Dynamic Tariff Adoption in Electricity Markets: Bridging Policy, Technology and Behavior

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)
von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation

von
M.Sc. Leo Semmelmann

Tag der mündlichen Prüfung: 17.04.2025
Referent: Prof. Dr. Christof Weinhardt
Korreferent: Prof. Dr. Wolf Fichtner
Prüfer: Prof. Dr. Clemens van Dinther

Karlsruhe, 2025

ACKNOWLEDGEMENTS

Diese Dissertation wurde von einer Vielzahl von Menschen positiv beeinflusst, denen ich zutiefst dankbar bin.

An erster Stelle möchte ich mich bei meinem Doktorvater Prof. Dr. Christof Weinhardt bedanken, für seine fortwährende Unterstützung, sein Input aus spannenden Ideen und hilfreichem Feedback, und das Schaffen einer zugleich angenehmen und produktiven Forschungsumgebung. Weiterhin danke ich meinem Korreferenten Prof. Dr. Wolf Fichtner sowie Prof. Dr. Clemens van Dinther und Prof. Dr. Andreas Geyer-Schulz für die konstruktiven Diskussionen im Rahmen des Prüfungsausschusses meiner Promotion.

Ein besonderer Dank geht auch an Prof. Dr. Philipp Staudt, von dem ich ungemein viel in den Jahren meiner Promotion und zahlreichen Forschungsdiskussionen lernen durfte und der viele meiner Forschungsprojekte mit Verve unterstützt hat. Ich möchte mich auch bei der Smart Grid and Energy Markets Forschungsgruppe bedanken, sowohl für die akademische Unterstützung, auch als für die gute Zeit und die Wohlfühlatmosphäre, die sich immer mehr nach einem Forschen mit Freunden als mit Kollegen angefühlt hat. Danke an Kim, Christina, Oliver, Philipp, Laura, und die Alumnis Sarah, Frederik, Marc, Bent, Armin, Patrick, die mich mit offenen Armen in die Gruppe aufgenommen haben. Ich möchte mich auch bei dem gesamten Lehrstuhl- und Lehrstuhlteam bedanken, für die Unterstützung und die angenehme Atmosphäre. Zudem möchte ich mich bei meinen Student:innen bedanken, mit denen ich viele spannende Projekte im Rahmen von Masterarbeiten und Seminaren angehen konnte.

A big thank you also goes to Kevin, for hosting me at the Herrick Labs at Purdue, giving me a warm welcome at the group and a unique practical view on many heat pump-related topics that I previously rather considered from a theoretical perspective. Also, thanks to Levy, Elias, Priyadarshan for the fun joint projects.

Auch bei meinen internationalen Ko-Autorinnen Katharina und Anya möchte ich mich für die angenehme und produktive Zusammenarbeit bedanken.

Zudem geht ein großes Dankeschön an meine Freunde und Familie, die mir immer ein Rückhalt und Unterstützung waren, danke Pascal, Nox, Kenny, Stella, Kim, Sophia, Sara, Jule, Jutta, Charmin und Max. Danke Jan-Erik, dass du mich in den letzten Jahren stets ermutigt hast, mir den Rücken frei gehalten hast und immer ein guter Freund bist. Auch ein besonderer Dank geht an meine Eltern, meine Geschwister, und Großeltern, die mich stets gefördert und unterstützt haben und mir ein großer Rückhalt waren. Zuletzt möchte ich mich bei Tamina bedanken - deine Unterstützung und Liebe waren ein positiver Einfluss, den ich weder in Worte beschreiben kann, noch für den ich mich genug bedanken könnte.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ARIMA	Seasonal Auto-Regressive Integrated Moving Average
BESS	Battery Energy Storage System
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
COP	Coefficient of Performance
EPEX	European Power Exchange
EV	Electric Vehicle
HH	Household
HP	Heat pump
kWh	Kilowatt-hour
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
PV	Photovoltaics
RMSE	Root Mean Square Error
RTP	Real-time pricing
SLP	Standard Load Profile
SOC	State of Charge
SVM	Support Vector Machine
TOU	Time of Use
TSO	Transmission System Operator
XGB	XGBoost

Part I.

Foundations

CHAPTER 1

INTRODUCTION

Global energy markets are experiencing a significant transition, shifting from centralized, fossil-fuel-dominated generation systems to more decentralized and renewable energy-focused ones (Gielen et al., 2019). The process of energy transition can be described in two phases (Markard, 2018). The first phase centers on introducing and initially developing innovative yet immature technologies like renewable energy sources, which rely heavily on public policy support and are implemented in pilot projects and niche markets. In contrast, the second phase marks these technologies' maturation and large-scale adoption, driving significant transformations in industries, business models, institutional frameworks, and complementary systems such as energy storage and smart grids. With renewables already accounting for 30% of global electricity production in 2023—a figure projected to rise to 46% by 2030—the energy transition is entering a critical stage (International Energy Agency (IEA), 2024b). The main challenges now arise in the second phase, where the focus transitions from introducing new technologies to scaling them up and integrating these mature solutions into existing systems while considering their broader impact on households, industries, business models, and infrastructure.

One of these challenges is the intermittent nature of renewable energy sources like wind and solar, which creates significant difficulties in maintaining the real-time balance between the supply and demand of electricity (Hirth and Ziegenhagen, 2015). This shift highlights the need for greater system flexibility and dynamic mechanisms to manage energy consumption patterns, aligning electricity use with scarcity signals from the market, which is also called demand response (Klein et al., 2019).

Dynamic tariffs are seen as a key tool to enable demand response in modern

power systems (Faruqui et al., 2010). By motivating electricity consumers to switch from static electricity tariffs to time-varying ones, peak demand and overall system costs can be decreased since the use of expensive peak power plants can be reduced (Faruqui et al., 2010). Dynamic electricity tariffs can be implemented in different forms, ranging from Time-of-Use tariffs, where prices vary by the time of day to encourage off-peak usage (Nicolson et al., 2018), to Real-Time Pricing, where prices are based on wholesale market rates and change hourly or more frequently (Doostizadeh and Ghasemi, 2012), to Critical Peak Pricing, which imposes higher rates during peak demand events (Herter, 2007), to further forms. Although the implementation of dynamic tariffs varies, their underlying rationale remains the same: consumers should have an economic incentive to shift their electricity consumption in response to power system signals. Specifically, they should reduce consumption during periods when peak power plants are needed (Faruqui et al., 2010) and increase it when there is surplus electricity from renewable sources, such as during times of high solar or wind generation that would otherwise be curtailed.

1.1 Motivation

Although dynamic tariffs have been extensively researched, their implementation and uptake remain limited. For example, as of 2024, only 7% of households in Germany have subscribed to dynamic tariffs (Verbraucherzentrale Bundesverband e.V., 2024), despite their availability for several years and a government mandate requiring utilities with more than 100,000 customers to offer dynamic tariffs starting in 2025 (Bundesministerium für Wirtschaft und Klimaschutz, 2023). Low adoption rates of time-varying rates can also be observed in several other countries, i.e., France (Cabot and Villavicencio, 2024) or the United States (Pereira and Marques, 2023).

On the way to a higher share of households subscribing to dynamic tariffs, there are various challenges and barriers. First, utilities offering dynamic tariffs must enhance their understanding of customer load patterns and improve their ability to forecast future electricity consumption. This is essential for minimizing energy costs and making informed trading decisions (Khan et al., 2016). As discussed in Khan et al. (2016), predicting household electricity consumption under dynamic tariff-induced demand response presents several challenges. These include ensuring user data privacy, developing scalable approaches for large datasets, and effectively interpreting

smart meter data. Additionally, novel forecasting methods are needed to improve accuracy, especially when operators must make decisions without historical data. Introducing new high-energy-demand devices, such as heat pumps (Love et al., 2017) and electric vehicles (Babrowski et al., 2014), further exacerbates these challenges by significantly altering household load profiles. Second, household loss aversion poses a significant barrier to the adoption of dynamic tariffs. A UK study demonstrated this, revealing that two-thirds of households were reluctant to adopt time-varying electricity tariffs (Nicolson et al., 2017). In particular, loss-averse households were substantially less willing to switch to dynamic tariffs. Third, while being beneficial on a power system level, the introduction of dynamic tariffs can also impose technical challenges on the distribution grid, leading to so-called avalanche effects when a high number of flexible devices react to price signals in the same way and, thereby, lead to higher peak loads and additional stress on the grid (Kühnbach et al., 2021).

Uncertainty lies at the heart of the challenges surrounding the widespread adoption of dynamic tariffs: utilities and operators face uncertainty about future household loads and the impact of new devices like heat pumps, households are uncertain about the price risks associated with dynamic tariffs, and grid operators are uncertain about how increasing adoption rates will affect stress on the distribution grid.

In this uncertain environment, lawmakers and practitioners must make critical decisions regarding investments, strategies, and policies to pave the way for greater demand response in power systems through the implementation of dynamic tariffs.

This thesis aims to reduce the uncertainty surrounding dynamic tariffs by integrating empirical studies, introducing novel algorithms, advancing operational approaches for dynamic tariff design in households, and developing frameworks to evaluate their impact on distribution grids. In doing so, it contributes to the aforementioned second phase of the energy transition, potentially enhancing system benefits through tariffs aligned with system signals and promoting the effective use of expanding flexibility resources and renewable generation capacity. Through the outcomes of this dissertation, practitioners, lawmakers, and researchers are enabled to make more informed decisions regarding strategic planning and policy design for the regulation of dynamic tariffs. This thesis establishes policy recommendations for managing flexibility potential, highlights the potential of novel guarantees that can alleviate households' loss aversion towards time-varying tariffs, and devises approaches for

designing grid charges to mitigate potential avalanche effects.

This thesis is structured into six major parts. While the first part of this thesis describes the fundamentals of dynamic tariffs, the second part lays the groundwork for the subsequent contributions by examining how household battery storage systems and their flexibility potential are regulated globally. It then empirically investigates the system-wide impact of widespread self-consumption-favoring policies combined with static electricity tariffs. Through a first-of-its-kind large-scale field study and subsequent simulation analyses, this part demonstrates the economic benefits of dynamic tariffs for the power system.

The third part of this thesis addresses the operational uncertainties practitioners face when implementing dynamic tariffs. This section explores forecasting challenges related to household loads, particularly in the presence of novel heat pump installations and state-of-the-art machine learning models. It also advances methods for peak load and peak time forecasting while incorporating privacy-preserving techniques and examines the generation of synthetic heat pump data. Although these contributions are not exclusively tailored to households adopting dynamic tariffs, they provide valuable insights into predicting and interpreting household loads—an essential step for effectively managing flexibility potentials within the framework of dynamic tariffs.

The fourth part of this thesis addresses households' aversion to dynamic tariffs by mitigating the associated cost risks. It proposes household-specific electricity price guarantees offered by aggregators in exchange for the right to manage the households' flexibility potential. This approach is enabled by a detailed formulation of the cost minimization problem for households equipped with battery storage systems and heat pumps under time-varying electricity prices, considering the thermal inertia of buildings, their flexibility potential, behavioral constraints, and the uncertainties of future price curves and weather profiles. These guarantees are evaluated from the economic perspective of aggregators alongside a proposed decision support system to aid in their formulation. The proposed guarantees enable the demand response potential of households while shielding them from cost risks.

In the scope of the fifth part, the technical uncertainty of increasing shares of dynamic tariff adoption - in terms of unclear associated grid stress and reinforcement costs - is addressed with an open-source simulation framework that enables

policymakers to evaluate the impact of different grid charge policy options.

In summary, this dissertation addresses key challenges on the path to higher dynamic tariff adoption by providing tools, methods, and analyses to mitigate operational, behavioral, and technical uncertainty. By doing so, it contributes to a greater utilization of demand response potential in modern power systems, ultimately fostering a greener, more cost-effective, and efficient energy landscape.

1.2 Research questions

The previously described challenges related to the adoption of dynamic tariffs are addressed through 11 specific research questions, which are outlined in the following.

Although various studies highlight the positive effects of dynamic tariffs on power systems, these studies are mostly simulation-based. Therefore, Research Question 1 refers to an empirical investigation of battery storage systems operated under self-consumption promoting regulation with static tariffs, which then serves as a foundation for a simulative assessment of more dynamic-tariff-oriented regulatory approaches in Research Question 2.

Research Question 1 *What are the empirically observed effects of self-consumption-promoting regulation on the operation of battery energy storage systems and their economic impact on the power system?*

Research Question 2 *How do dynamic regulatory approaches and alternative influencing factors enhance the value of battery energy storage systems for both the energy system and their owners?*

The actual implementation of dynamic tariffs from an operational perspective in combination with household energy management systems requires accurate and privacy-preserving forecasting methods, as well as available historical data. This is especially important for new large, electricity-consuming devices like heat pumps. Hence, Research Question 3 investigates effective load forecasting algorithms for heat pump loads, such as Long Short Term Memory (LSTM) neural networks and Transformer models. Research Question 4 investigates the load forecasting task from an aggregator perspective with smart meter data and focuses on peak loads. When aggregators or utilities operate households' home energy management systems, ensuring secure and privacy-preserving handling of household load data is essential. Hence, Research Question 5 investigates how privacy-preserving peak time forecasts

can be conducted with a Learning-to-Rank model. To improve the data availability for large and flexible heat pump loads, Research Question 6 investigates the accuracy of a novel method to generate synthetic heat pump load profiles.

Research Question 3 *To what extent do the same forecasting methods used for day-ahead predictions of traditional household loads also perform effectively for heat pump loads?*

Research Question 4 *How can smart meter data and hybrid LSTM-XGBoost models improve day-ahead aggregated load forecasts for energy communities, particularly in addressing insufficient peak load predictions?*

Research Question 5 *How can a Learning-to-Rank XGBoost model improve peak time forecasting accuracy while offering privacy-preserving advantages compared to conventional XGBoost models?*

Research Question 6 *How can synthetic heat pump load profiles be generated using a k-means clustering-based approach?*

While the research questions above focus on operational aspects of dynamic tariffs, in particular the forecasting and data availability of loads, the following questions focus on an integral barrier to dynamic tariff adoption: the loss aversion of households. Therefore, a novel scheme where aggregators control the dynamic tariff operation of households and offer household-individual price guarantees in return is suggested. Therefore, Research Question 7 tackles the formulation of the operation of home energy management systems with respect to the flexibility of battery storage systems, heating requirements, and the thermal inertia of buildings. In Research Question 8, the economic value of aggregators taking over the control of the dynamic tariff-based operation of household flexibility potentials is identified, while Research Question 9 analyzes the quality of a quantile regression-based decision support system that suggests household individual price guarantee levels.

Research Question 7 *How can the operation of household home energy management systems under dynamic tariffs be modeled by incorporating flexibility potentials like battery storage, heat storage, heat pumps, and building thermal inertia, alongside behavioral constraints such as thermostat setpoint profiles?*

Research Question 8 *What is the economic value of managing household flexibility potentials for aggregators, and how can it be quantified under the proposed price guarantee mechanism?*

Research Question 9 *How accurate are quantile regression-based predictions of household-level electricity price guarantees?*

On the path to a higher share of dynamic tariff adoption, there are also technical constraints. As described before, the simultaneous reaction of households and their flexibility potential can lead to new peak loads, also often denoted as "avalanche effects". Hence, Research Question 10 investigates the impact of increasing shares of dynamic tariff adoption on grid reinforcement costs, and Research Question 11 investigates the impact of regulatory options on these.

Research Question 10 *How does increasing adoption of dynamic tariffs by residential households impact distribution grid reinforcement costs across different grid topologies?*

Research Question 11 *How effective are alternative regulatory options for grid charges and PV feed-in remuneration in alleviating grid reinforcement costs given increasing dynamic tariff adoption?*

These research questions are addressed through a combination of empirical and simulation-based studies, emphasizing the open-source publication of datasets and code. The goal is to promote the adoption of dynamic tariffs in modern power systems. The contributions focus on mitigating uncertainties related to regulatory frameworks, technological advancements, and behavioral responses associated with dynamic tariffs.

1.3 Thesis structure

The structure of this thesis is depicted in Figure 1.1. In the scope of the first part of the thesis, the fundamentals of dynamic tariffs are outlined following this introduction (Chapter 2). The second part of the thesis includes an international analysis of regulations governing household battery storage systems. An empirical field evaluation of the impact of battery storage systems on the power system under self-consumption policies follows this. Additionally, the study explores the effects of alternative regulatory approaches that focus on more dynamic, tariff-oriented options (Chapter 3).

