifying and Engaging Prospective EC Members:*
The first crucial step involves identifying users who are willing to form an EC. These could be individuals, organizations, or businesses with synergistic value propositions and a shared sense of collective responsibility towards renewable energy. It is vital to comprehend their individual goals [Gjorgievski et al. 2021] for community participation and the means of interaction they can have with other EC users. Similar structures of user participation, such as those in community renewable energy projects, face challenges in identifying stakeholders due to diverse motives and engagement levels, as discussed in [Bauwens 2016].

(2) *Establishing the Collective Goals and Value Propositions:*
As a people-led initiative, the goals of each EC may vary based on the users' value propositions and aims [Hoffman and High-Pippert 2010]. Collective goals can range from reducing energy bills and increasing the share of renewable energy to lowering carbon emissions, combating energy poverty, fostering community engagement, increasing energy independence, and improving the local economy.

(3) *Conducting Socio-Technical Feasibility Assessments for Prospective Energy Communities:*
In the planning phase, it is crucial to understand the current energy consumption and production patterns of prospective EC users. Often, complementarity (for example, [Lowitzsch et al. 2020]) is evaluated between users' energy profiles and installed RES that can bring a symbiotic relationship within the EC. Conducting such studies aids in resource planning, grid connectivity, exploring the potential for providing grid services, and identifying opportunities for optimized renewable energy integration. Furthermore, the EC topology should also be determined at this stage based on user types and energy consumption profiles, ranging from centralized to distributed or decentralized, as denoted in [Gui and MacGill 2018]. This exercise of conducting socio-technical studies leads to the identification of feasible clusters for prospective EC deployment. This step heavily relies on the availability of the metering data of energy consumption and production, also known as Historical Validated Data (HVD), of the users. While the Clean Energy Package [European Commission 2019] establishes customer rights to access energy data and share it with chosen eligible parties, fostering new data-driven services, the lack of standardized procedures across countries poses a significant obstacle to implementation, as observed in [Karagiannis et al. 2023].

(4) *Identifying Relevant Stakeholders Based on User Clusters and Community Goals:*
Once feasible clusters are identified through empirical studies, a mechanism is needed to identify other stakeholders most relevant to the identified users from Step 1, the collective goals of the EC from Step 2, and the feasibility test results from Step 3. The aim of this step is to identify prospective EC operator(s).

(5) *Negotiating and Finalizing Operational, Business, and Financial Models:*
With all prospective stakeholders and goals identified, this phase involves negotiating the operational model and business relations between EC users. Financial models should be based on the potential gauged from preliminary studies and the estimated costs, including technological installations, infrastructure development, grid connections, and operational maintenance. Based on the value proposition and goals, as mentioned in [F.G. Reis et al. 2021], various business model archetypes are possible, such as energy cooperatives, community prosumerism, local energy markets, community collective generation, third-party-sponsored communities, community flexibility aggregation, community energy service companies (ESCOs), and e-mobility cooperatives. Prospective EC operators can negotiate with EC users the technological, administrative, and financial models and explain their approach for optimizing parameters aligned with individual user goals. There should be flexibility to adopt the EC operator and optimization algorithm, fostering innovation and adoption of new mechanisms, such as the one mentioned in [Manuel De Villena et al. 2022]. The outcome of this step includes selecting the appropriate EC operator, negotiating and finalizing the terms of operations, and adopting the modus operandi and relevant business model.

(6) *Establishing Contractual Agreements and Regulatory Compliance:*
This step involves establishing contracts between the EC operator and users, as well as the EC and the national permitting authority. Relations within the group of users and their collective relation with the EC operator should be clearly defined. Users (consumers, prosumers, producers) should be at the center of the community, with the contract specifying the scope and length of the power purchasing agreement (PPA). For formalities with regulatory authorities, procedures differ by country. The broad scope includes signing installation and operation agreements, registering the EC, reaching agreements with system operators and participants, connecting to the grid, and integrating with designated national infrastructure. Furthermore, this step includes the decision on whether the EC should adopt static or dynamic distribution. Static distribution means each participant is always allocated the agreed share generated, with surplus fed into the grid. Dynamic distribution relates to the

case when energy generated by facilities is allocated among participants based on demand to increase self-sufficiency, with individual surpluses going back into the “community surplus system." In case of Dynamic distribution, the energy allocation share can change based on the individual demands of the users within the EC.

(7) *Deployment and Operationalization of the Energy Community:*
Once the community receives the desired permits and has fully installed the technological and RES assets, it can legally commence operations. Community prosumers can begin producing and consuming energy generated at their properties and share excess energy with community consumers. The setup of the EC in the deployment phase requires steps such as recording master data of generation facilities and participants, transmitting metering points from participants to system operators, and transmitting the allocation mode of generation facilities to system operators.

(8) *Continuous Monitoring, Optimization, and Billing Operations:*
Regular monitoring and optimization are essential to ensure the performance of an EC. This requires near real-time metering data access from users, which can be extracted by establishing a consent-based, GDPR-compliant platform for accessing data from the standardized interface of smart meters. Additionally, this step encapsulates billing and operations that require transmitting energy data from all generating facilities and participants, processing and preparing data for participant billing, and informing users about Service Level Agreements (SLAs). Furthermore, users should also be empowered to monitor their own participation in the EC and P2P relationships they are part of.

(9) *Long-term Maintenance and Asset Management:*
Finally, ECs require continuous controls and maintenance to ensure installed assets last for the community’s full duration. The terms of maintenance are committed to in Steps 5 and 6.

The establishment and operation of an EC is a multi-faceted process that requires a systematic approach spanning from initial planning to long-term maintenance. To address the intricate socio-technical aspects inherent in the development of energy communities, we propose a novel end-to-end one-stop solution for fostering the planning, deployment, and operation of ECs. This solution empowers EC stakeholders by methodically addressing each of the requisite aforementioned steps. The proposed solution adopts an ecosystem approach, following a modular architecture by incorporating various building blocks. The following section provides a detailed description of these building blocks and their functionalities within the proposed ecosystem.

3 PROPOSED SOLUTION

The proposed ecosystem is meticulously designed to facilitate all the necessary steps required for establishing an Energy Community (EC). To implement the proposed solution, we developed a robust system comprising several building blocks, each with specific functions. These blocks are exhibited in Figure 1 and are briefly explained in subsequent subsections. The building blocks are categorized into four categories based on their functions. Additionally, this section describes how each block positively impacts the prerequisite steps for forming and operating an EC or P2P energy-sharing relationship.

3.1 Access to Historical Validated Data

As the name suggests, the building blocks categorized under this module serve the crucial and important function of providing access to metering and consumption data i.e. Historical Validated Data (HVD). As one can relate to the prerequisite steps required for ECs, HVD plays a pivotal role. The components marked with an asterisk (*) in Figure 1 are parts of the HVD supply chain.

Typically, smart meters transfer energy data to the Meter Data Administrator (MDA), which in many cases is the Distribution Network Operator (DNO). This collected data is validated by the MDA and made available for sharing based on customer requests. The European Union’s Clean Energy for All Europeans Package [European Commission 2019] empowers consumers by granting them the right to access their own energy-related data, including metering, production, and consumption information. The legal landscape allows users to share their energy data with third-party Eligible Parties. These third parties act as parties requesting or processing data shared by the customer and can leverage this data to provide valuable services such as identifying prospective EC clusters and finding synergies between users.

Data can be requested from country-specific existing Data Infrastructure that consists of national energy data management environments and online data hubs where historical metering and consumption data is collected, validated, and stored. This data must then be made available to established actors or eligible parties as required. However, data sharing is currently carried out in various ways across different countries, involving diverse players, processes, formats, and schemas. To address this challenge, the EDDIE (European Distributed Data Infrastructure for Energy) framework interfaces with these data-sharing infrastructures, offering a streamlined consent management user flow. By doing so, it simplifies the process of obtaining user consent for data sharing and ensures that data is accessed and shared securely and transparently. Additionally, the EDDIE Framework provides a transformation towards a common pivotal format, enabling seamless integration and compatibility of data across different Member States and diverse data-sharing infrastructures. Therefore, by leveraging EDDIE framework, our proposed solution facilitates the HVD, based on user consent, which in turn is stored in the cloud.

3.2 Access to in-house data sources

The building blocks associated with near real-time data access are indicated with a circumflex (ˆ) symbol in Figure 1. In-house data sources pertain to near real-time data obtained from the standardized interface on smart meters in most Member States of EU, given that the smart meter has been ordered and installed after July 4th, 2019. Accessing and processing this data can be challenging for customers due to its in-house availability and the need for transformation into a common format for further use. To address this challenge, the Administrative Interface for In-house Data Access

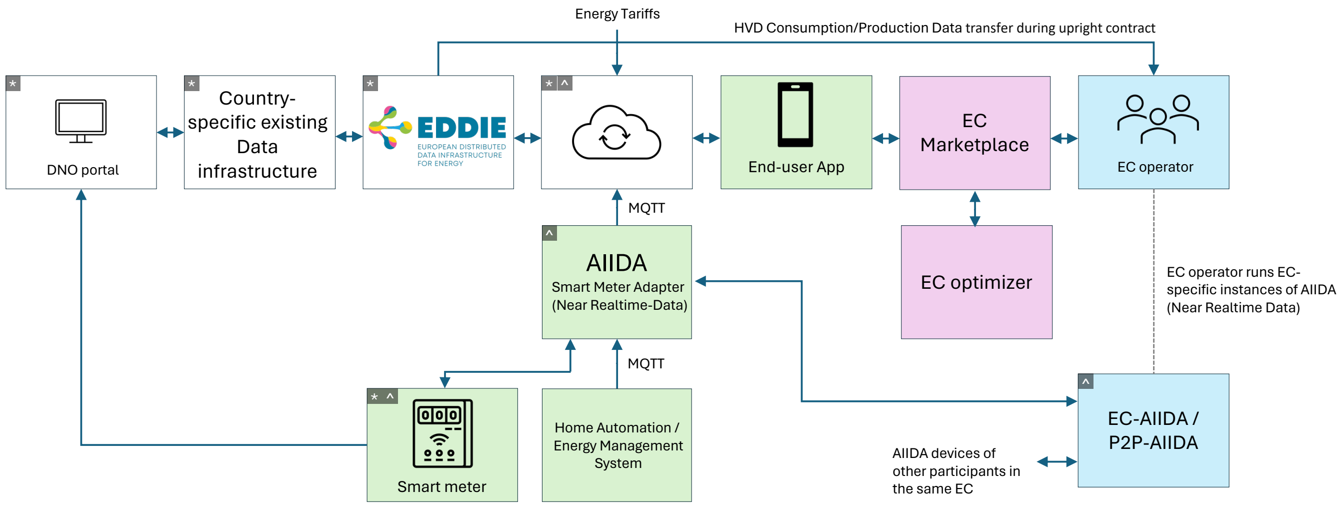


Fig. 1. Architecture of the proposed novel ecosystem solution depicting interactions between various developed building blocks. The building blocks shaded green are executed in the end-user environment, while the pink ones reside in the EC marketplace environment. The blue shaded blocks are hosted by the EC operator. Building blocks marked with an asterisk (*) correspond to the historical metering data supply chain, and those with a circumflex (ˆ) symbol correspond to the near real-time data supply chain.

(AIIDA) is designed to read data from various meter models, standards, and configurations, making it available through an online, consent-based mechanism.

The AIIDA component interacts with the cloud using MQTT protocol to send the requested data. To ensure user privacy and security, data can only be sent once users give explicit permission to share their data. EC operators execute an AIIDA instance in their environment, which communicates with the AIIDA instances of the users participating in the same EC. Based on user consent, the AIIDA instances of the users share their near real-time data with the EC operator's AIIDA instance and receive external signals simultaneously. This bidirectional real-time communication capability of AIIDA between users and EC operators additionally empowers all stakeholders to make informed decisions. Moreover, AIIDA can connect to other energy assets, such as Energy Management Systems and Home Automation devices, to control and inform these in-home assets about external signals like energy tariffs or signals from the EC operator once the user is part of an active EC. Therefore, AIIDA acts not only as the broker for near real-time data access but also as the means of communication between the external signals and the in-house assets for operational control and maintenance. Consequently, the AIIDA component plays a crucial role during the operational phase of an EC.

3.3 End-user App

We have developed an innovative, user-centric mobile application designed to empower end-users to initiate, deploy, operate, and monitor various ECs or P2P energy relations. This app enables users to identify a beneficial EC or P2P candidate to establish an energy-sharing relationship. It also helps in estimating the potential of establishing an EC or P2P with their preferred group of users. The application offers an intuitive interface where users can generate their QR codes and share with other user(s) they would like to share energy with or make an EC with, as shown in Figure 2(a). By scanning the QR codes, a reference to other users of the app is established. The QR code serves as an anonymous reference for the energy profiles used in the EC-Marketplace or EC-Optimizer to identify suitable partners.

To evaluate the feasibility of a potential EC or P2P relationship, a user can select other users and click on the 'Evaluate' button. When a user selects potential collaborators (individuals they are interested in forming an EC or engaging in P2P trading with), they initiate the data exchange process. When multiple potential participants in an energy community or P2P trading have been captured through the scanning of QR codes in the app, the list of anonymized IDs of potential candidates is transferred to the EC Marketplace. The unique identifier embedded within the QR code facilitates the linkage of datasets from these users at the EC marketplace. Subsequently, each end-user app instance (corresponding to a specific user) transfers relevant data, including energy consumption and generation profiles, to the EC marketplace. EC marketplace, using EC optimizer, evaluates various scenarios to help users determine the (i) optimal configurations in P2P scenarios and (ii) whether participation in an energy community would be more efficient.

In cases where no potential participants are transferred to the EC Marketplace, such as when a user does not scan any other QR code from other users, an independent evaluation is conducted to identify feasible EC or P2P relations and to determine the conditions under which participation in an energy community would be possible. The marketplace may present multiple unique energy communities for selection in this scenario. Consequently, the end-user app provides an innovative way to address the first step *'Identifying and Engaging Prospective EC Members'*. Furthermore, by facilitating the evaluation studies by providing the required HVD through EDDIE framework,

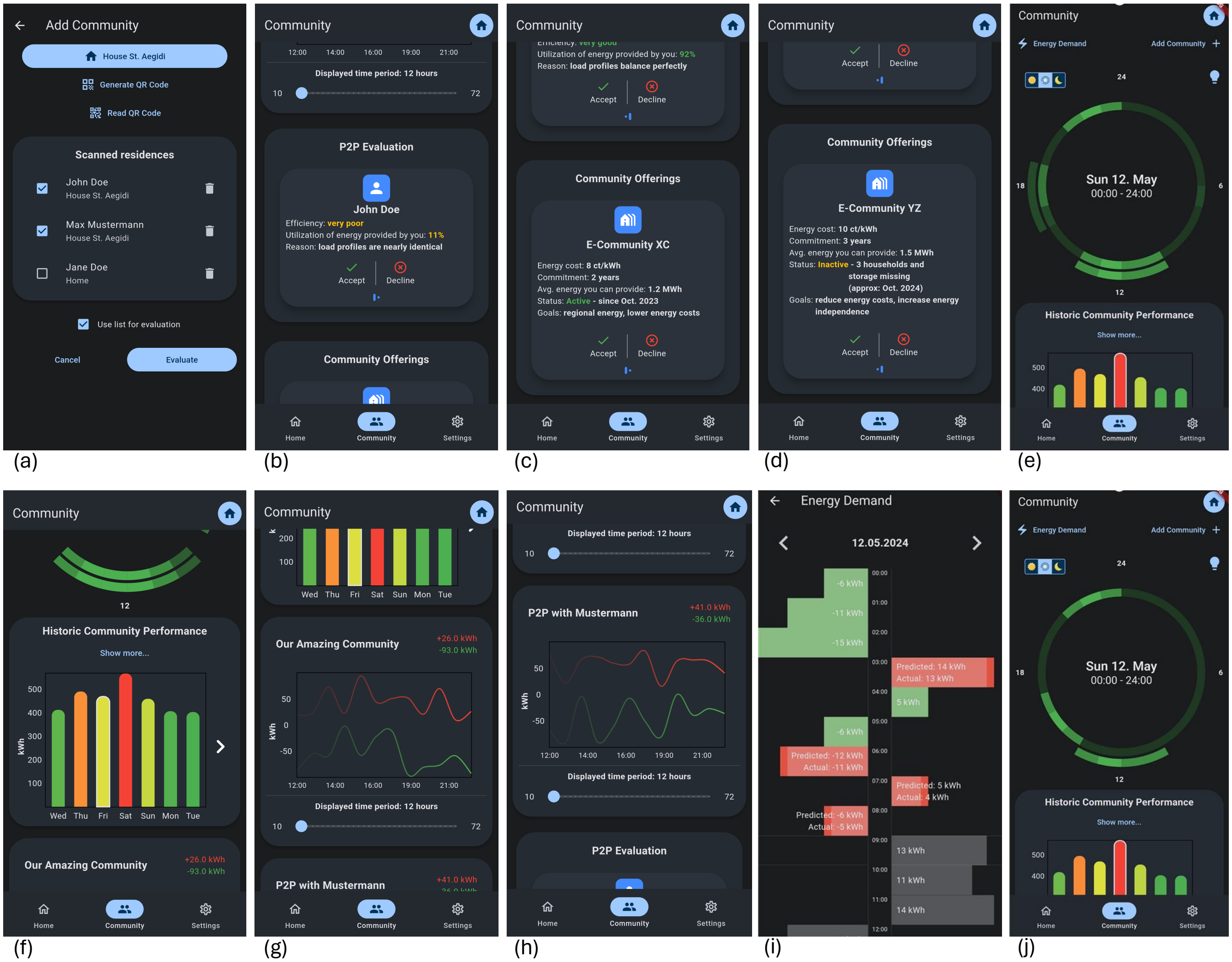


Fig. 2. Screenshots of the end-user mobile application for facilitating EC and P2P energy trading. (a) Interface for generating and scanning QR codes to form an energy community. (b) Results of the feasibility studies for P2P. (c) Details of an offer from functional EC for the user. (d) Identification of a planned EC. (e) Visual overview of personal demands vis-a-vis EC demands. (f) Historical EC performance of a week. (g) Monitoring of EC performance based on near real-time data. (h) Performance of P2P energy sharing based on near real-time data. (i) Interface for predicting demand and surplus energy by end-users and monitoring the accuracy of their predictions. (j) Example of the relation of the user with another EC where the symbiotic relation is more prominent.

it supports the third step *'Conducting Socio-Technical Feasibility Assessments for Prospective Energy Communities'*.

The results of these feasibility studies are shown to the users, as demonstrated in Figure 2(b), including whether the user would benefit from performing a P2P energy trade with the intended user and the reasoning behind the recommendation, which can pertain to societal, economic, regulatory, or technical constraints. Based on the data from users, EC marketplace gauges which offerings from existing ECs might be beneficial for them, as depicted in Figure 2(c). EC optimizer is responsible for identifying optimal configurations of participants in an energy community or in a P2P scenario.

Moreover, users can evaluate the scope of opportunities for an EC. For example, as shown in Figure 2(d), the energy community 'YZ' is inactive, and an asset for energy storage could benefit the EC. It refers to the scenario that the EC is planned and will be operational once enough suitable participants with desired assets have been identified and motivated to participate in the particular EC. Therefore, a user can upfront gauge benefits and consider installing a home storage solution accordingly.

The offers from prospective EC operators, based on the interactions on the EC marketplace, are communicated to the users through their app. In conjunction with the EC marketplace, the app provides an innovative method to fulfil the fifth step, *Negotiating and Finalizing Operational, Business, and Financial Models*. The user app offers functionalities to accept the negotiated terms of operation and grant permission for near real-time data access to the selected eligible party by the user. Furthermore, the app interacts with the regulatory authority to transmit the required information about the EC through the already-established link via the EDDIE. Therefore, the same app facilitates the prerequisite steps, *'Establishing Contractual Agreements and Regulatory Compliance'* (Step 6) and *'Deployment and Operationalization of the Energy Community* (Step 7)',.

Once the user is part of an EC or in a P2P energy-sharing relationship, they can monitor the performance of these interactions based on the real-time data received from the AIIDA and historical data fetched from EDDIE interfaces. One such user-friendly element is the donut chart across a time scale, as showcased in Figure 2(e), which conveys the user's energy needs (inner ring) performance vis-à-vis the energy demands of the EC (outer ring) over a period of 24 hours. Similarly, another EC with a relatively more prominent symbiotic relation with the user can be observed in Figure 2(j). This provides visual feedback to the users about the symbiotic relationship existing between the user and the community as a whole. Users can also visualize the overall energy demand overview of the preceding week as depicted in Figure 2(f).

Additionally, users can monitor the performance of their ECs based on near real-time data fetched through AIIDA. One can visualize the energy demand and generation in the EC or the P2P setup, as exhibited in Figure 2(g). This helps users to apprehend the available surplus energy in the community and the amount of energy they consumed at the same time. Similarly, they can visualize the performance of the P2P energy trade with other user(s), as shown in Figure 2(h). The sliding bar allows users to choose the granularity of data with respect to the time domain.

The user-centric approach of the proposed ecosystem is further emphasized by an additional functionality in the app that allows users to indicate their anticipated demand and generation. This feature enables users to actively participate in demand-side management by directly indicating their energy needs or surplus in the app. For instance, if a user anticipates an additional demand of 14 kWh from 03:00 to 04:00, they can indicate the availability of energy in the app, as shown in Figure 2(i). The EC operator can approve the additional demand or quote the actual amount of energy it can provide, here 13 kWh. This enables users to improve their forecasting accuracy over time. This feature enhances the overall user experience and promotes active engagement in ECs and P2P energy relations. The predicted changes that are approved by the EC operator are depicted with green color. Conversely, if a user expects to require less energy (e.g., 6 kWh at 5:00), they can indicate the upcoming demand in the end-user app. This bid for surplus or deficit energy is communicated to the AIIDA instance of the EC operator where the operator aggregates all bids from users. Through their respective AIIDA instances, users receive real-time information about feedback from EC operators. This transparency empowers all stakeholders of EC to make informed decisions. By directly participating in the bidding process, users actively shape the demand side management within the EC, fostering a more flexible and efficient system.

Therefore, the app further supports effective monitoring at the user level contributing towards the eighth step concerning continuous monitoring. By providing EC operators an interface for required data, it facilitates *'Continuous Monitoring, Optimization, and Billing Operations'*. This user-centric mobile application streamlines the process of forming and managing ECs, providing users with a convenient and accessible tool to engage in energy communities and contribute to the energy transition. By facilitating data sharing, cluster identification, and communication with EC operators, the app plays a crucial role in empowering end-users and promoting the growth of Energy Communities and P2P energy sharing.

3.4 Energy Community Marketplace

The end-user app and the EC marketplace work together in close coordination to streamline the formation and management of ECs. The EC marketplace functions as a platform to facilitate interactions between different feasible clusters and EC operators as service providers. The potential of prospective ECs can be calculated, and the identified clusters can be published on the EC Marketplace. EC operators participating in the EC marketplace can present their EC models and terms to the users of the identified cluster. The EC marketplace then acts as a broker between the users of the prospective EC cluster and willing EC operators, helping to negotiate offer terms, the business model of the EC, and the mode of operation, among other aspects. Once all stakeholders are finalized, the EC marketplace further facilitates the connection between the EC operator and the users' near real-time data access through AIIDA.

By providing a unified platform for interactions between users, clusters, and operators, the EC marketplace plays a vital role in promoting the growth and development of Energy Communities. Specifically, it heavily contributes to facilitating *'Identifying Relevant Stakeholders Based on User Clusters and Community Goals'* (Step 4), and *'Negotiating and Finalizing Operational, Business, and Financial Models'*. It simplifies the process of establishing and managing ECs, ensuring that users can easily find suitable operators. As the optimization approaches are very dynamic with the advancement of AI, the platform should be able to accommodate the multiple optimization models. Therefore, we have designed it such that there lies an opportunity to involve any third-party EC optimizers fostering innovation and efficiency in the complete ecosystem.

4 CONCLUSION

This publication presents a novel end-to-end ecosystem for facilitating the planning, deployment, and operation of energy communities (ECs) and peer-to-peer (P2P) energy relations. The typical prerequisite steps, outlined in Section 2, are systematically addressed by the diligently designed building blocks of the proposed solution, detailed in Section 3. The end-user mobile application developed serves as the cornerstone for the entire framework, enabling seamless communication between users and other stakeholders within the ecosystem. The EC marketplace acts as a pivotal platform, bridging the gap between prospective user clusters and EC operators. The ecosystem's modular architecture, comprising robust building blocks, ensures seamless interaction between all stakeholders. The integration of secure and GDPR-compliant data access components facilitates data-driven decision-making, while the user-centric design empowers end-users to actively initiate, negotiate, and monitor their involvement in energy communities. The ecosystem's modular architecture ensures streamlined interaction between the end-user, EC marketplace, and EC operator environments, fostering a more sustainable, democratized, and decentralized energy future.

ACKNOWLEDGMENTS

Co-funded by the European Union

This work is conducted under the ambit of EDDIE — 'European Distributed Data Infrastructure for Energy', co-funded by the European Union's Horizon Innovation Actions under grant agreement No. 101069510. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

Thomas Bauwens. 2016. Explaining the diversity of motivations behind community renewable energy. *Energy Policy* 93 (2016), 278–290. https://doi.org/10.1016/j.enpol.2016.03.017

Thomas Bauwens and Patrick Devine-Wright. 2018. Positive energies? An empirical study of community energy participation and attitudes to renewable energy. *Energy policy* 118 (2018), 612–625.

Mehmet Efe Biresselioglu, Siyami Alp Limoncuoglu, Muhittin Hakan Demir, Johannes Reichl, Katrin Burgstaller, Alessandro Sciullo, and Edoardo Ferrero. 2021. Legal Provisions and Market Conditions for Energy Communities in Austria, Germany, Greece, Italy, Spain, and Turkey: A Comparative Assessment. *Sustainability* 13, 20 (2021). https://doi.org/10.3390/su132011212

Directorate-General for Energy European Commission. 2019. *Clean energy for all Europeans*. Publications Office. https://doi.org/doi/10.2833/9937

Inês F.G. Reis, Ivo Gonçalves, Marta A.R. Lopes, and Carlos Henggeler Antunes. 2021. Business models for energy communities: A review of key issues and trends. *Renewable and Sustainable Energy Reviews* 144 (2021), 111013. https://doi.org/10.1016/j.rser.2021.111013

Vladimir Z. Gjorgievski, Snezana Cundeva, and George E. Georghiou. 2021. Social arrangements, technical designs and impacts of energy communities: A review. *Renewable Energy* 169 (2021), 1138–1156. https://doi.org/10.1016/j.renene.2021.01.078

Emi Minghui Gui and Iain MacGill. 2018. Typology of future clean energy communities: An exploratory structure, opportunities, and challenges. *Energy Research & Social Science* 35 (2018), 94–107. https://doi.org/10.1016/j.erss.2017.10.019 Energy and the Future.

Steven M Hoffman and Angela High-Pippert. 2010. From private lives to collective action: Recruitment and participation incentives for a community energy program. *Energy Policy* 38, 12 (2010), 7567–7574.

Gianpaolo Iazzolino, Nicola Sorrentino, Daniele Menniti, Anna Pinnarelli, Monica De Carolis, and Luca Mendicino. 2022. Energy communities and key features emerged from business models review. *Energy Policy* 165 (2022), 112929.

Vasileios Karagiannis, Shievam Kashyap, Nikolas Zechner, Oliver Hödl, Georg Hartner, Manuel Llorca, Tooraj Jamasb, Stefan Grünberger, Marc Kurz, Christoph Schaffer, and Stefan Schulte. 2023. A Framework for Enabling Cloud Services to Leverage Energy Data. In *2023 IEEE International Conference on Cloud Engineering (IC2E)*. 43–50. https://doi.org/10.1109/IC2E59103.2023.00013

Shievam Kashyap, Christoph Schaffer, Christoph Muck, and Christoph Plank. 2021. Intelligent Web-platform for Enabling Microgrids and Energy Sharing. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. 1–6. https://doi.org/10.1109/ISGTEurope52324.2021.9640134

J. Lowitzsch, C.E. Hoicka, and F.J. van Tulder. 2020. Renewable energy communities under the 2019 European Clean Energy Package – Governance model for the energy clusters of the future? *Renewable and Sustainable Energy Reviews* 122 (2020), 109489. https://doi.org/10.1016/j.rser.2019.109489

Miguel Manuel De Villena, Samy Aittahar, Sebastien Mathieu, Ioannis Boukas, Eric Vermeulen, and Damien Ernst. 2022. Financial Optimization of Renewable Energy Communities Through Optimal Allocation of Locally Generated Electricity. *IEEE Access* 10 (2022), 77571–77586. https://doi.org/10.1109/ACCESS.2022.3191804

Rogério Rocha, Ricardo Silva, João Mello, Sérgio Faria, Fábio Retorta, Clara Gouveia, and José Villar. 2023. A Three-Stage Model to Manage Energy Communities, Share Benefits and Provide Local Grid Services. *Energies* 16, 3 (2023). https://doi.org/10.3390/en16031143

Rossella Roversi, Andrea Boeri, Serena Pagliula, and Giulia Turci. 2022. Energy Community in Action—Energy Citizenship Contract as Tool for Climate Neutrality. *Smart Cities* 5, 1 (2022), 294–317. https://doi.org/10.3390/smartcities5010018

Gianluca Ruggieri, Rebecca Gambassi, Paolo Zangheri, Matteo Caldera, and Stefano F Verde. 2023. Key Economic Drivers Enabling Municipal Renewable Energy Communities' Benefits in the Italian Context. *Buildings* 13, 12 (2023), 2940.

Marianne Ryghaug, Tomas Moe Skjølsvold, and Sara Heidenreich. 2018. Creating energy citizenship through material participation. *Social studies of science* 48, 2 (2018), 283–303.

European Union. 2018. Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018L2001

European Union. 2019. Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L0944#:~:text=Directive%20(EU)%202019%2F944,(Text%20with%20EEA%20relevance.)

......

Poster Abstract: A Digital Twin Platform Applied to Hydrogen Electrolyzers

AMIT KUMAR SINGH, OFFIS – Institute for Information Technology, Energy Division, Germany
JELKE WIBBEKE, OFFIS – Institute for Information Technology, Energy Division, Germany
AMIN RAEISZADEH, OFFIS – Institute for Information Technology, Energy Division, Germany
NILS HUXOLL, OFFIS – Institute for Information Technology, Energy Division, Germany
MICHAEL BRAND, OFFIS – Institute for Information Technology, Energy Division, Germany

Digital Twin (DT) technology emerges as a promising solution for addressing the challenges in electrolyzer operation, resulting from scalability issues, component degradation, and limited lifespan. This paper presents a service-oriented DT architecture and platform applied to an industrial-scale hydrogen electrolyzer. By effectively mimicking the behavior of the physical object, the DT enhances predictive diagnosis and supervision of its operation. Initially, a generic DT framework is realized, followed by a demonstration of its application using an electrolyzer use case. The approach combines digital the object with services, facilitating comprehensive monitoring, fostering proactive maintenance, and performance enhancements. Condition monitoring and anomaly detection strategies are implemented to enhance real-time supervision. The overarching objective of this platform is to facilitate the creation of scalable, extendable, interoperable, and dynamic DTs for electrolyzers, thereby advancing their performance and efficiency in hydrogen production systems. Moreover, the DT architecture is platform-agnostic, allowing for the seamless extension to other use cases.

CCS Concepts: • **Software and its engineering** → *Software notations and tools*; • **Applied computing** → *Service-oriented architectures*; • **Computing methodologies** → *Modeling and simulation.*

Additional Key Words and Phrases: Renewable Energy, Hydrogen, Electrolyzer, Maintenance, Data-driven method, Machine Learning, Graphical User Interface

1 INTRODUCTION

Hydrogen is widely acknowledged as a pivotal component towards sustainable energy sources. The water electrolysis process for hydrogen production is highly esteemed within the hydrogen production sector, due to its straightforward operation, environmental cleanliness, and high purity of product [Zhang et al. 2024]. To produce hydrogen through the electrochemical process of electrolysis applied to water, a commonly employed device is known as an electrolyzer [Folgado et al. 2022]. The industrial-scale electrolyzer stands as a crucial component essential for the scalability of hydrogen production, exerting a profound influence on decreasing overall production costs. Moreover, electrolyzers are intricate devices, and integrating them into a system or process requires careful consideration. Conducting an analysis of their behavior is essential to understand their interaction with other components of the system. Additionally, the operation of an electrolyzer can be compromised by various factors like cell degradation, process temperature and fluctuation in input power. Thus, the expansion to hydrogen mass production necessitates electrolyzers that are efficient, resilient, cost-effective, and adaptable to scale [Federal Ministry of Education and Research 2021].

In this context, DT technology has emerged as a viable solution to address these challenges by strengthening the system's predictive diagnosis, and performance optimization. A DT embodies the virtual representation of a physical object (PO) or system, crafted to replicate its operation in real-time [Folgado et al. 2022]. The DT consists of a digital object (DO) mirroring the physical counterpart facilitating a collection of real-time measurements obtained from the PO. These real-time input measurements form the foundation of the DO, enabling it to accurately simulate the corresponding output values. Consequently, the DT mimics the real-time functionality of the PO virtually.

DTs can provide real-time situational and operational status data to a cyber-physical manufacturing system [Tao et al. 2018a]. This can significantly bolster the industrial system's intelligence in terms of performance optimization, predictive diagnosis, mitigation of unexpected events and analytical evaluation.

This paper demonstrates a service-oriented and event-driven DT platform applied to an industrial-scale hydrogen electrolyzer. The platform seamlessly integrates a DT of the electrolyzer, utilizing real-time measurements with a User Interface (UI) and corresponding services. The proposed platform considers the inherent variability of hydrogen electrolyzer and allows the utilization of both historical and real-time data to perform monitoring, estimation, and reliability of DT services.

2 DIGITAL TWIN FRAMEWORK

The DT architectural framework draws its inspiration from [Tao et al. 2018b], facilitating the 5-Dimensional DT concept. A DT, in its simplest form, can be associated to a virtual replica of a real PO. It represents the integration of a digital counterpart of a PO, functioning independently yet closely linked to its real-world counterpart. The depiction of the PO may take the shape of a DT, model, or shadow [Kritzinger et al. 2018]. The direction of data exchange between the digital and PO determines these different levels of integration. When the PO is in the planning stages and has yet to be realized, there is no data exchange, leading to what is termed a digital model. When the PO is present, and data unidirectionally moves from the physical to the DO, it is referred to as a digital shadow. Conversely, if data exchange occurs bidirectionally, enabling the DO to influence the physical one, it is categorized as a DT.

Authors' addresses: Amit Kumar Singh, amit.kumar.singh@offis.de, OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany; Jelke Wibbeke, jelke.wibbeke@offis.de, OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany; Amin Raeiszadeh, amin.raeiszadeh@offis.de, OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany; Nils Huxoll, nils.huxoll@offis.de, OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany; Michael Brand, michael.brand@offis.de, OFFIS – Institute for Information Technology, Energy Division, Escherweg 2, Oldenburg, Germany.

To effectively capture the PO and its behavior, the DT comprises of five essential components. According to the proposed framework, these components include the PO, the DO, the object data, the services, and the connections (CN) among them. The PO encompasses the complex equipment, functional subsystems, and sensory devices. During operation, the primary goal of the PO is to execute predetermined tasks, while sensory devices are responsible for gathering information on the states or malfunctions within the subsystem.

The DO functions digital representation of PO to mirror its behavior and dynamics. It contains models to depict the PO's functionality, which includes vital operations like state estimation, model calibration, or noise reduction from the measurement data. Furthermore, it facilitates object data storage capabilities ensuring seamless information storage within the DO, as illustrated in Fig. 1. The data storage encompasses one or more databases used to store various types of information within the digital twin, including measurement data, object metadata, model calibration parameters and service results.

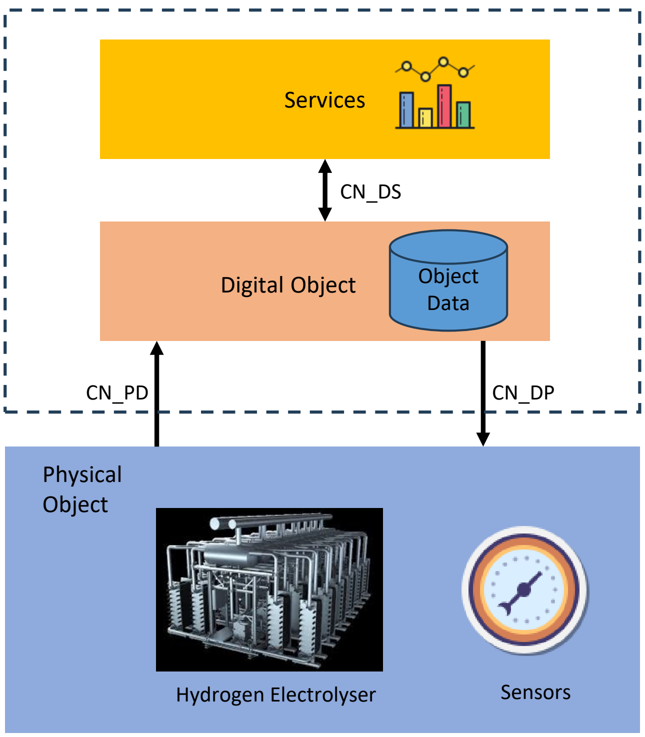


Fig. 1. Typical DT representation

Services are standardized applications that enhance the operations of the DO without having an impact on the digital representation. Common services include predictive maintenance, state estimation, anomaly detection, or monitoring. Consequently, even without specific functionalities tailored to a particular use case, a DT lacking all services remains a valid representation of the PO. The exchange of data between the physical and DO occurs in real-time. The connections CN_PD and CN_DP serve as data exchange linking the PO and DO, with CN_PD delivering measurement data and CN_DP conveying control actions. Additionally, the connection CN_DS links the DO with various services, delivering data required by the services and returning the results they produce. For instance, CN_DS may handle time-series data for anomaly detection service. The nature of the exchanged data is dependent on the specific use case. A practical application of the proposed DT framework utilizing a specific use case centered around a hydrogen electrolyzer is showcased in the subsequent section.

3 DIGITAL TWIN ARCHITECTURE FOR ELECTROLYSERS

The primary objective of the DO is confined to depicting the PO. Additional purposes or tasks of the DT are delegated to the supplementary services which are positioned external to the DO. As illustrated in Fig. 2 the DO encompasses a publish-subscribe communication interface, state estimator, service management, and data storage.

The publish-subscribe method's primary objective is to decouple the services and DO functionality enabling them to operate independently. Thus, it readily accesses the data upon request by the services and facilitates scalable, flexible, high-speed, high-volume, and high-fidelity communication between DO and supplementary services. The state estimator's role is to collect measurements and deduce any missing values that cannot be measured by the sensors, providing a comprehensive overview of the electrolyzer's condition. A central data storage is used for storing and analyzing time-series data. The platform efficiently ingests all data via the publish-subscribe mechanism, storing it in the database for historical record-keeping. This enables the utilization of historical data by services for various analysis and visualization purposes. Furthermore, the service management function within the DO monitors the operational status of all services. This includes operational metrics like CPU usage and service-dependent performance indicators like execution time.

Connected to the DO are real-time services which include condition monitoring and anomaly detection. The condition monitoring continuously monitors the electrolyzer's health in real-time, accurately predicting its performance under different conditions. Additionally, anomaly detection detects potential outliers and deviations in the measurement data. These proactive approaches enable timely maintenance and adjustments to enhance the electrolyzer's operational efficiency, supervision, and longevity. The final component of the services is the user interface (UI), realized with an emphasis on intuitive visualization of the services. The UI serves as a modular visualization of other services complementing the core functionality of DO. Through an interactive web application, it functions as a visualization tool for monitoring relevant DT states, including trends, estimation, condition monitoring, and anomalies, enabling comprehensive system supervision and tracking.

4 CURRENT IMPLEMENTATION

As an initial validation for the proposed DT platform, we designed the core DT architecture comprising of DO, and services interconnecting with the PO unidirectionally. Within the DO framework, a robust, scalable and elastic publish-subscribe method has been realized. These capabilities are fulfilled by Apache Kafka [Apache Software Foundation 2011], which captures data in the form of

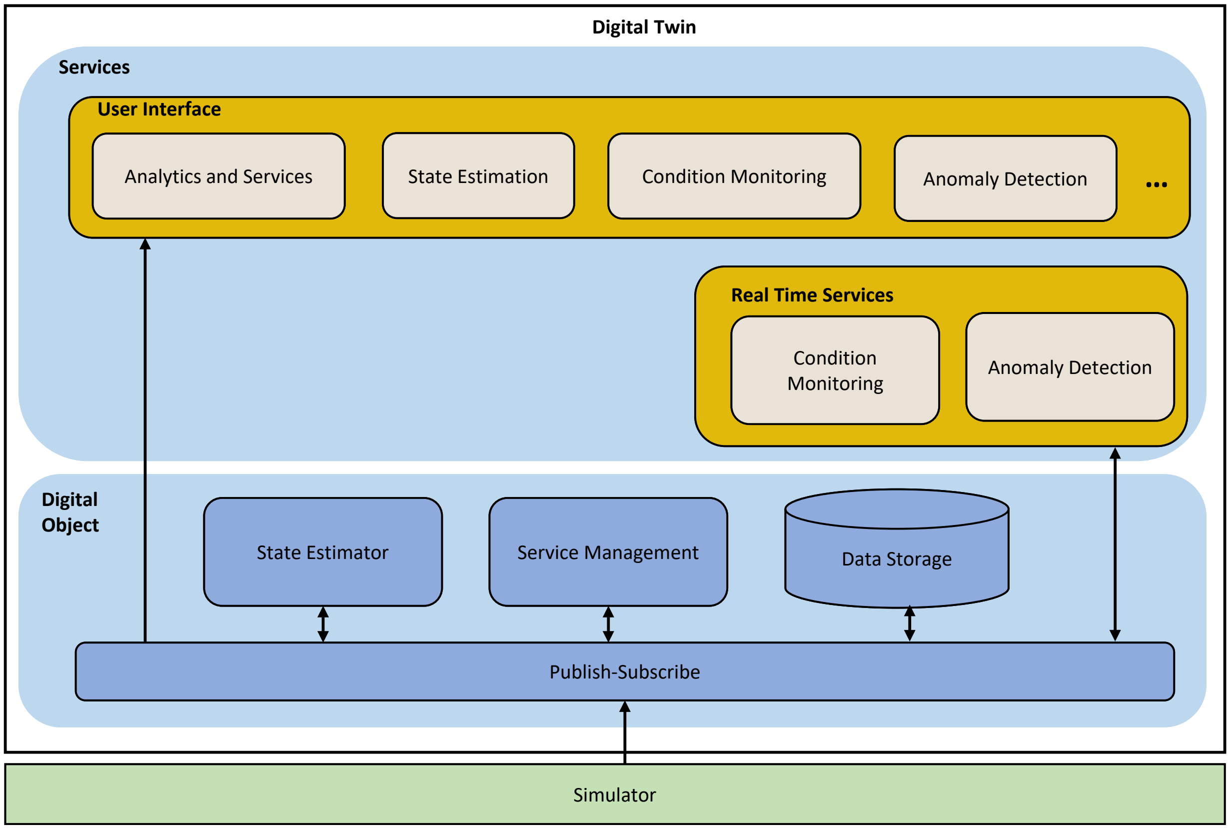


Fig. 2. DT platform for the hydrogen electrolyzer

events or event streams. Producers publish data into topics, while consumers read and process it accordingly. The electrolyzer serving as the PO, which is simulated for the scope of this study, is one of the producers. The data on the publish-subscribe infrastructure is stored in an InfluxDB database [InfluxData 2013] for historical record keeping. Moreover, the stored data can be queried by services using a predefined API.

In the demo, two real-time services make use of the publish-subscribe communication. First, the condition monitoring employs a pre-trained multi-layer perceptron model to consistently assess the health of the electrolyzer across diverse operational conditions. Second, a linear kalman filter is used as an algorithm for anomaly detection facilitating the mahalanobis distance.

TypeScript [Microsoft 2012] and Angular [Google 2016] are employed to craft dynamic and interactive UI components that offer customization and scalability based on services. The findings, illustrated in Fig. 3, serve as empirical evidence to deduce insights into various operational aspects of the electrolyzer.

The platform's components operate as containerized docker modules, providing not only enhanced flexibility and scalability for the DT core, but also enabling seamless addition, removal, or updating of services without interrupting its operation. Moreover, an automated software process has been integrated through continuous integration and continuous delivery. Thus, creating a DT platform that is fast, reliable, and modular using technologies like Docker, Kafka, and InfluxDB. The DT platform has been developed as a generic research and development platform utilizing an electrolyzer as a use case.

5 FUTURE WORK

As a part of future work, several avenues for research and development can be explored to enhance the capabilities and applications of the DT framework. We are currently progressing towards substituting the simulator with a real electrolyzer. Likewise, data structures based on the common information model (CIM) standard are tested to store the object data of the DO in addition to the InfluxDB for time series data. The transferability of DT technology to diverse systems, including power systems or industrial fleets could be explored.

The real-time services could be expanded to include advanced forecasting techniques to predict the state of health and remaining useful life of the electrolyzer. This could involve analyzing operational data and detecting degradation patterns, and thus enabling proactive maintenance strategies. Additionally, further development

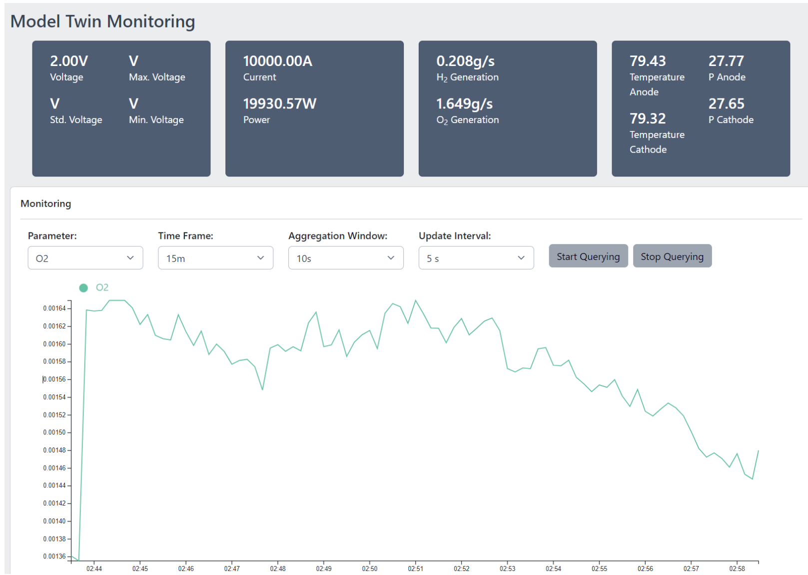


Fig. 3. UI of the DT platform

of existing services and the introduction of new ones can significantly enhance our knowledge and insights into electrolyzer behavior tailored to specific applications. For example, the introduction of Mosaik as a service into the ecosystem could enable co-simulation competence [Steinbrink et al. 2019].

6 CONCLUSION

This paper presents a service-oriented DT platform applied to an industrial-scale hydrogen electrolyzer. By seamlessly integrating the DO with UI and corresponding services, the platform facilitates comprehensive monitoring, estimation, and reliability assessment. This study makes use of the measurements, performs the core functionality of DO like state estimation, data storage, service management, and finally integrates it with supplementary services and UI. The real-time services like anomaly detection and condition monitoring algorithms are established to monitor and identify the electrolyzer's condition and abnormal behavior. By leveraging both historical and real-time data, the platform offers a robust DT framework for addressing the challenges of hydrogen production electrolyzers in the pursuit of a greener future.

ACKNOWLEDGMENTS

This research was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) within the H_2Giga Initiative under Grant Number 03HY122F.

REFERENCES

Apache Software Foundation. 2011. Apache Kafka. https://kafka.apache.org/. Accessed: Feb 1, 2023.

Federal Ministry of Education and Research. 2021. Wie das Leitprojekt H2Giga Elektrolyseure zur Wasserstoff-Herstellung in die Serienfertigung bringen will. https://www.wasserstoff-leitprojekte.de/projects/h2giga. Accessed: Jan 2, 2024.

Francisco Javier Folgado, Isaías González, and Antonio José Calderón. 2022. PEM electrolyser digital twin embedded within MATLAB-based graphical user interface. *Engineering Proceedings* 19, 1 (2022), 21.

Google. 2016. Angular. https://angular.dev/. Accessed: June 1, 2023.

InfluxData. 2013. https://www.influxdata.com. Accessed: June 2, 2023.

Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. 2018. Digital Twin in manufacturing: A categorical literature review and classification. *Ifac-PapersOnline* 51, 11 (2018), 1016–1022.

Microsoft. 2012. TypeScript. https://www.typescriptlang.org/. Accessed: June 2, 2023.

Cornelius Steinbrink, Marita Blank-Babazadeh, André El-Ama, Stefanie Holly, Bengt Lüers, Marvin Nebel-Wenner, Rebeca P. Ramírez Acosta, Thomas Raub, Jan Sören Schwarz, Sanja Stark, Astrid Nieße, and Sebastian Lehnhoff. 2019. CPES Testing with mosaik: Co-Simulation Planning, Execution and Analysis. *Applied Sciences* 9, 5 (2019). https://doi.org/10.3390/app9050923

Fei Tao, He Zhang, Ang Liu, and Andrew YC Nee. 2018a. Digital twin in industry: State-of-the-art. *IEEE Transactions on industrial informatics* 15, 4 (2018), 2405–2415.

Fei Tao, Meng Zhang, Yushan Liu, and Andrew YC Nee. 2018b. Digital twin driven prognostics and health management for complex equipment. *Cirp Annals* 67, 1 (2018), 169–172.

Zhe Zhang, Wenbo Yang, Liwei Wang, Tao Liang, and Zicong Liu. 2024. Construction and Simulation of Digital Twin Model and Electrolyzer Wide Power Adaptation Model for Alkaline Electrolytic Water Hydrogen Production. In *Proceedings of the 10th Hydrogen Technology Convention, Volume 1*. Springer Nature Singapore, Singapore, 281–290.

Poster Abstract: Peer-to-Peer Communication Using Enhanced German Smart Meter Gateway Infrastructure

EIKE NIEHS, JULIEN ESSERS, JOHANNES ROTHERT, and BERND ENGEL, elenia Institute for High Voltage Technology and Power Systems, TU Braunschweig, Germany

Additional Key Words and Phrases: smart grid, energy management, power grid

1 INTRODUCTION

Power grids are rapidly moving towards so-called smart grids. Smart grids describe the networking and control of different actors and components in the energy system using information and communication technology. Households are evolving from consumers to prosumers, as they no longer just consume electricity, but can also produce energy and feed it into the public grid. Additionally, coupling mobility and heating sectors, as well as the increasing number of battery storage systems, further increase the system complexity. The efficient monitoring and control of such systems requires the use of energy management systems. Recently, the German government accelerated the rollout of smart metering systems with the Act to Restart the Digitization of the Energy Transition [8]. The smart meter gateway (SMGW) was introduced to collect metering data within the local metrological network (LMN) and, to establish communication between controllable local systems (CLS) in the home area network (HAN) with external market participants (EMP) in the wide area network (WAN). Using smart metering systems for energy management purposes may potentially increase their value. However, the coordination of and communication between energy management systems using SMGW can be challenging.

We show how the communication scenarios in the German smart meter infrastructure can be used to enable peer-to-peer communication within the existing regulatory framework. In addition, we show hurdles and opportunities for the extension of the smart meter infrastructure.

There is a lot of existing research regarding energy management approaches on different levels in the energy system. Zafar et al. [9] provides a comprehensive overview of different concepts and configurations in home energy management systems. Celik et al. [1] present different coordination structures and objectives, including concepts for decentralized coordination. In contrast, there is limited work on the possibilities of implementing energy management concepts in the German smart meter infrastructure. Heimgaertner and Menth [5] discuss the implementation of a virtual power plant. Kroener et al. [6] show an implementation of a traffic light concept to take the network status into account.

2 GERMAN SMART METER INFRASTRUCTURE

Smart meters in Germany are highly standardized. Use cases and applications are defined on a high level within the stage model for the further development of standards for the digitalization of the energy transition [3] by the Federal Ministry for Economic Affairs and Climate Action with the involvement of further state actors and the industrial associations. They also define a set of relevant actors and roles within this infrastructure. Figure 1 shows the networks and actors associated with the SMGW of which we will subsequently describe the most relevant WAN and HAN actors for this work. The functional requirements for smart meters to implement the different use cases are defined by the Federal Office for Information Security in the Technical Guideline BSI TR-03109 [4].

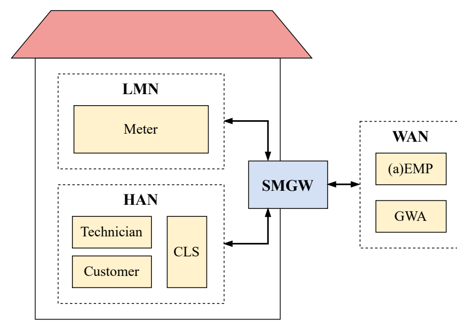


Fig. 1. Actors connected to Smart-Meter-Gateway.

The *Gateway-Administrator* (GWA) is responsible for the configuration, parameterization, and interaction with the SMGW. The GWA is the only actor in the WAN who can write and alter information on the SMGW, and the only one who can call remote procedures from the WAN. Despite this possibility, even the GWA is not able to actively initiate a connection to the SMGW as a client. Instead, the GWA offers certain endpoints for different kinds of communication, where the GWA acts as a server and the SMGW as a client.

External Market Participants (EMP) include all kinds of actors with a reasonable interest in relevant meter data. They can receive metering and tariff data sent by the SMGW via the *Info-Report*. The EMP becomes an *active External Market Participant* (aEMP) when the proxy functionality of the SMGW is used to communicate with controllable local systems, like EMS, in the Home-Area network. Therefore, the aEMP provides an additional server endpoint for the *CLS-Proxy*. After setting up this transparent communication proxy, the aEMP can send and receive arbitrary messages to and from the EMS in the HAN.

Controllable-Local-Systems (CLS) provide the communication endpoint in the HAN of the SMGW when an aEMP wants to interact with such components for controlling or monitoring purposes. The CLS functionality can be implemented by systems like control boxes, energy-management systems, or even electrical components such as inverters, or charging points. In difference to the proxy endpoint provided by the aEMP in the WAN, the CLS may act as both, server

Authors' Contact Information: Eike Niehs, e.niehs@tu-braunschweig.de; Julien Essers; Johannes Rothert; Bernd Engel, elenia Institute for High Voltage Technology and Power Systems, TU Braunschweig, Braunschweig, Germany.

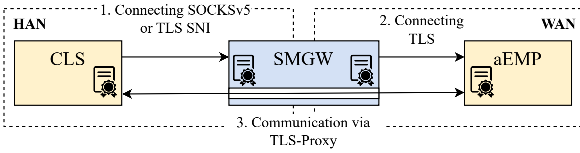


Fig. 2. HAN communication scenario 3.

or client, to establish a transparent communication proxy via the SMGW.

3 COMMUNICATION SCENARIOS

The *CLS-Proxy* as mentioned in Section 2 provides a transparent proxy functionality to establish general communication between aEMP in the WAN and CLS in the HAN. The proxy opens two independent TLS encrypted socket connections, both terminating at the SMGW and forwards messages between these two sockets. There are two relevant communication scenarios to establish the *CLS-Proxy*, HAN communication scenarios 3 and 4, which are introduced in the Technical Guideline [4]. These communication scenarios will then be used to establish peer-to-peer communication within the current technical regulatory framework.

3.1 Home-Area Network Communication Scenario 3

HAN communication scenario 3 (HCS3) as defined in [4] defines the initialization of a communication channel from a CLS in the HAN to an aEMP in the WAN. The CLS acts as a client and the SMGW as a server in the HAN. In the WAN, the SMGW acts as a client and the aEMP as a server. Both connections are established using TLS encrypted and client-authenticated socket communication. As shown in Figure 2 the CLS connects to the SMGW and requests a proxy to a certain predefined aEMP using SOCKSv5 or server name identification (SNI). If there is a proxy channel configured for this pair of CLS and aEMP, the SMGW acts as a client and establishes a connection to the aEMP. After both connections are successfully established, the CLS and aEMP can do arbitrary communication as long as the connection is alive.

3.2 Home-Area Network Communication Scenario 4

If the aEMP needs to communicate to an associated CLS in the HAN, the only way to establish the proxy is via the GWA using HAN communication scenario 4 (HCS4), shown in Figure 3. The aEMP requests a *CLS-Proxy* to a certain CLS from the GWA. The GWA sends the command for connecting the aEMP with the corresponding CLS via the management communication channel to the SMGW. The actual socket initiation involves two new sockets, each having the SMGW as client and CLS, resp. aEMP, as server.

3.3 Peer-to-Peer Communication

The German Federal Office for Information Security defines the security objectives and corresponding requirements within the Smart Meter Gateway Protection Profile (PP) [2]. The Gateway is considered the Target of Evaluation (TOE) which connects the WAN, LMN, and HAN while preserving defined security functionalities. One of the security objectives of the SMGW is the firewall functionality

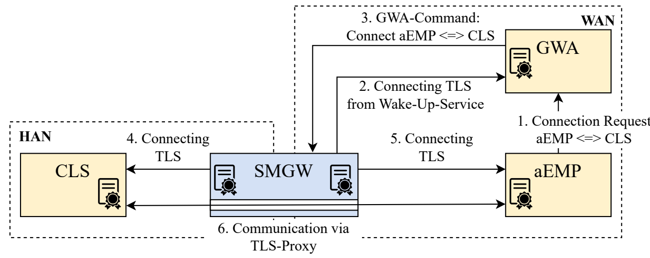


Fig. 3. HAN communication scenario 4.

to protect devices in the LMN and HAN, as well as itself against threats from the WAN. The PP states *"The firewall shall allow only connections established from HAN or the TOE itself to the WAN"* [2, p. 36]. The Act on Metering Point Operation and Data Communication in Smart Energy Networks (MsbG) §24 requires the general certification of communication units in an intelligent metering system in accordance with the PP [7]. Hence, it is not possible to establish a direct connection between two SMGWs within the current regulatory framework.

We introduce the following concept of peer-to-peer communication using the existing scenarios HCS3 and HCS4. Due to the limitations in the WAN, there is no client access to a server endpoint on the SMGW. This requires the introduction of an aEMP, which coordinates the communication as a proxy between the CLS'.

The concept for communication between two CLS, here *CLS:A* and *CLS:B* is shown in Figure 4. The upper part of the process shows the connection establishment from *CLS:A* to *aEMP* using HCS3. Using the established proxy on *SMGW:A*, the *CLS:A* requests a connection to *CLS:B* from *aEMP*. Since there is no active connection between *aEMP* and *CLS:B*, we use HCS4 to initiate the proxy by requesting it from the GWA. HCS4 triggers *SMGW:B* to open socket connections to *CLS:B* and *aEMP*. Finally, *CLS:A* and *CLS:B* can communicate with two SMGW and one aEMP in between. Where each intermediary terminates the connected TLS socket connections.

While this concept enables the communication between two CLS, the introduction of an aEMP forwarding messages adds a level of

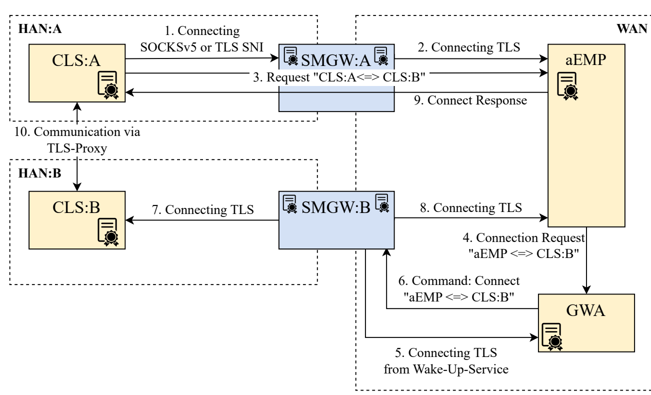


Fig. 4. Establishing communication between two controllable local systems using HCS3 and HCS4.

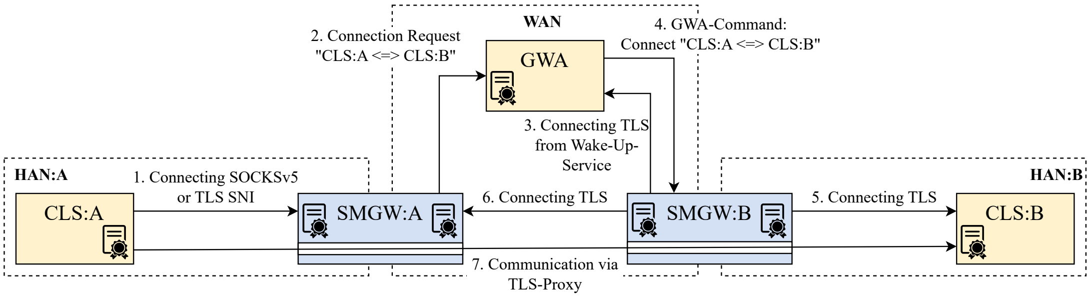


Fig. 5. Necessary extension to enable direct communication between different SMGW.

complexity and lowers the advantages of distributed systems. The aEMP is a central unit of possible failure. However, this concept may benefit from privacy improvements, because we can add a layer of end-to-end encryption on the application layer.

4 EXTENDING GERMAN SMART METER INFRASTRUCTURE

Figure 5 shows a possible future evolution to enable direct communication between two SMGWs. Instead of omitting the firewall at all, we are introducing a time-limited announcement phase in which the SMGW accepts a certain predefined external connection to be established from another SMGW device in the WAN.

After an initial connection and SOCKSv5 or SNI request from *CLS:A* to *SMGW:A*, *SMGW:A* sends a connection request to the *GWA* in the WAN. This command also changes the state of *SMGW:A* to the time-limited announcement phase for an incoming connection from *SMGW:B* which is the associated Gateway to *CLS:B*. The *GWA* uses the connection from *SMGW:B* to itself to send the connect command for *CLS:A* with *CLS:B*. Using a predefined communication profile configured in *SMGW:B*, the connection to *CLS:B* is established. Finally, step 6 establishes a connection from *SMGW:B* to *SMGW:A*. If the connection is in time, *CLS:A* and *CLS:B* can communicate via their connected SMGW, omitting the otherwise necessary aEMP.

4.1 Discussion

The process of establishing a connection between two CLS can be simplified compared to the description in accordance with the current regulatory framework in 3.3. Authentication when establishing TLS socket connections within the German smart meter infrastructure always uses both, server and client-side authentication. The authentication for *CLS-Proxy* communication scenarios is in general via a certificate. The SMGW must use certificates issued from the Smart Metering Public Key Infrastructure (SM-PKI) for communication in the WAN [4]. For each scenario, the certificates of the respective counterpart must be made known in advance. Implementing this extension requires adjustments in the Protection Profile as well as in the Technical Guideline. This requires the addition of a communication scenario where the SMGW acts as a server in the WAN. Since the current regulatory framework already requires certificate-based client authentication for the corresponding TLS connections, and the certificates used for WAN communication must be issued from within SM-PKI, it should be audited to include such a communication scenario.

ACKNOWLEDGMENTS

Funded by the German Ministry for Economic Affairs and Climate Action and the Projektträger Jülich within the research project KEMAL under grant number 03EI6064C.

REFERENCES

[1] Berk Celik, Robin Roche, Siddharth Suryanarayanan, David Bouquain, and Abdellatif Miraoui. 2017. Electric energy management in residential areas through coordination of multiple smart homes. *Renewable and Sustainable Energy Reviews* 80 (2017), 260–275. https://doi.org/10.1016/j.rser.2017.05.118
[2] Federal Office for Information Security (BSI). 2014. Protection Profile for the Gateway of a Smart Metering Systems (Smart Meter Gateway PP).
[3] Federal Office for Information Security (BSI). 2021. Stage model for the further development of standards for the digitalisation of the energy transition.
[4] Federal Office for Information Security (BSI). 2021. Technical Guideline BSI TR-03109.
[5] Florian Heimgaertner and Michael Menth. 2018. Distributed Controller Communication in Virtual Power Plants Using Smart Meter Gateways. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. https://doi.org/10.1109/ICE.2018.8436311
[6] Nils Kroener, Kevin Förderer, Manuel Lösch, and Hartmut Schmeck. 2020. State-of-the-Art Integration of Decentralized Energy Management Systems into the German Smart Meter Gateway Infrastructure. *Applied Sciences* 10 (2020). https://doi.org/10.3390/app10113665
[7] Parliament of the Federal Republic of Germany. 2023. Act on Metering Point Operation and Data Communication in Smart Energy Networks (MsbG). BGBl. I S. 2034.
[8] Parliament of the Federal Republic of Germany. 2023. Act to Restart the Digitization of the Energy Transition. BGBl. 2023 I No. 133.
[9] Usman Zafar, Sertac Bayhan, and Antonio Sanfilippo. 2020. Home Energy Management System Concepts, Configurations, and Technologies for the Smart Grid. *IEEE Access* 8 (2020), 119271–119286. https://doi.org/10.1109/ACCESS.2020.3005244

Poster Abstract: Forecasting and Optimization as a Service for Energy Management Applications with the Energy Service Generics Framework

DAVID WÖLFLE, FZI Research Center for Information Technology, Germany
HARTMUT SCHMECK, Karlsruhe Institute of Technology FZI Research Center for Information Technology, Germany

Energy management, in sense of computing optimized operation schedules for devices, will likely play a vital role in future carbon neutral energy systems, as it allows unlocking energy efficiency and flexibility potentials. However, energy management systems need to be applied at large scales to realize the desired effect, which clearly requires minimization of costs for setup and operation of the individual applications. In order to push the latter forward, we promote an approach to split the complex optimization algorithms employed by energy management systems into standardized components, which can be provided as a service with marginal costs at scale. In this work we contribute a practical example how such services can be implemented utilizing the *Energy Service Generics* framework, a software explicitly designed to allow scientists and practitioners to derive fully functional web services from forecasting or optimization algorithms. To this end the implementation of typical components of forecasting or optimization services for energy management applications is discussed based on functional source code listings of a simple photovoltaic (PV) power generation forecast service.

CCS Concepts: • **Software and its engineering → Development frameworks and environments**.

Additional Key Words and Phrases: building energy management, building control, model predictive control, framework, forecast, smart building

1 INTRODUCTION

Global scale efforts are required to mitigate the most severe consequences of climate change including a significant increase in energy efficiency of consumers as well as the decarbonization of energy supply [5]. The vast utilization of renewable energy sources required for the latter will additionally likely induce an increased demand for energy flexibility by consumers [1, 6, 9]. Energy Management Systems (EMSs), in a sense of software computing optimized operational schedules and executing these on devices, have been demonstrated to be capable of reducing energy demand, lowering CO_2 emissions and/or unlocking flexibility [2, 3, 8, 10, 11]. However, energy management solutions will be required at scale in order to achieve the desperately needed global impact, like e.g. applied to thousands of buildings.