The third part focuses on mitigating operational uncertainty by improving forecasting and synthetic data generation methods. In particular, Chapter 4 investigates the impact of heat pump installations on day-ahead load forecasting; Chapter 5 in-

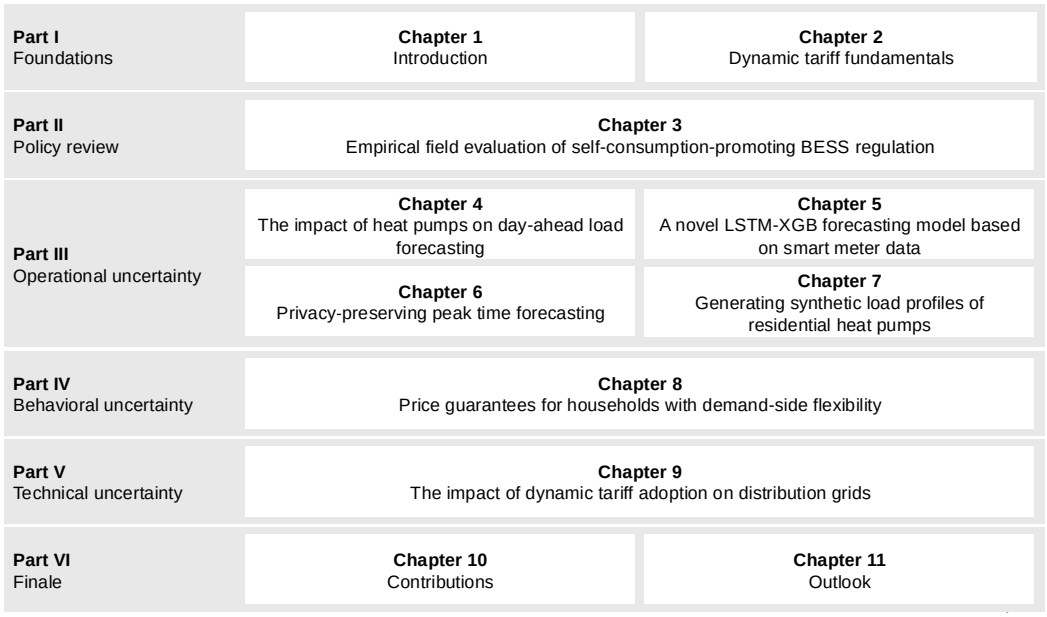


Figure 1.1.: The structure of this thesis

troduces a novel algorithm that improves day-ahead load forecasts on the basis of smart meter data and a special peak load handling algorithm; Chapter 6 focuses on privacy-preserving peak time forecasting and Chapter 7 introduces a novel method to generate synthetic heat pump load data.

The fourth section addresses households' loss aversion towards dynamic tariffs, which could hinder their widespread adoption. It introduces and evaluates household-level price guarantees provided by aggregators as a potential solution (Chapter 8).

The fifth part of this thesis examines the potential technical uncertainties faced by grid operators due to the additional stress on the distribution grid caused by simultaneous household responses to dynamic tariff price signals. Furthermore, it explores the impact of various grid charge and feed-in remuneration design options on grid reinforcement costs (Chapter 9).

Finally, the sixth part summarizes the answers to the main research questions addressed in this thesis and highlights the key contributions (Chapter 10). The concluding Chapter 11 offers an outlook on promising directions for future research.

Chapters 3, 4, 5, 6 and 7 are based on published articles, while Chapter 8 is a working paper and Chapter 9 is forthcoming. In Chapters 3 to 9, I consistently use "we" to reflect the collaborative nature of the research, as the underlying articles were developed in partnership with fellow researchers.

The chapters of this dissertation are based on the following articles:

Chapter 3: L. Semmelmann, M. Konermann, D. Dietze, and P. Staudt. *Empirical field evaluation of self-consumption promoting regulation of household battery energy storage systems*, Energy Policy, 2024.

Chapter 4: L. Semmelmann, M. Hertel, KJ. Kircher, R. Mikut, V. Hagenmeyer and C. Weinhardt. *The impact of heat pumps on day-ahead energy community load forecasting*, Applied Energy, 2024.

Chapter 5: L. Semmelmann, S. Henni and C. Weinhardt. *Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data*, Energy Informatics, 2022.

Chapter 6: L. Semmelmann, O. Resch, S. Henni and C. Weinhardt. *Privacy-preserving peak time forecasting with Learning to Rank XGBoost and extensive feature engineering*, IET Smart Grid, 2024.

Chapter 7: L. Semmelmann, P. Jaquart, and C. Weinhardt. *Generating synthetic load profiles of residential heat pumps: a k-means clustering approach*, Energy Informatics, 2023.

Chapter 8: L. Semmelmann, S. Kimbrough, and P. Staudt. *Price guarantees for households with demand-side flexibility potential and thermal building inertia under dynamic electricity tariffs*, Working Paper, 2025.

Chapter 9: L. Semmelmann, K. Kaiser, A. Heider, K. Kircher, G. Hug, C. Weinhardt. *Analyzing the Impact of Dynamic Tariff Adoption and Regulatory Options on Distribution Grids with an Open-Source Framework*, forthcoming in: Proceedings of the Sixteenth ACM International Conference on Future Energy Systems, 2025.

CHAPTER 2

DYNAMIC TARIFF FUNDAMENTALS

In this chapter, an overview of the most important aspects of dynamic tariffs, as discussed in both literature and practice, is presented. First, a detailed overview of various forms of dynamic tariff design is provided. Second, findings related to the impact of dynamic tariffs on the power system are examined. Third, the role of home energy management systems and flexibility potentials in enabling dynamic tariff adoption is explored. The insights presented in this chapter serve as a foundation for the subsequent parts of this dissertation.

2.1 Types of dynamic tariffs

Traditional electricity tariffs are designed to be simple and easy to understand: households pay a flat volumetric rate, with a fixed price per kWh of electricity consumed each month (Matisoff et al., 2020). These tariffs are typically adjusted periodically, such as once per year. While this structure is straightforward to explain to customers, it fails to reflect the varying costs of electricity supply. Besides that, volumetric electricity tariffs can also raise concerns about fairness. For instance, when some households consume disproportionately large amounts of electricity during high-demand hours—when expensive gas plants must be operated—the associated costs are socialized among all electricity consumers (Johnson et al., 2017).

Hence, policymakers are considering more cost-reflective, time-varying rates, also referred to as dynamic tariffs. Below is an overview of the most prevalent forms of dynamic tariff designs, explained in the context of their cost-causality, derived from Matisoff et al. (2020). An illustration of the different described tariff designs is depicted in Figure 2.1.

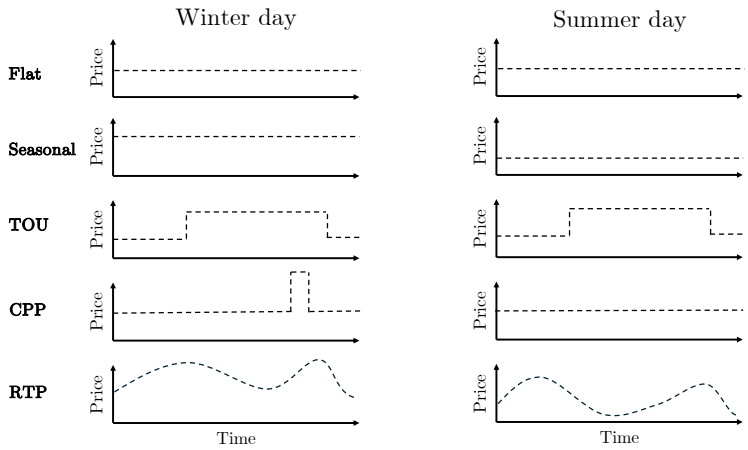


Figure 2.1.: Schematic illustration of electricity tariff designs

Seasonal Pricing: Seasonal differentiation of flat volumetric rates adjusts electricity prices based on the season, reflecting the higher demand during winter and summer periods or changes in fuel costs. While relatively simple, this approach introduces a basic level of temporal variation to account for predictable seasonal cost differences. However, this simple approach neglects potential diurnal patterns of supply and demand in power systems.

Time-of-Use Pricing: Time-of-Use (TOU) pricing establishes regular rate schedules with a limited number of price fluctuations throughout the day based on historical supply and demand patterns. This method provides predictable price signals, helping consumers shift their consumption to off-peak hours. TOU pricing has demonstrated that residential demand is responsive to such temporally varying prices, effectively reducing peak loads, as shown in empirical evidence from Italy (Torriti, 2012) or Spain (Enrich et al., 2024).

Critical Peak Pricing: Critical Peak Pricing (CPP) introduces significantly higher rates during critical peak demand hours, often triggered by extreme system conditions. While these price changes are not announced far in advance, they are typically limited in duration and frequency, with pre-set price levels. Variations, such as Variable Peak Pricing and Critical Peak Rebates, further refine this ap-

proach by either charging different rates at peak times or rewarding consumers for reducing demand during critical events. Empirical evidence from California shows that households react to these critical peak price signals by adjusting their consumption (Herter and Wayland, 2010). Figure 2.1 illustrates an exemplary critical peak event on a winter day, accompanied by a higher price, in contrast to a summer day without such an event.

Real-Time Pricing: Real-Time Pricing (RTP) links electricity prices directly to time-varying marginal costs of electricity provision. Prices adjust dynamically in response to changing system conditions, offering the highest level of temporal disaggregation. A typical implementation of RTP forwards day-ahead wholesale market electricity prices to users (Guo and Weeks, 2022; Häseler and Wulf, 2024). This implementation can already be

In the remainder of this dissertation, the term "dynamic tariffs" refers to the day-ahead market RTP implementation, which is already being offered in practice by some German electricity retailers (Tibber, 2023).

2.2 Home energy management systems and flexible devices

For dynamic tariffs to positively influence the energy system, it is essential that household consumption[1] aligns with price signals (Matisoff et al., 2020). This alignment can be achieved through either human or technical responses. While some empirical evidence suggests that humans do react, at least to some extent, to dynamic price signals (Torriti, 2012; Herter and Wayland, 2010; Enrich et al., 2024), the literature generally assumes that household loads remain inflexible even under dynamic tariffs. Instead, research primarily focuses on automated devices that respond to dynamic prices (Stute and Kühnbach, 2023; Avau et al., 2021; Miletić et al., 2022). Consequently, this dissertation adopts the common assumption of inflexible household loads and emphasizes the flexibility potential of automatable devices. While first studies in the field were focused on load shifting of household appliances under variable prices (e.g., dishwashers and washing machines, as in Gottwalt et al.

[1]While commercial and industrial customers can also subscribe to dynamic tariffs, this thesis focuses on households for the sake of conciseness.

(2011)), recent studies are instead focusing on novel devices, such as heat pumps and electric vehicles (Stute and Klobasa, 2024). The following sections provide an overview of four key technologies for demand response under dynamic tariffs, which are emphasized throughout this thesis.

Battery storage systems: Battery storage systems, especially when integrated with PV systems, are widely acknowledged for their ability to enhance household energy flexibility (Parra and Patel, 2016). By storing surplus PV-generated electricity during periods of low demand, these systems allow households to use the stored energy during times of high demand, adverse weather conditions, or elevated electricity prices. This functionality facilitates energy time-shifting, helping households reduce reliance on grid imports while increasing self-consumption (Luthander et al., 2015). Accordingly, Chapter 2 of this dissertation explores the effects of self-consumption policies on the utilization of household battery storage systems and compares these outcomes with operational strategies aligned with dynamic tariffs.

Electric vehicles: EV charging has inherent flexibility since the charging process is not strictly tied to the plug-in and plug-out times (Sørensen et al., 2023). Within the available time window, and considering the technical constraints of the EV and the charging device, the desired charging energy can be distributed flexibly, enabling the use of dynamic tariffs to reduce household electricity costs and shift load away from peak periods (Daneshzand et al., 2023).

Heat pumps: For the transition of the heating sector away from fossil fuels to electricity—potentially powered by renewable, emission-free sources—international policymakers have prioritized the adoption of heat pumps (Kou et al., 2024). While heat pumps significantly increase household electricity demand, particularly during winter, they also offer inherent flexibility. Buildings can be pre-heated in advance, or indoor temperatures can be allowed to temporarily decrease during peak price periods. This flexibility can be further enhanced by integrating a buffer tank, which acts as thermal storage (Hedegaard and Balyk, 2013). Given the relatively recent push for large-scale heat pump adoption and associated policies, uncertainty surrounds their integration (Chaudry et al., 2015). This dissertation addresses various aspects of heat pump integration: Chapter 4 examines their impact on load forecasting techniques, Chapter 7 introduces a novel method for generating synthetic heat pump load profiles, Chapter 8 presents a home energy management problem

focusing on the operation of heat pumps under dynamic tariffs while incorporating building thermal mass as additional flexibility, and Chapter 9 analyzes their role in the context of increased dynamic tariff adoption and its effects on distribution grid reinforcement costs.

Home energy management systems: Home energy management systems play a central role in controlling and scheduling devices in response to price signals from dynamic tariffs (Yousefi et al., 2020). These systems rely on actual or forecasted prices, as well as predictions of household energy consumption and generation, to make decisions about charging or discharging batteries, pre-heating buildings, managing thermal storage, and coordinating EV charging. Accordingly, Part II of this dissertation focuses on the role of home energy management systems combined with battery storage systems in enabling dynamic tariff-based operation strategies and emphasizes the importance of users adopting these strategies. Meanwhile, Part III addresses the improvement of load forecasting techniques, which are critical for the efficient operation of these systems.

2.3 System impact of dynamic tariffs

While there is empirical evidence for household reactions on dynamic tariff-based price signals (Herter and Wayland, 2010; Torriti, 2012; Enrich et al., 2024), the large-scale impact has to be investigated in simulative studies, given the yet low adoption of dynamic rates. In the following, an overview of relevant findings about the system impact of households adopting dynamic tariffs is presented.

In Katz et al. (2016), a partial equilibrium model is employed to simulate the effects of dynamic tariffs on electricity systems by capturing interactions between demand-side responses and supply-side dynamics. The study highlights how time-varying tariffs reduce wholesale price peaks, increase the value of wind power by improving system flexibility, and drive long-term adjustments in generation capacities, including shifts from peak-load to base-load generation. These system-wide benefits, however, depend on adoption rates and behavioral responses, which are shown to influence the redistribution of costs and benefits among consumers.

In McKenna et al. (2021), the demand side flexibility potential and price elasticity of Austrian households are estimated through an experimental field study. Consecutively, the inferred hourly price elasticities are then applied to an Austrian energy

system model, modelling future power system scenarios. The study sees system cost reductions between 4-7%.

In another notable study, Guo and Weeks (2022) model customer behavior and power system impact of dynamic tariff adoption in the scope of a game theoretic model. The authors find that dynamic electricity pricing leads to increased retailers' profit and consumer welfare. An interesting finding of the study is that the introduction of dynamic tariffs does not necessarily lead to reduced electricity consumption, which is in line with empirical evidence from (Torriti, 2012), underlining that they do not necessarily have a positive impact on energy conservation.

In Faruqui et al. (2010), the authors estimate the economic impact of dynamic tariffs in the EU on peak demand reduction potential. Their findings suggest that time-varying tariffs, combined with household responsiveness enabled by widespread smart meter adoption, can substantially lower peak electricity demand and reduce infrastructure costs for peaking plants.

The introduction of dynamic tariffs can also have socioeconomic implications, as highlighted by (Burger et al., 2020). If not carefully designed, they may result in higher electricity bills for lower-income households. However, in the context of the study, implementing two-part TOU tariffs has been able to help mitigate these negative effects while maintaining system efficiency benefits. These findings emphasize the importance of considering energy equity when adopting dynamic tariffs to prevent adverse societal impacts.

In summary, these studies demonstrate that dynamic tariffs have the potential to lower energy system costs, enhance welfare, and reduce peak loads. However, achieving a positive system impact requires the widespread adoption of dynamic tariffs. This, in turn, depends on addressing critical challenges and uncertainties related to policies, operational aspects, household behavior, and the technical effects on distribution grids. In this dissertation, these challenges are addressed in the following chapters.

2.4 Stakeholders on the path to dynamic tariff adoption

Supporting power systems on the path to increased dynamic tariff adoption requires a thorough understanding of the key stakeholders involved and their interactions. The following sections aim to identify these stakeholders and illustrate their interplay.

Households: At the core of dynamic tariff adoption are humans and households, which ultimately make essential decisions about adopting dynamic tariffs and utilizing the household's flexibility potential (Nicolson et al., 2018). As previously discussed, household responses to price signals can be behavioral adjustments (e.g., reducing appliance usage during peak hours), technological responses (e.g., activating a price-responsive mode in the home energy management system), or both. Key motivators for these responses include potential cost savings, perceived risks associated with dynamic tariffs, and the ease of adapting to them (Nakai et al., 2024; Nicolson et al., 2018; Buryk et al., 2015). Policymakers play a central role in shaping these factors, which are explored in the following.

Policymakers: The laws and regulations that shape the operation of power markets are defined by regulators and policymakers, making them influential actors in the integration of dynamic tariffs (Pereira and Marques, 2023; Faruqui et al., 2010). Policymakers can promote dynamic tariff adoption by obligating utilities to offer these tariffs, making them more economically attractive, or by discontinuing incentives for alternative tariff structures and flexibility strategies, such as those favoring self-consumption. For instance, from 2025 onward, utilities in Germany with more than 100,000 customers are obligated to offer dynamic tariffs (Bundesministerium für Wirtschaft und Klimaschutz, 2023). Policies, including taxes, grid charges, and feed-in remuneration for PV systems, are critical factors that determine how home energy management systems are operated (Stute et al., 2024). Consequently, the decisions made by policymakers can have a direct impact on households' technological responses to dynamic price signals.

Utilities and aggregators: Utilities serve as the critical link between electricity wholesale markets and consumers by providing dynamic tariff schemes (Guo and Weeks, 2022). Therefore, it is essential to incentivize utilities to offer dynamic tar-

iffs, ensuring that it becomes an economically viable choice for them. Aggregators, which pool and manage the flexibility of multiple consumers to participate in electricity markets, can also play a significant role in facilitating the adoption and effective implementation of dynamic tariffs (Ayón et al., 2017). The potential role of aggregators in the adoption process of dynamic tariff schemes is discussed in Chapter 8, where the concept of household-individual electricity price guarantees is introduced.

Grid operators: Also grid operators have an interest in the adoption of dynamic tariff schemes by households. Household responses to price signals can significantly alter load profiles, which in turn impacts stress on power lines and transformers (Stute et al., 2024). Chapter 9 of this thesis specifically examines how dynamic tariff adoption affects distribution grid stress and associated reinforcement costs and proposes policy options to mitigate these effects.

For completeness, it should be noted that the provided overview of stakeholders in the dynamic tariff adoption process is particularly applicable to European countries with unbundled utilities and grid operators, where consumers can freely choose their electricity tariff (Imran and Kockar, 2014). However, the stakeholder landscape may differ in other regions.

2.5 Dynamic tariff implementation in households

The following offers an overview of the interplay between policies, technology, and household behavior in the context of dynamic tariff adoption. While not exhaustive, the overview provided in Figure 2.2 highlights the key relationships and interactions that are explored throughout the chapters of this dissertation.

To begin with, household attitudes and preferences play a key role in determining whether and which type of dynamic tariff is adopted (Freier and von Loessl, 2022; Nicolson et al., 2017). These preferences, combined with household behavior, shape electricity consumption patterns (Katz et al., 2016) and heating behavior (Haas et al., 1998), ultimately influencing the household's load profile and the potential profitability of dynamic tariffs.

Although not explicitly shown in Figure 2.2, consumer attitudes also affect decisions regarding the adoption of technologies such as rooftop PV (Mundaca and Samahita, 2020). Regulatory factors, including PV feed-in remuneration policies and incentives for dynamic tariffs, can impact household decisions and behavior

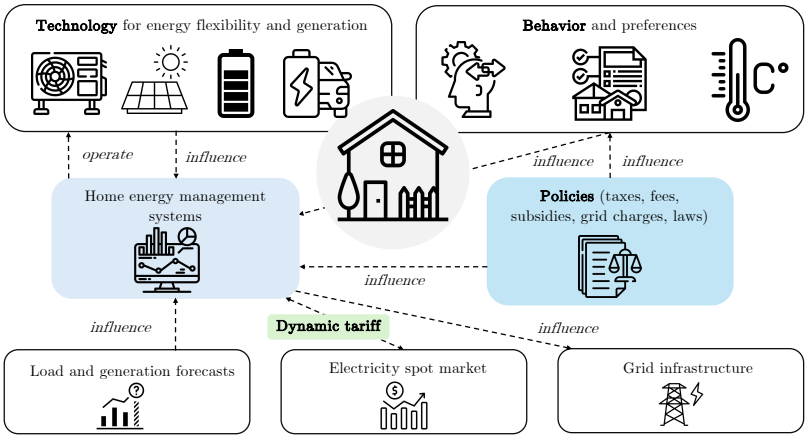


Figure 2.2.: Interplay of dynamic tariffs and households

(Klein et al., 2019).

If a household possesses flexibility assets such as a Battery Energy Storage System (BESS), their operation and scheduling are managed through a home energy management system (Luo et al., 2019). These play an integral role in implementing dynamic tariffs by optimizing the operation of large consumer devices such as heat pumps and electric vehicles while managing flexibility resources like BESS and thermal storage (Oldewurtel et al., 2011). Regulatory frameworks, i.e., grid charge structures, also influence the operational decisions of home energy management systems (Stute and Klobasa, 2024).

The effectiveness of home energy management systems in cost minimization under dynamic tariffs relies on accurate load forecasts (Colmenar-Santos et al., 2022), which enable scheduling of flexibility in response to projected demand. Widespread adoption of home energy management systems and household responses to dynamic electricity prices do more than just react to spot market signals—they can actively influence them. When adoption rates are high, these systems can alter demand patterns in a way that affects market prices (Faruqui et al., 2010; Katz et al., 2016; McKenna et al., 2021).

Ultimately, the collective response of geographically clustered households to dy-

namic tariffs can impact grid infrastructure. Simultaneous reactions to price signals may lead to new peak transformer and line loads, potentially straining the distribution network and leading to costly reinforcement measures (Stute and Klobasa, 2024).

2.6 Summary

To summarize this chapter's findings, dynamic tariffs offer a promising pathway to improving energy system efficiency by aligning electricity prices with real-time supply and demand conditions. Various dynamic tariff designs, including seasonal pricing, time-of-use pricing, critical peak pricing, and real-time pricing, reflect different levels of temporal granularity and cost-reflectiveness. The success of dynamic tariffs, however, relies on the overall adoption rate among households and how effectively they adjust their energy consumption patterns in response to price signals.

Households themselves play a central role in dynamic tariff adoption, with their behavioral and technological responses driving system-wide impacts. While in this thesis behavioral response to dynamic tariffs is rather neglected as household loads are seen as inflexible, the technological, automated response potential is emphasized and thoroughly investigated. Home energy management systems, along with flexible technologies like battery storage, EVs, and heat pumps, are critical enablers of this transition, allowing for automated and efficient responses to dynamic tariffs.

The chapter also emphasizes the influence of policymakers, utilities, aggregators, and grid operators in shaping the adoption and system impacts of dynamic tariffs. Policymakers can drive adoption by promoting tariff schemes and removing conflicting incentives, while utilities and aggregators play a vital role in offering and managing dynamic tariffs. Grid operators, in turn, must address the challenges of altered load profiles and increased grid stress resulting from dynamic tariff adoption.

Finally, the chapter highlights findings from empirical and simulation studies, demonstrating that dynamic tariffs can reduce peak demand, enhance the integration of renewable energy, and lower system costs. However, these benefits depend on widespread adoption, effective policies, and robust grid management strategies, which are explored in greater detail in the subsequent chapters of this dissertation, thereby bridging **policy, technology and behavior** by tackling uncertainty about *regulatory, operational, behavioral and technical* issues.

Part II.

Policy review

INTRODUCTION TO PART II

Policies play a significant role in determining how flexibility potentials—such as battery storage systems and home energy management systems—operate within households. The operation of these systems, in turn, shapes the dynamics of the overall power grid. Part I introduced this relationship, and this section focuses specifically on the regulation of household batteries, which are increasingly widespread in many international energy markets. A common regulatory approach incentivizes households to prioritize self-consumption of electricity generated by their own PV installations. However, this operational strategy often disregards market signals, such as day-ahead electricity prices. Part II expands on household battery storage regulation by reviewing international policies, conducting empirical investigations and simulations of alternative policies, and analyzing homeowner preferences.

CHAPTER 3

EMPIRICAL FIELD EVALUATION OF SELF-CONSUMPTION-PROMOTING BESS REGULATION

In this chapter, the widely adopted self-consumption-oriented regulation of BESS is examined. To begin, a framework is developed to illustrate how household BESS interacts with energy markets based on an extensive review of the literature. Following this, an overview of international policies is presented, highlighting how these systems are incentivized or regulated in different contexts. Next, a unique dataset of BESS in Germany is analyzed to assess their potential market impact. Building on the sample, alternative regulatory and operational approaches for BESS are explored through a simulative study. Lastly, a survey is conducted to identify the factors influencing perceptions of BESS effectiveness, offering valuable insights for shaping future policy design.

This chapter comprises the following article: L. Semmelmann, M. Konermann, D. Dietze, and P. Staudt. *Empirical field evaluation of self-consumption promoting regulation of household battery energy storage systems*, Energy Policy, 2024.

3.1 Introduction

The growing share of renewables in modern energy systems leads to an increasing need for flexibility on the demand side (Palensky and Dietrich, 2011; Strbac, 2008; Pedro et al., 2023). One promising technical solution for demand-side flexibility are BESS (Wu et al., 2015). The latest international statistics show that corresponding installations are on the rise: In Germany, the country from which we draw our data for this study, the total number of installed BESS was 320,000 in 2021, with a third

added that year. Only 0.8% of those BESS were industrial-scale BESS with capacities over 30 kWh, while the remaining installations were household BESS capacity (Peper et al., 2022). However, this is an international trend: 227,477 BESS systems (ANIE, 2023) are installed in Italy, Australia has 180,000 installations (SunWiz, 2023) and Austria reports 17,111 installed BESS (Bundesverband Photovoltaic Austria, 2023) in 2022. In total, there were over 1 million households in Europe with Photovoltaic (PV)-BESS systems with an aggregated capacity of 9.3 GWh in 2021 (SolarPower Europe, 2022). Globally, BESS installation capacity was 43 GWh in 2022. It is assumed that by 2030, 400 GWh will be reached worldwide (Rystad Energy, 2023). However, the impact of regulation on household BESS and its subsequent effect on the wider energy system has yet to be empirically evaluated. The term "household storage regulation" refers to the policies and rules governing the use of household energy storage systems, including whether dynamic tariffs are encouraged, the allowance for batteries to be charged from the grid, and the structure of grid charges (Fett et al., 2019).

Many of the globally installed household BESS are embedded in a regulatory framework that promotes self-consumption of generated PV power (Mateo et al., 2018), even though, various studies indicate that the self-consumption-oriented regulatory pattern is counterproductive for the system overall (Green and Staffell, 2017; Moshövel et al., 2015; Aniello and Bertsch, 2023; Tidemann et al., 2018). The major caveat of these findings is that they are almost exclusively based on simulation studies or only feature very small sample sizes. These studies are based on the assumption that batteries are operated as permitted by regulation while households do not change their behavior once a BESS is installed. However, studies on the rebound effect seem to suggest that energy efficiency investments affect energy consumption behavior (Deng and Newton, 2017), even though other authors contest its generality (Brockway et al., 2021; Rajabi, 2022). It is therefore necessary to evaluate the regulatory implications of BESS regulation to derive corresponding policy implications. To this end, we introduce and analyze the largest empirical dataset on BESS usage, with 947 BESS profiles from households measured over one year. We confirm that operating BESS for self-consumption can actually lead to welfare losses, i.e., costs that are socialized among energy

consumers. Based on these results, we propose and evaluate the impact of alternative regulatory approaches based on simulation using the same data that are better suited to incentivize a system-friendly battery management system operation.

Another often neglected aspect in literature is the influence of regulation on preferences and behavior of the households in which BESS are embedded (Ambrosio-Albala et al., 2020). The charging and discharging behavior of households (i.e., the consumption behavior) determines the performance of the battery against the market, even in the absence of any market signals directly to consumers. On the other hand, households that choose to subscribe to a dynamic tariff might reap additional benefits from their BESS by shifting demand to low-price periods (Zakeri et al., 2021; Green and Staffell, 2017). Dynamic tariffs lead to time-varying consumer electricity prices according to signals from the power system instead of fixed electricity rates (Dutta and Mitra, 2017). Proliferated variants of dynamic tariffs include time-of-use pricing (varying prices on- and off-peak), critical peak pricing (higher prices during system peaks) and real-time pricing (prices reflect wholesale market prices) (Dutta and Mitra, 2017). We base our simulative evaluation of alternative regulatory schemes on real-time pricing, a rate design that is widely available in Europe (Tibber, 2023).

Although potentially beneficial for the power system as a whole, such dynamic tariffs are not broadly popular (Schittekatte et al., 2023). It has been shown that this is partly caused by low energy literacy levels of households (Brounen et al., 2013; Reis et al., 2021).

Consequently, the interaction of households with their BESS has important implications for the market overall. It is, therefore, a valuable research direction to better understand how households perceive the effectiveness of their battery energy management systems. Beyond our empirical analysis of the impact of battery regulation and the proposal of alternative regulatory approaches, we further contribute to this question with an exploratory survey approach intended to explain the perceived effectiveness of BESS within a regulatory framework. We conducted a corresponding survey among 196 BESS owners. We find that perceived effectiveness is correlated with trust in the BESS management system and not correlated with any measures of perceived knowledge regarding energy consumption, the energy

system or energy markets. This has implications for the impact of corresponding BESS regulation.

In summary, this paper makes the following three contributions based on three distinct methodological approaches(empirical analysis, simulation, survey):

- Based on the first-of-its-kind empirical sample of 947 households over one year, we show the observed effects of a self-consumption promoting regulation on the impact of BESS operation on energy markets and provide insights on the corresponding costs incurred by the system.
- Using the same sample, we use optimization modeling to show how opting for gradually more dynamic regulatory approaches improves the value of these BESS for the system and its owners.
- Using a different sample of battery owners, we use survey results to show the impact of various sociological constructs on the perceived effectiveness of battery energy management systems and find that trust is more influential than perceived knowledge on energy consumption, the energy system and energy markets, which has implications for the impact of corresponding regulation.

These contributions hold various implications. First, the impact of regulation applied to a large fleet of battery energy management systems shows the danger of regulating decentral energy resources that are not aligned with overall power system objectives. Second, small adjustments to the regulatory environment in alignment with overall system efficiency objectives can greatly increase the value of resources for the system. Finally, our results suggest that households judge the effectiveness of their battery energy storage based on trust in third parties rather than their own system understanding of the energy system. Regulators are, therefore, well-positioned to incentivize more system-beneficial choices that can also benefit them and their customers.

Our study is based on an extensive review of related studies and international regulation of household battery storage systems. The review serves as a basis for our three-part methodology. First, empirical data from storage systems operated under

self-consumption-promoting policies is analyzed. Second, based on the presented sample, alternative policy options are investigated through an optimization-based simulation. Third, a survey is conducted to contribute a better understanding of customer attitudes towards their battery storage systems and energy management systems. Addressing the research questions at hand with three different methods enables a socio-technical perspective on the regulation of household storage systems.

3.2 Background

In this section, we review the relevant literature along a proposed framework shown in Figure 3.1. The installation of household BESS in combination with PV systems has been widely discussed in the literature (Luthander et al., 2015). We use this variety of studies to build a framework that depicts the interaction of these systems and their socio-technical embedding within regulatory and market frameworks in Figure 3.1. Zakeri et al. (2021) and Londo et al. (2020) show that the regulatory environment heavily influences household decisions to invest in PV-BESS systems. Besides regulation, the actual sizing decision is further influenced by consumer preferences and practical considerations (Agnew and Dargusch, 2017). The installed PV-BESS system is operated through a battery energy management system that schedules charging and discharging decisions of the BESS (Wu et al., 2022). The implementation of the energy management system again depends on the regulatory environment, the household's general preferences and the electricity tariff (Aniello and Bertsch, 2023; Wu et al., 2022; Zhou et al., 2018). The operation of PV-BESS systems impacts the energy spot market and power system as a whole (Fett et al., 2021). The described relationships are depicted in Figure 3.1 and further discussed and substantiated in the following using related literature.

Impact of regulation on PV-BESS investment decisions: The initial decision whether to install a PV-BESS system depends on the existing regulatory environment (Avilés et al., 2019). For instance, net metering policies, where the electricity costs of a customer are calculated after deducting its generation from its consumption, make BESS investments economically unreasonable as the grid serves as a virtual battery to the PV prosumer (Abdin and Noussan, 2018; Londo et al., 2020). Self-consumption-focused regulation, in which self-generated energy is

exempt from taxes and levies and excess production fed to the grid is remunerated by a feed-in-tariff (which is usually lower than the household tariff), promotes BESS (Castaneda et al., 2020; Zakeri et al., 2021). Consequently, it is economically reasonable for households to increase the self-consumption quota by installing a BESS (Aniello and Bertsch, 2023). For instance, self-generated electricity might be exempt from network charges, making BESS more attractive (Aniello and Bertsch, 2023). Dynamic tariffs paid by the grid operator for excess PV generation or BESS subsidies can further influence the sizing decision of PV and BESS systems from a regulatory perspective (Castaneda et al., 2020; Zakeri et al., 2021). In Section 3, we describe this in more detail and give an overview of international regulatory frameworks of household BESS.

Impact of consumer preferences on BESS investment decisions: While the regulatory environment mainly influences the profitability of different sizing options, the user makes the final decision incorporating factors like a desired safety from blackouts or maximizing self-sufficiency (Agnew and Dargusch, 2017). Further, the trust level in BESS technologies and the regulatory environment can be integral factors in the PV and BESS investment decision (Ambrosio-Albala et al., 2020). Consumer preferences can lead to a deviation from optimal economically induced PV and BESS sizes. Furthermore, the financial endowment and consumption behavior of households, such as ownership of an electric vehicle or roof area, can influence the PV and BESS sizing decision (Linssen et al., 2017; Wang et al., 2020a).