Forecasting and optimization algorithms are essential parts of EMSs and have traditionally been developed for a single specific target (e.g. for one particular building) [14], like in [2, 3, 8, 10, 11] or the publications reviewed by [12]. Usually, an optimized schedule for the operation of the target is the desired output of the optimization algorithm which may require one or more forecasts to compute the optimized strategy. For example, in order to compute the optimized operation strategy for a battery storage system in a private household with local energy production from a PV plant in combination with a dynamic electricity price, the optimization requires predictions of future development of the energy price, the electric load, and the power generation of the PV system.

Economic viability is certainly a key factor for the widespread adoption of EMSs, in particular as it has been shown that the development costs of target specific EMSs are higher than the monetary savings, even for medium sized commercial buildings [4]. This becomes even more obvious, if one considers the complexity of the EMS required for optimizing the battery storage system outlined in the example above and compares it to the typical energy costs of private households.

A potential way out of this dilemma, as promoted in a recent paper [13], is to split the complex optimization procedure into standardized parts and provide these parts as a service. In this context, a service refers to a web-based program that provides a limited functionality required for energy management applications via a standardized interface. Regarding the example above one could consider four dedicated services: The first for computing load forecasts. The second for computing the PV power generation forecast. The third service provides the electricity price forecast and finally the fourth service, computing the optimized schedule for the battery, based on the former three forecasts as input. This service concept allows the replacement of custom tailored optimization algorithms with standardized services, which can be utilized by potentially thousands of EMSs, thus enabling a dramatic cost reduction especially in combination with highly automated Machine Learning (ML) approaches.

This paper aims at demonstrating how the recently published Energy Service Generics (ESG) open source framework[1] can simplify the provisioning of forecasting and optimization algorithms for energy management. We contribute a practical example how a fully functional service can be implemented using the ESG framework. In this way, we intend to aid the widespread adoption of EMSs at scale.

2 A SIMPLE PV FORECAST SERVICE EXAMPLE

In order to illustrate the intended usage and provided functionality of the ESG framework this section presents an example of a PV power generation forecast service, i.e. a program that allows users to retrieve predictions of future power production of PV systems by interacting with an Application Programming Interface (API). The following subsections present and discuss the implementation of the components of the service. More details and reasoning about the design concept on which the components are based on are provided in Section 3 of [13]. It is worth noting that the code listings are excerpts, i.e. reduced to the essential parts regarding the scope of this work. The complete code is provided in the ESG repository[2].

Authors' addresses: David Wölfle, FZI Research Center for Information Technology, Karlsruhe, Germany, woelfle@fzi.de; Hartmut Schmeck, Karlsruhe Institute of Technology and FZI Research Center for Information Technology, Karlsruhe, Germany, hartmut.schmeck@kit.edu.

[1]https://github.com/fzi-forschungszentrum-informatik/energy-service-generics
[2]https://github.com/fzi-forschungszentrum-informatik/energy-service-generics/tree/main/docs/examples/basic_example

2.1 The Forecasting or Optimization Code

The forecasting or optimization code is the payload of the service, i.e. the goal of applying the ESG framework is to make this component accessible to users. The ESG framework has been designed to make integration of existing forecasting or optimization code simple, which is demonstrated in the example by utilizing the popular pvlib[3] for computing the PV power production forecast. In order to allow pvlib to compute the forecasts it is necessary to provide the corresponding input data to the library. The first part of this input data is the specification of the PV system for which the forecast should be computed. We assume that the PV system is sufficiently described by the geographic position, i.e. latitude and longitude, as well as the geometry of the PV system, i.e. azimuth and inclination, and the peak power. All other options to describe the PV system offered by pvlib[4] are neglected to make this example not more complex than necessary. The second part of the required input to compute PV power prediction consists of meteorological forecast data, especially forecasts of solar irradiance.

Before we proceed with the discussion about the handling of input data by the ESG framework it is worth to recall that the framework supports two types of endpoints, these are `/request/` and `/fit-parameters/`. This differentiation arises from the requirement that the ESG framework should allow forecasting or optimization code that uses ML approaches, which usually requires that some parameters of a model must be fitted utilizing observations of the target system, see Section 3.1 in [13] for details. Regarding the example above we assume that the service is intended to produce PV power generation forecasts for systems for which geometry and peak power values may be unknown and need to be estimated from power production measurements. For the purpose of this example this is implemented with a simple least squares approach in the code below. Thus, the input data necessary to obtain a forecast is separated into two groups: latitude and longitude are *arguments* while azimuth, inclination, and peak power are *parameters*. It is worth noting that parameters are not necessarily interpretable as in this example, e.g. weights and bias terms of a neural network could be parameters too. Finally, it should be considered that it may not be reasonable to demand all input data as user input. In the present example the service fetches the meteorological data automatically from a third party web service, which would in practice make the interaction with the service more convenient and less error prone for the user. Finally, the following points concerning the code listing below should be noted:

(1) The format of `input_data` and `output_data` is implicitly defined in the corresponding data models which are introduced in the following section.
(2) The functions `predict_pv_power`, `fetch_meteo_data` as well as `fit_with_least_squares` have been omitted from the listing, as the practical implementation details of those are not of particular interest for the scope of this work. However, the code of the omitted functions can be found in the repository of the ESG framework.
(3) Implementing `fit_parameters` is optional and can be omitted for services without fitable parameters. An example without fitable parameters could be a service wrapping the AutoPV algorithm proposed in [7].

```python
from esg.utils.pandas import series_from_value_message_list
from esg.utils.pandas import value_message_list_from_series

def handle_request(input_data):
    arguments = input_data.arguments
    parameters = input_data.parameters
    meteo_data = fetch_meteo_data(
        lat=arguments.geographic_position.latitude,
        lon=arguments.geographic_position.longitude,
    )
    pv_power = predict_pv_power(
        lat=arguments.geographic_position.latitude,
        lon=arguments.geographic_position.longitude,
        azimuth=parameters.pv_system.azimuth_angle,
        inclination=parameters.pv_system.inclination_angle,
        peak_power=parameters.pv_system.nominal_power,
        meteo_data=meteo_data,
    )
    output_data = {
        "power_prediction": value_message_list_from_series(
            pv_power
        )
    }
    return output_data

def fit_parameters(input_data):
    arguments = input_data.arguments
    measured_power = series_from_value_message_list(
        input_data.observations.measured_power
    )
    meteo_data = fetch_meteo_data(
        lat=arguments.geographic_position.latitude,
        lon=arguments.geographic_position.longitude,
        past_days=90,
    )
    measured_power = series_from_value_message_list(
        input_data.observations.measured_power
    )
    fitted_pv_system = fit_with_least_squares(
        lat=arguments.geographic_position.latitude,
        lon=arguments.geographic_position.longitude,
        meteo_data=meteo_data,
        measured_power=measured_power,
    )
    return fitted_pv_system
```

2.2 The Data Model

In addition to the forecasting or optimization code introduced above, the data model is the second component that is service specific. The data models define the format of the data the user exchanges with the service. For a service without fitable parameters, i.e. a service with `/request/` endpoints only, it is sufficient to define the arguments required for computing the request as well as the result of the computation. The corresponding data models are called `RequestArguments`

[3] https://github.com/pvlib/pvlib-python
[4] https://pvlib-python.readthedocs.io/en/stable/user_guide/modelchain.html

and `RequestOutput`. In case of a service with fitable parameters it is additionally necessary to define the data format for the input and output data for the `/fit-parameters/` endpoints. The data models specifying the input for the fitting process are referred to as `FitParameterArguments` and `Observations` and the corresponding output is `FittedParameters`. As the simple PV power generation forecast service used as example is designed to provide functionality to fit parameters, it is necessary to define all five data models introduced above. The corresponding implementation is shown in the listing below.

```
from esg.models.base import _BaseModel
from esg.models.datapoint import ValueMessageList
from esg.models.metadata import GeographicPosition, PVSystem
from pydantic import Field

class RequestArguments(_BaseModel):
    geographic_position: GeographicPosition

class RequestOutput(_BaseModel):
    power_prediction: ValueMessageList = Field(
        description="Prediction of power production in W"
    )

class FittedParameters(_BaseModel):
    pv_system: PVSystem

class Observations(_BaseModel):
    measured_power: ValueMessageList = Field(
        description="Measured power production in W"
    )

class FitParameterArguments(_BaseModel):
    geographic_position: GeographicPosition
```

At this point it is worth noting that the ESG package provides ready-to-use building blocks for data models. For example, in the code above `GeographicPosition` is imported from ESG. The former is a data model too, which defines that a geographic position consists of latitude and longitude as well as an optional height above ground. Furthermore it is crucial to note that the data models serve additional functionality beyond the definition of the data format. Particularly important is the provisioning of documentation of the format in human readable form (e.g. the `description` in the example above) as well as defining valid ranges for values. An example for the latter could be to enforce that values for latitude must be in the range of $-90° \ldots 90°$. This rule is in fact implemented in `GeographicPosition`, although this is not directly visible in the code example above. However, it can be perceived by inspecting the source code of the `GeographicPosition` data model[5].

2.3 The Worker

The worker component is responsible for executing the tasks, i.e. computing requests or fitting parameters by invoking the forecasting or optimization code, as well as task scheduling. Further details are provided in Section 3.2 and 3.3 of [13]. The ESG framework utilizes the Celery library[6] for implementing the worker, but extends the latter with functionality to make the implementation of services more convenient, for example by utilizing the data models for de-/serialization of input and output data. Thus, the main objective for implementing a worker is to wire up the data models with the forecasting or optimization code, which should usually require a rather simple code as displayed in the listing below for the PV power generation forecast example service.

```
from esg.service.worker import celery_app_from_environ
from esg.service.worker import invoke_fit_parameters
from esg.service.worker import invoke_handle_request

from data_model import RequestArguments, RequestOutput
from data_model import FittedParameters, Observations
from data_model import FitParameterArguments
from fooc import fit_parameters, handle_request

app = celery_app_from_environ()

@app.task
def request_task(input_data_json):
    return invoke_handle_request(
        input_data_json=input_data_json,
        RequestArguments=RequestArguments,
        FittedParameters=FittedParameters,
        handle_request_function=handle_request,
        RequestOutput=RequestOutput,
    )

@app.task
def fit_parameters_task(input_data_json):
    return invoke_fit_parameters(
        input_data_json=input_data_json,
        FitParameterArguments=FitParameterArguments,
        Observations=Observations,
        fit_parameters_function=fit_parameters,
        FittedParameters=FittedParameters,
    )
```

2.4 The API

The API component connects the worker with the user by allowing the latter to trigger the computation of requests or fitting of parameters as well as retrieving the corresponding results. To this end the API component has to check the user input for validity and create the computation tasks. Furthermore, the API component handles authentication and authorization of users. More details about the API design are provided in Section 3.1 of [13].

The implementation of the API component is available ready-to-use in the ESG framework. However, in order to operate the API it is necessary, similar to the worker, to wire up the API with the other components, in particular with the data model and the worker. Furthermore, some information like name and version number must be provided too. Nevertheless, the necessary code to instantiate an API component is trivially simple and shown in the following listing:

[5] https://github.com/fzi-forschungszentrum-informatik/energy-service-generics/blob/main/source/esg/models/metadata.py

[6] https://docs.celeryq.dev/

```
from esg.service.api import API

from data_model import RequestArguments, RequestOutput
from data_model import FittedParameters, Observations
from data_model import FitParameterArguments
from worker import request_task, fit_parameters_task

api = API(
    RequestArguments=RequestArguments,
    RequestOutput=RequestOutput,
    FittedParameters=FittedParameters,
    Observations=Observations,
    FitParameterArguments=FitParameterArguments,
    request_task=request_task,
    fit_parameters_task=fit_parameters_task,
    title="PV Power Prediction Example Service",
    version=Version("0.0.1"),
)

if __name__ == "__main__":
    api.run()
```

2.5 The Service

Following the operation concept for services (as given in Section 3.4 of [13]) the last remaining step to derive functional services is to build docker images that can be run, e.g. on Kubernetes. It is necessary to build two distinct images, one for the API (which includes the data model) and one for the worker (which includes the data model and the forecasting or optimization code). The following Dockerfile[7] (i.e. build instructions) is suitable to build a functional image for the API:

```
FROM energy-service-generics:latest-service
COPY api.py data_model.py worker.py /source/service/
CMD ["/source/service/api.py"]
```

The corresponding Dockerfile for the worker is:

```
FROM energy-service-generics:latest-service-pandas
RUN pip install pvlib scipy
COPY data_model.py fooc.py worker.py /source/service/
ENTRYPOINT [ "celery" ]
CMD [ "--app", "worker", "worker", "--loglevel=INFO" ]
```

3 CONCLUSION

The provisioning of forecasting or optimization components as web services could support the adoption of EMSs at scale. One possibility to derive such services is the utilization of the recently published Energy Service Generics (ESG) framework. In this paper we have provided a practical example how a simple PV power generation forecast service can be realized using the ESG framework. To this end we have discussed and demonstrated the necessary steps to implement the four components of the service, i.e. forecasting or optimization code, data model, worker, and API, thus demonstrating the practical usability of the ESG framework. The value of the framework can be perceived by realizing how short the code listings above are. Besides the forecasting or optimization code, which must be implemented anyway, deriving the functional example service required less than 100 lines of simple to implement Python code.

[7]https://docs.docker.com/reference/dockerfile/

ACKNOWLEDGMENTS

This research has partly been funded by the German Federal Ministry for Economic Affairs and Climate Action within the project FlexBlue.

REFERENCES

[1] M.I. Alizadeh, M. Parsa Moghaddam, N. Amjady, P. Siano, and M.K. Sheikh-El-Eslami. 2016. Flexibility in Future Power Systems with High Renewable Penetration: A Review. *Renewable and Sustainable Energy Reviews* 57 (May 2016), 1186–1193. https://doi.org/10.1016/j.rser.2015.12.200

[2] Bingqing Chen, Zicheng Cai, and Mario Bergés. 2019. Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19)*. Association for Computing Machinery, New York, NY, USA, 316–325. https://doi.org/10.1145/3360322.3360849

[3] Xianzhong Ding, Wan Du, and Alberto Cerpa. 2019. OCTOPUS: Deep Reinforcement Learning for Holistic Smart Building Control. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19)*. Association for Computing Machinery, New York, NY, USA, 326–335. https://doi.org/10.1145/3360322.3360857

[4] Markus Gwerder, Dimitrios Gyalistras, Carina Sagerschnig, Roy S Smith, and David Sturzenegger. 2013. *Final Report: Use of Weather And Occupancy Forecasts For Optimal Building Climate Control Part II: Demonstration (OptiControl-II)*. Technical Report. ETH Zürich.

[5] IPCC. 2022. Summary for Policymakers. In *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1 ed.), P.R. Shukla, J. Skea, A. Reisinger, R. Slade, R. Fradera, M. Pathak, A. Al Khourdajie, M. Belkacemi, R. van Diemen, A. Hasija, G. Lisboa, S. Luz, J. Malley, D. McCollum, S. Some, and P. Vyas (Eds.). Cambridge University Press, 3–48.

[6] Hendrik Kondziella and Thomas Bruckner. 2016. Flexibility Requirements of Renewable Energy Based Electricity Systems – a Review of Research Results and Methodologies. *Renewable and Sustainable Energy Reviews* 53 (Jan. 2016), 10–22. https://doi.org/10.1016/j.rser.2015.07.199

[7] Stefan Meisenbacher, Benedikt Heidrich, Tim Martin, Ralf Mikut, and Veit Hagenmeyer. 2023. AutoPV: Automated Photovoltaic Forecasts with Limited Information Using an Ensemble of Pre-Trained Models. In *Proceedings of the 14th ACM International Conference on Future Energy Systems*. ACM, Orlando FL USA, 386–414. https://doi.org/10.1145/3575813.3597348

[8] Frauke Oldewurtel, Alessandra Parisio, Colin N. Jones, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, and Manfred Morari. 2012. Use of Model Predictive Control and Weather Forecasts for Energy Efficient Building Climate Control. *Energy and Buildings* 45 (Feb. 2012), 15–27. https://doi.org/10.1016/j.enbuild.2011.09.022

[9] G. Papaefthymiou and Ken Dragoon. 2016. Towards 100% Renewable Energy Systems: Uncapping Power System Flexibility. *Energy Policy* 92 (May 2016), 69–82. https://doi.org/10.1016/j.enpol.2016.01.025

[10] Jyri Salpakari and Peter Lund. 2016. Optimal and Rule-Based Control Strategies for Energy Flexibility in Buildings with PV. *Applied Energy* 161 (Jan. 2016), 425–436. https://doi.org/10.1016/j.apenergy.2015.10.036

[11] Luigi Schibuola, Massimiliano Scarpa, and Chiara Tambani. 2015. Demand Response Management by Means of Heat Pumps Controlled via Real Time Pricing. *Energy and Buildings* 90 (March 2015), 15–28. https://doi.org/10.1016/j.enbuild.2014.12.047

[12] Pervez Hameed Shaikh, Nursyarizal Bin Mohd Nor, Perumal Nallagownden, Irraivan Elamvazuthi, and Taib Ibrahim. 2014. A Review on Optimized Control Systems for Building Energy and Comfort Management of Smart Sustainable Buildings. *Renewable and Sustainable Energy Reviews* 34 (June 2014), 409–429. https://doi.org/10.1016/j.rser.2014.03.027

[13] David Wölfle, Kevin Förderer, Tobias Riedel, Lukas Landwich, Ralf Mikut, Veit Hagenmeyer, and Hartmut Schmeck. 2024. Open Energy Services – Forecasting and Optimization as a Service for Energy Management Applications at Scale. https://doi.org/10.48550/ARXIV.2402.15230 arXiv:2402.15230

[14] David Wölfle, Arun Vishwanath, and Hartmut Schmeck. 2020. A Guide for the Design of Benchmark Environments for Building Energy Optimization. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '20)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3408308.3427614

Poster Abstract: Interactive Simulation of Swiss Energy and Environment Designed for Gaming and Choice Experiments

TOBY SIMPSON, MICHAEL MULTERER, and ROLF KRAUSE, Università della Svizzera italiana, Switzerland

CCS Concepts: • **General and reference**; • **Applied computing** → **Law, social and behavioral sciences**;

Additional Key Words and Phrases: energy transition, energy balance, choice experiment, serious gaming, research computer game

Availability of Data and Material:
The web interface for the model can be found at https://sure.euler.usi.ch

1 INTRODUCTION

1.1 Overview

We introduce a simple, low-latency model for the simulation and analysis of energy transition pathways in Switzerland. It is designed as the calculation and data storage layer of the "Ensured Energy" game (Figure 1), developed as part of the SFOE SWEET-SURE project.

The model presents users with the current and historic national energy balance of Switzerland based on BFE data [Bundesamt für Energie 2022a] along with future trends for energy consumption. Users can allocate capacity to different energy technologies in order to meet energy requirements. In this way they explore various pathways for energy transition. The model uses the resulting energy system to generate a set of indicators which assess the economic and environmental impact of their decisions. The simulation also allows changes to both the supply and demand side of the energy economy, thus testing resilience to various shocks. All instances of the model are recorded in a database and cross-referenced with survey data for further analysis, which in turn is used to inform policy makers as part of the SWEET-SURE project objectives.

1.2 Interactive Models

In the evaluation of energy transition pathways [Panos et al. 2023], decisions of individual citizens act directly through consumer preferences and indirectly as political choices. We gain understanding of citizen behaviour by creating an interactive environment for problem-solving which is engaging and informative, and embeds choice experiments to acquire survey information.

The use of games is recognised in business and the public sector [Hamari et al. 2014] for promoting favorable outcomes, such as increasing motivation [Cechanowicz et al. 2013; Sailer et al. 2017] and changing behaviour [Kazhamiakin et al. 2015; Schiele 2018]. When games are used as a tool for surveys [Harms et al. 2015], self-induced feedback results in rich data that allow for complex analysis, comparing favourably with traditional methods [Aubert and Lienert 2018]. Creative user input may provide new solutions, as in [Cooper et al. 2010] where players discovered new protein structures.

Authors' address: Toby Simpson, toby.simpson@usi.ch; Michael Multerer, michael.multerer@usi.ch; Rolf Krause, rolf.krause@usi.ch, Università della Svizzera italiana, Via Buffi 13, Lugano, Switzerland, 6900.

Fig. 1. Screenshot from *Ensured Energy* game [Jones and Martinez Oltra 2024].

1.3 SWEET-SURE Programme

The Swiss Energy research for the Energy Transition (SWEET) is a funding programme of the Bundesamt für Energie (BFE) that promotes innovation to support Switzerland's energy and climate strategy for 2050. The Sustainable and Resilient Energy for Switzerland (SURE) project considers pathways for the transition of the energy system towards a higher share of renewables, making an integrated assessment of sustainability, resilience, environmental resources, economics and public health.

A collection of quantitative models and data-driven approaches generate transition pathways [Panos et al. 2023], including disruptive future events and use them to inform stakeholder decisions, which in turn are tested for their resilience and recovery. Ultimately the project provides information to policy makers, technology developers and businesses on the design of their respective strategies towards a more sustainable and resilient energy future.

1.4 Objectives

We complement the large interdisciplinary studies of the SURE project by providing a simplified version of the problem of energy transition and resilience in the form of an interactive web-based model. The simplicity of the model is key for two reasons: firstly it must be accessible to non-specialists, giving an overview of the problem of energy substitution upon which they can base their decisions; secondly it must re-calculate with near-zero latency to allow for real-time interaction, something that the richer models in the SWEET-SURE framework can not achieve. The approach reflects and extends the work of [Berntsen and Trutnevyte 2017; Trutnevyte 2016], adding a description of the energy balance and emissions of Switzerland over time.

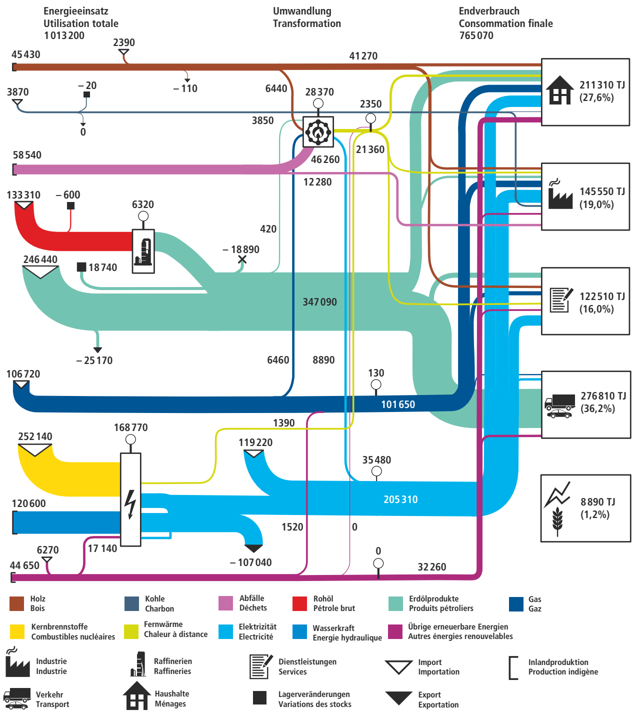


Fig. 2. Excerpt from [Bundesamt für Energie 2022a] showing a schematic representation of the energy balance for 2022.

2 MODEL DESCRIPTION

2.1 Energy Balance

The structure and calculation of the model derive from the energy balance for Switzerland [Bundesamt für Energie 2022a], a schematic version is shown in Figure 2. The energy balance provides a full description for all energy entering and leaving the national system on an annual basis. Its calculation, via a series of flows, is summarised below:

(1) Primary energy flows into the system either from domestic production, imports or stock, or out via exports, The primary subtotal is the total raw energy available to the system.
(2) Energy in its raw state may then be transformed into other types, most importantly electricity, or remain unchanged. After losses associated with transformation and delivery are deducted, the remaining subtotal represents energy available for final consumption.
(3) Final consumption accounts for the use of all raw and converted energy forms that were initially available. The system is therefore balanced with respect to input and output.

2.2 Calibration

The calibration process makes use of the historic energy balance dataset [Bundesamt für Energie 2022b] which gives time series for all data fields in the energy balance for traditional sources since 1980. Similar data relating to renewables since 1990 was made available by the BFE.

Firstly, a linear regression is performed on the prior 8 years of time series data. This generates a time-averaged value for the present day, removing the effects of the COVID pandemic, and generates a linear trend, which is used to extrapolate future values of final consumption. Values for transformation capacity are held constant, allowing the user to make decisions about levels of future capacity. This is the main interactive component of the game.

Values for efficiency and losses can be derived from the regressed time series data and are held constant at current levels. The historic data are available to view as part of the web interface along with the calculated regression and forecasts.

2.3 Scenario Construction

The model reconstructs a complete energy balance scenario for each year from 2022 to 2050. The calculation is as follows:

(1) Given quantities for final consumption the total for each energy type required for delivery is calculated.
(2) Losses associated with storage and delivery are added to give the total required after transformation.
(3) The quantities of each energy type available from transformation deducted from the energy required.
(4) Any remaining unfulfilled demand is satisfied by imports, with any excess of supply removed via exports.

The model therefore presents the user with the energy system in its current state, which includes estimates for future demand. The user selects capacities for each transformation process and thus manages the transition of energy supply from its current to a future state.

2.4 Metrics

The performance of the energy system is measured with respect to a set of economic and environmental metrics, which can be viewed as time series or ranked for comparison with other instances.

Some metrics are intrinsic to the energy balance itself, specifically net imports of energy and most importantly the net import electricity itself. This gives a measure of the level of self-sufficiency in the energy economy and a proxy measure of cost.

Import costs of raw energy depend upon market conditions and are beyond the scope of the model. It is however possible to estimate the total cost of electricity generation using figures from the literature [Ghodsvali et al. 2022; Holzer et al. 2023; Xexakis et al. 2020].

Other key metrics relate to the environmental impact of the energy system such as CO_2 emissions and land use. These again are calibrated to official data [Bundesamt für Umwelt 2024a,b] and calculated as a weighted sum. Emissions data are available as tonnes of equivalent CO_2/TJ for each form of electricity generation [Bundesamt für Umwelt 2023].

We must also calculate equivalent CO_2 emissions for final consumption. The are not directly available but can be calibrated from official data [Bundesamt für Umwelt 2024a,b]. Annual emissions by industry sector are regressed and the value is divided by final consumption of combustible energy per sector, giving the required value as tonnes/TJ of equivalent CO_2 emissions.

The metric for land use for each of the transformation processes is based on values in km^2/TJ from the literature [Ghodsvali et al. 2022;

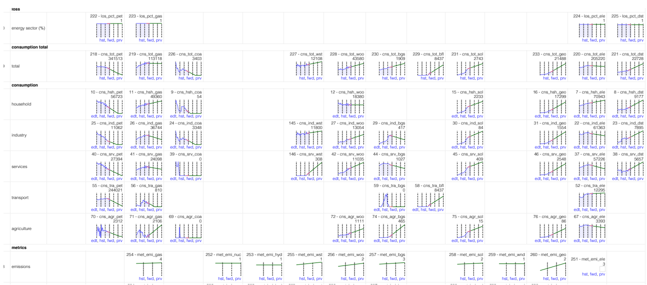


Fig. 3. Screenshot from the web interface, reflecting the layout of the energy balance table.


```
▼<root>
 ▼<tbl>
    <row res_id="1" res_name="one" res_gam=""/>
  </tbl>
 ▼<tbl>
    <row prm_id="163" grp_id="2" grp_ord="3" grp_code="cnv" grp_name="transformation"
    prc_id="8" prc_ord="6" prc_code="res" prc_name="reservoir" prd_id="4" prd_ord="6"
    prd_code="hyd" prd_name="hydro electric" reg_a="-522.42" reg_b="74838.7" prm_grw="0"
    prm_eff="-1" prm_cal="0" prm_emi="0.00000333333" prm_lnd="0.00113889"
    prm_cst="0.0222222" prm_avl="" prm_code="cnv_res_hyd" prm_desc="transformation,
    reservoir, hydro electric"/>
  </tbl>
 ▼<tbl>
    <row prm_id="163" yr="1990" tj="61610.4" v="91556.16796875"/>
    <row prm_id="163" yr="1991" tj="69062.4" v="91033.74768066406"/>
    <row prm_id="163" yr="1992" tj="66621.6" v="90511.32739257812"/>
    <row prm_id="163" yr="1993" tj="74887.2" v="89988.90710449219"/>
    <row prm_id="163" yr="1994" tj="82677.6" v="89466.48681640625"/>
    <row prm_id="163" yr="1995" tj="70016.4" v="88944.06652832031"/>
    <row prm_id="163" yr="1996" tj="57704.4" v="88421.64624023438"/>
```


Fig. 4. Screenshot from API, showing an excerpt from the XML document returned in response to an HTTP request.

Holzer et al. 2023; Trutnevyte 2016; Xexakis et al. 2020] multiplied by electricity output.

Seasonal availability at an annual level is calculated from energy market data [Swiss Energy-Charts 2024]. The ratio of energy produced during summer and winter months is calculated, giving a single scalar value per process. The model does not yet function at an intraday level and this a key simplification.

2.5 HTML and XML/HTTP Interfaces

The model is available at https://sure.euler.usi.ch either via a browser as shown in Figure 3, or application programming interface (API) as in Figure 4.

2.6 Acknowledgement

The research published in this article was carried out with the support of the Swiss Federal Office of Energy (SFOE) as part of the SWEET project SURE. The authors bear sole responsibility for the conclusions and the results presented in this publication.

REFERENCES

Alice Aubert and Judit Lienert. 2018. Gamified online survey to elicit citizens' preferences and enhance learning for environmental decisions. *Environmental Modelling & Software* 111 (09 2018). https://doi.org/10.1016/j.envsoft.2018.09.013

Philip B. Berntsen and Evelina Trutnevyte. 2017. Ensuring diversity of national energy scenarios: Bottom-up energy system model with Modeling to Generate Alternatives. *Energy* 126 (2017), 886–898. https://doi.org/10.1016/j.energy.2017.03.043

Bundesamt für Energie. 2022a. *Schweizerische Gesamtenergiestatistik 2022*. https://www.bfe.admin.ch/bfe/en/home/supply/statistics-and-geodata/energy-statistics/overall-energy-statistics.html/

Bundesamt für Energie. 2022b. *Switzerland energy balance*. https://opendata.swiss/en/dataset/energiebilanz-der-schweiz

Bundesamt für Umwelt. 2023. *Klimawirkung von Treibhausgasen und weiteren Substanzen*. https://www.bafu.admin.ch/dam/bafu/en/dokumente/klima/fachinfo-daten/CO2_Emissionsfaktoren_THG_Inventar.pdf

Bundesamt für Umwelt. 2024a. *Latest greenhouse gas inventory of Switzerland*. https://www.bafu.admin.ch/bafu/en/home/topics/climate/state/data.html

Bundesamt für Umwelt. 2024b. *Switzerland's Greenhouse Gas Inventory 1990–2022*. https://www.bafu.admin.ch/bafu/en/home/topics/climate/state/data/climate-reporting/ghg-inventories/latest.html

Jared Cechanowicz, Carl Gutwin, Briana Brownell, and Larry Goodfellow. 2013. Effects of gamification on participation and data quality in a real-world market research domain. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (Toronto, Ontario, Canada) *(Gamification '13)*. Association for Computing Machinery, New York, NY, USA, 58–65. https://doi.org/10.1145/2583008.2583016

Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (2010), 756–760.

Maryam Ghodsvali, Gamze Dane, and Bauke de Vries. 2022. An online serious game for decision-making on food-water-energy nexus policy. *Sustainable Cities and Society* 87 (2022), 104220. https://doi.org/10.1016/j.scs.2022.104220

Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences*. 3025–3034. https://doi.org/10.1109/HICSS.2014.377

Johannes Harms, Stefan Biegler, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. 2015. Gamification of Online Surveys: Design Process, Case Study, and Evaluation, Vol. 9296. 219–236. https://doi.org/10.1007/978-3-319-22701-6_16

Simona Holzer, Alexane Dubois, Julia Cousse, Georgios Xexakis, and Evelina Trutnevyte. 2023. Swiss electricity supply scenarios: Perspectives from the young generation. *Energy and Climate Change* 4 (05 2023), 100109. https://doi.org/10.1016/j.egycc.2023.100109

Saara Jones and Borja Martinez Oltra. 2024. *Ensured Energy*. https://astonedf.itch.io/ensured-energy

Raman Kazhamiakin, Annapaola Marconi, Mirko Perillo, Marco Pistore, Giuseppe Valetto, Luca Piras, Francesco Avesani, and Nicola Perri. 2015. Using gamification to incentivize sustainable urban mobility. In *2015 IEEE First International Smart Cities Conference (ISC2)*. 1–6. https://doi.org/10.1109/ISC2.2015.7366196

Vangelis Panos, Ramachandran Kannan, Stefan Hirschberg, and Tom Kober. 2023. An assessment of energy system transformation pathways to achieve net-zero carbon dioxide emissions in Switzerland. *Communications Earth & Environment* 4 (05 2023). https://doi.org/10.1038/s43247-023-00813-6

Michael Sailer, Jan Ulrich Hense, Sarah Katharina Mayr, and Heinz Mandl. 2017. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior* 69 (2017), 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Kristen Schiele. 2018. *Utilizing Gamification to Promote Sustainable Practices*. Springer International Publishing, Cham, 427–444. https://doi.org/10.1007/978-3-319-71312-0_16

Swiss Energy-Charts. 2024. *Public net electricity generation in Switzerland*. https://www.energy-charts.info

Evelina Trutnevyte. 2016. Does cost optimization approximate the real-world energy transition? *Energy* 106 (2016), 182–193. https://doi.org/10.1016/j.energy.2016.03.038

Georgios Xexakis, Ralph Hansmann, Sandra P. Volken, and Evelina Trutnevyte. 2020. Models on the wrong track: Model-based electricity supply scenarios in Switzerland are not aligned with the perspectives of energy experts and the public. *Renewable and Sustainable Energy Reviews* 134 (2020), 110297. https://doi.org/10.1016/j.rser.2020.110297

Poster Abstract: Controllability Architecture for Energy Devices

ALEXANDER HILL, JAN-HENRIK BRUHN, and CHRISTIAN PIEPER, OFFIS e. V., Germany

We present a novel control architecture to model Renewable Energy Resources. The proposed concept sets a focus on how to model Read and Control Capabilitiess (RCCs) of different energy devices, like Energy Storage Systems, Photovoltaic Systems or Heat Pumps. A second focus of the model is the self-description of the devices in a digitalized energy system. We explain how we combine RCCs with self-description into a compact model and show an example how to model a heat pump with this architecture. We believe that developing a more sophisticated RCC-model is useful for developing and comparing optimization algorithms for energy systems.