Impact of battery energy management systems on BESS operation: After the installation of a PV-BESS system, the daily charging and discharging decisions of the BESS are implemented through a battery energy management system (Angenendt et al., 2018; Lokeshgupta and Sivasubramani, 2019). The system uses various inputs, such as the current PV generation and the household load, but might also include actual or expected price curves (Mishra et al., 2012) and regulatory conditions (Young et al., 2019), which are described in the following.

Impact of regulation on battery energy management system operation: The battery energy management system operates the BESS to optimize household

utility given the regulatory environment (Zakeri et al., 2021). Important policy considerations that shape the way the battery management system is operated range from allowing or disallowing household BESS to charge from the grid (Bundesministerium der Justiz und für Verbraucherschutz, 2023), setting the remuneration for energy discharged from the BESS to the grid (Zakeri et al., 2021), exempting self-produced energy discharged from the BESS from taxes, fees and grid charges (Bundesnetzagentur, 2023b), to the promotion of dynamic tariff schemes (Parra and Patel, 2016). The impact of some of the aforementioned international regulatory approaches on operations of battery energy management systems is elaborated on in Section 3.

Impact of consumer preferences on battery energy management system operation: While regulation sets the boundaries in which a BESS can be operated and for its profitability, households control the selected tariff, pre-determining the way the battery energy management system schedules charging and discharging decisions (Zhou et al., 2018). Further, households are theoretically able to determine the operation schedule of a battery energy management system as they see fit. However, this requires technical capabilities and a market understanding that most households do not possess (Brounen et al., 2013). The decision for or against a dynamic tariff alters the scheduling of the battery energy management system (Zhou et al., 2018). The impact of consumer preferences and behavior on how their BESS is operated has not been investigated in detail yet.

Interaction of battery energy management systems with the spot market: Through the respective operation strategy, the BESS interacts with the spot market and has an impact on the overall power system (Fett et al., 2021). The share of households that operate their BESS with dynamic, price-responsive strategies influences the price level of the market and the power system as a whole. For instance, a responsive BESS fleet might reduce the curtailment of renewables (Fett et al., 2021) or change power system peak loads (Young et al., 2019). Household BESS can also influence wholesale prices for households without storage systems by impacting the dispatch of power plants (Say et al., 2020). An increased share of households with wholesale market-oriented tariffs can further

relieve the grid by reducing PV feed-in peaks (Günther et al., 2021).

Behavioral influence of consumers on battery usage: First studies indicate that household electricity consumption might change after the installation of a PV-BESS system, thereby altering the discharging requirements of the BESS. For instance, a recent empirical study from Arizona has shown that after co-installing electric vehicles, PV systems and a BESS, consumers have changed their daily load profiles (Shen et al., 2023). Al Khafaf et al. (2022) find that the installation of PV and BESS leads to behavioral changes of consumers in a case study that analyzes smart meter data of Australian energy consumers. This finding should be taken into account when analyzing different policy or tariff options. However, existing simulation studies in the field of household PV-BESS operation strategies assume constant household load profiles, neglecting the potential impact of behavioral changes (Say et al., 2020; Linssen et al., 2017).

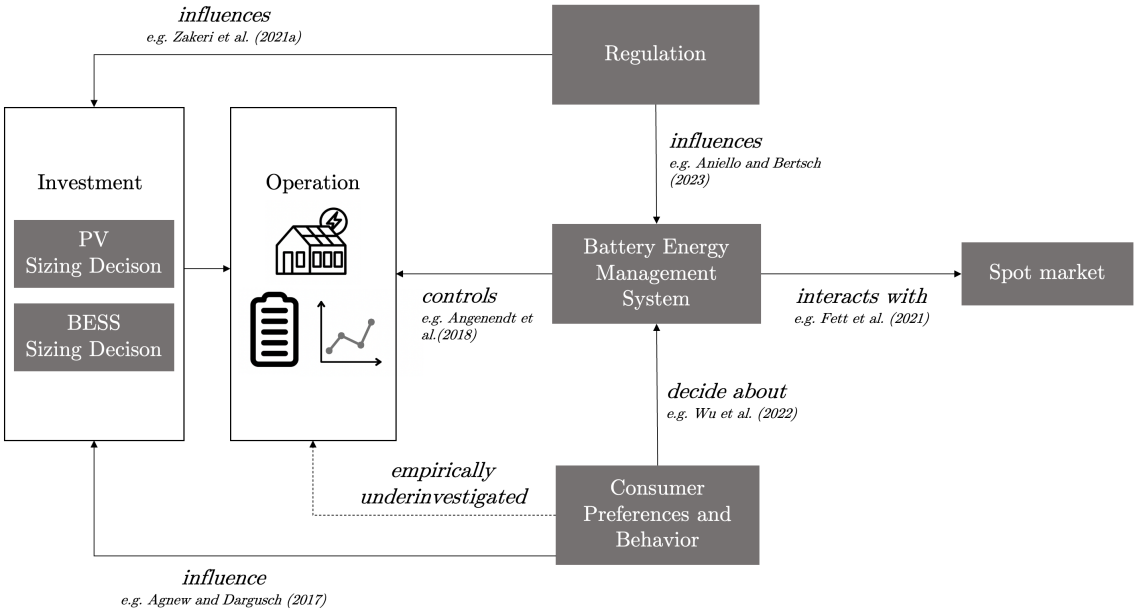


Figure 3.1.: Factors influencing battery installation decision and operation strategy

Due to a lack of empirically observed BESS load profiles, the studies mentioned in this section rely on various assumptions and corresponding optimization and simulation models (Angenendt et al., 2018; Linssen et al., 2017; Hesse et al., 2017; Naumann et al., 2015). This leads to certain shortcomings: First, the resulting BESS load profiles are dependent on the underlying household load profiles, which are often only

available in a low number or derived from standard load profiles (Linssen et al., 2017). As a consequence, the outcomes are based on a small sample and less varied. Second, the BESS sizing decisions are often directly derived from an optimization model, neglecting that BESS might only be available in certain sizes or that the investment decision is based on personal preferences, beliefs or expectations (Schopfer et al., 2018). Third, the naive simulation of battery energy management systems based on household load profiles without any consideration of changes in household behavior and decisions, neglects the possible influence of user perception and actions on the resulting BESS operation strategy and contradicts first empirical studies showing behavioral change after PV-BESS installations (Al Khafaf et al., 2022).

Given the rising number of household storage systems in Europe and globally (SolarPower Europe, 2022), it is becoming increasingly important to give battery operations a regulatory framework to align them with the overarching goals of the power system. This requires researchers to evaluate possible regulatory designs for residential BESS to enable policymakers to make informed decisions, especially about the promotion of the widely proliferated self-consumption operation strategy of PV-BESS systems. Although there are various studies discussing the impact of policy and tariff options (Parra and Patel, 2016; Green and Staffell, 2017; Zakeri et al., 2021), a thorough analysis of empirical data and its implications is still missing. Past studies are hence calling for a consideration of socio-technical and non-monetary factors for investment decisions (Say et al., 2020; Schopfer et al., 2018) and an analysis of heterogeneous, real-world household load profiles (Aniello and Bertsch, 2023).

We are contributing to this research gap with the first study that analyzes a large-scale sample of residential BESS operation profiles and evaluates consumer attitudes towards BESS in their regulatory framework. The analysis of the empirically observed data allows us to draw conclusions about the market performance of PV-BESS systems operating in a self-consumption-promoting regulatory environment with fixed household tariffs as the observed systems have been operated in this environment. Based on these results, we can draw conclusions about the efficiency of current regulation without the limitations of neglecting consumer behavior. Furthermore, we can explore the impact of possible alternative tariff options with an optimization-based simulation. Finally, we contribute to the overall understanding of household perspectives on their BESS, thus addressing the socio-technical per-

spective on BESS through a survey amongst battery owners.

3.3 International regulation of household BESS

In this section, we portray various international BESS regulation approaches to describe the international regulatory context of BESS. The two fundamental types of national regulatory policies of BESS are self-consumption promoting and net metering or net billing policies (Fett et al., 2019).

Self-consumption: In self-consumption-promoting regulation, households are encouraged to increase the consumption of self-generated energy, typically from PV making household BESS an attractive option (Castaneda et al., 2020; Zakeri et al., 2021; Angenendt et al., 2018). This is promoted, for instance, by exemptions from taxes and levies or a gradual decrease of feed-in tariffs (Fett et al., 2019; Castaneda et al., 2020) making feed-in less attractive. In Germany, self-generated energy from PV installations under 30 kW is exempt from most taxes and levies (Bundesregierung, 2023; Fett et al., 2019). At the same time, feed-in tariffs are reduced (to 0.086 Euros in 2023, compared to an average household electricity price of 0.452 Euros) (Bundesnetzagentur, 2024a, 2023a). In addition, charging the BESS through the grid leads to a loss of the previously listed benefits (Bundesministerium der Justiz und für Verbraucherschutz, 2023). A comparable regulatory framework is implemented in Austria, where self-generated energy up to 5,000 kWh is exempt from taxes and levies, thereby encouraging the installation of storage systems (Unternehmensserviceportal (USP), 2024). Similarly, in Croatia, self-generated PV is exempted from fees and network charges (Croatian Parliament, 2021). The United Kingdom (UK) is also supportive of batteries used to increase self-consumption in households (Department for Business, Energy and Industrial Strategy, 2021), further promoted by a VAT relief for battery installations introduced in 2024 (HM Revenue and Customs, 2024). In the UK, energy exports to the grid are compensated by a "Smart Export Guarantee", which replaced feed-in tariffs in 2019 (Department for Energy Security and Net Zero, 2020). Smart Export Guarantees are offered by a couple of utilities and can vary amongst them, in contrast to a nationwide uniform feed-in tariff. Contrary to Germany, the UK actively encourages households to charge their batteries from the grid and to be compensated for feed-in from batteries with the respective Smart Export Guarantee. Japan implements a self-consumption

scheme with no additional costs for self-generated PV electricity (International Energy Agency, 2021). In Australia, the exemption from relatively high volumetric grid charges (Say et al., 2019) for self-generation, sets a strong incentive for installing BESS (SunWiz, 2023). In Italy, the government offered feed-in tariffs until 2013, which led to a mandatory opt-out from the net metering scheme, thereby encouraging self-consumption (Abdin and Noussan, 2018), which is still the most attractive option for BESS owners even in the absence of feed-in tariffs.

Net metering and net billing: The general idea of net metering regulation is that electricity can be sold to the grid at the price for consuming electricity at that point in time (Londo et al., 2020; Abdin and Noussan, 2018). The easiest way to understand this regulation is that in the presence of flat volumetric tariffs, the electricity meter runs backward when electricity is provided to the grid. This makes the grid a virtual battery for prosumers. A more detailed classification of the regulatory framework depends on the way the PV feed-in is compensated. The policy scheme is then either called net metering (compensation at retail rates including taxes and levies) or net billing (compensation only at current wholesale prices). Italy introduced an optional net billing scheme in 2008, which is seen as the main driver of PV installations in the country after the end of feed-in tariffs (Autorità di Regolazione per Energia Reti e Ambiente, 2008; Abdin and Noussan, 2018). The scheme values PV feed-in weighted at the national market price, which is then deducted from the customer's electricity bill. In the United States, net metering and net billing policies are implemented on a state level in various states (Gregoire-Zawilski and Siddiki, 2023). Policy elements vary from state to state. For instance, in Texas, Oregon and Maine, the policy design is rather utility-favoring and implements a net billing scheme, including a valuation of feed-in at market prices rather than retail rates and no possibility to roll-over credits to upcoming periods (e.g., months or years). On the other hand, Florida, New York and New Mexico, pursue a customer-favoring design that resembles net metering with a valuation of excess generation at retail rates and a compensation for remaining credits at the end of the year (Gregoire-Zawilski and Siddiki, 2023). In Spain, a net billing scheme was introduced in 2018, which is based on a monthly balance of electricity consumption drawn from the grid and PV feed-in, whereas the feed-in is valued by a rate set by the system operator, which is marginally lower than the wholesale price (Ordóñez et al., 2022). Since negative

balances are not rewarded, the regulatory framework encourages proper sizing of the PV installation. Ecuador implements a net metering scheme with rather long credit transfer periods (Ordóñez et al., 2022). Negative balances can be used as rolling credits up to two years after being generated. Although the presented regulatory net metering policies incorporate different features, they all take away incentives to install household BESS as the grid serves as a virtual battery (Abdin and Noussan, 2018; Ordóñez et al., 2022). However, there are endeavors to incentivize household storage installations, even in regions with net metering or net billing policies in place. California, a state with a plain-vanilla net billing policy, moved away from fixed feed-in compensation to a new model called "NEM 3.0", which implements a time-variable compensation. This should encourage homeowners to install battery systems alongside their PV installation and to shift their feed-in away from peak feed-in periods (California Public Utilities Commission, 2023).

The different regulatory frameworks and the wide range of individually implemented policy features for BESS regulation show that a best practice for storage regulation has yet to emerge. This may be due to a lack of experience with large market penetration of battery storage or because of different national electricity market designs. In any case, our results can provide some guidance for making corresponding policy decisions. Our study is an important contribution to this policy discussion since we are the first to evaluate a large empirical dataset from a region promoting self-consumption to describe the impacts of the regulation framework. Furthermore, the study also allows us to evaluate alternative regulatory strategies.

3.4 Methodology

The review underlines the need for an empirical analysis of household battery storage systems operated in a self-consumption-focused regulatory setting. We tackle the identified research gap with an approach divided into three parts, as depicted in Figure 9.1.

First, we analyze spot market profits of an empirical year-long sample of 947 household battery storage system profiles. This analysis fills a research gap addressing the potential effects of customer behavior on the market performance of battery storage systems. We find that the system benefits of household BESS are currently markedly low. We therefore evaluate alternative regulation.

We thus extend our empirical perspective with an analysis of alternative policy options, derived from our review of international regulation and related work. While the empirical characteristics of the 947 systems at hand set the boundaries for the analysis of alternative regulatory options, the resulting charging and discharging profiles are obtained using optimization. The results of the simulation allow for a comparison of the empirical analysis with policy options, such as the promotion of dynamic tariffs or allowing household storage systems to charge from the grid. Given that households need to adapt to this changing regulation, we further analyzed the attitudes of households towards their BESS to better understand antecedents for household behavior.

We, therefore, conduct a survey amongst battery owners. Thereby, we also provide a behavioral perspective on the problem at hand, which is often neglected in related studies.

Overall, our three-part approach allows us to analyze household battery storage systems operated in a self-consumption regulatory framework as a socio-technical policy problem, rather than from a purely techno-economic perspective.

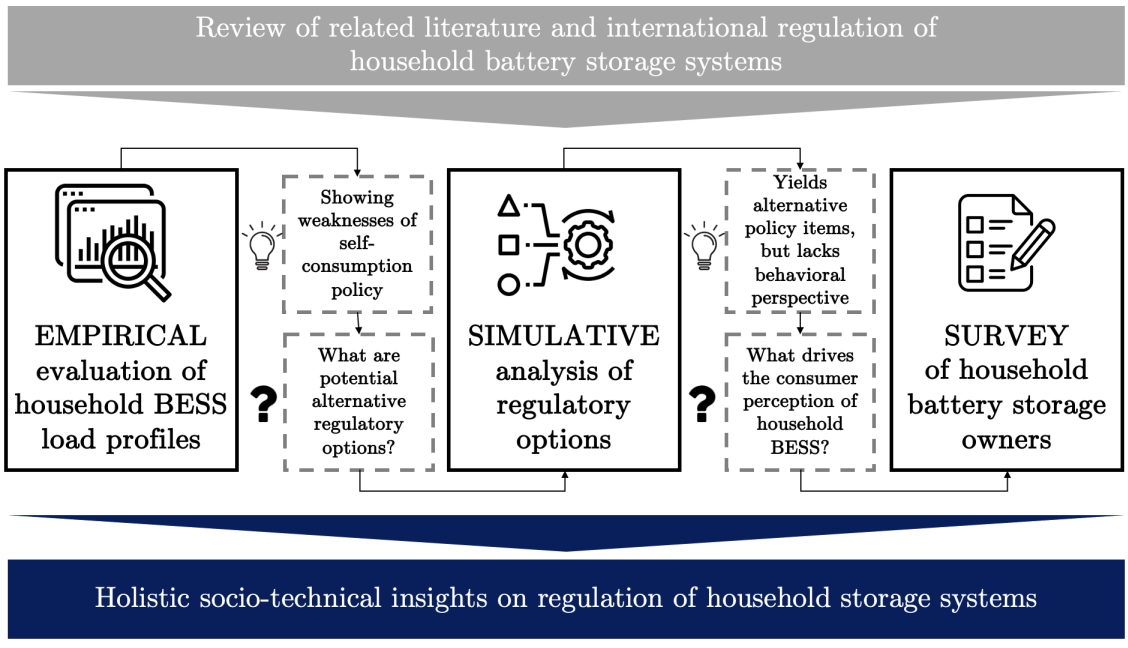


Figure 3.2.: Paper structure and methodological overview.