CCS Concepts: • **Information systems** → *Data layout*; • **Software and its engineering** → **Software architectures**; **Software system models**.

Additional Key Words and Phrases: energy system model, energy management systems, district energy management, EMS, energy community, renewable energy community, citizen energy community

1 INTRODUCTION

For an increasingly sustainable and decentralised energy system, more Renewable Energy Resources (RERs) are being deployed. In the building sector, these systems can supply both thermal and electrical energy, forming Multi Energy Systems (MESs). To optimize and potentially even reduce the use of energy of these MESs, Energy Management Systems (EMSs) are used [Hidalgo-Rodríguez and Myrzik 2018]. This is necessary with an increased use of RERs, due to their volatile nature, combined with irregular energy usage in single households. Due to this irregular usage, a coalition of households in close vicinity to share the energy in between them is beneficial [Li and Xu 2018]. The EU endorses these coalitions, furthermore called Renewable Energy Communities (RECs), which are already used in Austria [Biresselioglu et al. 2021].

A problem of EMSs for RECs compared to EMSs for single households is the heterogeneity of energy systems between the different participants. These heterogeneous MES in RECs are difficult to control and optimize. This is due to different energy devices having different Read and Control Capabilities (RCC). To describe these RCCs, a sensor-actuator model like the model described in IEC-61850 [Ustun et al. 2019] is often used. These models are very elaborate, which also brings complexity to describe the capabilities of a system. They also provide little abstraction about different reoccurring types of energy devices by operating on a sensor and actuator level for specific devices.

The complexity of different RCC in optimization algorithms and EMSs are often avoided by just assuming specific controllability for devices. For example, Hemm et al. give no information if or how their algorithms control their Photovoltaic (PV) plants and assume direct control of their electrical Energy Storage System (ESS) [Hemm et al. 2019]. Schönfeldt et al. built a Mixed Integer Linear Program (MILP) to optimize a district and assume that the used Heat Pump (HP) has a direct control and the PV plants are curtailable [Schönfeldt et al. 2022]. Gagin et al. developed a two-step optimization algorithm, where a general optimization is used and refined by sector-specific optimizers [Gagin et al. 2023]. Only the general optimization is described, no control over PV plants and direct control over electrical ESS is assumed. In [Wagner et al. 2022], no information about the controllability of the used devices is mentioned. Other algorithms [Finck et al. 2020] only decide to turn a HP on or off and again have no control over PV production.

Architectures like GAMES [Barbierato et al. 2020] mention control values as an important part of the integration of hardware and simulators with the help of Domain Specific Languages without going further into detail about how to model them. DEMKit [Hoogsteen et al. 2019] has a form of RCC, but they are fixed for a set of devices.

From another perspective, specifications to provide standardised controllability of devices, like the SmartGrid Ready (SGReady) interface, have been defined. This standard allows the same control over a pool of HPs, which are made from various different manufacturers, as demonstrated in [Fischer et al. 2018]. However, as shown by Fischer et al., this interface requires specific controllers/algorithms to be able to interface with the specific four-state interface of the SGReady standard, which makes it unsuitable for different schedulers for MESs. Additionally, laws like the German energy law §14a EnWG [Bundesgesetzblatt 2022, p. 1304] require energy devices like HPs and electric vehicle chargers to be curtailable by the Distribution System Operator (DSO) to ensure grid stability. For this, the DSO sends a control signal to the devices which can curtail for up to two hours a day. This curtailment must be modelled into the devices and optimization algorithms and EMSs must take the control signal into their calculations and the devices must on the other hand offer the curtailment capability.

We propose a general system to describe RCC for different energy devices with self-description capabilities while simultaneously combining the description with the sensor and control information. This proposed system is focused on a tertiary economic control layer, although it could be extended to do more fine-grained control of device-specifics.

2 READ AND CONTROL CAPABILITY CONCEPT

To be able to represent different RCC, we developed a schema that can be used both to describe the composition of different energy device features and properties and how those can be measured and controlled, as seen in Figure 1. This scheme was developed in the context of the EMS Quarter Energy Management System (QEMS)[1], which is a District Energy Management System (DEMS) used for the simulation of MES scheduling, using an optimization algorithm [Hill et al. 2023].

To describe the features and properties of an energy device, or *BaseEntity*, the energy-type dependent inputs, outputs, and storage and transformation capabilities between energy-types are described

Authors' address: Alexander Hill, alexander.hill@offis.de; Jan-Henrik Bruhn, jan-henrik.bruhn@offis.de; Christian Pieper, christian.pieper@offis.de, OFFIS e. V., Escherweg 2, Oldenburg, Germany, 26121.

[1] https://gitlab.com/qems

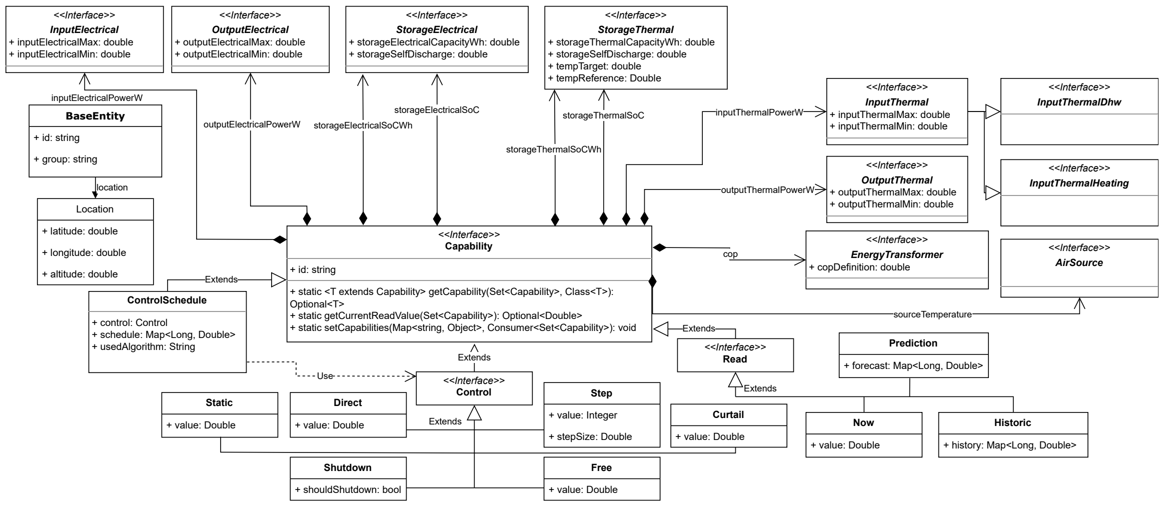


Fig. 1. Class diagram of the RCC concept. The upper part shows interfaces of different components of energy devices, like input and output, and the lower part shows the RCC which describes the capabilities of component sensors and actuators. Input, output, storage and energy types are separated interfaces. All interfaces contain static values like maximum input and a collection of RCC for attributes that represent sensors/actuators. The capabilities themselves are separated as read and control classes.

with their respective RCC. The model currently focuses on electrical and thermal energy, with transformation capabilities between those via i.e. heat pumps. Additionally, related information like the air source of an air-to-water HP or weather information is stored. The consumption of electrical power or an electrical demand can thus be described via a *InputElectrical*-Interface, while the heat production would be described through an *OutputThermal*-Interface.

The aforementioned interfaces proscribe the presence of a certain set of properties to describe the specific case, like the minimum and maximum power values. But they also require the implementation of a property which contains a set of RCC instances, which describe how the measurements can be read out, and how the property can be controlled. They are divided into control capabilities, the schedule capability and read capabilities: Control capabilities are i.e. fixed value, curtailment, shutdown/start or step-wise control. The schedule capability is directly related to a control capability, which simply describes at which point in time a referenced control capability should have which value. Read capabilities are divided into now (current), historic and prediction values.

With the combination of the two concepts, we can describe a multitude of energy devices, consumer, producer and prosumer. With additions, we also describe related weather information. This concept was then translated into an example Java-API[2], that contains the energy descriptions as interfaces which are then composed of different RCC.

For an example on how to implement a device with different RCC, see Figure 2. Using our concept, an air-to-water HP consists of *InputElectrical*, *AirSource*, *OutputThermal* thermal output and a power-to-heat *EnergyTransformer*, which describes the Coefficient of Performance (COP) from the input to the output. For the RCC of

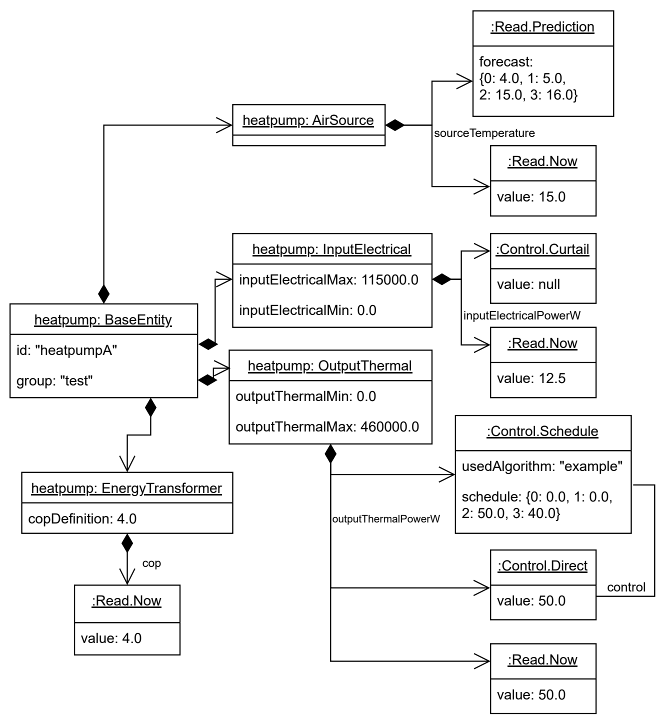


Fig. 2. Instance of a HP with the RCC concept. The described HP can be controlled directly by a direct control of heat production and/or indirectly by curtailing the electricity demand. Also, a schedule for the heat output can be set and the air source component provides a forecast of the air temperature.

the device, every RCC component can be read, for example, the current COP of the device or the current thermal output. Additionally,

[2]https://gitlab.com/qems/api/qems-api/-/tree/v0.1.0

the *AirSource* component also provides a temperature prediction, which can be used to calculate future COP. To control the device, two capabilities are shown: A direct control of the output in *OutputThermal* and curtailment of the electrical input in *InputElectrical*. The *Control.Direct* instance can be used by an EMS, while the curtailment on the *Control.Curtail* capability of the *InputElectrical* may be due to an external dimming signal from the DSO. The device also possesses an advanced control capability in the form of a *Control.Schedule* for the thermal output. Aside from the base readability of the sensors, more complex read capabilities like predictions are possible, which is useful in this case for the air source. The prediction provides data about the outside temperature, which can be used in scheduling algorithms.

3 LIMITATIONS AND FUTURE WORK

As we developed this system for tertiary economic control with constant grid connection, voltage or other metrics for stability are currently not represented in our system. This is the reason why we currently only consider most input, output and storage capacity in Watt or Watt-hour for simplification purposes. If needed, the RCC system can also be extended with additional sensors and other units like water temperature, flow rate or voltage. It then may be beneficial to alter the current naming convention of explicitly naming the attribute with their respective unit, energy type and direction and instead noting these properties elsewhere to reduce attribute name length.

Additionally, there is a multitude of possible new controls: especially more automated and indirect controls are currently not represented. For example, complex combinations of the presented controls like with SGReady [Fischer et al. 2018] are not representable. Other controls like automatic voltage or self-consumption control could be possible as capabilities for a battery system.

Furthermore, it is possible and explicitly allowed to define and use multiple controls for the same device and even attribute. This can make sense if you want to give the device some level of autonomy while still having some control, i. e. only curtailing the electrical input of a HP without forcing the device on how much it should consume to prevent accidental damage due to bad control signals. Through this, the presented scheme remains compatible with energy devices which are controlled via a secondary controller like a Programmable Logic Controller (PLC), which for example aggregates multiple different energy devices into one closed system. At the same time, constraints and interconnections are not possible to describe with such a system.

4 CONCLUSION

We here presented a novel approach to describe RCC of general energy devices, which are currently underutilized in energy informatics. Overall, our system can represent different energy devices, while at the same time delivering a fast way to grasp the self-description RCC of the device and also contains the respective values at the same time. This results in a compact abstraction of devices that is needed for an effective selection of usable optimization algorithms and EMSs.

We would like to invite the scientific community to discuss how RCCs can be best described and emphasize the need for such a description. Either for better algorithm comparisons and hardware integration into EMSs.

REFERENCES

Luca Barbierato, Daniele Salvatore Schiera, Edoardo Patti, Enrico Macii, Enrico Pons, Ettore Francesco Bompard, Andrea Lanzini, Romano Borchiellini, and Lorenzo Bottaccioli. 2020. GAMES: A General-Purpose Architectural Model for Multi-energy System Engineering Applications. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, Madrid, Spain, 1405–1410. https://doi.org/10.1109/COMPSAC48688.2020.00-59

Mehmet Efe Biresselioglu, Siyami Alp Limoncuoglu, Muhittin Hakan Demir, Johannes Reichl, Katrin Burgstaller, Alessandro Sciullo, and Edoardo Ferrero. 2021. Legal Provisions and Market Conditions for Energy Communities in Austria, Germany, Greece, Italy, Spain, and Turkey: A Comparative Assessment. *Sustainability* 13, 20 (Jan. 2021), 11212. https://doi.org/10.3390/su132011212 Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

Bundesgesetzblatt. 2022. Teil 1 Nr. 28: Gesetz zu Sofortmaßnahmen für einen beschleunigten Ausbau der erneuerbaren Energien und weiteren Maßnahmen im Stromsektor. http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl122s1237.pdf

Christian Finck, Rongling Li, and Wim Zeiler. 2020. Optimal control of demand flexibility under real-time pricing for heating systems in buildings: A real-life demonstration. *Applied Energy* 263 (April 2020), 114671. https://doi.org/10.1016/j.apenergy.2020.114671

David Fischer, Marc-André Triebel, and Oliver Selinger-Lutz. 2018. A Concept for Controlling Heat Pump Pools Using the Smart Grid Ready Interface. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 1–6. https://doi.org/10.1109/ISGTEurope.2018.8571870

Stepan Gagin, Michael Bettermann, and Hermann de Meer. 2023. Multi-vector optimization scheme for distributed components in energy islands. *e & i Elektrotechnik und Informationstechnik* 140, 5 (Aug. 2023), 460–470. https://doi.org/10.1007/s00502-023-01145-1

Regina Hemm, Stefan Mark, Michael Niederkofler, Andreas Schneemann, and Friederich Kupzog. 2019. Blockchain-based self-consumption optimization in local energy communities. AIT, Madrid, Spain. https://doi.org/10.34890/634

Diego I. Hidalgo-Rodríguez and Johanna Myrzik. 2018. Optimal Operation of Interconnected Home-Microgrids with Flexible Thermal Loads: A Comparison of Decentralized, Centralized, and Hierarchical-Distributed Model Predictive Control. In *2018 Power Systems Computation Conference (PSCC)*. 1–7. https://doi.org/10.23919/PSCC.2018.8442807

Alexander Hill, Christian Pieper, and Tobias Brandt. 2023. 4 District Energy Management Scheduling using a Participatory Local Energy Market. In *4 District Energy Management Scheduling using a Participatory Local Energy Market*. De Gruyter, 53–74. https://doi.org/10.1515/9783110777567-004

Gerwin Hoogsteen, Johann L. Hurink, and Gerard J. M. Smit. 2019. DEMKit: a Decentralized Energy Management Simulation and Demonstration Toolkit. In *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. 1–5. https://doi.org/10.1109/ISGTEurope.2019.8905439

Zhengmao Li and Yan Xu. 2018. Optimal coordinated energy dispatch of a multi-energy microgrid in grid-connected and islanded modes. *Applied Energy* 210 (Jan. 2018), 974–986. https://doi.org/10.1016/j.apenergy.2017.08.197

Patrik Schönfeldt, Adrian Grimm, Bhawana Neupane, Herena Torio, Pedro Duran, Peter Klement, Benedikt Hanke, Karsten von Maydell, and Carsten Agert. 2022. Simultaneous optimization of temperature and energy in linear energy system models. In *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*. 1–6. https://doi.org/10.1109/OSMSES54027.2022.9768967

Taha Selim Ustun, S. M. Suhail Hussain, and Hiroshi Kikusato. 2019. IEC 61850-Based Communication Modeling of EV Charge-Discharge Management for Maximum PV Generation. *IEEE Access* 7 (2019), 4219–4231. https://doi.org/10.1109/ACCESS.2018.2888880

H. Wagner, F. Peñaherrera, S. Fayed, O. Werth, S. Eckhoff, B. Engel, M. H. Breitner, S. Lehnhoff, and J. Rolink. 2022. Co-simulation-based analysis of the grid capacity for electric vehicles in districts: the case of “Am Ölper Berge” in Lower Saxony. In *6th E-Mobility Power System Integration Symposium (EMOB 2022)*, Vol. 2022. 33–41. https://doi.org/10.1049/icp.2022.2713

Poster Abstract: Hardware-in-the-Loop Simulation Environment for Validating Distribution System Control Applications

MARCEL OTTE, OFFIS - Institute for Information Technology, Germany
CARSTEN KRÜGER, OFFIS - Institute for Information Technology, Germany
PRATYUSH DAS, OFFIS - Institute for Information Technology, Germany
SEBASTIAN ROHJANS, Jade University of Applied Sciences & OFFIS - Institute for Information Technology, Germany
SEBASTIAN LEHNHOFF, OFFIS - Institute for Information Technology, Germany

Customer premises are becoming increasingly important for the power system as their flexibility potential is growing with the ongoing roll-out of electric vehicles, storage systems, photovoltaics, and heat pumps. However, the flexibility has the potential to be used not only within the customer's premises but also to serve different markets and grid operators, leading to use cases with multiple actors and systems. Thus, the validation of distribution system control applications with a focus on cross-system and cross-actor communication is needed. This work conceptualises a suitable hardware-in-the-loop simulation environment for validating distribution system control applications with the customer premises in a hardware-in-the-loop setup. The realisation shows how grid operators, field devices, a meter operator with smart meters, the customer, and market participants can be integrated into the simulation environment. Accordingly, suitable standards and market roles are identified and implemented. Derived from that concept, this work gives an outlook on future use cases to be validated using the simulation environment.

CCS Concepts: • **Hardware → Smart grid**; *Simulation and emulation*; *Renewable energy*; • **Computing methodologies → Model development and analysis**.

Additional Key Words and Phrases: hardware-in-the-loop, grid operator, smart meter, aggregator, customer

1 INTRODUCTION

The validation of distribution system control applications is not only complex due to the number of electric vehicles, heat pumps, photovoltaic systems, storage systems and the interactions among them but also due to the actors involved if their flexibility is used beyond the customer premises. This poses new challenges for applications that require cross-system and cross-actor communication, automation and coordination. Thus, before deployment, a suitable simulation environment has to consider how the interaction between the systems and actors has to be realised facilitating the basis for the integration of novel control applications.

State-of-the-art hardware-in-the-loop setups provide the foundation to validate control applications in a specific use case. By way of example, the work by [7] shows a hardware-in-the-loop validation of an energy management system for low-voltage distribution systems. More examples for integrating energy management systems are given in [4, 8]. In [1], an approach how to improve the controllability using smart meters is shown. Distribution control systems with a focus on grid constraints can apply virtual field devices in combination with a grid simulation [2]. Combining the approaches with the actors involved forms the motivation for the concept of this simulation environment.

Authors' addresses: Marcel Otte, marcel.otte@offis.de, OFFIS - Institute for Information Technology, Oldenburg, Germany; Carsten Krüger, carsten.krueger@offis.de, OFFIS - Institute for Information Technology, Oldenburg, Germany; Pratyush Das, pratyush.das@offis.de, OFFIS - Institute for Information Technology, Oldenburg, Germany; Sebastian Rohjans, sebastian.rohjans@jade-hs.de, Jade University of Applied Sciences & OFFIS - Institute for Information Technology, Oldenburg, Germany; Sebastian Lehnhoff, sebastian.lehnhoff@offis.de, OFFIS - Institute for Information Technology, Oldenburg, Germany.

2 CHALLENGES FOR VALIDATING DISTRIBUTION SYSTEM CONTROL APPLICATIONS

A hardware-in-the-loop-based simulation environment has to cope with activating flexibility at the customer premises that go along with complexities through interactions of load, storage and generation interconnected behind the customer's meter. The complexity arises not only with different technologies in place but also with individual user behaviours within the customer premises. Thus, for validating distribution system control applications, different user installations (e.g. electric vehicle fleet, photovoltaic with storage systems etc.) and optimisation goals (e.g. self-sufficiency) behind-the-meter can be in place and can vary within a grid area. One can also differ if flexibility is enabled directly by sending control signals to the flexibilities or indirectly by incentives (e.g. dynamic tariffs). Besides, the European Directive [3] fosters the customer to become more active in the power system leading to the demand for validating novel control applications.

Enabling the flexibility for use cases that are beyond the customer premises, requires the involvement of processes across multiple and various actors, which have to be validated. For the provision of flexibility (e.g. adapting the charging power of an electric vehicle), services can vary (e.g. congestion management or day-ahead optimisation), but also the actors involved (e.g. customer, distribution grid operator, aggregator etc.). In earlier work [9, 10], system architectures scope how processes for distribution systems spread over actors. Especially cross-system and cross-actor communication shift the focus on interoperability and automation and therefore has to be covered in the simulation environment. Use cases can spread over grid operator coordination or market-based optimisation through aggregators or considering local congestion management.

Market and grid operation-based control applications differ in their time behaviour that has to be managed within the simulation environment. While a day-ahead market has slower dynamics (e.g. hourly resolution), remedial actions of a grid operator are performed within a shorter time (e.g. in Germany within five minutes for local

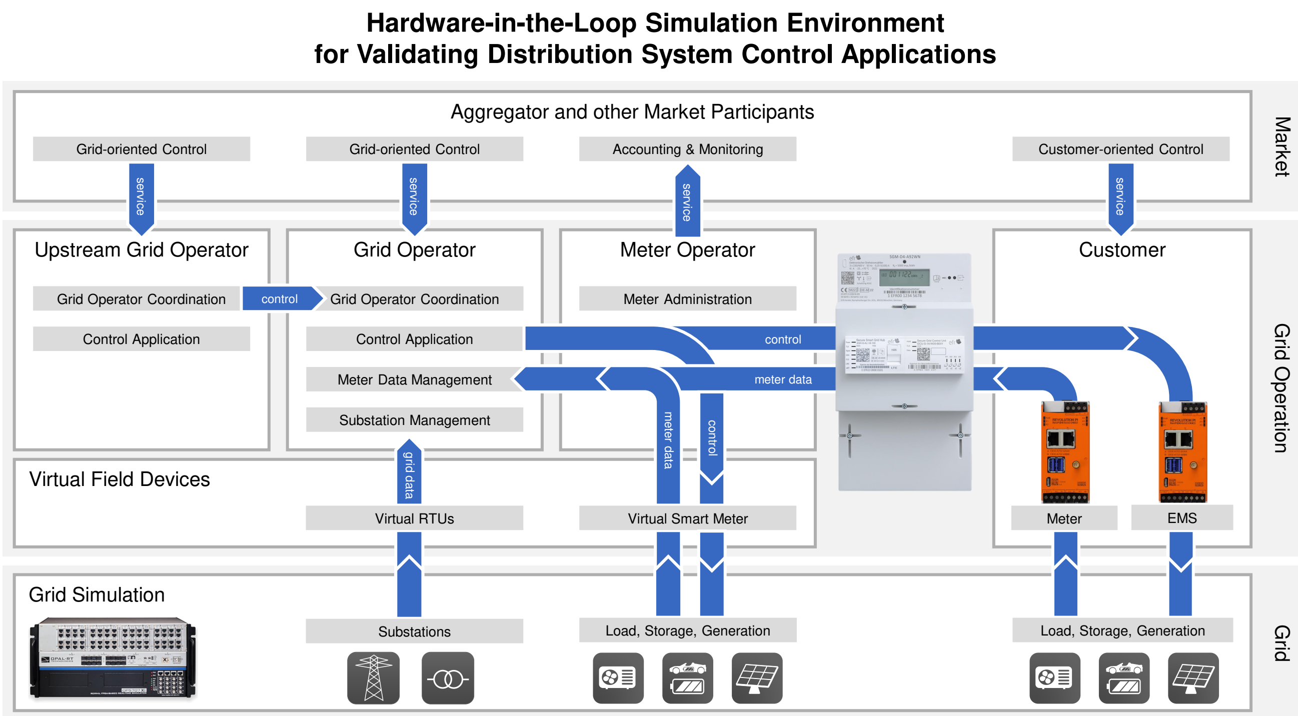


Fig. 1. Laboratory Architecture including the systems of actors (grey) and the communication among them (blue)

congestion management). For validating control applications in the cross-system and cross-actor simulation environment, steady-state conditions can be applied. However, for synchronising market and grid operation, there is no central co-simulation platform in place, instead, the Network Time Protocol (NTP) serves for time synchronisation and data has to be exchanged based on standardised protocols.

3 CONCEPT OF A VALIDATION ENVIRONMENT

Derived from the previous section, Figure 1 depicts the concept for the validation environment separated into the layers grid, grid operation and market.

The grid layer serves as a basis for the physical behaviour of the power system and for obtaining substation loading based on multiple customers' loads, storage and generation systems. Separating the grid layer from the grid operation layer ensures that only certain information about the power systems is directly available through measurements (e.g. through voltage or power meter). Thus, the amount of sensors, meters and their distribution within the grid can vary. In this setup, the simulator OPAL-RT is proposed for providing digital TCP/IP-based interfaces for each sensor or meter. It has to be acknowledged that other simulators could also perform load flow analysis and provide interfaces. Besides receiving grid state information, customer profiles and control signals impacting the grid are coupled to the grid simulator using as well the same digital TCP/IP-based interface approach.

For the grid operation, the separation of actors and systems results in data sovereignty and interface definitions across the simulation environment. To begin with the actors in place, the upstream grid operator (e.g. a distribution system operator or a transmission system operator) the grid operator (e.g. a distribution system operator), the meter operator and the customer are in place. The information about the power system from the grid layer is accessible only from the grid operator through virtual Remote Terminal Units (RTUs) using the standard [5] and smart meter using the standard [6] with the DLMS/COSEM specifications. Each actor operates on a virtual machine with its applications and databases. While the upstream grid operator can forward a control signal to the grid operator using a cascade, the grid operator can send control signals to the customer's flexibility through smart meters (e.g. enabling demand response). The customer is both, completely set up in a virtualised environment using Docker to scale the amount of customers in a grid area and once as a hardware-in-the-loop setup. The latter is realised using a German smart meter, a local Switch and two Revolution Pis with a Linux OS. One Revolution Pi operates an energy management system, while the other simulates the power values at the grid connection point for the smart meter. The administration of the smart meter is conducted by the meter operator.

The market layer portrays the actors impacting the power system through services to either grid operators and customers or both. The customer premises equipped with flexibilities have the potential to be optimised by the corresponding manufacturer of the components (e.g. providing solar-based charging opportunities) or even enabling participation in energy markets through the supplier or aggregator. The resulting behavioural change through direct or indirect control approaches leads for example to a high energy demand when prices in the energy market are low. However, using the customer's flexibility for ancillary services with the grid operator as a service user is feasible, too. Approaches on how to coordinate the activation of flexibility from distribution systems are conducted in earlier work [10].

4 CONCLUSION & OUTLOOK

To conclude, this work presented a hardware-in-the-loop simulation environment for the validation of distribution system control applications. The proposed environment covers the systems of market participants, a grid operator including its upstream grid operator for coordination, a grid area with virtual field devices, the meter operator including smart meter, and the customer premises. The latter was implemented both once on physical devices and also completely virtualised for scaling the number of customers. The relevant interfaces and standards are identified and actor separation is taken into account as a requirement. The simulation environment can be aligned to meet further requirements, standards and use cases.

As an outlook, the simulation environment will be used to validate different distribution system control applications. As the first use case, local congestion management in the low-voltage system will applied to investigate how customer flexibility can be used to relieve congestion in the distribution grid. Secondly, the same flexibility will be integrated into the coordination among grid operators, but for congestion management of an upstream grid operator. Lastly, coordination is not only crucial among grid operators but also with market participants. Thus, how market participants, such as aggregators, can enable flexibility considering the grid operation, will be tested in this environment, too.

ACKNOWLEDGMENTS

This research is funded by German Federal Ministry for Economic Affairs and Climate Action (BMWK) under agreement no. 03EI4043A (Redispatch 3.0).

REFERENCES

[1] Joshua Comden, Jing Wang, Subhankar Ganguly, Steven Forsyth, Reynaldo Gomez, and Andrey Bernstein. 2023. Hardware-in-the-Loop Evaluation of Grid-Edge DER Chip Integration Into Next-Generation Smart Meters. In *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 1–6.

[2] Pratyush Das, Anand Narayan, Davood Babazadeh, Payam Teimourzadeh Baboli, and Sebastian Lehnhoff. 2021. Real-time Context-Aware Operation of Digitalized Power Systems by Reporting Rate Control of PMUs. In *2021 IEEE Madrid PowerTech*. 1–6. https://doi.org/10.1109/PowerTech46648.2021.9494954

[3] European Union. 2022. Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU. *Official Journal of the European Union* (2022). http://data.europa.eu/eli/dir/2019/944/2022-06-23

[4] Kevin Förderer, Manuel Lösch, Ralf Növer, Marilen Ronczka, and Hartmut Schmeck. 2019. Smart Meter Gateways: Options for a BSI-Compliant Integration of Energy Management Systems. *Applied Sciences* 9, 8 (2019). https://doi.org/10.3390/app9081634

[5] IEC 60870-5-104 2016. *Telecontrol equipment and systems – Part 5-104: Transmission protocols*. Standard. International Electrotechnical Commission, Geneva, CH.

[6] IEC 62056 2023. *Electricity metering data exchange*. Standard. International Electrotechnical Commission, Geneva, CH.

[7] Paweł Kelm, Irena Wasiak, Rozmysław Mieński, Andrzej Wędzik, Michał Szypowski, Ryszard Pawełek, and Krzysztof Szaniawski. 2022. Hardware-in-the-Loop Validation of an Energy Management System for LV Distribution Networks with Renewable Energy Sources. *Energies* 15, 7 (2022). https://doi.org/10.3390/en15072561

[8] Nils Kroener, Kevin Förderer, Manuel Lösch, and Hartmut Schmeck. 2020. State-of-the-Art Integration of Decentralized Energy Management Systems into the German Smart Meter Gateway Infrastructure. *Applied Sciences* 10, 11 (2020). https://doi.org/10.3390/app10113665

[9] Carsten Krüger, Marcel Otte, Stefanie Holly, Saskia Rathjen, Arlena Wellssow, and Sebastian Lehnhoff. 2023. Redispatch 3.0–Congestion Management for German Power Grids–Considering Controllable Resources in Low-Voltage Grids. In *ETG Congress 2023*. VDE, 1–7.

[10] Marcel Otte, Jirapa Kamsamrong, and Sebastian Lehnhoff. 2024. Aggregator in Digitalised Power Systems. *ISGAN Annex* 6 (2024).

Poster Abstract: Optimizing the feature space for power system modeling

ALEKSANDR BEREZIN, OFFIS e.V., Germany

CCS Concepts: • **Computing methodologies → Model development and analysis**; *Artificial intelligence.*

Availability of Data and Material:
The source code for this project is available in the following repository: https://gitlab.com/funbotan/energeai

1 PROBLEM STATEMENT

This project attempts to break out of the limitations of currently existing approaches to problems such as power flow (PF), state estimation (SE), or optimal power flow (OPF), which I will collectively refer to as Inference tasks on power systems (ITPS). The existing methods can be categorized into the following three categories:

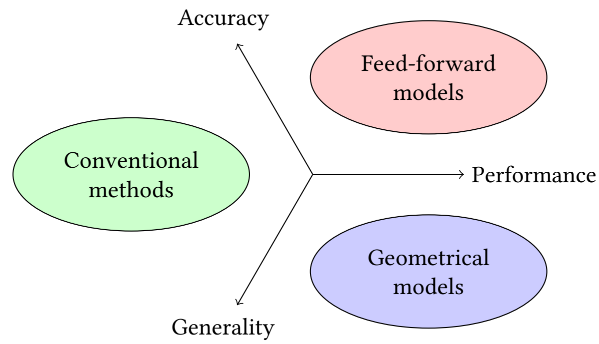
- **Conventional methods** are based on solving non-linear equations derived from the fundamental laws governing all power systems (Ohm's law, Kirchhoff's current law (KCL), Kirchhoff's voltage law (KVL)) with the Newton-Raphson or similar methods. These methods reliably produce accurate results as long as they converge, and they can be applied to any power system without restrictions. Their downside is computational complexity, which is $O(N^3)$ in the number of system elements for most methods.
- **Feed-forward models** are artificial neural networks (ANNs) built with multi-layer perceptrons (MLPs), convolution or attention layers, and any other feed-forward layers that work with a fixed number of inputs and outputs. Their advantage is the ability to achieve an arbitrarily high level of accuracy with a corresponding scale (number of trainable parameters and amount of training data). However, this accuracy can only be maintained if the power system remains static and drops drastically with any changes, even if most of the grid remains the same [1].
- **Geometric models** or graph neural networks (GNNs), such as graph convolutional networks (GCNs) and graph attention networks (GATs), utilize the graph geometry (which in our case is the topology of the power grid) to only learn local interactions between nods. This allows a model trained on one graph to perform inference on another without any adjustments. The downside of this approach is that GNNs cannot generally approximate an arbitrary function, which in practice limits their maximum accuracy [2].

The problem statement can be simplified to a three-way trade-off between conventional methods, feed-forward, and geometric models. There are three characteristics we would like to optimize at the same time:

- **Accuracy**: how closely a method can approximate the ground truth data.
- **Performance**: how fast can the evaluation of the method be completed, measured as the opposite of computational complexity.
- **Generality**: how well can the method adapt to changes in the grid topology without retraining.

Conventional methods provide generality and accuracy but are computationally inefficient. Feed-forward models are efficient and can be accurate but cannot generalize. Geometric models are comparably efficient and can generalize, but their accuracy is limited. This is illustrated in figure 1.

Fig. 1. The three-way trade-off between conventional methods, feed-forward and geometric models



2 ENGINEERING OBJECTIVE

The principal idea of this project is that we can optimize the feature space corresponding to power system states using geometric methods and then apply feed-forward models in the optimal space, thus combining the advantages of both methods while mitigating their disadvantages.

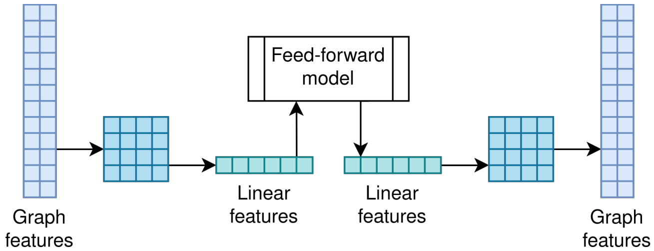
Using the compression heuristic, the Multiscale variational graph autoencoder (MVGAE) [4] model was determined as the most promising candidate for such an optimizer. This model iteratively reduces a graph by aggregating neighborhoods into nodes until only one node is left, and then the original graph is reconstructed from it. To then obtain an ITPS solver, I propose to insert a feed-forward model between the encoder and decoder of the MVGAE to translate from one optimal feature space to another. The proposed model architecture is represented graphically in figure 2.

The general algorithm for training a task-specific model for an ITPS is as follows:

(1) Train separate MVGAEs for both the input and output data of the problem (e. g. power and voltage values).
(2) Extract the encoder function from the MVGAEs trained on the input data and the decoder from one trained on output data.
(3) Compose a task-specific model as a sequence of the encoder, a feed-forward model, and the decoder.
(4) Train the resulting model, but only update the weights in the feed-forward model (weights of encoder and decoder are frozen at this step).

Author's address: Aleksandr Berezin, aleksandr.berezin@offis.de, OFFIS e.V., Oldenburg, Germany.

Fig. 2. Proposed model architecture



3 SCIENTIFIC OBJECTIVE

The deeper theoretical justification of the compression heuristic mentioned earlier is the Manifold hypothesis [3]. It posits that many high-dimensional datasets that occur in the real world intrinsically lie along low-dimensional latent manifolds inside that high-dimensional space. The effectiveness of deep neural networks (DNNs) is then explained through their ability to untangle and unfold (flatten) these nonlinear manifolds into linear low-dimensional latent spaces, where subsequent operations are much easier.

A metric to assess the extent of manifold disentanglement specifically for power systems data would aid in developing the method proposed in the previous section, e. g. in hyperparameter tuning. Even if the proposed solution does not work as expected, such a metric would still be a valuable contribution to the field as it could be used to evaluate other models.

To develop this metric, I will use the manifold learning method described in [3]. I will then compare the geodesic distances between datapoints to Euclidean distances between embeddings of those data points produced by the MVGAE-based model. A convergence of these two measurements will indicate a successful disentanglement process, enabling me to answer the main research question of this project:

RQ. *Can the MVGAE-based model reliably unfold the manifolds that occur in power system datasets?*

If successful, this project can pave the way for more general and powerful task-specific models, up to and including a foundation model for power systems.

REFERENCES

[1] Stephan Balduin, Tom Westermann, and Erika Puiutta. 2020. Evaluating Different Machine Learning Techniques as Surrogate for Low Voltage Grids. arXiv:2006.12389 [eess.SP]

[2] Aleksandr Berezin, Stephan Balduin, Thomas Oberließen, Sebastian Peter, and Eric MSP Veith. 2024. On zero-shot learning in neural state estimation of power distribution systems. arXiv:2408.05787 [cs.LG]

[3] Pratik Prabhanjan Brahma, Dapeng Wu, and Yiyuan She. 2016. Why Deep Learning Works: A Manifold Disentanglement Perspective. *IEEE Transactions on Neural Networks and Learning Systems* 27, 10 (Oct. 2016), 1997–2008. https://doi.org/10.1109/tnnls.2015.2496947

[4] Truong Son Hy and Risi Kondor. 2022. Multiresolution Equivariant Graph Variational Autoencoder. arXiv:2106.00967 [cs.LG]

Poster Abstract: learning techno-economic parameters of electricity generation units from their operational data

CHIARA FUSAR BASSINI, Hertie School, Germany

In an increasingly decarbonized energy market, conventional power plants must guarantee reliable and flexible dispatch to support phases of low RES generation. To model this behavior, we need an accurate calibration of the flexibility and efficiency of each individual power plant. However, in today's models, these parameters are often exogenous technology-level assumptions derived from technical reports and domain knowledge. This project proposes an alternative approach: a data-driven characterization of power plants, where parameters are directly derived from empirical data on plant operation. We deploy an interpretable Neural Network to learn unit-level techno-economic parameters of conventional power plants. We are particularly interested in flexibility parameters related to the cyclic operation of generation units, such as availability, efficiency and cold-start costs, and their relation to environmental and market variables. In a second step, we leverage the insights from this unit-specific technical parametrization to detect strategic unit dispatch and abusive output manipulation that exploit market power. Finally, we plan to perform a cross-unit parameter comparison to uncover anomalies in the operation of individual generation units and potential hints of market power abuse.

CCS Concepts: • **Computing methodologies → Machine learning**; **Modeling and simulation**.

Additional Key Words and Phrases: Energy modeling, Parameter estimation, Power plant generation, Power plant unavailability

Availability of Data and Material:
The original data is available on the ENTSO-E Transparency Platform [11] and EEX Transparency Platform [1].

1 INTRODUCTION

The introduction of more intermittent renewable energy in electricity systems is reshaping how conventional power generation assets are operated. In high RES systems, thermal power plants are no longer used to cover baseload demand, but rather for real-time output adjustment and to overcome phases of wind and solar scarcity [23]. Assessing the actual flexibility of the conventional supply fleet has therefore major relevance both for real-time grid stability and for the mid to long-term assessment of installed generation adequacy [17]. Making generic, technology-level assumptions on key flexibility parameters, such as start-up times and costs, is no longer sufficient: the actual operational history of each individual power plant becomes more and more relevant.

In an ideal world, a researcher would tackle this issue by incorporating more granular techno-economic information provided by power plant operators to their simulations, so that her model can better capture the flexibility of individual generation units. Decision makers would then use the insights from the energy model to make informed decisions about the design of electricity markets. However, public data on unit-level parameters is scarce, often not up-to-date or derived from unofficial sources. Therefore, many of the most common energy market models resort to technology-level exogenous assumptions [7, 13, 15]. Although such parametrization is the current standard practice in energy research, it can lead to misleading outcomes and overconfident estimates, as demonstrated by Ruhnau et al. in the MODEX-POLINS model comparison [21].

This research project proposes a data-driven approach which exploits open-sourced generation data on unit-level generated and available capacity. Learning techno-economic parameters of power plants directly from their operational data presents numerous advantages. On the one hand, a data-driven parameter estimation can yield a more realistic representation of power plants. Insights from the model can be leveraged to substitute current assumptions with more granular, unit-level empirical estimates and therefore contribute to the improvement of existing energy models. On the other hand, it can address the intransparency of dispatch decisions and discover potential unit-level anomalies and systematic market power abuse. The high spatio-temporal dimension of the data under consideration creates new opportunities for the deployment of Machine Learning (ML) methods, which can systematically process time series data (e.g. via RNNs), enable the estimation of probabilistic parameter distributions (e.g. via Bayesian layers) and allow for more flexibility when dealing with co-integrated panel data.

2 RESEARCH QUESTION

This research project tackles the overarching research question:

> **Can we deploy ML to construct a technical and economic representation of power plants that reproduce their behaviour on electricity markets?**

The main question can be divided into two subordinate research questions:

(1) *Can we derive a technical parametrization of individual plants that adequately fits their observed dispatch?*
(2) *Do we find evidence for the strategic deployment of units on energy markets through output manipulation?*

3 RELATED WORK

Despite being comparably new with respect to more established modeling frameworks, Machine Learning methods have already been successfully used for energy research; Rolnick et al. [20] provide an overview of ML applications by goal and area. Vom Scheidt et al. [24] review more than 200 energy-related papers that deploy data analytics and ML, based on area, application and methodology. The authors find a strong predominance of papers focusing on renewable generation forecasting, particularly of wind and solar. So far, comparably less attention has been devoted to the prediction of unit-level dispatchable generation.

For what concerns the first research question, a suitable approach

Author's address: Chiara Fusar Bassini, c.fusarbassini@hertie-school.org, Hertie School, Berlin, Germany.

for the empirical estimation of power plant parameters is provided by the technique of Bayesian calibration [16]. This technique updates prior beliefs on the probability distributions of model parameters based on evidence from empirical data, using a maximum likelihood estimation between the model outcomes and empirically observed data. Beykirch et al. [3, 4] apply this framework to learn parameter distributions for cold and warm-start costs, running costs and conversion efficiency. Boksteen et al. [6] use it to calibrate the plant efficiency of Combined Cycle Gas Turbines (CCGT). However, Bayesian calibration requires Monte-Carlo simulations with a significant number of iterations, which could nonetheless yield moderate results if the priors are poorly selected [18].

Also energy economists have attempted to estimate techno-economic parameters from power plants, using regression-based methods that exploit the statistical dependence between generated capacity and market variables. Nuno et al. [19] exploit the correlation between generation and wholesale electricity prices to estimate national, regional and unit-level variable costs for the French and German market. Unfortunately, as stressed out by the authors themselves, this methodology only provides an indication of the variable generation costs of individual units and cannot deliver in-depth insights into the actual dispatch strategy of power plant operators.

For what concerns the second research question, we will focus on capacity withholding, a strategy that can be detected using operational data, specifically the actual generation and reported availability of the unit. Capacity withholding refers to the physical (through output manipulation) or economic (through bid manipulation) reduction of the offered capacity on electricity markets in order to create supply scarcity and increase prices. In both cases, supporting empirical evidence has been identified in several European markets, including the German one [2, 8, 12]. Previous regression analyses have highlighted patterns compatible with capacity withholding at a market level: with the deployment of machine learning techniques, we aim at discovering market power strategies with further specificity at a unit or portfolio level.

4 METHODOLOGY

The primary data source for the project is the Transparency Platform of the European Network of Transmission System Operators for Electricity [11]. To comply with Regulation (EU) 543/2013 on submission and publication of data in electricity markets [9], generation units of at least 100 MWp operating in the European Union are required to publish information on generated capacity and major changes in available capacity. Data from ENTSO-E Transparency is complemented with data from EEX Transparency Platform [1], which publishes information about changes in the available capacity of generation units under Regulation (EU) 1227/2011 on Wholesale Energy Market Integrity and Transparency [10].

For our analysis, we primarily focus on the German market. On the one hand, the German electricity market is well-represented on EEX and ENTSO-E, so that the data can provide us with a full picture of the market. For most of the represented conventional technologies, the reporting units cover more than 80% of the total generated electricity by that technology during 2021. On the other hand, the German energy market has undergone major transformations over the course of the reported time period (2015 onwards). The splitting of the German-Austrian zonal market, the phase-out of nuclear energy and the rapid transition to RES have had a major impact on the operational scheduling of generation units [23], so that the data could deliver particularly interesting insights on the flexibility of the conventional fleet.

We conceptualize a multi-step prediction model to forecast the Day-Ahead generation output of single generation units. The model includes a customized layer representing the operating costs of the plant, which contains the parameter of interests. In this sense, the parameters of interest are treated equivalently to any other differentiable model parameters (e.g. weights and bias matrices). At each iteration, the model takes as input market variables, such as previous and forecasted load and RES generation, to compute unit-level generation for the next 24 hours. During training, the parameters of interest, which are initialized using realistic estimates from previous studies [22], are updated by backpropagation.

By embedding the optimal dispatch prediction into a differentiable NN pipeline, we learn the set of parameters that minimize the distance between the predicted and actual unit-level generation. Different functions for the distance, i.e. loss functions, can be applied. With continuous output predictions, loss functions such as the Mean Squared Error (MSE) or the Mean Absolute Error (MAE) can be used. It is also possible to discretize generated output to one or multiple plant states (e.g. *no generation, low generation, full generation*, as proposed by Beykirch et al. [3] [4]) and use categorical loss functions.

5 CONCLUSION AND OUTLOOK

This paper outlines a PhD research project, which has the goal of learning unit-specific techno-economic parameterizations of conventional generation units in the German market from their historical operational data. Its implementation integrates techno-economic parameters of interests, such as start costs and fuel conversion efficiency, into a Neural Network pipeline. By learning the parameters of a power plant from its historical dispatch data, we can derive more realistic estimates of unit-level parameters. This method could be particularly beneficial for power plants where little to no technical information is available.

The proposed approach is motivated by a lack of information on techno-economic parameters for conventional power plants from official sources. The empirically estimated parameters could be compared to technology-level estimates from scientific databases, such as OPSD conventional plants dataset [25] and PyPSA-EUR input repository [14] and used to increase their granularity, therefore improving existing energy models. Furthermore, a cross-unit comparison of the derived technical parameters can deliver insights on the operation and characteristics of the different units. Outliers and implausibilities can hint at market power abuse or operational anomalies. For example, the model might learn particularly high starting costs for a generation unit, reflecting inflexibility in its technical or operational setup or possible must-run obligations.

Future research could go one step further and explore intertemporal variations within the forecasted dispatch and the underlying parameters. For example, one could extend the model setup to learn probability distributions for the parameters (instead than deterministic values), in order to account for their temporal and situational variability. This could be implemented through the application of Bayesian layers [5]. Further work could go in the exploration of model mispredictions. By analyzing the subset of hours where the model is under- or overestimating generation, specific patterns and deviation from expected or rational behavior could be identified.

ACKNOWLEDGMENTS

The project was funded by the Federal Ministry of Education and Research (project code: 16DKWN102). The author thanks her supervisors, Prof. Dr. Lynn Kaack and Prof. Dr. Lion Hirth and fellow researchers Jorge Sánchez Canalez and Alice Lixuan Xu for their valuable input and feedback.

REFERENCES

[1] 2023. *EEX Transparency Platform.* https://www.eex-transparency.com/power Accessed on: 1 March 2024..

[2] Julian Bergler, Sven Heim, and Kai Hüschelrath. 2017. Strategic capacity withholding through failures in the German-Austrian electricity market. *Energy Policy* 102 (2017), 210–221.

[3] Mario Beykirch, Tim Janke, and Florian Steinke. 2018. Learning dispatch parameters of thermal power plants from observations. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 1–6.

[4] Mario Beykirch, Tim Janke, and Florian Steinke. 2019. Bayesian Inference with MILP Dispatch Models for the Probabilistic Prediction of Power Plant Dispatch. In *2019 16th International Conference on the European Energy Market (EEM)*. IEEE, 1–6.

[5] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight uncertainty in neural network. In *International conference on machine learning*. PMLR, 1613–1622.

[6] Sowande Z Boksteen, Jos P van Buijtenen, Rene Pecnik, and Dick van der Vecht. 2014. Bayesian calibration of power plant models for accurate performance prediction. *Energy conversion and management* 83 (2014), 314–324.

[7] Tom Brown, Jonas Hörsch, and David Schlachtberger. 2017. PyPSA: Python for power system analysis. *arXiv preprint arXiv:1707.09913* (2017).

[8] Tunç Durmaz, Sevil Acar, and Simay Kızılkaya. 2024. Generation failures, strategic withholding, and capacity payments in the Turkish electricity market. *Energy Policy* 184 (2024), 113897. https://doi.org/10.1016/j.enpol.2023.113897

[9] European Commission. 2013. Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009. Official Journal of the European Union. http://data.europa.eu/eli/reg/2013/543/oj Accessed on: 1 March 2024..

[10] European Commission. 2013. Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009. Official Journal of the European Union. http://data.europa.eu/eli/reg/2013/543/oj Accessed on: 1 March 2024..

[11] European Network of Transmission System Operators (ENTSO-E). 2023. *ENTSO-E Transparency Platform.* https://transparency.entsoe.eu/ Accessed on: 1 March 2024..

[12] Sara Fogelberg and Ewa Lazarczyk. 2019. Strategic withholding through production failures. *The Energy Journal* 40, 5 (2019), 247–266.

[13] Lion Hirth, Oliver Ruhnau, and Raffaele Sgarlato. 2021. The European Electricity Market Model EMMA-Model Description. (2021).

[14] Jonas Hörsch, Fabian Hofmann, David Schlachtberger, and Tom Brown. 2018. PyPSA-Eur: An open optimisation model of the European transmission system. *Energy strategy reviews* 22 (2018), 207–215.

[15] Mark Howells, Holger Rogner, Neil Strachan, Charles Heaps, Hillard Huntington, Socrates Kypreos, Alison Hughes, Semida Silveira, Joe DeCarolis, Morgan Bazillian, et al. 2011. OSeMOSYS: the open source energy modeling system: an introduction to its ethos, structure and development. *Energy Policy* 39, 10 (2011), 5850–5870.

[16] Marc C Kennedy and Anthony O'Hagan. 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 3 (2001), 425–464.

[17] Thomas-Olivier Léautier. 2019. *Imperfect markets and imperfect regulation: An introduction to the microeconomics and political economy of power markets*. Mit Press.

[18] Ralph T Muehleisen and Joshua Bergerson. 2016. Bayesian calibration-what, why and how. (2016).

[19] Marinho Nuno, Yannick Phulpin, Folliot Damien, and Martin Hennebel. 2016. Approaching Generation Variable Costs From Publicly Available Data. In *European Energy Market (EEM), 2016 13th International Conference on the.*

[20] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. 2022. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–96.

[21] Oliver Ruhnau, Michael Bucksteeg, David Ritter, Richard Schmitz, Diana Böttger, Matthias Koch, Arne Pöstges, Michael Wiedmann, and Lion Hirth. 2022. Why electricity market models yield different results: Carbon pricing in a model-comparison experiment. *Renewable and Sustainable Energy Reviews* 153 (2022), 111701.

[22] DU Sauer, K Görner, R Elsen, Karl-Josef Wolf, G Gasteiger, H Katzenberger, E Kakaras, A Kather, D Lindenberger, M Oechsner, et al. 2016. Konventionelle Kraftwerke—Technologiesteckbrief zur Analyse "Flexibilitätskonzepte für die Stromversorgung 2050". *Schriftenreihe Energiesysteme der Zukunft: Essen, Germany* (2016).

[23] Wolf-Peter Schill, Michael Pahle, and Christian Gambardella. 2017. Start-up costs of thermal power plants in markets with increasing shares of variable renewable generation. *Nature Energy* 2, 6 (2017), 1–6.

[24] Frederik vom Scheidt, Hana Medinová, Nicole Ludwig, Bent Richter, Philipp Staudt, and Christof Weinhardt. 2020. Data analytics in the electricity sector–A quantitative and qualitative literature review. *Energy and AI* 1 (2020), 100009.

[25] Frauke Wiese, Ingmar Schlecht, Wolf-Dieter Bunke, Clemens Gerbaulet, Lion Hirth, Martin Jahn, Friedrich Kunz, Casimir Lorenz, Jonathan Mühlenpfordt, Juliane Reimann, et al. 2019. Open Power System Data–Frictionless data for electricity system modelling. *Applied Energy* 236 (2019), 401–409.

Received 29 April 2024; revised September 5, 2024; accepted

Poster Abstract: Social Inclusion by Design - Policies for Bridging Electricity Prosumerism to Low-Income Households

CHRISTINA SPECK, Karlsruhe Institute of Technology, Germany

CCS Concepts: • **Applied computing** → *Law*; • **Human-centered computing** → *Visualization*; • **Information systems** → *Users and interactive retrieval.*

Additional Key Words and Phrases: policy design, energy modeling, urban building, electricity prosumer

1 INTRODUCTION

The shift towards renewable energy (RE) as a strategy to combat climate change requires substantial investments in multiple sectors, including heating and mobility, to ensure a successful energy transition [10]. However, public approval is crucial for advancing these investments [23]. Engaging residential households, especially tenants, in this transition is key, as they can contribute to solar electricity generation through rooftop surfaces, thus mobilizing private capital [22].

Enhancing active engagement in residential households can be achieved by enabling them to benefit from this transition. Prosumerism, a concept frequently discussed in literature and endorsed by the EU [5, 6, 19], represents a pivotal form of household engagement. Traditionally, prosumers are understood as "an energy user who generates [RE] in his/her domestic environment and either stores the surplus energy for future use or trades it to interested energy customers in a smart grid" [20, p. 1]. However, this conventional understanding of prosumerism is largely centered around individuals or entities that generate and sell electricity, typically through owning photovoltaic (PV) systems. Such a definition inherently excludes a significant portion of the population — tenants in multi-family buildings — who may not have the means or opportunity to own PV systems.

Tenants, in particular, are susceptible to rising energy prices and energy poverty due to the split incentives inherent in the landlord-tenant dilemma [15]. Overall around half of the EU population lives in multi-family buildings, predominantly renter-occupied [7], further emphasizing the vulnerability of low-income households. Research on RE communities in Europe underscores the limited engagement with vulnerable households, revealing significant challenges in achieving inclusivity [12]. Additionally, studies indicate that only a minority of community energy projects explicitly aim to address energy poverty, citing barriers such as high membership fees and limited savings [6].

This paper expands the definition of prosumerism to include tenants who actively participate in the energy market without owning RE supply entities (e.g., rooftop PV systems, balcony PV systems). In this broader context, prosumers are seen as "active energy citizens, willing to participate in energy markets" [5, p. 1], often organized within collectives such as energy communities. This expanded definition includes tenants who engage in demand response programs, variable pricing schemes, or utilize energy services that allow them to influence their consumption patterns based on real-time market conditions [19]. The enactment of policies establishing reliable standards is essential for realizing the benefits of prosumerism, as public sentiment towards RE depends on market maturity and consumer-friendliness [14]. However, existing policies often favor homeowners, exacerbating social disparities [3]. Given the challenges of including low-income tenants in the benefits of RE expansion, this PhD research project seeks to answer three research questions:

[RQ1] To what extent do tenancy status and access to RE, alongside other demographic variables, influence socio-political support for RE expansions?

[RQ2] How profitable are electricity prosumerism options for low-income tenants under diverse political constraints?

[RQ3] What design elements for interactive visualizations of energy policy comparisons can enhance decision-makers' understanding of policy implications?

2 RELATED WORK

2.1 Defining Electricity Prosumerism for Tenants

Electricity prosumerism has emerged as a critical component of the transition to RE [6], yet research has traditionally focused on homeowners as primary participants, neglecting tenants in multi-apartment buildings [21]. To expand the traditional electricity prosumerim view for tenant inclusion, I build upon the work of Pieńkowski [19], who emphasizes the importance of ownership in RE supply entities, and Campos et al. [5], who differentiates between onsite and offsite energy supply. I propose a refined categorization of tenant prosumers, focused on electricity prosumerism, designed to more accurately capture the diverse ways in which tenants can participate in the electricity market, as detailed below:

(1) **Individual residential prosumers:** Tenants who individually own an RE supply entity, typically a small-scale system such as a balcony PV panel. This category provides tenants with direct control over the electricity produced and consumed, allowing them to sell surplus electricity back to the grid under feed-in-tariff schemes [4] or other peer-to-peer sharing schemes [13].
(2) **Partnering residential prosumers:** Tenants who share ownership of a rooftop PV system with other tenants or residents within the same building. This model enables shared investment and collective benefits, reducing the financial burden on individual tenants while promoting collaboration. In this scenario, a private network of prosumers can be managed for example through a virtual energy company within a private wire scheme such as in the UK [4].
(3) **Partnering virtual prosumers:** Tenants who share ownership of a remote RE supply entity, such as a community

Author's address: Christina Speck, christina.speck@kit.edu, Karlsruhe Institute of Technology, Kaiserstraße 93, Karlsruhe, Germany, 76133.

solar project. In this model, tenants invest collectively in a system located outside their building, with financial returns managed among the partners. In the UK such a scenario is managed by a so called local energy company, sometimes providing dynamic price signals [4].

(4) **Collective residential prosumers:** Tenants who do not own an RE supply entity but benefit from systems owned by a homeowners association or other tenants within the building. These tenants engage through peer-to-peer sharing schemes [13], donations [16] or a subsidized prices for electricity provided by the homeowners association, such as the German landlord-to-tenant electricity premium [7], that provide improved prices for solar electricity generated and consumed within the building.

(5) **Collective virtual prosumers:** Tenants who benefit from surplus solar electricity generated by other households outside their multi-family building. These arrangements, often facilitated by peer-to-peer schemes, allow tenants to access RE at reduced costs without physical ownership of the generating equipment. One example is the sharing with friends scheme [13].

(6) **Service-integrated residential prosumers:** Tenants who participate in energy production through contracts with an external service provider that owns the RE supply entity, such as in a PV leasing business model[13]. The electricity is generated within the building, and tenants directly consume the solar electricity produced, benefiting from the provider's management of the system.

(7) **Service-distributed virtual prosumers:** Tenants who engage in virtual prosumerism through contracts with an external service provider that generates electricity remotely, connected by contracts such as power purchase agreements [5, 13].

<table>
<tr><th></th><th></th><th>Supply within multi-family building</th><th>Supply outside multi-family building</th></tr>
<tr><td rowspan="2">Traditional prosumer view</td><td>Self-owned (Tenant)</td><td>Individual residential prosumer</td><td></td></tr>
<tr><td>Co-owned (Tenant)</td><td>Partnering residential prosumer</td><td>Partnering virtual prosumer</td></tr>
<tr><td rowspan="3">Extended prosumer view</td><td>Homeowners association-owned</td><td rowspan="2">Collective residential prosumer</td><td></td></tr>
<tr><td>Residential household-owned</td><td>Collective virtual prosumer</td></tr>
<tr><td>External service provider-owned</td><td>Service-integrated residential prosumer</td><td>Service-distributed virtual prosumer</td></tr>
</table>

Table 1. Extended electricity prosumer categories for tenants, based on Pieńkowski [19] and Campos et al. [5]

2.2 Policies for Electricity Prosumerism

The successful integration of collective prosumerism into the energy system requires supportive policies. The European Union mandates member states to ensure RE community participation is accessible to all consumers, including low-income households [2]. However, current policies often marginalize tenants, especially in social housing sectors, due to power asymmetries in tenancy relationships [3]. The German "Mieterstrom" model, which allows landlords to sell PV-generated electricity to tenants, has faced regulatory and incentive challenges, hindering its potential [6]. Austria's energy communities are considered successful with the potential to combat energy poverty by offering stable tariffs for low-income households, however, participants formulate the need for policies that support funding of such energy communities [9].

2.3 Open Research Platforms for Local Political Engagement

Modeling approaches are crucial for understanding the policy implications of integrating tenants into prosumerism. However, existing models often lack accessibility for non-experts, including local decision-makers who could implement these policies. Interactive visualization tools can enhance understanding [17] by using techniques like selection, exploration, and reconfiguration [8]. However, it is yet to be examined if and how these techniques can be applied to the communication of complex energy system models. If these techniques prove to be successful, they can facilitate and enable non-experts, such as local decision-makers, to use energy model results, assess their potential impacts, and contribute to informed decision-making processes.

3 METHODOLOGY

In addressing **RQ1**, I utilize data from the "Social Sentiment in Times of Crises" panel study, which collects bi-weekly sentiment data from 1,500 participants in Germany. Key variables include tenancy status, access to RE, and support for RE expansion. I employ structural equation modeling, following the methodology outlined by [1], to analyze how tenancy status and access to RE influence support for political actions towards RE expansions.

In addressing **RQ2**, I first undertake a structured literature review [24] to identify existing business models and associated policies aimed at enabling tenants to become prosumers. Further, I plan interviews with local political decision-makers to identify current challenges in design and implementation of such policies, similar to Campos et al. [5]. Subsequently, I perform a policy scenario analysis through multi-agent simulation to explore the effect of tenant prosumerism options and associated policies on low-income tenants inclusion as well as the financial burden within a local electricity neighborhood in multi-apartment buildings.

In the search for appropriate interactive visualizations in the communication of policy scenario model results **[RQ3]**, current interactive visualization practices are being assessed through a structured review of interactive policy comparison tools for the public following the method of Gnewuch and Maedche [11]. This review is embedded within a design science research method by [18]. This review is then followed by the implementation of a policy comparison tool prototype in the use case of inclusive prosumer policy scenarios for low-income tenants. Following a full design science research cycle, the prototype is then being evaluated by non-expert users, such as local decision-makers.

4 CONCLUSION AND OUTLOOK

Overall, the proposed methodologies offer a comprehensive approach to investigate social inclusion in electricity prosumerism. By combining structural equation modeling, structured literature review, interviews with policy experts, multi-agent simulation, and design science, the study aims to advance understanding of policy implications for low-income tenants. With the proposed categorisation of tenant prosumers, I assess policy adequacy for new forms of prosumerism. The findings will inform the development of equitable policies that prioritize the involvement of vulnerable households in the shift to renewable electricity.

5 ACKNOWLEDGMENTS

Christina Speck thanks the German Federal Government, the German State Governments, and the Joint Science Conference (GWK) for their funding and support as part of the NFDI4Energy consortium. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 501865131.

REFERENCES

[1] David Bidwell. 2013. The role of values in public beliefs and attitudes towards commercial wind energy. *Energy Policy* 58 (July 2013), 189–199. https://doi.org/10.1016/j.enpol.2013.03.010

[2] Mehmet Efe Biresselioglu, Siyami Alp Limoncuoglu, Muhittin Hakan Demir, Johannes Reichl, Katrin Burgstaller, Alessandro Sciullo, and Edoardo Ferrero. 2021. Legal Provisions and Market Conditions for Energy Communities in Austria, Germany, Greece, Italy, Spain, and Turkey: A Comparative Assessment. *Sustainability* 13, 20 (Jan. 2021), 11212. https://www.mdpi.com/2071-1050/13/20/11212 Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[3] Annika Bode. 2022. To what extent can community energy mitigate energy poverty in Germany? *Frontiers in Sustainable Cities* 4 (Nov. 2022).

[4] Donal Brown, Stephen Hall, and Mark E. Davis. 2019. Prosumers in the post subsidy era: an exploration of new prosumer business models in the UK. *Energy Policy* 135 (Dec. 2019), 110984. https://doi.org/10.1016/j.enpol.2019.110984

[5] Inês Campos, Pontes Luz Guilherme, Marín-González Esther, Gährs Swantje, Hall Stephen, and Holstenkamp Lars. 2020. Regulatory challenges and opportunities for collective renewable energy prosumers in the EU. *Energy Policy* 138 (March 2020), 111212. https://doi.org/10.1016/j.enpol.2019.111212

[6] Inês Campos and Esther Marín-González. 2020. People in transitions: Energy citizenship, prosumerism and social movements in Europe. *Energy Research & Social Science* 69 (Nov. 2020), 101718.

[7] Christoph Domenig, Fabian Scheller, Phillipp Andreas Gunkel, Julian Hermann, Claire-Marie Bergaentzlé, Marta A. R. Lopes, Jake Barnes, and Russell McKenna. 2024. Overcoming the landlord–tenant dilemma: A techno-economic assessment of collective self-consumption for European multi-family buildings. *Energy Policy* 189 (June 2024), 114120.

[8] Ana Figueiras. 2015. Towards the Understanding of Interaction in Information Visualization. In *Proceedings of the 2015 19th International Conference on Information Visualisation*. IEEE Computer Society, Barcelona, Spain, 140–147.

[9] Helen Fischer, Reinhard Haas, Amela Ajanovic, and Frank Radosits. 2024. Energiegemeinschaften – eine Evaluierung bisheriger Erfahrungen und zukünftiger Perspektiven für Österreich. (2024).