3.4.1 Evaluating empirical BESS operation

First, we evaluate the effectiveness of the self-consumption-promoting regulatory framework through an empirical dataset from Germany. Germany implements a self-consumption promoting regulation that makes it financially attractive to consume self-generated energy. This evaluation is unique as it is the first to evaluate the empirical consumption behavior of households that own a BESS. This differentiates the results from simulation results that assume no change in behavior or optimal response to signals (Angenendt et al., 2018; Naumann et al., 2015; Aniello and Bertsch, 2023). We can calculate the effectiveness of household BESS by evaluating the hypothetical market performance on the spot market. This indicator is representative of the overall value the BESS is adding to the system. To do that, we multiply charging and discharging actions aggregated over an hour with the spot market price in the respective time interval on the day-ahead market. The underlying economic rationale is that high spot market profits for household BESS operation indicate a high overall power system utility generated from shifting loads from periods with high demand and costly generation (and high prices) to periods with lower demand and cheap generation (and correspondingly low prices) (Zafirakis et al., 2016; Lamp and Samano, 2022). We underline this point with a complementary analysis in the Appendix, showing that German day-ahead market prices are strongly positively correlated with balancing energy costs, fossil fuel-powered conventional production and residual load.

To analyze the profitability of the BESS charging and discharging decisions on the European EPEX spot day-ahead market within the German market zone (EPEX Spot, 2023), we aggregate the battery load profiles to an hourly resolution matching the day-ahead market contract duration (Märkle-Huß et al., 2018). We model the BESS as price takers, thereby having no impact on market prices and always finding a counterparty for possible trades. We assume that every sell- and buy-order could be conducted at market prices, which is a reasonable assumption given the liquidity of the market in question and the current capacity of market-oriented BESS operation strategies (Naseri et al., 2023; Wankmüller et al., 2017).

We evaluate the spot market profits (denominated as Π) of household BESS operation using Equation 3.1, which is similar to objective functions used in simulation

studies of market-focused battery operation (Schneider et al., 2020; Krishnamurthy et al., 2017). The same objective function is used in the following section to evaluate alternative tariff structures. In the evaluation of the empirical case, we ignore the inverter efficiency losses η_{inv} because the observed values already incorporate these losses.

$$\Pi = \sum_{t \in T} \quad (\eta_{inv} p_t P_t^d \tau - \frac{1}{\eta_{inv}} p_t P_t^c \tau) \tag{3.1}$$

When the BESS is charged with power P_t^c over a period τ, costs at the current market price p_t occur. This is irrespective of whether the battery is actually charged from the grid or from self-generation, as self-generation could be sold on the market, which, therefore, leads to opportunity costs. We proceed correspondingly for discharging of the BESS with power P_t^d, which leads to revenues either as market income or as foregone costs. The charge and discharge operations correspond to the empirically observed operations $P_t^{d,real}$ and $P_t^{c,real}$. We use this definition of BESS energy flows to the empirically observed behavior further in the following section to differentiate the empirical case from hypothetical alternative regulation approaches and dynamic tariff options.

$$P_t^d = P_t^{d,real}, \quad P_t^c = P_t^{c,real} \quad \forall t \in T. \tag{3.2}$$

3.4.2 Simulative analysis of regulatory alternatives

Second, given that our empirical analysis confirms the assumption that a BESS operation schedule prioritizing self-consumption does not necessarily align with the power system's needs (Green and Staffell, 2017), we propose several alternative approaches by building on more dynamic regulatory policies. We gradually increase the degrees of freedom with the following suggestions. We consider four different variations of the baseline optimization problem, varying from the actual, empirically observed BESS operation to a completely market-oriented operation. Through the cases, we explore different regulatory options for BESS and their influence on market profits on the day-ahead market. The cases increasingly deviate from the known, self-consumption-promoting empirically observed profile. Comparing different possible operation strategies to the empirical profiles

allows us to draw conclusions about the economics of current BESS operations. This perspective is again simulation-based, which we further discuss in Section 9.7.3.

Various constraints are accommodated to analyze different operation strategies derived from the empirical profiles. First, the maximum charge and discharge power P^{max} of the BESS P^{max} must never be exceeded for P_t^c and P_t^d (Equation 4.8b). Second, the SOC of the BESS is connected to the previous SOC and the current charging or discharging operation, as denoted in Equation 4.8c. Finally, the SOC_t has to be kept within the boundaries of the minimum SOC^{min} and maximum SOC^{max}, as in Equation 4.8d (Krishnamurthy et al., 2017). To realistically consider efficiency losses, the marketable discharge power has to be multiplied by the DC-AC inverter efficiency η_{inv} and the charging power has to be divided by it. A round-trip inverter efficiency of 90% is assumed, based on Soini et al. (2020). Since efficiency losses are considered for charging and discharging the BESS, η_{inv} is set at $\sqrt{90\%}$. Since the empirical SOC measurements and the charging power derived from it already include charging efficiency losses, we divide the charged energy by the efficiency rate of $\sqrt{90\%}$. Thereby, we prevent a double consideration of efficiency losses, which would potentially distort the comparison of the empirical case with alternative approaches.

$$\max \quad \sum_{t \in T} \left(\eta_{inv} p_t P_t^d \tau - \frac{1}{\eta_{inv}} p_t P_t^c \tau\right) \tag{3.3a}$$

$$\text{s.t.} \quad P^{max} \geq P_t^c, \quad P^{max} \geq P_t^d, \quad \forall t \in T, \tag{3.3b}$$

$$SOC_t = SOC_{t-1} + \tau(P_t^c - P_t^d), \quad \forall t \in T, \tag{3.3c}$$

$$SOC^{min} \quad \leq SOC_t \quad \leq SOC^{max}, \quad \forall t \in T \tag{3.3d}$$

Using this basic optimization problem and the notation of restricted charging and discharging powers in Equation 3.2, we can formulate alternative regulatory options and tariff designs mathematically. We establish the following alternative regulatory options based on our review of existing policies and studies from Section 3.

Case 1: Flexible Discharging We begin by only slightly deviating from the self-consumption promoting regulation by only allowing battery charging from self-

generated energy while subscribing the household to a real-time pricing tariff that sets the price for buying and selling energy to or from the grid. We model this mathematically by limiting the charging power P_t^c at time t to the respective empirical observation $P_t^{c,real}$, while P_t^d can be freely chosen within the given battery constraints.

$$P_t^c \leq P_t^{c,real} \quad \forall t \in T. \tag{3.4}$$

.

By replacing Constraint 3.2, which limits BESS operations to the underlying empirical observations, with the new Constraint 3.4, we ensure that the battery is only charged with available excess PV power of the household (Parra and Patel, 2016). Since only the upper bound for BESS charging is set, it is now possible to suspend charging operations at periods of high market prices and sell electricity to the grid, thereby reducing opportunity costs. It is now also possible to discharge the BESS to provide energy to the grid at the current market price. This operation strategy incorporates dynamic tariff elements (Parra and Patel, 2016), as well as a time-variable feed-in compensation component, comparable to the newly introduced variable "NEM 3.0" policy in California (California Public Utilities Commission, 2023).

Case 2: Calendaric limits *Case 1* offers more potential to react to market signals than the empirical self-consumption, but is still similar due to Constraint 3.4, which limits charging to the empirical observations. To enable a higher degree of market price exploitation, we replace Constraint 3.4 with Constraint 3.5a, which only limits the sum of all charging operations P_t^c to the sum of all empirically observed charging operations $P_t^{c,real}$. By employing Constraint 3.5a, we enforce that the battery is not discharged more than in the actual empirically observed case. *Case 2* can therefore be thought of as a compromise between real-time pricing with grid charges and net billing similar to the Texan regulatory approach (Gregoire-Zawilski and Siddiki, 2023). Under a net metering regime, the grid essentially serves as a BESS for prosumers with PV generation. Using the proposed approach, households are forgiven grid charges for the energy they provide to the system through their BESS. This incentivizes further flexibilization as households can provide energy when it is

expensive in the system and buy it back later as incentivized through real-time pricing without having to pay additional grid charges or levies. This right of buying back provided energy is temporally restricted to avoid the usage of the grid as long-term storage. For instance, one might want to avoid households providing energy during the summer to buy it back in the winter. We, therefore, differentiate between *Case 2Y*, *Case 2M* and *Case 2W*, where the limiting sum of charging operations is applied either for the whole year or within every month m or for every week w, for the set of all months M and respectively all weeks W, as depicted in Constraints 3.5b and 3.5c.

$$\sum_{t \in T} P_t^{c,real} \quad \geq \quad \sum_{t \in T} P_t^{c}. \tag{3.5a}$$

Since by omitting Constraint 3.4, the BESS can also be charged in times without PV excess power ($P_t^c \geq P_t^{c,real}$), potential differences in market prices can be exploited more flexibly.

$$\sum_{t \in T,m} P_t^{c,real} \quad \geq \quad \sum_{t \in T,m} P_t^{c}. \quad \forall \quad m \in M. \tag{3.5b}$$

$$\sum_{t \in T,w} P_t^{c,real} \quad \geq \quad \sum_{t \in T,w} P_t^{c}. \quad \forall \quad w \in W. \tag{3.5c}$$

Case 3: Market responsive Finally, in *Case 3*, all constraints on charging and discharging based on the empirically observed data are dropped. This enables a fully flexible BESS operation based on market prices and represents a pure real-time pricing tariff. *Case 3* resembles the optimization case from Schneider et al. (2020), but without considering battery degradation in the optimization problem. From a policy perspective, the implementation of *Case 3* requires allowing household battery charging from the grid, which is, for instance, in Germany, penalized by the loss of self-consumption tax and levy exemptions (Bundesministerium der Justiz und für Verbraucherschutz, 2023).

We further differentiate between *Case 3*, where BESS are assumed to be exempt from grid charges, and *Case 3NE*, in which charging the BESS from the grid causes additional charges p^{grid}. We add the grid charges p^{grid} to the current market price

p_t in time steps where the charging power P_t^c exceeds the empirically observed value $P_t^{c,real}$ as this indicates charging the BESS from the grid. We use hypothetical grid charges p^{grid} in the amount of 0.0735 Euros per kWh in our optimization model based on the actual charges of a Southern German distribution grid operator (Netze BW, 2021). While seemingly very different from the empirical case, *Case 3NE* represents households that subscribe to a real-time tariff within the context of self-consumption promoting regulation.

$$p'_t = \begin{cases} p_t + p^{grid} & P_t^c > P_t^{c,real} \\ p_t & P_t^c \leq P_t^{c,real} \end{cases} \tag{3.6}$$

Figure 3.3 provides a graphical overview of the presented cases. The cases represent an increasing degree of market flexibility, but also decreasing similarity to the original operation strategy from the *Empirical Case*.

Characteristics / *Cases*	Empirical	1	2	3	3NE
BESS discharging at *t* empirically observed	✓				
BESS charging at *t* empirically observed	✓				
Max. BESS charging at *t* from empirical observation	✓	✓			
Max. EFC from empirical observation	✓	✓	✓		
Discharging the BESS can be postponed		✓	✓		
The BESS can be operated with full flexibility on the market				✓	✓
Grid charges regarded in optimization					✓

Figure 3.3.: Properties of investigated cases

3.4.3 Survey of household storage owners

The results of our analysis and our proposed regulatory adjustments show the importance of the interaction between regulation and system. Besides regulation, the interaction of BESS and the energy spot market is also governed by personal preferences and behavior. For instance, households that decide to subscribe to a real-time pricing tariff already have an incentive to operate their BESS in a more system-friendly way as they are exposed to temporally differentiated external price signals. However, few consumers choose to do so. The reasons for this lack

of engagement with more dynamic tariff designs are unclear. Recent research has shown that consumers with a higher energy literacy are more likely to adopt time-of-use pricing (Reis et al., 2021). We, therefore, acquired additional sample data from a survey conducted by Bilendi[2] to better understand the attitude of household BESS owners towards their battery energy management systems. Due to data privacy regulation, we were unable to directly contact the consumers from the described empirical sample. We, therefore, acquired an additional sample to conduct the survey with.

We focus our analysis on understanding the determinants of the perceived effectiveness of household BESS. We choose this as our dependent construct as we aim to better understand how households perceive the interaction of their BESS with the external energy spot market. Perceived effectiveness is instrumentalized using the scales proposed by Luo et al. (2008). As potential determinants of perceived effectiveness, we choose trust in the system (Gefen et al., 2003), the perceived behavioral control (Sheeran and Orbell, 1999), the importance of financial profitability and the sustainability of BESS (Bucher et al., 2016), the overall satisfaction with the BESS (Liao and Chuang, 2004) and several indicators to measure the perceived self-rated individual knowledge on personal energy consumption, the energy system and the energy market (Schlösser et al., 2013). We use an Ordinary Least Squares (OLS) regression, a statistical method that estimates the relationship between independent variables and a dependent variable by minimizing the sum of the squared differences between observed and predicted values, to determine the impact of these variables on perceived effectiveness (Craven and Islam, 2011; Jia et al., 2021).

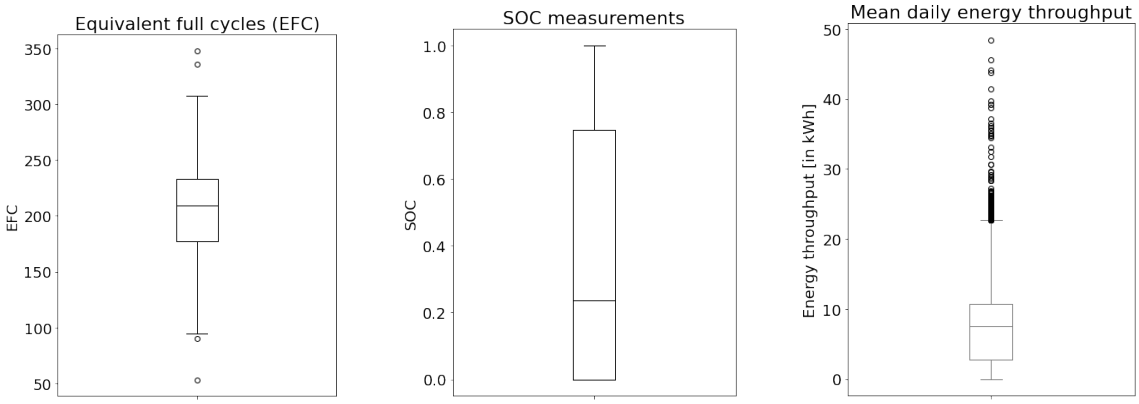
3.5 The dataset: Empirical BESS load profiles

We begin by introducing the empirical dataset of 947 battery load profiles. The BESS profiles were measured over the course of the year 2021 and include state of charge (SOC) measurements with a one-minute resolution. The profiles come from regionally distributed German households and were anonymized before being provided to us for this study. All the residential BESS within the study are installed with corresponding PV systems. They have an energy capacity of 2.5 kWh (6.7%), 5

[2] https://www.bilendi.de

kWh (37.2%), 7.5 kWh (31.5%), or 10 kWh (24.6%). The BESS' maximum discharge power P^{max} is either 1.25 kW (6.7%), 2.5 kW (68.7%) or 3.75 kW (24.6%). On average, the Power-to-Energy ratio, which is often used to set the power rating in relation to the energy capacity, lies at 0.41. These figures provide important insights into the current distribution of household BESS in Germany.

The provider of the BESS guarantees the nameplate energy capacities for ten years by oversizing the systems to account for degradation. It operates a dedicated battery management system that ensures the nameplate capacity even as the total (oversized) capacity diminishes (SENEC GmbH, 2024). As a result, the BESS is operated with the guaranteed nameplate capacity throughout the observed period.



(a) EFC (b) SOC (c) Energy Throughput

Figure 3.4.: Descriptive statistics of empirically observed BESS usage

The empirically observed battery charging or discharging power P_t^{real} is based on the SOC change between two time steps SOC_{t+1} and SOC_t, divided by the time resolution τ, as in Equation 3.7. We aggregate the values to an hourly time resolution to link the data to the contract duration of the European EPEX spot market (Märkle-Huß et al., 2018).

$$P_t^{real} = \frac{SOC_{t+1} - SOC_t}{\tau} \tag{3.7}$$

As battery cycling influences the battery degradation and, therefore, the battery's lifetime (Kucevic et al., 2020), we also analyze the BESS usage throughout this study. We, therefore, determine the number of Equivalent Full Cycles (EFC), which set the Energy Throughput (E^{tp}) of the BESS in relation to its nominal capacity.

The energy throughput E^{tp} is the sum of the absolute charging (P_t^c) and discharging power (P_t^d) of the BESS over time T measured in steps of the time resolution τ, as formulated in Equation 3.8 (Koltermann et al., 2023).

$$E^{tp} = \sum_{t=0}^{T} |(P_t^c + P_t^d)\tau| \tag{3.8}$$

The energy throughput of one entire charging and discharging cycle in the amount of the nominal BESS energy capacity E^{BESS} represents one EFC. Hence, we divide E^{tp} by two times E^{BESS} to calculate the resulting EFC (Kucevic et al., 2020; Maheshwari et al., 2020):

$$EFC = \frac{E^{tp}}{2E^{BESS}} \tag{3.9}$$

Cycle depth is also commonly mentioned as a source of battery degradation. Deep cycles lead to faster BESS degradation (Schimpe et al., 2018). We are neglecting an analysis of the cycle depth and focusing instead on EFC. We do so because the empirically observed self-consumption-oriented BESS operation strategy - charging until the battery is full, discharging until it is empty - already leads to the highest possible cycle depths. Hence, any alternative operation strategies could even lead to comparable or even lower cycle depths.

Figure 3.4 presents the introduced measures for the empirical sample. Most household BESS exhibit 198 EFC on average, with few outliers. The SOC measurements are distributed between 0 and 1 as the share of charged total capacity, with most measurements lower than 0.5, indicating that the BESS is more frequently empty than fully charged. The 75th percentile of the mean daily energy throughput of the individual BESS is below 10.7 kWh, while there are again a couple of outliers. Some of them are caused by the BESS still being charged from the day before, and then being discharged during the beginning of the day. Others are caused by multiple charging and discharging cycles per day. These patterns are only observed in a few households.

3.6 Results

In this section, we apply the first two steps of our methodology - an empirical analysis of BESS load profiles and the simulative analysis of alternative regulatory options - to the underlying dataset.

In Figure 3.5, we depict the cumulated empirical annual profits per individual household. The average profit per installed BESS is 5.0 Euros. This means that every installed BESS only contributes 5.0 Euros to the overall system welfare per year. In total, 23.02% of households operating a BESS exhibit negative spot market returns as depicted in Figure 3.5. This means that this BESS operation leads to additional costs for the system overall. More expensive power stations have to be operated because these BESS are operated within the system. This finding strongly calls into question the effectiveness of the corresponding regulation.

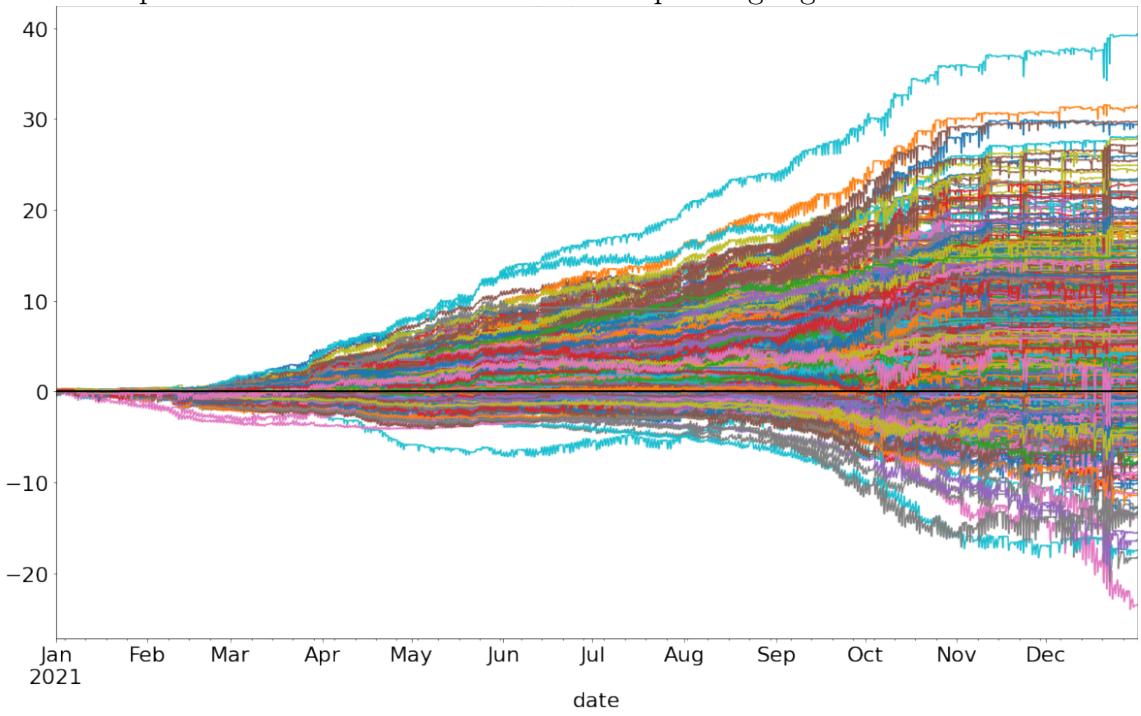


Figure 3.5.: Cumulated yearly household market profits in the empirical case (in EUR)

In the following, we compare the market results from the empirically observed BESS operation in a regulatory environment that promotes self-consumption with the introduced alternative regulatory policies. To do that, we assume perfect foresight for market prices and load values, which is common in studies focusing on BESS profitability (Olk et al., 2019; Wankmüller et al., 2017; Sioshansi et al., 2009).

Sioshansi et al. (2009) observe in a study about battery trading profitability on the PJM market that the perfect foresight assumption overestimates battery trading revenues by 10-15% compared to a backtesting-based trading strategy. In addition, we evaluate the efficiency of the respective BESS operation strategies in different regulatory environments by considering EFCs. We also relate the EFCs to the economic performance Π by calculating the Profit per Cycle (PPC), as in Equation 3.10. Assessing the efficiency of BESS operations is especially important in light of the scarce resources used to build lithium-ion BESS (Costa et al., 2021).

$$PPC = \frac{\Pi}{EFC} \tag{3.10}$$

To better understand the BESS charging and discharging operations induced by the presented strategies, we depict the BESS SOC of an exemplary household over a day in Figure 3.6. The y-axis represents the SOC and the x-axis shows the time of day. The lines represent the progression of the SOC over the day for the empirical case and for the simulated alternative regulatory approaches, while the green dots depict the hourly day-ahead spot market prices. We note that we only discuss an exemplary profile for the sake of comprehensibility. However, in the Appendix, we plot the SOC curves of all households and all cases, underlining that the exemplary Figure 3.6 is representative of the whole sample, exhibiting comparable patterns. Furthermore, we depict all diurnal day-ahead price curves and their average over the year in the Appendix, illustrating their resemblance with the price curve from Figure 3.6.

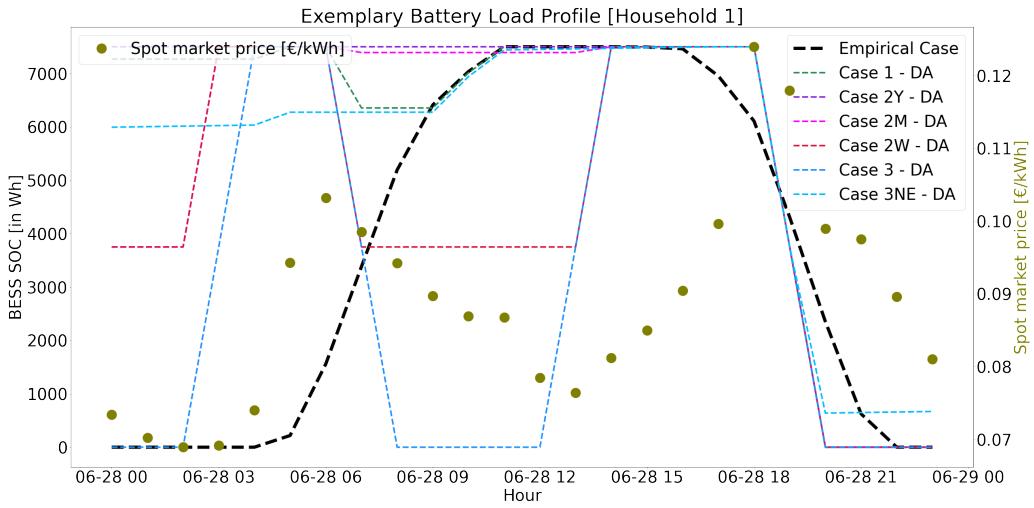
The empirically observed operations follow the expected self-consumption pattern. The BESS is charged in the morning until midday, when it is completely charged. In the afternoon and evening, the BESS is discharged. Our results are specifically important in light of the solar duck curve (California Independent System Operator, 2015). When prices are very low during the afternoon in spring or summer, battery storage are often already fully charged and cannot absorb energy during times of negative wholesale prices (Denholm et al., 2015). We can observe a deviation from the empirical profiles when we analyze the resulting alternative regulatory options. Two peak price hours in the evening are used to discharge the BESS. In *Cases 1, 2W and 3*, the battery energy management system also discharges the BESS during a

morning price peak at 6 AM. Most observed battery energy management strategies under alternative tariff designs directly feed morning PV generation into the grid due to a relatively high price level rather than using it to charge the BESS on this exemplary day. *Cases 2W and 3*'s midday charging decision at 1 PM coincides with the lowest daily electricity price of 0.08EUR/kWh. Thereby, we illustrate a general outcome of the more flexible operation strategies: The alternative BESS operations follow the overall market signals of the power system.

Case 2Y, 2M, 2W allow flexible BESS operation but restrict the maximum amount of EFC to the empirically observed values. By restricting battery cycles yearly, *Case 2Y* shifts most grid charging to the more lucrative winter months. In contrast, *Cases 2M and 2W* keep the cycle amount at the monthly and weekly level of the empirically observed values. From a household perspective, this might lead to months, weeks, or days without BESS usage, representing a significant difference to the empirically observed status-quo operation profiles from the *Empirical Case*.

Case 3 differs from *Case 3NE* because the latter incorporates grid charges in the optimization problem when the BESS is charged through the grid. Apart from that, both cases allow fully flexible BESS operations. As a result, in *Case 3*, the BESS is used most, but at the cost of a higher amount of cycles, whereas *Case 3NE* even reduces the number of cycles by weighing up grid charges against possible profits while considering direct feed-in of PV instead of charging from self-generation.

Table 3.1 provides an overview of market results and EFC for the considered regulatory and tariff cases and a comparison to the analysis of the empirical data. In the scope of the alternative tariff options, the market profits and hence, the added value to the power system from storage operation, increase significantly, compared to the empirical case. Even moderate changes to the regulatory environment as *Case 1 (flexible discharging)* would lead to a ten-fold increase of average welfare gain per storage system. *Case 1* even leads to less EFC. This means a Pareto improvement under such a regulatory regime as system benefits increase, while costs for the individual are decreased. In *Case 2 (calendric limits)*, system welfare can be further increased, also leading to the highest profits per cycle, indicating the most efficient use of the systems from an economic point of view. In *Case 3 (market responsive)*,



(a)

Figure 3.6.: BESS SOC of exemplary household and spot market price for different operation strategies over one day

the high flexibility and highest returns come at the price of the highest amount of EFC, possibly leading to faster degradation of the BESS (Schimpe et al., 2018; Kucevic et al., 2020). Implementing grid charges for charging the BESS from the grid as in *Case 3NE (market responsive with grid charges for grid charging)* leads to the lowest EFC. When we compare the actual implementation of the operation strategies in Figure 3.6, we can observe that *Case 3NE* resembles *Case 1*. Looking at profits per cycle in Table 3.1, we can see that the current operation leads to the worst efficiency. Given the scarcity of the materials used in lithium-ion batteries (Costa et al., 2021), these systems should be used more effectively and efficiently.

Case	Profit [EUR]	Equivalent Full Cycles	Profit per Cycle [EUR/EFC]
Empirical Case	5.0	198	0.03
Case 1	54.7	141	0.39
Case 2Y	159.5	208	0.76
Case 2M	118.0	208	0.57
Case 2W	113.7	208	0.55
Case 3	225.2	640	0.35
Case 3NE	44.7	87	0.51

Table 3.1.: Average annual results per household on the day-ahead market

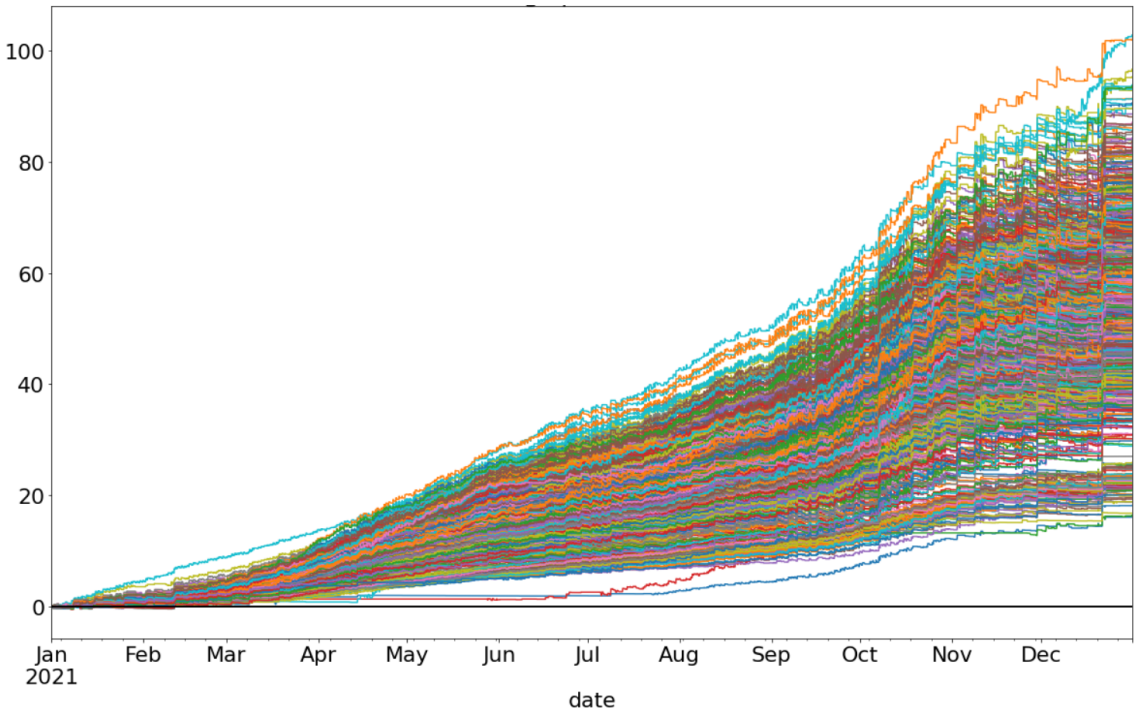


Figure 3.7.: Cumulated yearly household market profits in Case 1 (in EUR)

Figure 3.7 shows the cumulated annual profits per individual household for *Case 1 (flexible discharging)*. In contrast to the empirical case, the relatively constrained regulatory option of *Case 1* leads to consistently positive returns for every household in the dataset. This result shows that instead of mandating or clearly incentivizing specific behavior, regulators should focus on system objectives and provide guardrails to be respected that lead to those ends. However, in order to do that, a detailed understanding of household attitudes towards their BESS and its antecedents is necessary.

3.7 Consumer preferences and attitudes

The survey sample for this analysis was provided by Bilendi, a European sample provider and included 333 data points. In their internal characterization of participants, Bilendi keeps a flag for "owning a PV installation", which is unique and useful for our purposes since the installation of PV is an indicator for the installation of a BESS (Figgener et al., 2021). In addition to the final sample, 111 participants started the survey but failed an attention check and were, therefore, immediately

Construct	Reference	Items	Cronbach's α
Perceived effectiveness	Luo et al. (2008)	4	0.81
Trust in system	Gefen et al. (2003)	6	0.94
Perceived behavioral control	Sheeran and Orbell (1999)	4	0.84
Importance financial profitability	Bucher et al. (2016)	5	0.93
Importance sustainability	Bucher et al. (2016)	4	0.89
Satisfaction	Liao and Chuang (2004)	3	0.92
Knowledge energy consumption	(Schlösser et al., 2013)	2	0.55
Knowledge energy market	(Schlösser et al., 2013)	2	0.72
Knowledge energy system	(Schlösser et al., 2013)	2	0.72

Table 3.2.: Evaluation of constructs

screened out. Of the 333 valid completions, 79 participants were removed because they failed a comprehension check. Finally, another 59 participants were screened out because their answers to the questions about what BESS they used were incomprehensible or the provider did not exist. This led to a final cleaned sample of 195 BESS owners. 33% of those respondents are female, 79% stated a household income of more than 3000 Euro per month, 72% are older than 40 years and 99% have some form of advanced education. This shows that the sample consists of a group with high socio-economic advantage, which can be expected as investing in household BESS requires substantial financial resources.

Table 3.2 provides an overview of the used constructs, including the number of items, original reference, and Cronbach's alpha based on the resulting responses. All Cronbach's alpha values are in the acceptable range except for self-rated knowledge on energy consumption. In the following regression, the items of this construct are, therefore, used individually, while the other constructs are based on the average of the corresponding items.

The results of regressing the perceived effectiveness against the other discussed constructs are presented in Table 3.3. The adjusted R^2 value of 0.74 signals a reasonable explanatory power of the corresponding model.