[10] Dolf Gielen, Francisco Boshell, Deger Saygin, Morgan D. Bazilian, Nicholas Wagner, and Ricardo Gorini. 2019. The role of renewable energy in the global energy transformation. *Energy Strategy Reviews* 24 (April 2019), 38–50. https://doi.org/10.1016/j.esr.2019.01.006

[11] Ulrich Gnewuch and Alexander Maedche. 2022. Toward a Method for Reviewing Software Artifacts from Practice. In *The Transdisciplinary Reach of Design Science Research*, Andreas Drechsler, Aurona Gerber, and Alan Hevner (Eds.). 337–350.

[12] Florian Hanke, Rachel Guyet, and Marielle Feenstra. 2021. Do renewable energy communities deliver energy justice? Exploring insights from 71 European cases. *Energy Research & Social Science* 80 (Oct. 2021), 102244. https://doi.org/10.1016/j.erss.2021.102244

[13] Mahdi Karami and Reinhard Madlener. 2022. Business models for peer-to-peer energy trading in Germany based on households' beliefs and preferences. *Applied Energy* 306 (Jan. 2022), 118053. https://doi.org/10.1016/j.apenergy.2021.118053

[14] Serena Y. Kim, Koushik Ganesan, Princess Dickens, and Soumya Panda. 2021. Public Sentiment toward Solar Energy—Opinion Mining of Twitter Using a Transformer-Based Language Model. *Sustainability* 13, 5 (Jan. 2021), 2673. https://doi.org/10.3390/su13052673 Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[15] Michaela Lang, Ruth Lane, Kun Zhao, Stephanie Tham, Katrina Woolfe, and Rob Raven. 2021. Systematic review: Landlords' willingness to retrofit energy efficiency improvements. *Journal of Cleaner Production* 303 (June 2021), 127041.

[16] Esther Mengelkamp, Johannes Gärttner, Kerstin Rock, Scott Kessler, Lawrence Orsini, and Christof Weinhardt. 2018. Designing microgrid energy markets: A case study: The Brooklyn Microgrid. *Applied Energy* 210 (Jan. 2018), 870–880. https://doi.org/10.1016/j.apenergy.2017.06.054

[17] Mrinal Patwardhan and Sahana Murthy. 2015. When does higher degree of interaction lead to higher learning in visualizations? Exploring the role of 'Interactivity Enriching Features'. *Computers & Education* 82 (March 2015), 292–305.

[18] Ken Peffers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. 2007. A design science research methodology for information systems research. *JMIS* 24, 3 (2007), 45–77. Publisher: Taylor & Francis.

[19] Dariusz Pieńkowski. 2021. Rethinking the concept of prosuming: A critical and integrative perspective. *Energy Research & Social Science* 74 (April 2021), 101967. https://doi.org/10.1016/j.erss.2021.101967

[20] Dinusha Rathnayaka, Vidyasagar Potdar, Tharam Dillon, and Samitha Kuruppu. 2015. Framework to manage multiple goals in community-based energy sharing network in smart grid. *International Journal of Electrical Power & Energy Systems* 73 (Dec. 2015), 615–624. https://doi.org/10.1016/j.ijepes.2015.05.008

[21] Fabian Scheller, Karyn Morrissey, Karsten Neuhoff, and Dogan Keles. 2024. Green or greedy: the relationship between perceived benefits and homeowners' intention to adopt residential low-carbon technologies. *Energy Research & Social Science* 108 (Feb. 2024), 103388. https://doi.org/10.1016/j.erss.2023.103388

[22] Anne-Lorène Vernay, Carine Sebi, and Fabrice Arroyo. 2023. Energy community business models and their impact on the energy transition: Lessons learnt from France. *Energy Policy* 175 (April 2023), 113473. https://doi.org/10.1016/j.enpol.2023.113473

[23] Gordon Walker, Patrick Devine-Wright, Sue Hunter, Helen High, and Bob Evans. 2010. Trust and community: Exploring the meanings, contexts and dynamics of community renewable energy. *Energy Policy* 38, 6 (June 2010), 2655–2663.

[24] Jane Webster and Richard Watson. 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly* 26 (June 2002). https://doi.org/10.2307/4132319

Poster Abstract: Optimizing the provision of distributed energy resources operational load flexibility on local market platforms for grid stability

DANIEL BULL, Karlsruhe University of Applied Sciences, Institute of Refrigeration, Air-Conditioning and Environmental Engineering (IKKU), Germany and University of Freiburg, Department of Sustainable Systems Engineering (INATECH), Germany

CCS Concepts: • **Applied computing → Command and control**; • **Computing methodologies → Modeling methodologies**; **Model verification and validation**; **Simulation evaluation**; **Continuous simulation**.

Additional Key Words and Phrases: energy flexibility, flexibility market, energy market, local market platform, distributed energy resources, energy system simulation, energy system optimization, model predictive control, flexibility signal lead time, hydrogen hub

Availability of Data and Material:
The authors do not have permission to share the data presented in this work.

1 INTRODUCTION

The ongoing shift from centrally installed conventional energy generation to distributed, volatile renewable energy generation is leading to a growing number of grid challenges, such as grid congestions [5]. To address these challenges, local market platforms are currently investigated [1]. These markets aim to encourage small- and medium-sized prosumers to trade their available distributed energy resources (DER) flexibility locally, which can be used in case a grid congestion is being forecasted [6]. However, the market design and the trading processes highly vary among the platforms, as current reviews [1, 4, 6–9] show.

To contribute to the optimized participation of DERs in future local market platforms, the following perspectives of DER operation and load flexibility optimization are formulated and investigated in this work:

1.1 Local energy system optimization

What energy savings and flexibility provision increases can be achieved by optimizing the operation of commercial buildings heating and cooling systems compared to today's conventional controls?

To quantify the benefits of an optimized energy system control, a model predictive control (MPC) application for the energy system of a commercial building is being developed and compared to the existing conventional control in a simulation study. Cost and efficiency advantages in the operation and in the provision of flexibility on a simulated market platform are measured and evaluated.

Since some of the market parameters are still highly heterogeneous along the introduced platforms, an additional investigation of the optimal lead time between an accepted flexibility offer and delivery is conducted:

Author's address: Daniel Bull, daniel.bull@h-ka.de, Karlsruhe University of Applied Sciences, Institute of Refrigeration, Air-Conditioning and Environmental Engineering (IKKU), Moltkestraße 30, Karlsruhe, BW, Germany, 76133 and University of Freiburg, Department of Sustainable Systems Engineering (INATECH), Emmy-Noether-Straße 2, Freiburg, BW, Germany, 79110.

1.2 Market design parameters

Which lead time between an accepted flexibility offer and the delivery is optimal?

While the traded flexibility power is mainly described by consistent parameters along the proposed market platform designs, especially the signal lead time varies widely between a few minutes up to 24 h [9]. Since this parameter has a great influence on the available flexibility power and the corresponding cost from prosumers DER, an extensive investigation of the effects of changing signal lead times under different influencing factors and energy system designs is performed as presented in [2].

2 METHODOLOGY

2.1 Local energy system optimization

To quantify the benefits of optimized system operation and flexibility provision, a MPC application for an innovative energy system in a commercial building in Karlsruhe, Germany, is implemented and compared to the existing conventional control. The energy system consists of two heat pumps (HP) with a total maximum power of 120 kW_{el} (Fig. 1), the three thermal water storages: *heat storage* (HS), *cold storage* (CS) and *drying storage* (DS) which are used to buffer the thermal demands of the building, the two long term storages *intermediate slab* (IS) and *ground slab* (GS) in the form of concrete slabs containing water filled tubes, the two recooler (RC), the two heat exchangers *heat exchanger hot* (HXH) and *heat exchanger cold* (HXC), the two hydraulic gates *hydraulic gate slabs* (HGS) and *hydraulic gate cold* (HGC), several circulation pumps (CP), mixing vales (MV) and switching valves (SV). The colors of the pipes represent the temperature levels.

To benchmark the performance of the MPC in a safe and consistent environment, a simulation model of the energy system has been implemented, containing the fundamental thermodynamic correlations such as mass flows, temperature differences and heat transfer coefficients. Furthermore, in line with the control of the real energy system, the simulation model includes conventional control algorithms, which serve as a reference for comparing energy savings and the availability of power flexibility with the newly developed MPC.

The optimal control problem (OCP) of the MPC contains the same energy system formulation as the simulation model in the constraints, but with a reduced number of states within the components, to keep the OCP smaller and solvable within the requirements for a real time controlling application. The objective of the OCP consists of the total costs of the electricity demand of all components, the imputed switching costs of the HP, RC, SV and MV as well as the imputed slack costs of the thermal storages. For further complexity reduction, the OCP is divided into 3 parts: a 1 hour binary part with

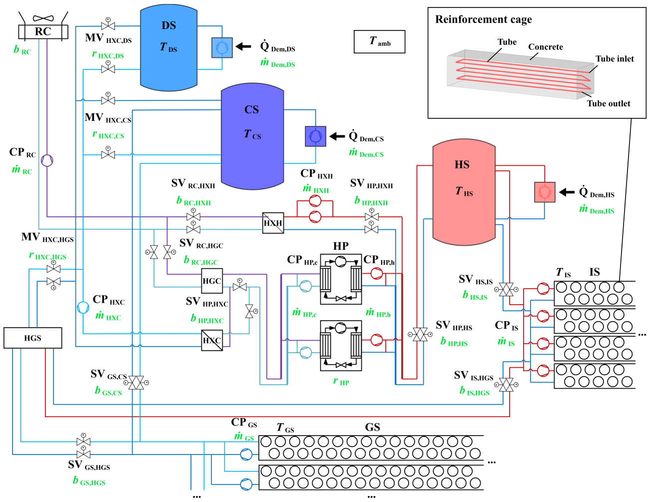


Fig. 1. Overview of the innovative energy system layout of a commercial building in Karlsruhe, Germany.

a resolution of 10 minutes which allows a control of the SV, MV, CP, RC and HP in a binary fashion, a 22 hour binary/linear part with a resolution of 1 hour approximated by using McCormick envelopes which allows linear energy flows between the components and a 6 day reduced linear part with a resolution of 6 hours which is activated when a frost period is detected to store sufficient thermal energy in the GS. To even further reduce the overall solving time of the OCP, a warm start is performed. The OCP is repeatedly solved with progressively advancing values from the simulation model, leading to a closed loop MPC. Flexibility requests are simulated by changing electricity costs during a flexibility demand. The thermal demands of the building stem from recordings of the year 2022.

2.2 Market design parameters

To quantify the influences of the signal lead time on cost and availability of flexibility, a generic energy system model of commonly used industrial energy system components, such as combined heat and power (CHP), boilers, HPs, PV-fields, heat storages (HS), and BES, as can be seen in Fig. 2 in the *Energy system components* section, has been implemented [2]. The components can be changed and parameterized according to the investigated energy system, allowing to evaluate today's commonly used designs and new planned HP-PV-BES designs. The components contribute to the thermal (red line), electric (green line) and fuel balance (brown line). Thermal and electric demands in the *Demands* section are modeled as fixed time-varying profiles, representing the time-dependent demands of connected industry processes or buildings. The price for the needed electricity from the grid and the price for the gas supply in section *Supplier*, are also modeled as fixed time-varying profiles, depending on the investigated tariff. On the far left in section *Market*, a market platform is attached to the electricity balance, requesting additional power. The energy system components are controlled by a MPC, which we assume to be connected to the market platform, taking its flexibility requests into account.

The energy system components are described on an energy flow level according to their physical relationships, as e.g., the formulation of the boilers, containing the gas consumption $\dot{F}_{b,t}$, the thermal production $\dot{Q}_{b,t}$, the thermal efficiency factor η_b, the maximum power $\dot{q}_{b,\max}$ and minimum part load $\lambda_{b,\text{part}}$.

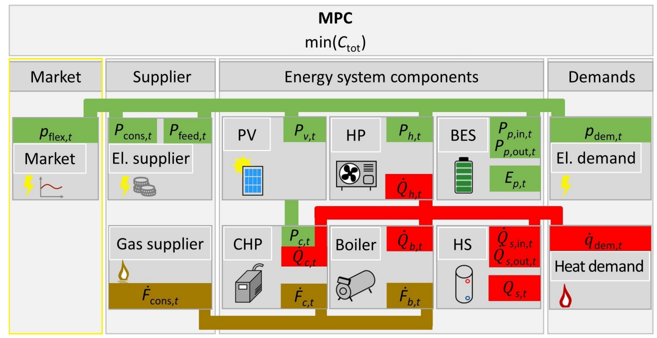


Fig. 2. Generic energy system layout, including the MPC, and the connected flexibility market.

To obtain an optimized operation of the energy system with minimized total costs, an OCP is formulated. The objective of the OCP contains the total costs C_{tot} of the energy system, consisting of the total gas cost, the electricity costs, the electricity earnings as well as the wear of the BES. To simulate a MPC-control of the energy system, the introduced OCP is repeatedly solved in progressing time steps, together with the time-related demands.

The requested flexibility power is described by two additional constraints, forcing the energy system to either decrease (positive flexibility $p_{\text{flex,t}}$) or increase (negative flexibility $p_{\text{flex,t}}$) its electricity demand. If a flexibility signal is received, the current OCP control schedule is recalculated together with the requested flex demand. Depending on the flex signal lead time, the recalculation is conducted at different iteration steps before the flex delivery.

To calculate the additional costs of providing flex power to the grid, a reference run without a flex request is carried out first. Then, a run with flexibility power is conducted and the costs subtracted by the costs of the reference run.

3 RESULTS

For the investigation of the costs of operational flexibility depending on the signal lead time, the energy system structure, as well as the thermal and electrical load profile from a company in southern Germany is used as presented in [2]. The company has an electrical demand of around 8 GWh per year and a thermal demand of around 7.5 GWh per year. The energy system includes three boilers, a CHP and a HS. The MPC of the energy system is assumed to have a forecast horizon of 48 hours, which is updated on an hourly basis. All forecast are assumed to be known under perfect foresight. The data shown stem from a typical spring evening, which is characterized by low thermal loads and warm outside temperatures.

To investigate the effects of changing signal lead times, multiple flex signal requests with a lead time between 0 and 24 hours are sent to the energy system. The requested flexibility power varies in 20

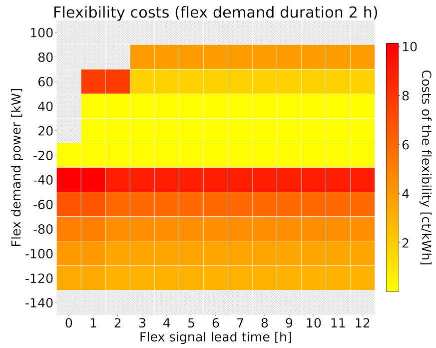


Fig. 3. Flexibility heatmap showing the flexibility costs of the sample energy system depending on flex power and signal lead time on a spring evening.

kW steps between -240 kW and 240 kW, which equals the maximum output power change of the CHP. The duration of the requested flexibility powers varies between one and four full hours.

To visualize the found results, a new developed flexibility heatmap (Fig. 3) is introduced. The cost development of rising flexibility demand can be found by reading the heatmap vertically. Looking at the heatmap horizontally, the cost effects of the flex signal lead time can be investigated. The prices are represented by a color scale from yellow (lower costs per kWh) to red (higher costs per kWh). Grey checked boxes show the limits of the available flex power. Reading the flexibility heatmap horizontally from left to right, it is noticeable how the flex power price stops changing at a certain lead time. In the present case of a flat electricity tariff, this stop in change happens at a signal lead time of 3 hours. The stop shows the time were the HS of the energy system are prepared as good as possible for all requested flex demands, leading to the lowest costs and the therefore optimum lead time for this case.

Investigating the effects of the signal lead time on new emerging system layouts, Fig. 4 presents a heatmap of the previously introduced industrial company with an adapted HP-PV-BES energy system design. As can be seen, the change to the new layout increases the possible flexibility power strongly to up to 400 kW positive and 1500 kW negative flex power. However, since the energy system highly depends on the electricity demand of the HP and the

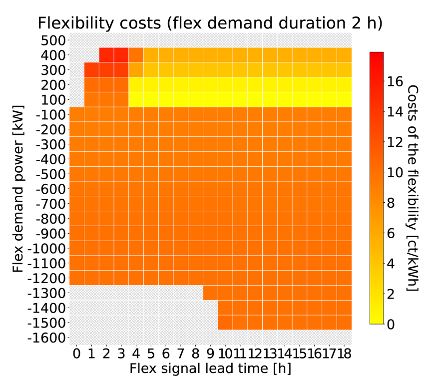


Fig. 4. Flexibility costs of a HP system with PV and BES on a spring evening.

supply from the PV and BES, the lead time also increases sharply to up to 10 hours. Furthermore, the prices of flexibility highly depend on the storage used for providing flexibility, leading to low costs by using the HS and predominantly high costs by using the BES.

4 CONCLUSION AND OUTLOOK

The signal lead time of the market platforms highly influences the costs and the amount of operational flexibility. As presented, the currently installed CHP based systems requires 3 hours of signal lead time, to offer flexibility at the lowest possible costs. This effect is further amplified by the latest transition towards an optimized electrified heating and cooling of buildings, as represented by the innovative energy system in Karlsruhe, Germany, or the introduced HP-PV-BES system, which requires 10 hours of lead time to provide its full flexibility potential.

Further ways of providing energy flexibility besides heating and cooling systems have been found in the use of hydrogen hubs. Due to their large long term storage capacities and the high electrical demand of their system components such as the electrolyzer, they serve as an ideal source of energy flexibility [3]. Additional research on optimizing their flexibility will be part of future work.

REFERENCES

[1] Ioannis Bouloumpasis, David Steen, and Le Anh Tuan. 2019. Congestion Management using Local Flexibility Markets: Recent Development and Challenges. In *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, Bucharest, Romania, 1–5. https://doi.org/10.1109/ISGTEurope.2019.8905489

[2] Daniel Bull, Adrian Bürger, Markus Bohlayer, Marco Braun, and Anke Weidlich. 2024. Optimizing the lead time of operational flexibility trading from distributed industrial energy systems in future energy and flexibility markets. *Preprint* (2024). https://optimization-online.org/2024/08/optimizing-the-lead-time-of-operational-flexibility-trading-from-distributed-industrial-energy-systems-in-future-energy-and-flexibility-markets/

[3] Elisa Ghirardi, Giovanni Brumana, Giuseppe Franchini, and Antonio Perdichizzi. 2023. H2 contribution to power grid stability in high renewable penetration scenarios. *International Journal of Hydrogen Energy* 48, 32 (2023), 11956–11969. https://doi.org/10.1016/j.ijhydene.2022.06.026 XII Edition of the International Conference on Hydrogen Production.

[4] Clara Gouveia, Everton Alves, José Villar, Ricardo Ferreira, Ricardo Silva, José Pablo Chaves, Tomás Gómez, Leslie Herding, Nicolás Morell, Michel Rivier, David Ziegler, Mathaios Panteli, Jan Budke, Katarzyna Zawadzka, and Catarina Augusto. 2019. *EUniversal UMEI MARKET ENABLING INTERFACE TO UNLOCK FLEXIBILITY SOLUTIONS FOR COST-EFFECTIVE MANAGEMENT OF SMARTER DISTRIBUTION GRIDS, Deliverable: D1.2 Observatory of research and demonstration initiatives on futureelectricity grids and markets.* Technical Report. https://euniversal.eu/wp-content/uploads/2021/02/EUniversal_D1.2.pdf (accessed on 23.02.2022).

[5] Shaojun Huang and Qiuwei Wu. 2018. Real-Time Congestion Management in Distribution Networks by Flexible Demand Swap. *IEEE Transactions on Smart Grid* 9, 5 (2018), 4346–4355. https://doi.org/10.1109/TSG.2017.2655085

[6] Xiaolong Jin, Qiuwei Wu, and Hongjie Jia. 2020. Local flexibility markets: Literature review on concepts, models and clearing methods. *Applied Energy* 261 (2020), 114387. https://doi.org/10.1016/j.apenergy.2019.114387

[7] Hosna Khajeh, Hannu Laaksonen, Amin Shokri Gazafroudi, and Miadreza Shafie-khah. 2020. Towards Flexibility Trading at TSO-DSO-Customer Levels: A Review. *Energies* 13, 1 (2020). https://doi.org/10.3390/en13010165

[8] Orlando Valarezo, Tomás Gómez, José Pablo Chaves-Avila, Leandro Lind, Mauricio Correa, David Ulrich Ziegler, and Rodrigo Escobar. 2021. Analysis of New Flexibility Market Models in Europe. *Energies* 14, 12 (2021). https://doi.org/10.3390/en14123521

[9] José Villar, Ricardo Bessa, and Manuel Matos. 2018. Flexibility products and markets: Literature review. *Electric Power Systems Research* 154 (2018), 329–340. https://doi.org/10.1016/j.epsr.2017.09.005

Poster Abstract: Novel Privacy Attacks on Load Data: Understanding their Impact and Mitigation Strategies

DEJAN RADOVANOVIC, Center for Secure Energy Informatics at the Salzburg University of Applied Sciences, Austria

CCS Concepts: • **Security and privacy → Privacy protections**; • **Computing methodologies → Machine learning**; *Knowledge representation and reasoning*; • **Information systems** → Information retrieval.

Additional Key Words and Phrases: Smart Meter Data, Data Analysis, Privacy, Machine Learning, Attacks, Data Anonymization, Privacy-Utility Trade-off

1 INTRODUCTION AND MOTIVATION

Smart meters enable utilities and customers to monitor energy consumption in real-time, optimize grid operations, and promote energy efficiency. These systems play a crucial role in the transition towards sustainable energy solutions by providing fine-grained data on electricity usage, often at 15 or 30 minute-intervals within the European Union [5, 6]. This level of detail allows utilities to implement sophisticated programs such as demand response, time-of-use pricing, and load forecasting, which are essential for managing the modern electrical grid effectively.

However, the fine-granular data collection raises significant privacy concerns. The detailed insights into household energy usage patterns can potentially expose personal lifestyles, behaviors, and household occupancy behavior [1, 7, 10–13]. Furthermore, when combined with socio-demographic data, these load profiles can potentially expose sensitive personal information, including income levels, family size, and lifestyle choices [3, 4, 9, 15].

Privacy risks in smart metering are particularly concerning under regulations like GDPR, which require strict data protection. The challenge of **de-anonymization**, where anonymized data can be cross-referenced to re-identify individuals, poses significant compliance issues. Even with pseudonyms, attackers can often link anonymized load profiles to specific households using external, publicly, available load profiles. These concerns are further intensified by the increasing use of advanced machine learning (ML) and deep learning (DL) techniques. These techniques can uncover complex patterns and relationships within the load profiles that were previously inaccessible, heightening the risk of privacy infringements. For instance, ML-based **consumer characterization** can unintentionally lead to profiling that reveals socio-demographic attributes, which could be exploited for targeted marketing, discriminatory practices, or even criminal activities.

Addressing these issues requires balancing data protection with maintaining utility for energy management. Traditional methods like data aggregation reduce utility, undermining smart metering applications. While pseudonymization preserves utility, it may not provide sufficient privacy protection in the face of increasingly sophisticated data analysis techniques [14]. Given these challenges, there is a critical need for strategies that effectively balance privacy with utility. This trade-off necessitates the development of more advanced methods that can dynamically adjust to the specific requirements of different smart meter applications, ensuring both effective energy management and robust privacy protection.

This Ph.D. project is driven by the need to address these pressing privacy issues in the context of smart metering. It aims to investigate the emerging privacy risks associated with load profiles and to introduce novel mitigation strategies that preserve the utility of the data for essential applications while preserving consumer privacy. The research will focus on two key privacy attack scenarios: de-anonymization and consumer characterization. It will examine the impact of multivariate data fusion and the application of ML/DL techniques on diverse datasets to provide a comprehensive understanding of these privacy challenges. Additionally, the study will explore advanced data anonymization methods, particularly generative models, which hold promise for balancing data utility with robust privacy protection in smart metering.

2 RESEARCH OBJECTIVES AND QUESTIONS

Aligned with the primary goal of this Ph.D. thesis, the research objectives are systematically formulated to address the identified privacy risks, evaluate the effectiveness of proposed mitigation strategies, and explore the balance between data utility and privacy preservation in smart metering. These objectives are expressed through five specific research questions:

(1) **Identification and Analysis of Privacy Risks:** How does analyzing load profile patterns contribute to the potential de-anonymization and re-identifiability of a household? What are the privacy implications of uniquely identifying households through load profiles, especially when tracking their changes across datasets from the same geographic region over different time periods? Additionally, how might such tracking expose personal life changes, such as relationship status, economic changes, family expansions, or specific health-related equipment usage?

(2) **Exploration of Privacy Risks in Multivariate Data Fusion:** How does the fusion of multiple datasets from different regions and cultures for socio-demographic prediction in smart metering amplify privacy risks, particularly when linking data points that were not originally intended to be connected, and what are the implications of enhanced consumer characterization for individual privacy across varied geographic and cultural contexts?

(3) **Development and Application of Advanced Representation Learning Techniques:** Can the application of self-supervised representation learning techniques from fields such as natural language processing or computer vision be effectively adapted to learn the structure of unlabeled load profile data? Furthermore, to what extent does the use of explainable supervised embeddings, incorporating elements like load profiles, weather data, the day of the week, and

Author's address: Dejan Radovanovic, dejan.radovanovic@fh-salzburg.ac.at, dejan.radovanovic@stud.plus.ac.at, Center for Secure Energy Informatics at the Salzburg University of Applied Sciences, Urstein Süd 1, Puch bei Hallein, Salzburg, Austria, 5412
.

holidays, affect the de-anonymization and consumer characterization when these embeddings are utilized as inputs for advanced deep learning models such as attention models, transformers, and residual neural networks?

(4) **Implementation of Transfer Learning Across Diverse Datasets:** Is it feasible to transfer robust and potentially explainable representations from one dataset to another, particularly when applying transfer learning across geographically independent datasets from different time periods and encompassing thousands of households? Additionally, how significantly does this transfer learning process improve the capabilities for de-anonymization and consumer characterization, especially when transferring knowledge from large labeled datasets to smaller or unlabeled datasets?

(5) **Evaluation of Advanced Data Anonymization Techniques:** How effective are advanced data anonymization techniques such as Autoencoders, Generative Adversarial Networks, and Diffusion Models in autonomously balancing privacy and utility trade-offs for critical smart metering applications such as demand-response programs, load forecasting, time-of-use pricing, theft detection, and grid stability, using a self-learning approach where the model itself learns the use-case-specific balance?

3 METHODOLOGY

The methodology is structured into two primary components: the **privacy attack scenario**, which seeks to identify and quantify the risks associated with smart meter data, and the **privacy-preserving methodology**, which focuses on mitigating these risks while maintaining the utility of the data for energy management applications.

3.1 Privacy Attack Scenario

This scenario is designed to explore two major privacy risks: de-anonymization and re-identifiability, and consumer characterization. The attack scenario begins with the collection and preprocessing of various datasets, including load profiles, socio-demographic data, and contextual information such as weather patterns and temporal markers (e.g., weekdays, holidays). Preprocessing involves aligning timestamps across datasets, standardizing measurement units, and handling missing data to ensure consistency and comparability. The complete methodology is depicted in Figure 1.

Once the datasets are prepared, the de-anonymization and re-identifiability attack is executed. This attack simulates an adversary attempting to re-identify individual households from anonymized load profiles by matching these profiles with publicly available load profiles. The methodology employs advanced indirect clustering techniques to reduce the dimensionality of the data, thereby facilitating the comparison of anonymized fingerprints with known profiles. The success of the attack is evaluated based on the accuracy of matching load profiles to specific households, highlighting the vulnerabilities in current anonymization practices.

In parallel, the consumer characterization attack is conducted to assess the potential for inferring socio-demographic characteristics from load profiles. This attack uses supervised machine learning models to predict attributes such as household size, income level, and appliance usage patterns. The models are trained on labeled data, enabling them to learn correlations between load profiles and socio-demographic features. The effectiveness of the attack is measured by the accuracy and precision of the model predictions, which are compared against ground truth data.

3.2 Privacy-Preserving Methodology

The second component of the methodology focuses on developing and evaluating privacy-preserving techniques that mitigate the risks identified in the attack scenario. Central to this approach is the use of advanced generative models, including Autoencoders, Generative Adversarial Networks (GANs), and Diffusion Models, to anonymize load profiles while preserving their utility for energy management applications [2, 8]. These models generate synthetic data that preserves the statistical properties of the original datasets, thereby enabling meaningful analysis without compromising individual privacy.

The privacy-preserving methodology starts with selecting generative models tailored to the data characteristics and identified privacy risks. Autoencoders learn compact representations of load profiles, which are then decoded into anonymized versions. GANs generate synthetic data that closely resembles the original, maintaining utility for tasks like demand response and load forecasting. Diffusion Models are also explored for their ability to iteratively refine data, balancing privacy and utility effectively.

The effectiveness of these privacy-preserving techniques is assessed through a detailed evaluation of the privacy-utility trade-off. This involves quantifying the extent to which the anonymized data can still support critical smart meter applications, while simultaneously reducing the risk of privacy infringement. The trade-off is measured using metrics such as Contrastive Loss (CL), Information Loss (IL), Privacy Gain (PG), and Re-identification Risk (RR). The models are iteratively optimized to enhance this balance.

Finally, the robustness of the privacy-preserving methods is validated through cross-validation techniques, including k-fold cross-validation, which ensure that the models generalize well across different subsets of the data. Benchmarking against existing methods is also conducted to verify the effectiveness and correctness of the proposed solutions.

4 RESULTS

The initial findings of this Ph.D. project address the first research question (RQ1) by examining whether individual households can be uniquely identified from their energy consumption data, even when only one week's worth of data is available. Using data from Upper Austria and applying standard algorithms, the study found that for certain households, the weekly consumption patterns are so distinct that they can be identified with an accuracy 16 to 35 times higher than random guessing. This is particularly evident in smaller residential areas, with sets of 25 and 100 households, where the ability to distinguish a household based on its energy usage remains consistently high [14].

In addition to the elaborated preprocessing and examination of the collected datasets a potential privacy mitigation technique has been explored. This study investigates the privacy influence of load

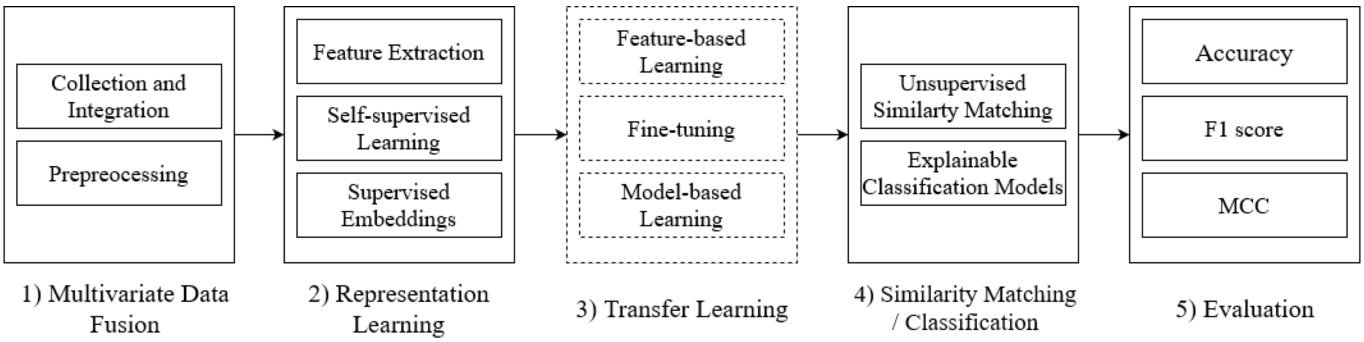


Fig. 1. Methodological approach for the attack scenarios of **de-anonymization** and **consumer characterization**, consisting of five steps. The integration of transfer learning (indicated by dashed lines) is not planned for addressing Research Questions 1 to 3.

profiles with varying time granularities in the context of predicting socio-demographic characteristics. The results indicate that as the time granularity of load profiles becomes coarser, the accuracy of these predictions generally decreases, with some stability observed at one-hour and daily intervals, depending on the desired use case.

5 CONCLUSION AND OUTLOOK

The research has partially addressed RQ1, with a preprocessing pipeline for multiple datasets already implemented. Initial findings indicate the potential to identify households through one-week energy consumption patterns, raising concerns about existing anonymization methods. Additionally, a paper on the privacy impact of varying time granularities on load profiles is nearing submission.

Moving forward, the research will focus on deepening the exploration of de-anonymization and consumer characterization (RQ2, RQ3, RQ4) across various datasets, utilizing advanced ML and DL techniques. Additionally, efforts will concentrate on refining privacy-preserving techniques, particularly by applying advanced generative models that anonymize data without compromising utility, as explored in RQ5.

REFERENCES

[1] Joana M Abreu, Francisco Câmara Pereira, and Paulo Ferrão. 2012. Using Pattern Recognition to Identify Habitual Behavior in Residential Electricity Consumption. *Energy and Buildings* 49 (2012), 479–487. https://doi.org/10.1016/j.enbuild.2012.02.044

[2] Matthew Baucum, Anahita Khojandi, and Rama Vasudevan. 2020. Improving deep reinforcement learning with transitional variational autoencoders: A healthcare application. *IEEE Journal of Biomedical and Health Informatics* 25, 6 (2020), 2273–2280.

[3] Christian Beckel, Leyna Sadamori, and Silvia Santini. 2013. Automatic socio-economic classification of households using electricity consumption data. In *Proceedings of the fourth international conference on Future energy systems*. 75–86. https://doi.org/10.1145/2487166.2487175

[4] Christian Beckel, Leyna Sadamori, Thorsten Staake, and Silvia Santini. 2014. Revealing household characteristics from smart meter data. *Energy* 78 (12 2014), 397–410. https://doi.org/10.1016/J.ENERGY.2014.10.025

[5] European Commission. 2012. Recommendation of 9 March 2012 on preparations for the roll-out of smart metering systems. Official Journal of the European Union. Available online: https://op.europa.eu/en/publication-detail/-/publication/a5daa8c6-8f11-4e5e-9634-3f224af571a6/language-en [2024-04-03].