The individual coefficients and p-values show a strong correlation between trust in the battery energy management system and its perceived effectiveness. Interestingly, self-rated knowledge is not correlated to perceived effectiveness across all constructs and items capturing this concept. This means that perceived effectiveness increases

Residuals:					
Min	Q1	Med	Q3	Max	
-2.16	-0.28	0.01	0.31	1.27	
Residual standard error:	0.47				
Degrees of freedom:	185				
Adjusted R^2	0.74				
F-statistic:	60.9				
Coefficients	Estimate	Std.Error	t-value	p-value	
(Intercept)	0.57	0.28	2.03	0.043	*
Energy Knowledge Consumption 1	-0.04	0.04	-1.03	0.306	
Energy Knowledge Consumption 2	-0.01	0.03	-0.27	0.790	
Energy Knowledge System	0.02	0.05	0.43	0.666	
Energy Knowledge Market	0.08	0.04	1.77	0.079	
Perceived Behavioral Control	0.06	0.03	2.06	0.041	*
Satisfaction with System	0.04	0.04	0.86	0.393	
Trust in System	0.56	0.04	12.70	0.000	***
Financial Importance	0.11	0.04	3.00	0.003	**
Sustainability Importance	0.08	0.04	2.36	0.020	*

Table 3.3.: Regression on Perceived Effectiveness (*p<0.05, **p<0.01, ***p<0.001)

with trust but not with increasing or decreasing self-perceived knowledge of energy consumption, the energy system, or the energy market. Additionally, perceived behavioral control positively affects perceived effectiveness. Similarly, the importance of financial performance and the importance of a sustainable operation of the BESS positively influence the perceived effectiveness. Interestingly, satisfaction with the energy system does not correlate with perceived effectiveness. The implications of this model are further discussed in the following discussion section.

3.8 Discussion

With this study, we contribute to an improved integration of household BESS into the energy system and a socio-technical understanding of the phenomenon BESS usage. To this end, we use empirical data to show the effect of regulation and corresponding consumption behavior on the effect of market integration.

Our results show that under the regulatory policy of self-consumption promotion and the resulting household behavior, the average welfare gain per BESS is virtually

zero. In other words, the installed BESS do, on average, not lead to any balancing benefit for the power system. This is not to say that they do not benefit the individual, but based on their operation and the corresponding household behavior, they do not lead to a benefit for the other market participants. In specific cases, they even add costs for other energy consumers. It is advisable that incentives would be designed so that individual benefits also lead to a global welfare gain. One might argue that the use of local renewable generation is a value in itself. However, this energy would be consumed in any case.

Our study has a few limitations. We are limited to the observations we report. We cannot say whether the absence of BESS would impact the price such that the outcome could be different. However, the results still show that the coordinated behavior of all household BESS is not beneficial or even detrimental to the energy system as a whole. Our sample might still be categorized as early adopters with corresponding biases in their behavior. We have no specifics on the households as the sample is anonymous. However, given the already considerable market penetration of BESS, it is unlikely that the demographics of battery owners will shift beyond single-family homes as of now. In any case, based on the data, households seem to behave according to the incentives given by the regulation and charge their storage fully during the day before depleting it in the early evening. These results show that uniform and homogeneous regulation of BESS without specific economic signals is likely to yield suboptimal results.

We only analyze empirical BESS load profiles without considering connected PV and household load profiles, as this data was not available due to privacy reasons. Although this does not change the overall direction of our results, since the BESS load profiles are directly connected to household loads and PV generation, we see potential for more granular analyses in future studies that also have access to the corresponding data.

Given that current regulation leads to seemingly suboptimal results, we propose alternative regulation and tariff structures utilizing price signals with a higher resolution derived from existing designs. This approach is, of course, based on the simulation of behavior, which, as we argue throughout this paper, does not necessarily represent empirical behavior. However, the simulation results in this study

represent behavior based on empirical data of battery-owning households that would not be impacted. Therefore, there is no reason to believe that they would change their behavior only because their batteries are operated differently. Corresponding battery energy management systems could redirect power flows without any effect on comfort or change in the behavior of households, leading to lower cycle numbers and more income as in Cases 1 and 3NE. The alternative regulation is, therefore, a Pareto improvement compared to the current regulation.

Although our study was solely conducted with a sample of German households, it yields important insights for lawmakers internationally. As described in Section 3.3, various countries, such as the UK, Australia or Japan, are also promoting battery installations to increase self-consumption of household PV generation in a comparable way. Hence, we argue that the results of our study can be generalized to some extent to other countries. Nonetheless, we call for a comparable empirical investigation in other countries under self-consumption-promoting regulation.

To round off our research on household behavior and BESS, we conduct a survey to better understand the perception of battery owners of their BESS effectiveness. We conducted this study with a sample different from that used for the empirical BESS usage evaluation. This was necessary because this sample is anonymous to us due to data privacy reasons. Again, the survey sample is not representative of the general population, but we are unaware of any study that describes the representative demographics of a battery owner. This survey should be understood as an exploratory study that can only provide some indication in regard to the relationship of the perceived effectiveness of BESS with other constructs. Yet, the results are interesting and open avenues for further research. They seem to indicate that trust in the system supplier is linked to perceived effectiveness, while different types of self-perceived knowledge are not. This suggests that the perception of effectiveness currently is independent of the perception of knowing what is effective but rather linked to a general positive feeling towards the BESS. It is also noteworthy that having an objective that is pursued with the BESS (be it financial or sustainable) positively influences perceived effectiveness. This seems reasonable as judging effectiveness is easier when it is clearly understood to what end one expects the system to be effective. Interestingly, satisfaction is not correlated with perceived effectiveness, while perceived behavioral control over personal energy consumption is. This might

be caused by the fact that those who feel that they would have an idea of how to act in the absence of the storage feel that the storage does what they would otherwise do. These results highlight the importance of proper regulation as households trust their supplier who will adapt operation to the corresponding regulatory policies.

The results indicate the need for further research. For instance, it is unclear how potential additional market profits should be optimally distributed between the aggregator and BESS owner to incentivize market-friendly behavior or to maximize perceived fairness. Furthermore, we only considered the participation of BESS in energy spot markets. We, therefore, ignored the potential of residential BESS to participate in reserve markets. This could further increase the revenue potential of market-oriented BESS operation strategies (Naseri et al., 2023). Finally, our findings on the perceived effectiveness of battery energy management systems lead to further questions in regard to the perception of information systems in household appliances.

3.9 Conclusion and Policy Implications

Distributed battery energy storage systems are an important asset for future energy systems that will continuously rely more on intermittent renewable generation. Therefore, the regulation and resulting battery energy storage system operation are highly relevant. However, there is currently no international consensus on an optimal regulatory framework, as shown by the variety of international policies. One widely proliferated regulatory framework incentivizes households to increase the self-consumption of electricity generated by their PV installations. Although previous research questions the benefits of self-consumption promoting regulation, this has never been analyzed empirically. To this end, we analyze an empirical sample of 947 year-long load profiles of household battery energy storage systems. We find that a self-consumption promoting regulation causes an operation of battery energy storage systems that leads to virtually no additional welfare for the energy system overall, while it does benefit battery owners. In individual cases, this regulation even leads to additional costs for the system that are socialized among energy consumers. These results hold important implications for policymakers worldwide. Given our results based on empirical field behavior, we show that self-consumption regulation needs to be carefully designed in order to contribute to the overall optimization of the energy system.

We, therefore, move on to simulate alternative regulatory approaches and tariff designs and show that these may lead to a Pareto improvement. This highlights the positive impact of slightly adjusting regulatory policies. We propose only slight adjustments to a self-consumption promoting regulation such as delayed feed-in, time-varying feed-in compensation, dynamic tariffs and structures from net billing. Our analysis shows that these adjustments lead to universally system-beneficial household battery energy storage systems. The paper is, therefore, a valuable point of reference for energy regulators and academics in the field.

Furthermore, we present a framework to describe the relationships between household battery regulation, battery energy management systems, household preferences, and the energy market, contributing to the understanding of these systems. We find that household preferences and behavior are often neglected when analyzing regulatory options. To contribute to this research gap, we complete the research with a third methodological approach with an exploratory perspective on the perceived effectiveness of battery energy storage systems. To this end, we conduct a survey among battery owners. The results indicate that trust in the system operator rather than self-perceived competency correlates with perceived effectiveness. This shows the important role of regulation in the operation of household battery energy storage systems as it shapes behavior. In conclusion, our study suggests that regulatory approaches towards household BESS operation should employ carefully designed and temporally differentiated signals to be more aligned with overall energy policy objectives.

Part III.

Operational Uncertainty

INTRODUCTION TO PART III

The operational performance of home energy management systems plays a critical role in the successful implementation of dynamic tariffs. These systems manage the scheduling of appliances such as battery storage systems, heat pumps, and electric vehicle charging. To leverage the flexibility potential of these appliances, accurate forecasting of future loads and access to comprehensive data for optimization and planning are necessary. Accurate load forecasting is essential for households to optimize the scheduling of their flexibility potentials, for aggregators to manage the flexibility of multiple households, and for grid operators to predict stress on the network infrastructure and take necessary measures in response.

The following part of this dissertation aims to address operational uncertainties in the operation of home energy management systems, ultimately enhancing their performance in the context of dynamic pricing. The contributions include an analysis of forecasting methods in light of the integration of novel appliances such as heat pumps (Chapter 4), the development of improved forecasting techniques to accurately predict peak loads (Chapter 5), the exploration of privacy-preserving approaches for peak time load forecasts (Chapter 6), and the proposal of a method to generate synthetic heat pump load profiles based on weather data (Chapter 7).

CHAPTER 4

THE IMPACT OF HEAT PUMPS ON DAY-AHEAD LOAD FORECASTING

This chapter examines the impact of heat pump installations on the selection of forecasting methods for predicting day-ahead electricity loads for groups of households, which are referred to as energy communities, throughout this chapter. With the growing number of heat pump installations and their significant potential for flexibility, accurate load forecasting becomes increasingly important for scheduling flexibility options and aligning electricity price signals with household energy demand in the context of dynamic tariffs.

Furthermore, this chapter investigates the influence of having access to separate historical data on heat pump electricity consumption on the quality of forecasts. This analysis provides valuable insights into key operational data retrieval considerations for aggregators managing home energy management systems under dynamic pricing schemes. Additionally, a battery-based peak shaving use case is implemented using the forecasted day-ahead electricity loads to evaluate the practical applicability of the forecasts in real-world scenarios.

This chapter comprises the following article: L. Semmelmann, M. Hertel, KJ. Kircher, R. Mikut, V. Hagenmeyer and C. Weinhardt. *The impact of heat pumps on day-ahead energy community load forecasting*, Applied Energy, 2024.

4.1 Introduction

Many European countries plan to install hundreds of thousands of heat pumps annually over the coming decades (Bundesverband Wärmepumpe, 2023). This leads to additional loads and burdens for distribution grids, for instance, through the over-

loading of transformers and power lines (Protopapadaki and Saelens, 2017; Çakmak and Hagenmeyer, 2022). To postpone heat pump-induced grid reinforcement measures, alternatives such as Demand Side Management or Battery Energy Storage Systems (BESS) can be used (Logenthiran et al., 2012; Stecca et al., 2020). One widely discussed concept for managing low voltage nodes are so-called energy communities, which combine tens to hundreds of households in a neighborhood to manage electricity needs (Barone et al., 2023) collectively. Operators of energy communities have to plan supply and demand under grid constraints to minimize purchase costs for the community members. A critical aspect of managing energy communities and distribution girds is scheduling flexibility measures. This requires an accurate forecast of upcoming and day-ahead loads (Coignard et al., 2021).

Although several studies discuss different methods to forecast day-ahead loads in energy communities and distribution grids, most focus on traditional load patterns, mainly dominated by conventional household appliances (Kuster et al., 2017; Coignard et al., 2021). These traditional load patterns will change in many countries by transforming the heating sector towards heat pumps (Love et al., 2017). This development has a severe impact on the operators of energy communities. Previous studies have not addressed two main questions: First, it is unclear if the same forecasting methods perform well for traditional household loads and heat pump loads. Second, the potential impact of the aggregation level on energy community load forecasts has not been investigated: it is unclear if operators of energy communities should directly forecast the whole load of the energy community, consisting of heat pump and traditional household loads, or if separate forecasts for the household and heat pump loads should be conducted and then aggregated. Several past studies underline that a higher aggregation level improves the quality of the forecast (Shaqour et al., 2022; Sevlian and Rajagopal, 2018). However, it has not been investigated if this holds true for forecasting different types of loads that follow distinct distributions.

In summary, this paper addresses the following research questions:

- Do the same methods perform well for forecasting traditional household loads and heat pump loads?
- Does the aggregation level of energy community loads – in particular, the

decision between directly forecasting the whole energy community load vs. forecasting heat pump loads and household loads separately – have an impact on the forecasting quality?

- How are the presented forecasting methods and aggregation strategies performing in an actual battery-based peak shaving use case of energy community operators?

We answer these questions by suggesting a state-of-the-art methodology, including feature engineering, feature selection, sophisticated Bayesian hyperparameter optimization, sliding window forecasting, and detailed benchmarking. The presented methodology is applied to a recent dataset of heat pump and household loads from an energy community in Hamelin, Germany (Schlemminger et al., 2022). We publish our pre-processing approach, the feature-engineered data, our results and the best-performing methods open-source.

The present paper is structured as follows. In Section 4.2, state-of-the-art related work is presented. Section 4.3 covers our methodology, focusing on the investigated models. Section 4.4 depicts the researched case study with energy community load data from Hamelin, Germany (Schlemminger et al., 2022). Section 9.7 presents our results in light of the previously introduced research questions. In Section 4.6, the results are discussed in detail. Finally, Section 9.8 presents the conclusion.

4.2 Related work

A wide range of studies discuss potential methods for load forecasting (Kuster et al., 2017; Coignard et al., 2021; Wang et al., 2022c). The overarching goal of these methods is to forecast upcoming loads based on previous observations. The time horizon of the load forecast can range from the next minutes to the next day, up to several days, months, and years (Kuster et al., 2017; Khuntia et al., 2016; Hahn et al., 2009). Also, the time resolution of the underlying data can range from a few minutes to a single hour, multiple hours, and whole days. All these factors play a role in the resulting quality of the forecast and the selection of the best-performing methods (Kuster et al., 2017). Our study mainly focuses on a day-ahead forecast of hourly loads, which is especially relevant for operational aspects in energy communities such as energy trading or scheduling flexibilities, as applied in several studies (Xu et al.,

2015; Javadi et al., 2022; Anees et al., 2021; Parra et al., 2017).

A broad spectrum of possible methods for day-ahead load forecasting tasks is discussed in the literature. The first advances in the field were made through statistical models such as the Autoregressive Integrated Moving Average (ARIMA) method (Hagan and Behr, 1987). The Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) (Tarsitano and Amerise, 2017) is an extension of the method. A different approach for load forecasting is using tree-based methods such as random forests (Dudek, 2015) or XGBoost (Tarsitano and Amerise, 2017). Tree-based methods use decision trees at their core to split the input data to make predictions over upcoming loads. Advantages of tree-based methods, for instance, XGBoost, are a high computational efficiency, good performance, and easy handling of multivariate data (Chen and Guestrin, 2016). For multivariate load forecasting, further input features like temperature measurements can be used, which can also be an important factor of electrical load forecasts (Semmelmann et al., 2023b). Over recent years, also neural networks have been increasingly used for load forecasting tasks, such as Long Short-Term Memory neural networks (Kong et al., 2017) or Transformers (Gao et al., 2023; Hertel et al., 2023). The same methods are analogously commonly used for heat load forecasts (Gong et al., 2022; Chung et al., 2022).

Most papers on load forecasting strictly differentiate between forecasts for traditional household loads and heat loads, which are either based on district heating systems (Gong et al., 2022; Powell et al., 2014), radiators installed at single-family houses (Bacher et al., 2013) or individual heat pumps (Xu et al., 2019; Song et al., 2023). However, the effective and reliable management of distribution grids and energy communities of the future requires consideration of heat pump-induced loads, which will lead to significant additional loads (Fischer and Madani, 2017). This leads to several practical considerations. First, it is unclear if the same methods that perform well for the forecasting of common household loads also perform well for the task of heat pump load forecasting. Through an increased share of heat pumps and thereby, a change of load structures, the recommended forecasting methods might also change. Second, whether the aggregation type impacts the load forecasting quality has not been investigated. Although many studies have shown that the higher the aggregation level, the better the forecasting quality due to stochastic smoothing

(Shaqour et al., 2022; Sevlian and Rajagopal, 2018), it is unclear if this holds true for aggregating household and heat pump loads. Hence, we investigate if household and heat pump loads should be directly aggregated and then forecasted or if the different loads should be individually forecasted, and then the forecasts should be aggregated. This also has practical implications for the management of the energy community: If individually-aggregated forecasts perform better than a directly-aggregated forecast, it might be worthwhile to advocate for data sharing of heat pump loads of households (Semmelmann et al., 2022). Fourth, most forecasting studies are decoupled from the actual use case in energy communities and distribution grids. The quality of the presented forecast methodologies is solely measured in terms of metrics such as the Mean Absolute Percentage Error (MAPE) or Root Mean Squared Error (RMSE) (vom Scheidt et al., 2020), without a thorough discussion of the metrics and its applicability for distribution grid-related tasks, such as peak shaving at transformers (Reihani et al., 2016).