[6] European Commission. 2014. *Cost-benefit analyses and state of play of smart metering deployment in the EU-27*. SWD(2014) 189 final. European Commission. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1403084595595&uri=SWD:2014:189:FIN [2024-04-03].

[7] Zhong Fan, Parag Kulkarni, Sedat Gormus, Costas Efthymiou, Georgios Kalogridis, Mahesh Sooriyabandara, Ziming Zhu, Sangarapillai Lambotharan, and Woon Hau Chin. 2013. Smart Grid Communications: Overview of Research Challenges, Solutions, and Standardization Activities. *IEEE Communications Surveys and Tutorials* 15 (2013), 21–38. https://doi.org/10.1109/SURV.2011.122211.00021

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.

[9] Konstantin Hopf, Mariya Sodenkamp, Ilya Kozlovkiy, and Thorsten Staake. 2016. Feature extraction and filtering for household classification based on smart electricity meter data. *Computer Science - Research and Development* 31 (8 2016), 141–148. Issue 3. https://doi.org/10.1007/S00450-014-0294-4/TABLES/3

[10] Wilhelm Kleiminger, Christian Beckel, Thorsten Staake, and Silvia Santini. 2013. Occupancy Detection from Electricity Consumption Data. *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13*, 1–8. https://doi.org/10.1145/2528282.2528295

[11] J Zico Kolter and Tommi Jaakkola. 2012. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. *Journal of Machine Learning Research - Proceedings Track* 22 (4 2012), 1472–1482.

[12] Mikhail Lisovich and Stephen Wicker. 2008. Privacy concerns in upcoming residential and commercial demand-response systems. *IEEE Proceedings on Power Systems* 1 (2008), 1–10.

[13] Andrés Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin. 2010. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building* (New York, NY, USA). ACM, 61–66. https://doi.org/10.1145/1878431.1878446

[14] Dejan Radovanovic, Andreas Unterweger, Günther Eibl, Dominik Engel, and Johannes Reichl. 2022. How unique is weekly smart meter data? *Energy Informatics* 5 (2022), 1–13. Issue 1.

[15] Yi Wang, Qixin Chen, Dahua Gan, Jingwei Yang, Daniel S. Kirschen, and Chongqing Kang. 2019. Deep learning-based socio-demographic information identification from smart meter data. *IEEE Transactions on Smart Grid* 10 (5 2019), 2593–2602. Issue 3. https://doi.org/10.1109/TSG.2018.2805723

Poster Abstract: Load Forecasting in Energy Communities for Flexibility Services Provision

DEMETRIOS N. PAPADOPOULOS*, INATECH-University of Freiburg, Germany

This doctoral project presents a methodology for improving flexibility provision within energy communities through advanced forecasting techniques. The methodology encompasses community-level forecasting for peak shaving and household-level forecasting for grid congestion alleviation. Considering the varying spatiotemporal resolutions in power data, the project focuses on optimizing the forecasts by investigating approaches such as tree-based models, recurrent neural networks, transformer models, and ensembles while using feature engineering, feature selection, and hyperparameter optimization. The project also introduces novel forecasting metrics specifically tailored to measure flexibility provision. These metrics are then compared for models trained locally versus globally (cross-learning), providing a thorough evaluation of forecasting techniques. Lastly, a spatio-temporal analysis examines the direct impact of forecasting models on flexibility provision. This structured approach allows for a comprehensive analysis across different granularities, prediction horizons, and spatial sizes, providing insights into the effectiveness of forecasting models in supporting flexibility provision within energy communities.

CCS Concepts: • **Applied computing → Forecasting**; • **Computing methodologies → Modeling and simulation**; **Machine learning**.

Additional Key Words and Phrases: energy communities, time-series forecasting, spatio-temporal forecasting, grid services

Availability of Data and Material:
No data or materials were used

1 MOTIVATION

The Intergovernmental Panel on Climate Change (IPCC) highlights the urgency to reduce the use of fossil fuels, especially in the energy sector [10]. In response, policymakers worldwide have expedited the installation of renewable energy (RE) systems while also transitioning the transportation and heating sectors towards electrification [11]. The increasing installation of RE systems and the electrification of previously non-electrified loads suggest that the power grid infrastructure will likely experience significant stress in the years ahead. For instance, Germany aims to achieve an 80% share of RE in electricity generation, along with the deployment of 15 million electric vehicles (EVs) and six million heat pumps (HPs) by 2030 [15].

To address these challenges, flexibility services at the distribution system operators (DSO) are becoming vital. These services such as congestion management and peak shaving help alleviate the stress on the grid and in parallel defer the investments on the grid infrastructure. Currently, considerable research is directed towards enabling end-user resources to provide flexibility services [1].

However, while these end-user initiatives signal progress towards flexibility enhancement, their emphasis on individual appliances limits the margins of improvement. In contrast, community-based provision of flexibility services utilizes a broader array of resources, thereby broadening the sources from which flexibility originates and increasing the quantity of available flexibility [12]. This approach also fosters collaboration, distributing the responsibility of flexibility provision among the participants [2].

Nonetheless, to effectively provide these services, it is essential to determine the timing and the extent of the flexibility needed, especially during peak times. This can only be achieved through accurate forecasting models, which will optimize the energy patterns of communities and deliver flexibility to the public grid. This work aims to assess intelligent solutions in which energy communities can provide peak shaving and congestion alleviation to the local DSO under different forecasting architectures. The key research questions are:

(1) **How can community peak load forecasting be leveraged to alleviate congestion at the grid-transformer?**
(2) **Given household data, how can global and local forecasting strategies in energy communities help alleviate grid congestion?**
(3) **What is the effect of different forecasting horizons and spatio-temporal resolutions in energy communities for preventing grid congestion?**

2 RELATED WORK

2.1 The State of Forecasting in Flexible Energy Communities

The majority of studies on communities that offer flexibility services focus on optimizing flexibility provision and rarely discuss the deployment of forecasting models. For example, the EcoGrid 2.0 project utilized the HPs from 800 houses in Denmark for peak shaving and grid-investment deferral. In their project, the forecasting implementation is omitted and it is assumed that the local DSO can provide the timing and the magnitude of the grid overloading [8]. Similarly, in a German project involving more than 853 households, peak shaving was provided by utilizing the flexibility of HPs and batteries. In the suggested methodology, a perfect forecast is assumed [17]. The Serene project, which presented a case study of 23 houses in the Netherlands and tested their ability to provide peak shaving from EVs and batteries, also utilised perfect forecasts [9].

In contrast, other studies that focus on forecasting do not address the impact of forecasting on flexibility services. For example, [16] proposed an LSTM-XGBoost hybrid model for accurate day-ahead load forecasting of community load, while [14] compared various forecasting algorithms for energy prosumption within communities.

2.2 Advanced Forecasting Techniques and Evaluation Frameworks in Energy Research

To offer peak shaving services to the local DSO, it is essential to accurately forecast peak demand. This is why we explore advanced forecasting techniques, specifically peak forecasting. A study on ultra-short-term peak demand forecasting suggested a novel methodology

Author's address: Demetrios N. Papadopoulos, dimitrios.papadopoulos@inatech.uni-freiburg.de, INATECH-University of Freiburg, Emmy- Noether-Str. 2, Freiburg im Breisgau, Baden-Württemberg, Germany, 79112.

of bagging, random subspacing, boosting and pruning on LSTM [18]. The authors in [7] have forecasted the peak load by combining conventional and peak forecasts (data fusion), producing a probabilistic forecast model. For their study, smart meter data and a hypothetical distribution network were used. To evaluate peak forecasts, traditional forecasting metrics can be used. There are some specific metrics for peak forecasting which can be used such as the relative peak demand error (RPDE)[6], the peak absolute percentage error (PAPE) and hit rate (HR) [4].

Another advanced technique is cross-learning, also known as local and global forecasting. The local approach treats each time series independently, similar to individual regression tasks. In contrast, the global approach uses a single forecasting model across multiple time series, extracting information from all of them. An example in the energy sector is the study by [5], which applied global forecasting in a micro-grid setup. The authors forecasted energy demand for various buildings in three countries using a feed-forward neural network and an RNN. The study found that global models performed competitively and had superior training times, as only one model was needed for multiple time series.

Finally, it is crucial to carefully consider the evaluation of forecasting models beyond forecasting metrics. Just as financial asset prediction models are assessed based on their ability to generate profits, energy forecasting models should be evaluated in alignment with their intended outcomes [13]. In the context of energy communities, accuracy alone does not guarantee improved outcomes. In the study of [3] the authors followed a similar approach, creating a community that optimizes itself when forecasts are received as inputs. The objective of the optimization is to control appliances in a way that self-sufficiency and self-consumption are maximized. Their results indicate that forecasting metrics do not necessarily correlate with decreased self-sufficiency. Hence, evaluating forecasting models solely based on forecasting errors may not capture their true value.

In summary, while many forecasting studies focus on traditional model evaluation and peak forecasting, they often overlook flexibility provision and real outcomes like grid congestion. More comprehensive frameworks and advanced metrics are needed to better link forecasting performance with practical applications. Simulation-based evaluations and comparisons of forecasting models, including underexplored techniques like cross-learning and peak forecasting, are crucial for optimizing flexibility.

3 METHODOLOGY

The methodology employed in this doctoral project is structured into three main sections, each addressing distinct aspects: (i) Community-level forecasting for peak shaving provision, (ii) household-level forecasting for congestion management alleviation, and (iii) the interconnection between spatio-temporal dimensions and prediction horizons.

3.1 Community-level forecasting for peak shaving

Challenge: Currently, there is a trend towards utilizing communal flexibility resources for providing flexibility services at the DSO level. Since the DSO measures the load on their transformers, they can request flexibility from the community when needed. However, to effectively do so, the DSO must accurately predict peak loads to signal the community, located at the distribution level, to provide these flexibility services. To achieve this, we propose the following approach.

Proposal: To conduct peak forecasting for the load of the entire community, a measurement apparatus from the local DSO will be positioned at the lower voltage side of the transformer, gathering minute-by-minute data of instantaneous power. These data will be used to train ML models and make predictions. However, establishing an ML pipeline for peak forecasting entails several crucial steps: feature engineering, model selection and HPO. Initially, feature engineering plays a pivotal role, where domain knowledge is leveraged to identify significant features such as flagging historical peak demand periods or anticipated PV production times. Furthermore, traditional time series features, including lag features, time-related features, Fourier features, and rolling or expanding window features, will be integrated into the pipeline. To evaluate the forecasts, various peak forecasting metrics will be used.

3.2 Household-level forecasting for congestion allievation

Challenge: Given that the distribution network where the energy community is situated has successfully rolled out smart meters to all residents, the DSO can now conduct power flow analysis to identify grid congestion both at the transformer and along the distribution lines. To effectively signal each community member to provide flexibility services, the DSO needs to accurately predict load prosumption at each node/household. To achieve this, we propose the following comprehensive approach to ensure the best possible outcomes.

Proposal: To examine forecasting at the prosumer level, a monitoring and control system will be installed by one of the project's collaborating partners. The system will monitor and control flexible appliances, namely, HPs, EVs, batteries and PVs. Similar to the previous methodologies, features derived from domain knowledge will be constructed. For example, flagging of working hours, on/off status of their HP and others. Additionally, feature creation from power flow analysis will be investigated, leading to the creation of physics-informed features, such as a time series depicting the volume of maximum congestion observed in the community grid.

However, in comparison with the peak forecasting pipeline, this model will have to be versatile enough to adapt to the dynamic changes of each household. Following the local forecasting approach, individual models will be trained and optimized for a specific household profile. This entails, that the local forecasting model will receive inputs only on the behaviour of a single household delving deeper into the behaviour of each community participant. On the contrary, following the global modelling approach, a single model will train on multiple time series profiles of similar appliances across the community. In the end, the results of these two architectures will be compared.

3.3 Spatio-temporal analysis of the community level forecasting models on flexibility provision

Challenge: Given that the DSO has a full picture of the prosumption across all nodes within its distribution network, and with an energy community actively engaged in activities such as peak shaving and congestion alleviation, it becomes imperative to evaluate the effects of adjusting the size of the energy community, the granularity of data measurements, and the effectiveness of the prediction horizon. Understanding the impact of changes in community size on the efficiency of flexibility services is crucial for optimizing their delivery. Similarly, investigating the advantages of adjusting data granularity and prediction horizons can offer valuable insights into improving the overall efficiency and precision of these services.

Proposal: To assess the provision of flexibility services, which depends critically on the accuracy of forecasting models, it is essential to use evaluation metrics that directly measure and compare the performance of these models based on community outcomes. Thus, an evaluation metric is essential to directly measure and compare forecasting models based on community outcomes. To facilitate this examination, the methodology will be divided into two blocks. The first block, the forecasting block, will involve predictive models generating forecasts of instantaneous power. The chosen metric for assessment will be the Weighted Absolute Percentage Error (WAPE).

The second one will be the evaluation block which will represent a grid model. By receiving the inputs of the forecasting block (consumption and production profiles) and solving a power flow, congestion in the grid will be identified via appropriate congestion metrics. This framework enables comparing community sizes, granularity and prediction horizon with their impact on congestion and flexibility provision.

4 CONCLUSION AND OUTLOOK

In conclusion, this research focuses on forecasting energy community load for flexibility services provision, investigating different forecasting architectures and evaluation frameworks. It aims to advance forecasting pipelines, establish robust evaluation schemes, and examine forecasting errors across spatio-temporal resolutions. The findings will benefit practitioners by providing transparent, replicable methodologies and real-world insights. Policymakers, DSOs, and community members will also find value in the results, which advocate for flexibility provision and a sustainable energy future.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Professor Anke Weidlich and to Dr. Nicole Ludwig for their invaluable support. Additionally, this PhD project is integrated within the broader context of the Grid-sensitive Energy Community Coordination (GrECCo) which has received funding from the German Ministry of Economic Affairs and Climate Action.

REFERENCES

[1] Akintonde O. Abbas and Badrul H. Chowdhury. 2021. Using customer-side resources for market-based transmission and distribution level grid services – A review. *International Journal of Electrical Power & Energy Systems* 125 (2021), 106480. https://doi.org/10.1016/j.ijepes.2020.106480

[2] Alexandros-Georgios Chronis, Foivos Palaiogiannis, Iasonas Kouveliotis-Lysikatos, Panos Kotsampopoulos, and Nikos Hatziargyriou. 2021. Photovoltaics Enabling Sustainable Energy Communities: Technological Drivers and Emerging Markets. *Energies* 14, 7 (2021). https://doi.org/10.3390/en14071862

[3] Jonathan Coignard, Maxime Janvier, Vincent Debusschere, Gilles Moreau, Stéphanie Chollet, and Raphaël Caire. 2021. Evaluating forecasting methods in the context of local energy communities. *International Journal of Electrical Power & Energy Systems* 131 (2021), 106956. https://doi.org/10.1016/j.ijepes.2021.106956

[4] Shuang Dai, Fanlin Meng, Hongsheng Dai, Qian Wang, and Xizhong Chen. 2021. Electrical peak demand forecasting- A review. (2021). arXiv:2108.01393

[5] Evgenii Genov, Stefanos Petridis, Petros Iliadis, Nikos Nikopoulos, Thierry Coosemans, Maarten Messagie, and Luis Ramirez Camargo. 2021. Short-Term Load Forecasting in a microgrid environment: Investigating the series-specific and cross-learning forecasting methods. *Journal of Physics: Conference Series* 2042, 1 (nov 2021), 012035. https://doi.org/10.1088/1742-6596/2042/1/012035

[6] Georgios Giasemidis, Stephen Haben, Tamsin Lee, Colin Singleton, and Peter Grindrod. 2017. A genetic algorithm approach for modelling low voltage network demands. *Applied Energy* 203 (2017), 463–473. https://doi.org/10.1016/j.apenergy.2017.06.057

[7] Ciaran Gilbert, Jethro Browell, and Bruce Stephen. 2023. Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks. *Sustainable Energy, Grids and Networks* 34 (2023), 100998. https://doi.org/10.1016/j.segan.2023.100998

[8] Carsten Heinrich, Charalampos Ziras, Angeliki L.A. Syrri, and Henrik W. Bindner. 2020. EcoGrid 2.0: A large-scale field trial of a local flexibility market. *Applied Energy* 261 (2020), 114399. https://doi.org/10.1016/j.apenergy.2019.114399

[9] Gerwin Hoogsteen, Aditya Pappu, Bahman Ahmadi, Johann L. Hurink, Edmund W. Schaefer, Cihan Gercek, and Richard P. van Leeuwen. 2022. On the Effects of Active Energy Community Participation in the Energy System. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 1–5. https://doi.org/10.1109/ISGT-Europe54678.2022.9960552

[10] IPCC. 2022. The evidence is clear: the time for action is now. We can halve emissions by 2030. https://www.ipcc.ch/2022/04/04/ipcc-ar6-wgiii-pressrelease/

[11] IRENA. 2022. *World Energy Transitions Outlook: 1.5°C Pathway*. IRENA.

[12] Rolf Naber, Rob Raven, Matthijs Kouw, and Ton Dassen. 2017. Scaling up sustainable energy innovations. *Energy Policy* 110 (2017), 342–354. https://doi.org/10.1016/j.enpol.2017.07.056

[13] Peng Peng, Yuehong Chen, Weiwei Lin, and James Z. Wang. 2024. Attention-based CNN–LSTM for high-frequency multiple cryptocurrency trend prediction. *Expert Systems with Applications* 237 (2024), 121520. https://doi.org/10.1016/j.eswa.2023.121520

[14] Aida Mehdipour Pirbazari, Ekanki Sharma, Antorweep Chakravorty, Wilfried Elmenreich, and Chunming Rong. 2021. An Ensemble Approach for Multi-Step Ahead Energy Forecasting of Household Communities. *IEEE Access* 9 (2021), 36218–36240. https://doi.org/10.1109/ACCESS.2021.3063066

[15] Ricardo Reibsch, Jakob Gemassmer, and Tabea Katerbau. 2024. Low voltage grid resilience: Evaluating electric vehicle charging strategies in the context of the grid development plan Germany. *eTransportation* 20 (2024), 100323. https://doi.org/10.1016/j.etran.2024.100323

[16] Leo Semmelmann, Sarah Henni, and Christof Weinhardt. 2022. Load forecasting for energy communities: a novel LSTM-XGBoost hybrid model based on smart meter data. *Energy Informatics* 5 (2022). https://doi.org/10.1186/s42162-022-00212-9

[17] Nahal Tamadon, Ebrahim Shayesteh, Marco Cupelli, and Antonello Monti. 2018. Local Balancing of Low-Voltage Networks by Utilizing Distributed Flexibilities as Part of the InterFlex Field Trial. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*. 3568–3575. https://doi.org/10.1109/IECON.2018.8591450

[18] Mao Tan, Siping Yuan, Shuaihu Li, Yongxin Su, Hui Li, and Feng He. 2020. Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. *IEEE Transactions on Power Systems* 35, 4 (2020), 2937–2948. https://doi.org/10.1109/TPWRS.2019.2963109

Poster Abstract: Towards Online Meta-Modeling of Communication Networks in Energy Systems

MALIN RADTKE, University of Oldenburg, Germany

The integration of intelligent, partly agent-based control structures into the energy system is transforming the electricity grid into a cyber-physical system (CPES). This integration is deepening the mutual dependencies between communication and electric power systems, which requires systematic testing to ensure functionality, stability, and security. However, traditional simulation methods for cyber-physical energy systems (CPES), especially those involving integrated communication simulation frameworks, are challenged by significant computational demands. The proposed research addresses the trade-off between simulation performance and accuracy through a novel approach. The primary objective is to approximate communication simulation behavior online during simulation execution using a meta-model. The accuracy of the meta-model is continuously evaluated by comparing it to the original simulation. When the meta-model reaches a pre-defined accuracy threshold, it replaces the simulation, offering a significant computational advantage without significantly affecting the accuracy of the results.

CCS Concepts: • **Computing methodologies → Simulation types and techniques**.

Additional Key Words and Phrases: simulation, energy systems, CPES, communication networks, surrogate models

1 INTRODUCTION

The integration of the energy system with intelligent control structures, automated devices and Information and Communication Technologies (ICT) imparts a cyber-physical character to the smart grid [10, 11, 15]. The properties and behavior of the ICT system, such as the topology of the communication network, protocols, communication latency, bandwidth, information security, and reliability issues, influence the interconnected power grid. Conversely, the power grid also influences the ICT system, for instance, through disturbances [11].

As critical infrastructure, systematic testing of new technologies is essential to ensure the functionality, stability and security of Cyber Physical Energy Systems (CPES). However, the high complexity of these systems complicates this process [15]. Analytical testing on a formal basis is not feasible due to the complexity and interdisciplinary nature of these systems [15]. Therefore, the examination of CPES behavior as a system-of-systems is mostly conducted through simulations [10, 11, 15]. Using (co-)simulations, it is possible to efficiently and flexibly investigate dynamic and interdisciplinary interactions between subsystems, such as the previously separate areas of physical energy systems (e.g., power grid, distributed energy resources, heat and gas network, ICT) and therein implemented partially agent-based decentralized applications [15].

Especially for the dimensions and system complexities relevant in the context of coupled energy systems, the overhead is enormous due to the use of models that are not optimized for specific use cases and the need to run custom solvers or execution environments [2]. This is the case, for example, if the communication network is to be simulated using an integrated framework as presented by Frost et al. [3], Oest et al. [9]. Here, messages sent between agents in a multi-agent system (MAS) are simulated in an external network simulator. This results in a performance overhead due to the connection between the frameworks (e.g., over a Transmission Control Protocol (TCP) socket) and might also lead to time synchronization errors.

Building on the challenges of simulating complex CPES, in particular the simulation of communication networks within these systems, a novel approach is proposed to improve performance. The approach is to develop a so-called meta-model or surrogate to approximate the original communication simulation during the simulation. A pre-training is carried out on previously generated training data and the model is then improved online in the simulation. This model would be based on analyzing and learning communication patterns between end devices (such as software agents), with the aim of optimising the simulation process by reducing the need for custom solvers or execution environments. The ability to evaluate the accuracy of these learned patterns is crucial, as it allows to demonstrate the improvement of the meta-model over time during the training process. The hypothesis is that there is an accuracy threshold above which the communication simulation can be replaced by the meta-model. This could lead to a significant increase in simulation speed.

2 RELATED WORK

To highlight the research gap addressed by this project, this section presents related and previous work (an overview can be seen in Figure 1). Subsequently, the points that are not yet covered by the existing work are presented.

Analytical methods, valued for their performance and computational efficiency, simplify complex network interactions into mathematical formulas, allowing rapid assessment of network behavior and performance metrics [16]. Approaches like integrating communication performance indicators [1] and using graph-based models [14] are common. Queueing Petri Nets have also been used for performance modeling [12]. The potential of analytical models for realistically simulating Vehicle-to-Vehicle (V2V) communication has been shown [8], introducing the analytical model CoDiPy to enhance simulation performance and enable large-scale and long simulation runs. Network calculus, another distinct analytical approach, provides a theoretical framework for analyzing performance guarantees in computer networks by computing worst-case bounds on delay and buffer requirements [5].

Detailed Simulations offer high accuracy by modeling intricate system components but are computationally intensive. Frameworks like OMNeT++ and ns-2 (newer version: ns-3) are widely used [10]. Co-simulations, such as cosima integrating OMNeT++ into mosaik [9] and cosima-mango [3], enable comprehensive analysis of CPES.

Author's address: Malin Radtke, malin.radtke@uni-oldenburg.de, University of Oldenburg, Ammerländer Heerstraße 114-118, Oldenburg, Nidersachsen, Germany, 26129.

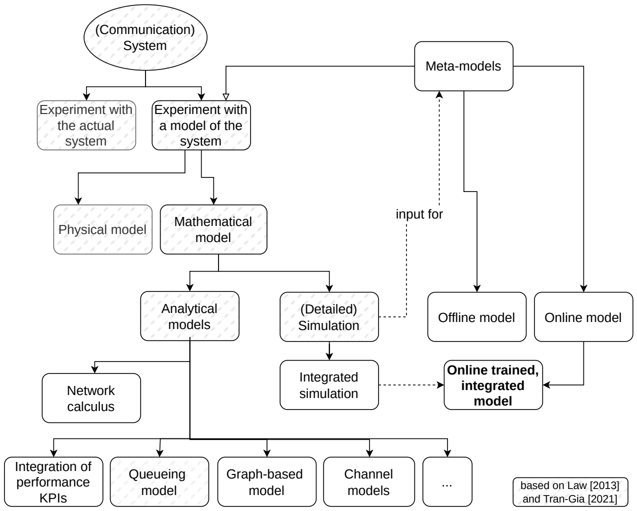


Fig. 1. Overview on related work, based on Law [6], Tran-Gia and Hoßfeld [16]

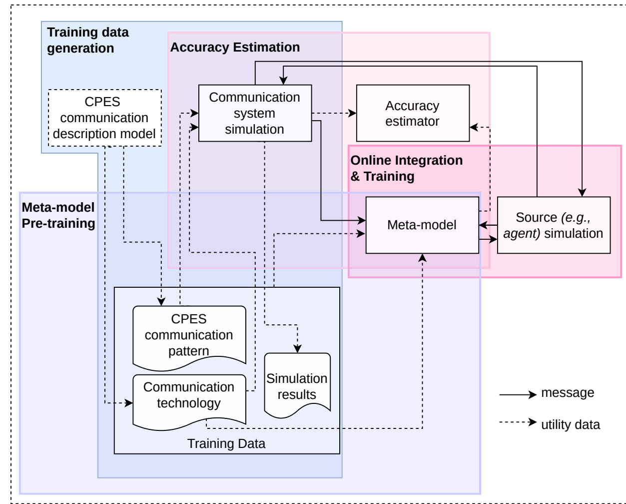


Fig. 2. Schematic overview on Research objectives

Meta-Models approximate the input-output response of the original model to enhance computational speed, either trained offline or online [4]. Offline meta-models have been employed to predict network performance metrics under various conditions [7, 17]. Examples include the RouteNet GNN Challenge, which uses graph-based networks to learn network properties [17], and models for Vehicular Ad-Hoc Networks (VANET) [13].

Although analytical methods offer improved simulation speed, they often sacrifice accuracy. Detailed simulations provide high accuracy but are resource-intensive. Offline meta-models enhance speed but lack dynamic adaptability. The proposed research combines these strengths by developing an online trained, integrated meta-model for communication network simulations in CPES.

3 APPROACH

The overall objective of this research is to develop a meta-model that can dynamically approximate communication network simulations during simulation execution, significantly improving computational efficiency while maintaining high accuracy. The system architecture and research objectives are depicted in Figure 2.

The approach begins with the **training data generation** using a comprehensive communication description model tailored to CPES. This model captures various use cases, input-output combinations, and communication behaviors pertinent to CPES. The generation of realistic and representative training data is crucial as it forms the foundation for training effective meta-models. **Suitable meta-models** are identified through an extensive literature review. These models are then pre-trained using the generated data. The pre-training phase involves comparing the performances of different meta-models to select the most effective ones. The selected meta-models should be capable of handling graph-structured network data and adapting to dynamic network states, which are typical in CPES environments. The **integration of the meta-model** into the simulation environment involves several critical steps. Firstly, the meta-model is formalized as a simulation model to ensure it can seamlessly interact with the ongoing simulation. This formalization involves defining the meta-model's structure, input-output relationships, and state transition mechanisms. The next step is integrating the meta-model into the simulation process. This integration requires the development of synchronization methods to align the meta-model with the simulation's time and events, ensuring consistent and coherent operation. An **accuracy estimation method** is developed to continuously monitor the meta-model's performance during simulation. This method evaluates the meta-model's predictions against the actual simulation data. When the meta-model achieves a predefined accuracy threshold, it replaces the original, more computationally intensive simulation. This substitution significantly enhances the overall simulation speed without sacrificing the accuracy of results.

To evaluate the effectiveness of the proposed approach, a systematic evaluation is conducted. This evaluation compares the integrated meta-model's performance against traditional simulation methods across several metrics, including computational efficiency, accuracy, scalability, and portability. Various scenarios are simulated to demonstrate the practical benefits and robustness of the approach. The scenarios include different CPES configurations, communication network topologies, and operational conditions to ensure a comprehensive assessment.

4 CONCLUSION AND OUTLOOK

The computational overhead associated with the simulation of communication networks within integrated communication simulations has been identified as a major challenge in the testing of new technologies and algorithms in CPES.

To address this challenge, a novel approach is proposed to balance the trade-off between simulation performance and accuracy. The core of this approach is the development of a meta-model to approximate the behavior of communication simulations online during simulation execution. This meta-model is continuously evaluated for accuracy against the original simulation. When a pre-defined accuracy threshold is reached, the meta-model is set to replace the original simulation. This method is expected to provide computational savings while maintaining a high level of accuracy in the simulation results, thus improving the simulation of agent-based communication in CPES. The development of the proposed approach leads to a number of contributions. Some of these are summarized below.

Enabling Preliminary Reliability Analysis. When facilitating the implementation of new use cases through agent-based control structures, conducting preliminary reliability analysis of the MAS under the influence of the communication network is enabled. Scenarios in relevant scales and complexities can be conducted due to the acceleration of the communication simulation.

Advancing Communication Technology Evaluation. The results of this work also contribute to the investigation and advancement of communication technologies and networks, since developers can evaluate the performance of different design configurations without the need for extensive and time-consuming simulations.

General Insights into Online Meta-Modeling and Integration in Agent-Based Simulations. In addition, the intended results provide general insights into online modeling of meta-models and their integration into agent simulations, which can be applied in various domains.

Transferability of the approach approximate communication simulations. The approach may have broader applications than just energy systems and could be applied to any type of end device, not just agents in MAS.

ACKNOWLEDGMENTS

I would like to thank Prof. Dr. Sebastian Lehnhoff for his supervision of my doctoral project, and Prof. Dr. Reinhard German for his advice during the shepherding process.

REFERENCES

[1] Ola Ali, Ahmed Aghmadi, and Osama A. Mohammed. 2024. Performance evaluation of communication networks for networked microgrids. *e-Prime - Advances in Electrical Engineering, Electronics and Energy* 8 (2024), 100521. https://doi.org/10.1016/j.prime.2024.100521

[2] Marita Blank, Malin Gandor, Astrid Nieße, Stefan Scherfke, Sebastian Lehnhoff, and Michael Sonnenschein. 2015. Regionally-Specific Scenarios for Smart Grid Simulations. *2015 IEEE 5th International Conference on Power Engineering, Energy and Electrical Drives* (2015).

[3] Emilie Frost, Malin Radtke, Marvin Nebel-Wenner, Frauke Oest, and Sanja Stark. 2024. cosima-mango: Investigating Multi-Agent System robustness through integrated communication simulation. *SoftwareX* 26 (2024), 101667. https://doi.org/10.1016/j.softx.2024.101667

[4] Ping Jiang, Qi Zhou, and Xinyu Shao. 2020. Surrogate Model-Based Engineering Design and Optimization. In *Surrogate Model-Based Engineering Design and Optimization*. Springer Singapore, Singapore, 1–5. https://doi.org/10.1007/978-981-15-0731-1_1 Series Title: Springer Tracts in Mechanical Engineering.

[5] Wolfgang Kellerer and Amaury Van Bemten. 2016. *Network Calculus: A Comprehensive Guide.* Technische Universitat M ¨ unchen ¨ Lehrstuhl fur Kommunikationsnetze.

[6] Averill M. Law. 2013. *Simulation modeling and analysis* (4. ed., international ed., [nachdr.] ed.). McGraw-Hill, Boston.

[7] Albert Mestres, Eduard Alarcón, Yusheng Ji, and Albert Cabellos-Aparicio. 2018. Understanding the Modeling of Computer Network Delays using Neural Networks. In *Proceedings of the 2018 Workshop on Big Data Analytics and Machine Learning for Data Communication Networks*. ACM, Budapest Hungary, 46–52. https://doi.org/10.1145/3229607.3229613

[8] Michael Niebisch, Daniel Pfaller, and Anatoli Djanatliev. 2022. CoDiPy: Performance Evaluation of Vehicular Cooperative Downloading in Python. In *2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. 461–465. https://doi.org/10.1109/WiMob55322.2022.9941695

[9] Frauke Oest, Emilie Frost, Malin Radtke, and Sebastian Lehnhoff. 2023. Coupling OMNeT++ and Mosaik for Integrated Co-Simulation of ICT-Reliant Smart Grids. *SIGENERGY Energy Inform. Rev.* 3, 1 (jun 2023), 14–25. https://doi.org/10.1145/3607120.3607123

[10] Peter Palensky, Arjen van der Meer, Claudio Lopez, Arun Joseph, and Kaikai Pan. 2017. Applied Cosimulation of Intelligent Power Systems: Implementing Hybrid Simulators for Complex Power Systems. *IEEE Industrial Electronics Magazine* 11, 2 (June 2017), 6–21. https://doi.org/10.1109/MIE.2017.2671198

[11] Peter Palensky, Arjen A. Van Der Meer, Claudio David Lopez, Arun Joseph, and Kaikai Pan. 2017. Cosimulation of Intelligent Power Systems: Fundamentals, Software Architecture, Numerics, and Coupling. *IEEE Industrial Electronics Magazine* 11, 1 (March 2017), 34–50. https://doi.org/10.1109/MIE.2016.2639825

[12] Piotr Rygielski and Samuel Kounev. 2014. Data Center Network Throughput Analysis Using Queueing Petri Nets. In *2014 IEEE 34th International Conference on Distributed Computing Systems Workshops (ICDCSW)*. 100–105. https://doi.org/10.1109/ICDCSW.2014.11

[13] Christina Stadler, Xenia Flamm, Thomas Gruber, Anatoli Djanatliev, Reinhard German, and David Eckhoff. 2017. A stochastic V2V LOS/NLOS model using neural networks for hardware-in-the-loop testing. In *2017 IEEE Vehicular Networking Conference (VNC)*. 195–202. https://doi.org/10.1109/VNC.2017.8275597

[14] Sanja Stark, Anna Volkova, Sebastian Lehnhoff, and Hermann de Meer. 2021. Why Your Power System Restoration Does Not Work and What the ICT System Can Do About It. *Association for Computing Machinery* Proceedings of the Twelfth ACM International Conference on Future Energy Systems (2021), 269–273. https://doi.org/10.1145/3447555.3465415

[15] Cornelius Steinbrink, Marita Blank-Babazadeh, André El-Ama, Stefanie Holly, Bengt Lüers, Marvin Nebel-Wenner, Rebeca Ramírez Acosta, Thomas Raub, Jan Schwarz, Sanja Stark, Astrid Nieße, and Sebastian Lehnhoff. 2019. CPES Testing with mosaik: Co-Simulation Planning, Execution and Analysis. *Applied Sciences* 9, 5 (March 2019), 923. https://doi.org/10.3390/app9050923

[16] Phuoc Tran-Gia and Tobias Hoßfeld. 2021. Performance Modeling and Analysis of Communication Networks. *Würzburg University Press* (2021). https://doi.org/10.25972/WUP-978-3-95826-153-2

[17] Junior Momo Ziazet, Charles Boudreau, Brigitte Jaumard, and Huy Duong. 2022. ADDRESSING ROUTENET SCALABILITY THROUGH INPUT AND OUTPUT DESIGN. *ITU Journal on Future and Evolving Technologies* (2022). https://doi.org/10.52953/GIOD4389

Poster Abstract: Coordinating the Heterogeneity of Aggregators in Digitalised Power Systems

MARCEL OTTE, OFFIS - Institute for Information Technology, Germany

The flexibility potential of decentralised energy resources in the distribution grid raises not only the question of how flexibility will be coordinated between grid operators, but also which role aggregators will have in coordinating flexibility. Aggregators can incentivise customers to provide flexibility, by offering this flexibility in markets or to different service users. However, the aggregated resources (e.g. electric vehicles, photovoltaic systems, battery storage system), possible services (e.g. constraint management, access to wholesale markets or balancing services), service users (e.g. grid operator, balance responsible party) and optimisation goals (e.g. self-sufficiency, sustainability, costs or promotion of communities) can vary, resulting in heterogeneity of aggregators. The European Union also strengthens the role of aggregators in their directive, thus facilitating multiple aggregators for one customer. However, the coordination in a digitalised power system lacks a suitable framework to unfold their full potential. Thus, this project aims to solve the research question: "How can heterogeneous aggregator types be coordinated in a digitalised power system?"

CCS Concepts: • **Hardware → Smart grid**; *Simulation and emulation*; *Renewable energy*; • **Computing methodologies → Model development and analysis**.

Additional Key Words and Phrases: Aggregator, Coordination, Customer, Distribution System Operator, Transmission System Operator, Smart Meter

1 INTRODUCTION

Latest European regulations in combination with the increasing amount of flexibility at the customer premises are the main driver motivating this project. The flexibility potential of electric vehicles, heat pumps, photovoltaic and storage systems at the customer low-voltage level are not fully utilised yet. Nonetheless, this potential is in the interest of various actors, such as grid operators and market participants. On higher voltage levels, aggregators facilitate access to larger flexibilities by combining decentralised energy resources into assets, also known as a virtual power plant. Thus, aggregators have the potential to integrate as well the flexibility from low-voltage systems for suiting it to the grid operator and energy markets needs, and therefore provide an added value to the customer.

The ongoing digitalisation of the power system facilitates flexibility provision from customer premises, but their flexibility activation deviates from conventional and schedule-driven power plants. Enabling flexibility at customer premises can be either conducted through direct or indirect control approaches [9]. For indirect control, a smart meter in place is a prerequisite for monitoring consumption or feed-in in a higher resolution and therefore allows time-of-use tariffs. Controlling the flexibility directly (e.g. reducing the charging power of an electric vehicle) requires a digital communication channel to the resource. While conventional power plants operate on economically optimised schedules, customers on the household level have many different priorities and approximating their behaviour is more difficult [9].

The European directive 2019/944 [4] strengthens the aggregator along with the customer role. In that directive, the regulation defines the term *active customer*, *independent aggregator* and the term *aggregation*. The active customer "[...] consumes or stores electricity generated within its premises located within confined boundaries or,[...] within other premises, or who sells self-generated electricity or participates in flexibility or energy efficiency schemes[...]" Thus, customers can be integrated into aggregation, which is defined as "a function performed by a natural or legal person who combines multiple customer loads or generated electricity for sale, purchase or auction in any electricity market". Moreover, to prevent, that customers are bound to the services of their energy supplier, the independent aggregator was defined, which is "a market participant engaged in aggregation who is not affiliated to the customer's supplier". The independent aggregator is not yet widely implemented on a national level in the European States [3].

Existing research on integrating and coordinating aggregators in the power system focuses primarily on energy economics and policies. However, applying that research area on energy informatics reveals the technical challenges for coordinating the heterogeneity of aggregators.

2 RELATED WORK

Aggregators can act as an intermediary for customers, who want to offer flexibility from their low-voltage systems. Figure 1, which relies on [16], lists the service opportunities of an aggregator along with the target groups. These service opportunities fall into the categories of constraint management, adequacy, wholesale and balancing, in which service users vary from Transmission System Operator (TSO) to Distribution System Operator (DSO) and Balancing Responsible Parties (BRP). An aggregator is not limited to only one service, thus facilitating multiple services to different service users. Activating the flexibility can affect the grid operator coordination, as a service for the TSO may cause congestion for the DSO, where the flexibility is located.

From a socio-technical & policy perspective, the role of the aggregator is recognised in both, research and in regulation. Aggregators are differentiated according to the roles, they can take in (e.g. being a balancing responsible party (BRP)). An overview of aggregator implementation is depicted in Figure 2.

The differentiation according to the roles, which an aggregator ingests, is commonly found in research with just minor differences in the labelling of the setups [1, 6, 10, 11, 16]. The depicted decision

Author's address: Marcel Otte, marcel.otte@offis.de, OFFIS - Institute for Information Technology, Oldenburg, Germany.

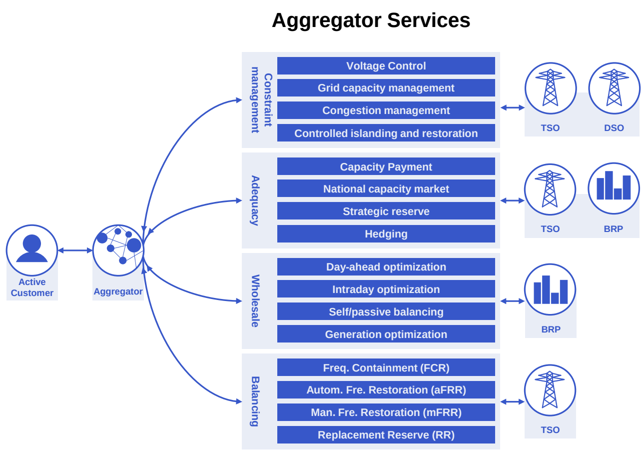


Fig. 1. Aggregator services along with the Customer, TSO, DSO and BRP (adopted from [16]).

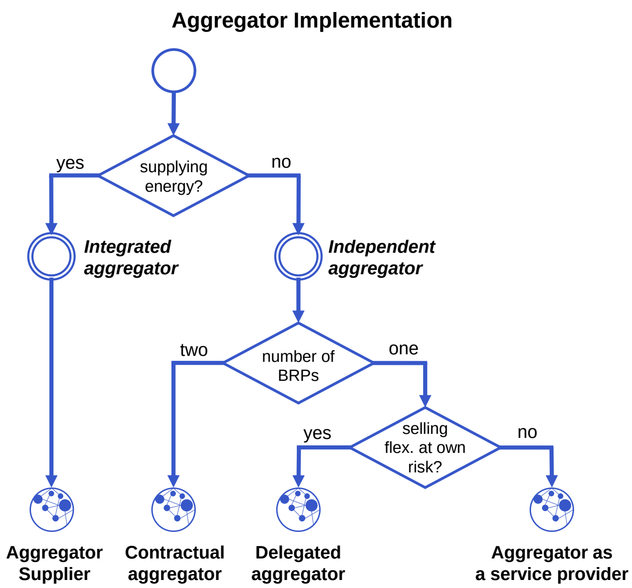


Fig. 2. Aggregator implementation models [1].

tree is taken from [1] and determines on the first branch if the aggregator takes in the role of the supplier. If not, the aggregator is a so-called independent aggregator, which was already outlined in the introduction as one market player, that is strengthened in the European Directive [4]. The independent aggregator facilitates that "customers should be allowed to have multiple contracts with different market participants without one foreclosing the other." [3]. Thus, customers should be able to have a contract with a supplier and an independent aggregator, that enables the flexibility provision. The mechanisms in case of imbalances of the balancing group and the further branches of the decision tree can be found in the cited papers.

Coordination of flexibilities is commonly discussed for grid operator coordination. In a decentralised power system, the flexibility is not any more centred at the transmission level, but also on low-voltage systems, which shifts the focus on grid operator coordination. An overview of different coordination approaches is given by [13]. In their work, coordination mechanisms are analysed, which are centralised flexibility markets, local (DSO) & global (TSO) flexibility markets with either shared resources or shared responsibilities, and common TSO-DSO flexibility markets. In each model, the aggregator is depicted, but the coordination approaches mainly focus on the grid operator perspective. In [2, 15] it is stated that aggregator reduces coordination issues.

Interoperability in a digitalised power system remains an ongoing challenge and therefore spreads out over different research areas. Even if many standards exist, interoperability often remains challenging [12]. An aggregator modelled in a standardised way is not identified in the literature. However, in their work, [11] state that 85 % of the aggregators, which they analysed, use proprietary soft- and/or hardware for the operation of virtual power plants. An approach that tackles interoperability in power systems is the Common Information Model (CIM) [5]. Among others, CIM has the capability to depict domain ontology, communication and serialisation [18]. Shifting the focus to other domains, industrial concepts that tackle similar issues are the asset administration shell or generally the concept of digital twins.

3 CONCEPT & METHODOLOGY

Phase 1: System Design & Requirements: Before developing the software framework for coordinating and integrating aggregators, the aggregator has to be placed into the digital power system. Thus, during the first phase, the project scopes the system perspective by applying the so-called Smart Grid Architecture Model (SGAM) [14], which is well suited for design, development, and validation [17]. SGAM spans over the domains from large-scale generation and the transmission grid to the distribution grid, DER and the customer premises. Moreover, the model differentiates between zones to depict the systems from the process level to the market level. With the aim of a holistic view, one defines the component, communication, information, function and business layer. Thus, using the SGAM, this work foresees a system design that covers the placement of the involved actors, such as TSO, DSO, meter operator, customer and other aggregator. As part of this project, these coordination approaches are incorporated in a publication [7] with a focus on the aggregator perspective.

Phase 2: Aggregator Software Framework Development: After the system design reveals the relevant actors and systems along with different coordination approaches, the software framework can be developed. Firstly, the heterogeneity of aggregators has to be fully explored by deriving an ontology and taxonomy. The proof of heterogeneity is required as their characteristics have to be considered for the coordination within the framework. Accordingly, an aggregator model will be derived based on that heterogeneity using the

Common Information Model [5]. Secondly, the conflicts among actors for flexibility activation and procurement have to be identified and suitable algorithms explored. Afterwards, combining the derived aggregator model with their placement in the power system and the conflicts results in the overall framework that facilitates the coordination of heterogeneous aggregator types.

Phase 3: Laboratory Validation: The aggregator model and the coordination algorithms will be validated in a hardware-in-the-loop-based simulation environment. In another work, which will be published in [8] the concept of the validation environment is derived with a focus on cross-actor and cross-system use cases, which considers among others, the actors from Phase 1, physical/virtual smart meter and a grid simulation. The description of the setup goes beyond the scope of this section, but more details can be found in that submission. The laboratory setup will provide the environment of the system designs from Phase 1 and evaluate how the software framework is suitable for coordination in terms of scalability, execution time, adaptability, expandability and interoperability.

4 CONCLUSION

This project is motivated by the increasing flexibility potentials at the customer premises and the European Directive 2019/944 [4], which both promote aggregators to enable these potentials. However, in this work, it is argued that the lack of a suitable framework for coordination hinders aggregator from unfolding their full potential. Thus, the research question is defined as "How can heterogeneous aggregator types be coordinated in a digitalised power system?" To answer the research question, this project is structured into three phases. Firstly, the system design will reveal the involved actors for coordination and specifies the resulting requirements. Secondly, the framework will be derived to place aggregators in the power system environment. During that phase, one core aspect will be interoperability, and the heterogeneity of aggregators, which will be addressed by modelling them in a standardised way, followed by coordination algorithms. Lastly, a validation, based on a hardware-in-the-loop setup, will be conducted to prove the concept of coordination.

REFERENCES

[1] Fabio Bignucolo, Arturo Lorenzoni, and Jan Marc Schwidtal. 2019. End-users aggregation: a review of key elements for future applications. In *2019 16th International Conference on the European Energy Market (EEM)*. 1–6. https://doi.org/10.1109/EEM.2019.8916520

[2] Scott Burger, Jose Pablo Chaves-Ávila, Carlos Batlle, and Ignacio J Pérez-Arriaga. 2017. A review of the value of aggregators in electricity systems. *Renewable and Sustainable Energy Reviews* 77 (2017), 395–405. https://doi.org/10.1016/j.rser.2017.04.014

[3] European Commission and Joint Research Centre. 2022. Explicit demand response for small end-users and independent aggregators: status, context, enablers and barriers. *Publications Office of the European Union* (2022). https://doi.org/10.2760/625919

[4] European Union. 2022. Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU. *Official Journal of the European Union* (2022). http://data.europa.eu/eli/dir/2019/944/2022-06-23

[5] IEC 61970 2024. *Energy management system application program interface.* Standard. International Electrotechnical Commission, Geneva, CH.

[6] Selina Kerscher and Pablo Arboleya. 2022. The key role of aggregators in the energy transition under the latest European regulatory framework. *International Journal of Electrical Power & Energy Systems* 134 (2022), 107361. https://doi.org/10.1016/j.ijepes.2021.107361

[7] Marcel Otte, Jirapa Kamsamrong, and Sebastian Lehnhoff. 2024. How can Aggregators Improve the TSO-DSO-Customer Coordination in Digitalised Power Systems? *ISGAN* (2024).

[8] Marcel Otte, Carsten Krüger, Sebastian Rohjans, and Sebastian Lehnhoff. 2024. Hardware-in-the-Loop Simulation Environment for Validating Distribution System Control Applications. *[submitted and not yet published]* (2024).

[9] Pierre Pinson, Henrik Madsen, et al. 2014. Benefits and challenges of electrical demand response: A critical review. *Renewable and Sustainable Energy Reviews* 39 (2014), 686–699.

[10] Ksenia Poplavskaya and Laurens De Vries. 2018. A (not so) independent aggregator in the balancing market theory, policy and reality check. In *2018 15th International Conference on the European Energy Market (EEM)*. 1–6. https://ieeexplore.ieee.org/document/8469981

[11] Ksenia Poplavskaya and Laurens de Vries. 2020. Chapter 5 - Aggregators today and tomorrow: from intermediaries to local orchestrators? In *Behind and Beyond the Meter*, Fereidoon Sioshansi (Ed.). Academic Press, 105–135. https://doi.org/10.1016/B978-0-12-819951-0.00005-0

[12] Johann Schütz, Mathias Uslar, and Jürgen Meister. 2021. A case study research on interoperability improvement in Smart Grids: state-of-the-art and further opportunities. *Open Research Europe* 1, 33 (2021). https://doi.org/10.12688/openreseurope.13313.1

[13] Ricardo Silva, Everton Alves, Ricardo Ferreira, José Villar, and Clara Gouveia. 2021. Characterization of TSO and DSO Grid System Services and TSO-DSO Basic Coordination Mechanisms in the Current Decarbonization Context. *Energies* 14, 15 (2021). https://doi.org/10.3390/en14154451

[14] Smart Grid Mandate. 2011. Standardization mandate to European standardisation organisations (ESOS) to support European smart grid deployment. *European Commission: Brussels, Belgium* (2011). https://www.cencenelec.eu/media/CEN-CENELEC/AreasOfWork/CEN-CENELEC_Topics/Smart%20Grids%20and%20Meters/Smart%20Grids/m490_smart-grids_mandate.pdf

[15] Jan Stede, Karin Arnold, Christa Dufter, Georg Holtz, Serafin von Roon, and Jörn C. Richstein. 2020. The role of aggregators in facilitating industrial demand response: Evidence from Germany. *Energy Policy* 147 (2020), 111893. https://doi.org/10.1016/j.enpol.2020.111893

[16] The Universal Smart Energy Framework (USEF). 2021. The Framework Explained. (2021). https://www.usef.energy/news-events/publications/

[17] Mathias Uslar, Sebastian Rohjans, Christian Neureiter, Filip Pröstl Andrén, Jorge Velasquez, Cornelius Steinbrink, Venizelos Efthymiou, Gianluigi Migliavacca, Seppo Horsmanheimo, Helfried Brunner, and Thomas I. Strasser. 2019. Applying the Smart Grid Architecture Model for Designing and Validating System-of-Systems in the Power and Energy Domain: A European Perspective. *Energies* 12, 2 (2019). https://doi.org/10.3390/en12020258

[18] Mathias Uslar, Michael Specht, Sebastian Rohjans, Jörn Trefke, and José M González. 2012. *The Common Information Model CIM: IEC 61968/61970 and 62325-A practical introduction to the CIM.* Springer Science & Business Media.

Poster Abstract: Procedural Generation of Communication Networks in Power Systems

XAVIER WEISS, LARS NORDSTRÖM*, PATRIK HILBER*, and EMRE SÜREN, KTH, Sweden

1 INTRODUCTION

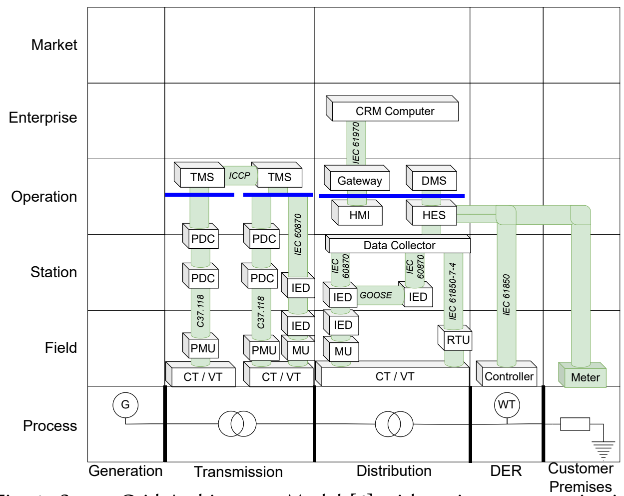


Fig. 1. Smart Grid Architecture Model [4] with various communication protocols

Communication networks enable grid operators to infer and influence the state of the electrical grid, as captured in Fig.1. The design of a network can vary widely depending on the:

- type of physical equipment being monitored.
- type of communication components being used.
- technical, physical and social constraints.

Procedural generation is a method that can generate a large population of power system communication networks. By maintaining a sufficient level of abstraction, it can enable:

- improved generalizability of state-of-the-art models for analyzing, controlling or compromising power system communication networks by validating them on more examples.
- improved performance of data-driven / machine-learning models by generating more diverse and a greater amount of training data.
- synthetic communication network designs to be shared without confidentiality concerns.
- promising candidate network topologies to be discovered when designing new communication systems.
- the general susceptibility of a network to cyberattacks to be estimated without modelling the network in detail.

This work therefore introduces a procedural generation algorithm for synthesizing tree-like communication networks for use in power systems through the following contributions:

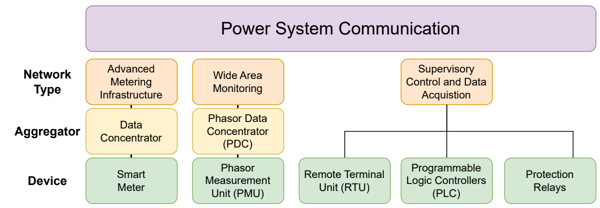
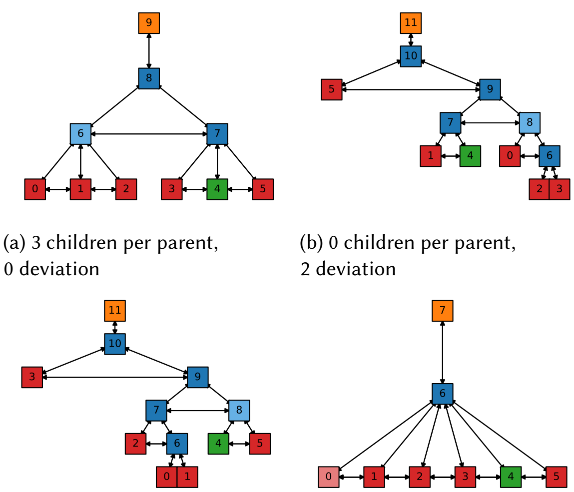


Fig. 2. Taxonomy of Communication Networks

(1) Generation of tree-like communication networks for Supervisory Control and Data Acquisition (SCADA), Wide Area Monitoring systems (WAMs) and Advanced Metering Infrastructure (AMI). As illustrated in Fig.2.
(2) Extensibility to other communication network types through the use of human-readable JavaScript Object Notation (JSON) specification files that encapsulate knowledge specific to the network type.
(3) Integration of communication networks with power systems by assigning communication devices to specific physical equipment based on specified conditions.
(4) Applicability to higher voltage grids by assigning multiple communication devices to aggregated equipment based on specified splitting criteria.

2 PROCESS



(a) 3 children per parent, 0 deviation

(b) 0 children per parent, 2 deviation

(c) -1 children per parent, 3 deviation

(d) Flat hierarchy

Fig. 3. Small communication network topologies with 6 devices, modelled in terms of devices, aggregators, and a control center.

Our procedural generation algorithm, summarized in Alg.1, can produce a range of power-system communication network topologies, as illustrated in Fig.3, based on the following inputs:

(1) A set of hyperparameters:

---

*Both authors contributed equally to this research.

Authors' address: Xavier Weiss, xavierw@kth.se; Lars Nordström, larsno@kth.se; Patrik Hilber, hilber@kth.se; Emre Süren, emsuren@kth.se, KTH, Stockholm, Sweden, 100 44.

Sibling To Sibling Communication: If enabled, lateral communication connections exists between children of the same parent node.

Flat: If enabled, treats the communication network hierarchy as flat, where all devices connect to a single aggregator.

Children Per Parent: Average number of child nodes per parent node.

Children No. Deviation: Random deviation in the number of children per parent.

Proportion: What proportion of the population of devices belong to each category.

Seed: A random number generation seed, used for reproducibility.

(2) A physical **grid** model (optional): a PandaPower [3] network containing a representation of the physical equipment in an electrical grid.

(3) A JSON **spec**ification file: provides information on the possible types of device, aggregator and root nodes. For physical grids it also determines what equipment can be connected to by what devices and how this is split if the equipment is an aggregate (e.g. a load).

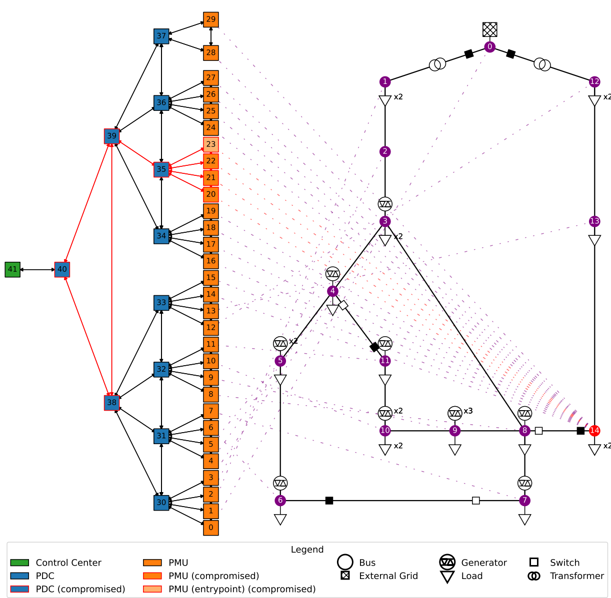


Fig. 4. Example WAMS communication network for CIGRE MV grid [1]

Its output is a set of Python objects that have pointers to any children, siblings or parents they may be connected to. In Fig.4, a procedurally-generated WAM network is shown. Here the devices are Phasor Measurement Units (PMUs) and are connected to buses in the physical grid. The aggregators are Phasor Data Concentrators (PDCs) which can have either PMUs, or other PDCs, as children. Finally, the control center serves as the root node.

3 VALIDATION

Insights into the validity of Algorithm.1 can be obtained by comparing its output structures to a real-world IEC 61850 communication network in Fig.5:

Algorithm 1 Pseudocode of procedural hierarchical network generation algorithm

```
1:  components = [ ]
2:  if Length(components) = 0 then                              ▷ Leaf Nodes
3:      for i = 0, . . . , N do
4:          category = Random(spec[“device”][“categories”])
5:          nSplits = get_n_splits(grid, category, i)
6:          for split = 0, . . . , nSplits do
7:              device ← create_device(category)
8:              components.append(device)
9:  else if Length(components) > 1 then                         ▷ Internal Nodes
10:     n = 0
11:     remaining = [ ]
12:     while n ≠ Length(components) do
13:         nChildren = get_n_children()
14:         if nChildren ≤ 1 then                               ▷ Skip some child nodes
15:             remaining.append(components[n : n + nChildren])
16:         else
17:             category = Random(spec[“aggregator”][“categories”])
18:             aggregator ← create_aggregator(category)
19:             for i = n, . . . , n + nChildren do             ▷ Edges
20:                 connect(aggregator, components[i])
21:                 if i ≥ 1 & sib2sib = “adjacent” then        ▷ Siblings
22:                     connect(components[i − 1], components[i])
23:                 else if i ≥ 1 & sib2sib = “all” then
24:                     for j = n, . . . , n + nChildren do
25:                         connect(components[j], components[i])
26:             remaining.append(aggregator)
27:         n = n + nChildren
28:     components = remaining
29: else if Length(components) = 1 then                         ▷ Root Node
30:     root ← create_root(spec[“root”])
31:     connect(root, components[0])
32:     STOP                                                    ▷ End Recursion
33: else
34:     GOTO LINE 2                                             ▷ Recursion
```

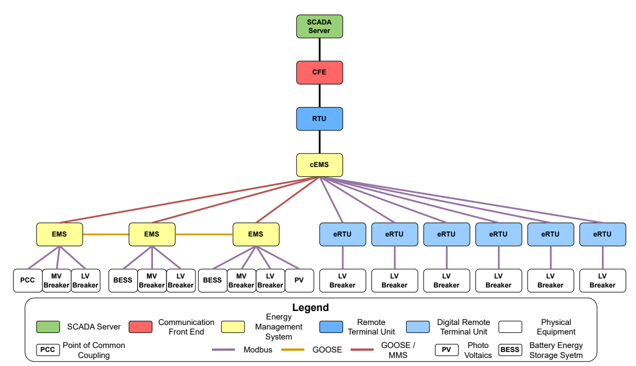


Fig. 5. Diagram of a real-world LV/MV Communication Network [5]

- Real and synthetic communication networks are tree-like.
- Real networks use communication protocols and communicate over physical mediums which may cause delays and data loss. The procedurally generated networks do not model this.
- The population of procedurally generated networks are generated through the sampling of normal distributions, while the real-world population may be skewed.

One application for procedurally generated communication networks is cybersecurity. Here the effect of network generation hyperparameters on the vulnerability to cyberattacks are checked to see whether changes to network structure have a sensible impact.

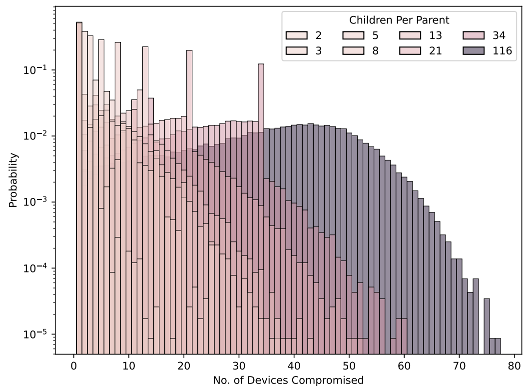


Fig. 6. Variation in compromise distribution with differing number of children per aggregator.

Fig.6 shows varying levels of redundancy in the number of children per parent. For each case a total of 116,000 Monte Carlo simulations are run, 1000 for each possible entrypoint. The results show greater redundancy yields a more cyber-resilient network.

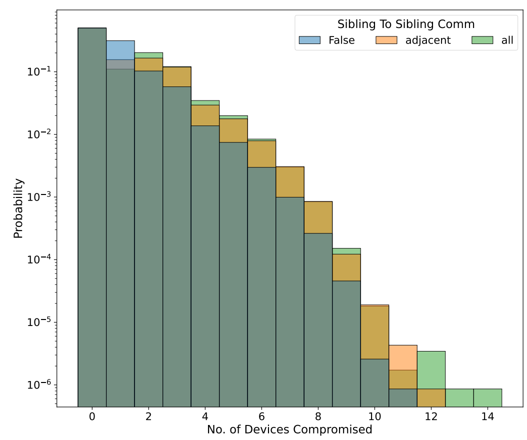


Fig. 7. Variation in compromise distribution with different types of sibling-to-sibling communication.

Fig.7 shows that increased sibling-to-sibling communication leads to a rightwards shift in the compromise distribution. Since the parent can act as a bottleneck for reaching other parts of the network, the effect of lateral movement is, however, more limited.

In Fig.8 a single CPU core was used to generate 10,000 networks with worst case (most redundant), average case (randomized no. of children per aggregator) and best case (least redundant) network complexities. Evidently, there is a linear increase in computation time with network size. Based on the trends, it would take approximately 80 seconds to generate a network containing 1 million devices (a quick test of this took ~ 77.1 seconds). Given that such networks can be generated offline and typically only need to be generated once, the authors' consider this to be sufficiently efficient [2] for topology generation - especially if an additional speed-up from parallelization is included.

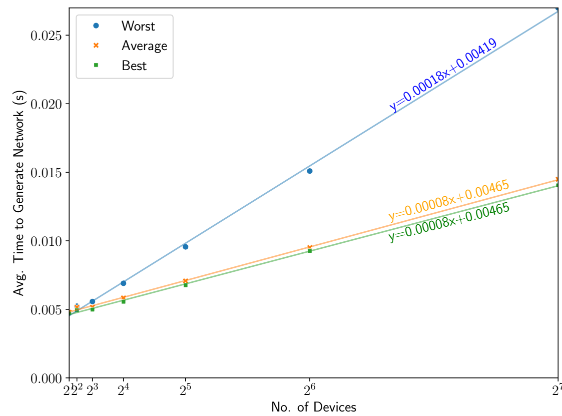


Fig. 8. Average computational time taken to generate one procedural network, based on 10,000 single-thread trials.

4 CONCLUSION

The proposed procedural generation algorithm is able to generate a variety of tree-like power system communication networks. These networks are designed to be abstract representations of the layout of a typical communication network, since they are used to evaluate cybersecurity. Domain experts can specify new network types through a specification file that specifies the cybersecurity characteristics of components in the network, the proportion of different device types, and the rules for determining where and how many devices are placed. Example network types are provided for SCADA, AMI, WAM and generic networks. The generic network type was then checked in the cybersecurity domain by varying the hyperparameters of the algorithm. The results match expectations, with increased redundancy yielding networks with fewer compromises. Future work may validate this algorithm through expert feedback and potentially extend it to satisfy reliability, cost, or geographical constraints. In addition, by including communication network protocols it would become possible to simulate network activity, information loss, and latency for research testbeds.

5 ACKNOWLEDGMENTS

Funded by Energimyndigheten, project number: P2022-00692

REFERENCES

[1] Stefano Barsali et al. 2014. *Benchmark systems for network integration of renewable and distributed energy resources.*

[2] A. Medina, A. Lakhina, I. Matta, and J. Byers. 2001. BRITE: An Approach to Universal Topology Generation. In *MASCOTS 2001, Proceedings Ninth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems.* 346–353. https://doi.org/10.1109/MASCOT.2001.948886

[3] Leon Thurner, Alexander Scheidler, Florian Schäfer, Jan-Hendrik Menke, Julian Dollichon, Friederike Meier, Steffen Meinecke, and Martin Braun. 2018. Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. *IEEE Transactions on Power Systems* 33, 6 (2018), 6510–6521. https://doi.org/10.1109/TPWRS.2018.2829021

[4] Mathias Uslar and et al. 2019. Applying the Smart Grid Architecture Model for Designing and Validating System-of-Systems in the Power and Energy Domain: A European Perspective. *Energies* 12 (01 2019), 258. https://doi.org/10.3390/en12020258

[5] Yiming Wu and et al. 2023. An implementation of IEC 61850 for microgrid control. In *27th International Conference on Electricity Distribution.* IET.