The present paper fills these research gaps with a state-of-the-art methodology that considers the latest developments in load forecasting research (Hou et al., 2022). We also aim to address some common pitfalls in forecasting and machine-learning-based science itself. Recent studies found that many machine learning results are not reproducible due to a lack of transparency (Kapoor and Narayanan, 2023; Pfenninger et al., 2017), which is further aggravated through scarce open-source datasets (Gilbert et al., 2023). We address that by making our underlying data, results, and evaluation methodology open-source, thereby enabling researchers to easily build upon our results and benchmark their results against this study. In addition, for the sake of reproducible results (Pfenninger et al., 2017), we avoid the use of complex hybrid models and instead focus on base models. However, we encourage researchers to use our open-source data set, results, and benchmarking methodology to show possible advances of sophisticated hybrid models over the presented models.

4.3 Methodology

Our methodology follows the latest advances and common practices in load forecasting literature (Hou et al., 2022; Wang et al., 2022c). In the first step, additional input features are engineered to enrich the dataset. Subsequently, the feature set is reduced through a feature selection technique. Then, several forecasting mod-

els are presented. In the next step, we introduce another recent advance in load forecasting, the decomposition technique Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) (Ran et al., 2023). We conclude our methodology by introducing the two investigated aggregation levels, the underlying metrics, and our investigated peak shaving application of the presented forecasting strategies.

4.3.1 Feature engineering

An integral step in load forecasting is to create additional features that might help to capture additional patterns underlying the data (Zhu et al., 2022). We create the following features, in addition to given load and perfect foresight weather data, based on previous studies:

Type-of-day features: Several load forecasting related studies create additional features for type-of-day variables (Ziel, 2018). These binary features indicate if the given observation lies on a weekday, weekend, or holiday. We create the corresponding binary type-of-day features based on the timestamps of given observations.

Cyclical calendric features: Features such as the hour or month have a cyclical character, which might not be captured by representing them with their actual values (Haben et al., 2021). For instance, hour *0* and hour *24* would be interpreted as far away through a regression model, although they are the same value. This misinterpretation can be avoided by applying a *sine* and *cosine* transformation to the *day* and *month* observations, as described in Haben et al. (2021). We also create features for a twofold and fourfold *sine* and *cosine* transformation for possible consideration of patterns that occur with a higher frequency.

Rolling average of apparent temperature: It has been shown that rolling averages of the observed temperature are important input features for load forecasts due to the thermal inertia of buildings (Wang et al., 2016; Semmelmann et al., 2023b). Hence, we create a rolling average for the apparent outdoor temperatures' last 24 and 48 hours. We selected these intervals based on a pre-evaluation of correlations between temperatures and loads. We created the rolling temperature features based on the apparent temperature instead of the actual temperature due to the higher correlation between loads and apparent temperature observations. The ac-

tual temperature is the objective air temperature measurement, while the apparent temperature includes factors that affect the perception of temperature, for instance, humidity, wind speed and solar radiation.

Average load at same time step: For a better capture of the medium-term effects on (heat) load, Chung et al. (2022) suggest creating an additional feature with the average load at the same hour over the last week. Considering the load of the previous seven days goes beyond including lagged variables, which are only provided for two days in this study for computational reasons.

Past loads: Previous studies showed that past loads are amongst the most important load forecasting features (Hou et al., 2022). How these previous loads are given as input features to the model depends on the type of the model. While the following neural network-based methods, such as LSTMs, can handle whole feature vectors as input (Gers et al., 2000), classical machine learning methods, such as Random Forests, require a tabular representation of the data (Chen and Guestrin, 2016). This means a one-dimensional feature vector is used to predict one target value. Past load features are included as lagged to accommodate the tabular representation. On day d, the feature vector includes 48 past loads x_t based on the first timestep $t_0(d)$: $\{x_{t_0(d)-1}, x_{t_0(d)-2}, ..., x_{t_0(d)-48}\}$. Given the hourly time resolution, we consider 48 past loads, which equal two days, based on literature and initial experiments (Fan and Chen, 2006; Liang and Cheng, 2002). We take lagged features in relation to the first timestep of the respective day, $t_0(d)$, to ensure that the classical machine learning methods are working with the same input features as the neural network-based methods, which receive the two day-before past loads as an input vector.

4.3.2 Feature selection

We select the most relevant features through a filter and an embedded feature selection method (Pudjihartono et al., 2022) for computational efficiency and a reduction of potential overfitting. We separately conduct the feature selection process for the household-only, heat pump-only and aggregated energy community datasets, to ensure a fair comparability of the aggregation levels, which is described later in further detail. First, we filter out irrelevant features with a Pearson correlation lower than

0.1 (Koprinska et al., 2015). Then, the Random Forest algorithm is used as an additional embedded method to rank the potential features based on their predictive power (Pudjihartono et al., 2022). We only consider the ten features with the highest Random Forest feature ranking (Koprinska et al., 2015). The Random Forest method itself is explained in detail in the following "Models" subsection. Thereby, we combine the advantages of the correlation-based filter method (quickly reducing the search space) and the random forest-based embedded method (identifying features with high predictive power through a forecasting model) (Hu et al., 2015; Pudjihartono et al., 2022). The final feature set includes only features that pass both feature selection methods. The feature selection process is applied before enriching the dataset with the lagged features, to ensure comparability between neural network-based and classical machine learning methods.

We note that we include the cyclical hourly features, based on the previously described *sine* and *cosine* transformation, independently from the feature selection results, to maintain a relationship between past loads and the current observation for the classical machine learning methods. Since the past loads are based on the first daily time step $t_0(d)$ for the classical machine learning methods, the timesteps of the observations are essential to capture the relationship with past loads.

4.3.3 Models

In the following section, we introduce the investigated models in our study. We selected the underlying models based on a thorough analysis of benchmarking studies, identifying the most common and latest methods used for load forecasting tasks (Chung et al., 2022; Hou et al., 2022). We note that we excluded hybrid models for the sake of reproducibility of our results and that there are several further potential candidate models whose evaluation would go beyond the scope of this study.

Random Forests: Random forests are a machine learning method that combines an ensemble of decision tree predictors with random sampling (Breiman, 2001). In the first step, random samples are drawn from the underlying dataset used to build decision trees. The splitting of these trees is based on a random subset of features from which the best split is used. Possible splitting decisions are evaluated according to decision tree algorithms such as the Classification and Regression Tree (CART)

method. Finally, an ensemble of a large number of trees is created, which is then used to make its prediction as the average of the included trees. Random forests have been applied in several studies for day-ahead load forecasting tasks due to their high computational efficiency, rather low overfitting, and good quality of forecasts (Lahouar and Slama, 2015; Fan et al., 2022).

XGBoost: The XGBoost algorithm, introduced by Chen and Guestrin (2016), is a highly efficient machine learning algorithm applied in various forecasting tasks. Comparable to the previously introduced Random Forest method, XGBoost utilizes an ensemble of CART models. During the model's training, the loss function's gradient is constantly calculated. At the same time, new tree learners are added iteratively to the model to reduce the error of the model. The optimization function of the model includes a regularization term, which helps the model to prevent overfitting. Additional measures to prevent overfitting are the "shrinkage method" which reduces the influence of individual trees in the model, and column subsampling, which also increases the computational speed of the model. In general, the high computational efficiency and strong prediction accuracy make XGBoost a popular model for load forecasting studies (Abbasi et al., 2019).

LSTMs: A highly popular method for time series forecasting problems are Long Short-Term Memory networks (LSTMs) (Hochreiter and Schmidhuber, 1997; Gers et al., 2000). LSTMs are a special form of Recurrent Neural Networks, that use gate units and memory cells to "forget" irrelevant information over long-term patterns but remember important information. The ability to recognize patterns and to capture long-term dependencies makes LSTMs a popular choice for time series forecasting problems, amongst other load forecasting tasks. The original study from Hochreiter and Schmidhuber (1997) only contains one input layer, one hidden layer (which includes the memory cells and gate units and can be called "LSTM layer") and one output layer. Based on recent studies that apply LSTM networks for load forecasting, we include an additional hidden layer (Kong et al., 2017) and the option to include a dropout layer to prevent overfitting (Lin et al., 2022).

Transformers: A novel neural network architecture was introduced by Vaswani et al. (2017), which is increasingly used for natural language processing and computer vision tasks. Recently, the Transformer architecture was successfully applied to short-term load forecasting problems, due to its good performance in handling long-

term patterns (Hertel et al., 2022; Ran et al., 2023). In a more recent study, the applicability of Transformer models for long-term time series forecasting was debated, given that simple linear models were outperforming them on several datasets (Zeng et al., 2023). However, with the right training strategy and enough training data, Transformers outperform linear models and other baselines for short-term and long-term load forecasting (Hertel et al., 2023; Emami et al., 2023).

The standard Transformers are based on an encoder-decoder structure, although other variants exist (Nie et al., 2023). The load, calendar and weather features for the past time steps are fed into the encoder, and the calendar and weather features for the time steps to predict are fed into the decoder. The encoder layer consists of a stack of identical layers which in turn include two sublayers: a multi-head self-attention mechanism and a fully connected feed-forward network. The multi-head attention layer allows the model to access information from various representation subspaces at varying positions. The decoder contains, in addition to the multi-head self-attention layers, multi-head cross-attention layers accessing the output of the encoder. Overall, the Transformer architecture incorporates attention mechanisms at three different points: self-attention in the encoder, self-attention in the decoder and cross-attention that allows the decoder to access the output of the encoder. A final linear layer transforms the decoder output into the predicted load values.

4.3.4 Bayesian hyperparameter selection

An essential part of setting up machine learning models is the selection of the right model parameters, so-called hyperparameter tuning. A novel method for the optimal selection of hyperparameters is based on a Bayesian Optimization model (Wu et al., 2019), which especially comes with the benefit of high computational efficiency and fast convergence times. The model utilizes a Gaussian Process probabilistic model to map hyperparameters to an underlying optimization function, which aims to minimize the forecasting error of the model. The model uses an acquisition function to determine new hyperparameters, as a trade-off between exploration of new areas in the space of possible hyperparameters and the exploitation of existing well-performing observations. Our Bayesian hyperparameter model is initialized with 10 randomly drawn hyperparameter sets, the κ value of the model is set at 3, determin-

ing the trade-off between exploration and exploitation. We run the Bayesian model for 100 iterations. For every iteration, we run the target model with the respective hyperparameters given by the Bayesian model, which is then updated with the Root Mean Squared Error (RMSE) achieved by the target model. The data used for the hyperparameter tuning is not included in the test dataset, which is later used for the evaluation. We note that we discretize the hyperparameter search space, which is commonly done, but comes with some drawbacks, since the parameter spaces of categorical variables, such as the activation function of neural network methods, are disjoint (Wistuba et al., 2015; Lévesque et al., 2016). We accept this drawback for the sake of the computational efficiency of the method and given the fairness since all models are using the same approach.

4.3.5 CEEMDAN Decomposition

An increasing number of load forecasting studies is applying decomposition techniques to improve the model performance (Ran et al., 2023; Li et al., 2023a; Karijadi and Chou, 2022). Decomposition techniques decompose a given signal - such as a time series of loads - into subcomponents for a better understanding of underlying patterns and trends. One recent advance is the so-called Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method (Torres et al., 2011). The method first adds white noise to the target signal. Then, the signal is decomposed into different Intrinsic Mode Functions (IMFs) and the respective residue is calculated. The process is repeated and the IMFs are re-calculated until the residue cannot be decomposed anymore. For a thorough description of the method we refer to Torres et al. (2011) and Ran et al. (2023). Aggregating the resulting IMFs and residue yields the underlying signal.

The CEEMDAN algorithm has several advantages over alternative decomposition methods like the Empirical Mode Decomposition: it exhibits an improved handling of the mode mixing problem (having similar oscillations in different modes), it is more robust to noise, as well as being non-stationary (Torres et al., 2011). We first decompose the target load time series into IMFs with the CEEMDAN method on a monthly rolling basis (Duan et al., 2023). Then, we train a dedicated model for every IMF. Finally, we aggregate the forecasts of the forecasted IMFs to get the resulting

forecast for the target load. We compare the CEEMDAN method extension with the respective base models, taking over the same Bayesian-optimized hyperparameters from the base model.

4.3.6 Aggregation levels

Our study aims at forecasting loads at a low-voltage transformer (which would be for our German case between the 400V low-voltage level and the 20kV medium-voltage level (Bayer et al., 2018)), to which multiple households of an energy community are connected. Given a dataset of multiple individual household loads, we retrieve the aggregated energy community household load P_t^{HH} by aggregating the individual loads of all N households for each time step t:

$$P_t^{HH} = \sum_{i=1}^{N} P_t^{i,HH} \qquad \forall t \in T \tag{4.1}$$

The same procedure is repeated for the energy community heat pump load P_t^{HP}:

$$P_t^{HP} = \sum_{i=1}^{N} P_t^{i,HP} \qquad \forall t \in T \tag{4.2}$$

The final transformer power P_t^{Comb} consists of the sum of the household and heat pump load:

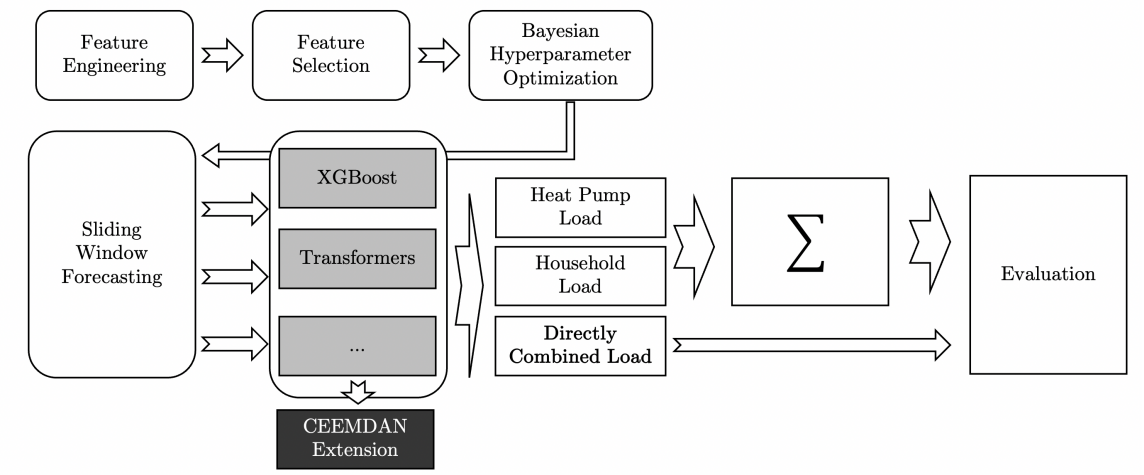
$$P_t^{Comb} = P_t^{HH} + P_t^{HP} \qquad \forall t \in T \tag{4.3}$$

From the perspective of the energy community or distribution grid operator, it remains unclear if household and heat pump loads should be:

- **Separate:** equaling to an individual forecast of the aggregated heat pump and household loads, and then summing the forecasts up to get the whole transformer load. This approach is reasoned by the different underlying distributions of heat pump and household load data (as shown later in the paper), which might make it reasonable to train distinct models for more accurate forecasting results (Kell et al., 2018).
- **Combined:** equaling to aggregating first all heat pump and household loads,

and then predicting the whole transformer load. This approach is reasoned by the frequently observed pattern that the higher the aggregation level, the better the forecasting results (Sevlian and Rajagopal, 2018).

Analyzing the effects of the aggregation level on forecasting quality has practical implications: predicting household and heat pump loads individually requires the operator to be able to separately access them, which might be challenging given the currently low level of observability in distribution grids (Bhela et al., 2017). Our study investigates if having additional, separate heat pump load data leads to a respective improvement in forecasting accuracy, justifying additional efforts for data retrieval.



(a)

Figure 4.1.: Forecasting methodology.

We illustrate the overall methodology in Figure 4.1. To enable a fair comparison of methods, we conduct a separate feature engineering, feature selection and Bayesian hyperparameter optimization for every dataset (Households (HH), Heat Pumps (HP) and Combined Load ($Comb$)). Thereafter, we conduct a sliding window forecast over one year for the presented forecasting models, either as a standalone model or combined with the CEEMDAN decomposition. Based on the sliding window forecast, we train the models on every first day of the investigated months based on the data from the past year. Then, the day-ahead loads of every day in the investigated month are forecasted. The process is repeated for every month in the test set. Then, the forecasting results are evaluated based on the metrics depicted in the

following.

4.3.7 Metrics

We evaluate the forecast quality by widely proliferated evaluation metrics, such as the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R^2 (vom Scheidt et al., 2020). The presented metrics combine advantages like intuitive interpretability (MAPE, MAE, R^2), share of explained variance through the forecasting model (R^2), appropriate consideration of large errors (RMSE) and reduced sensitivity to outliers (MAE) (Willmott and Matsuura, 2005; vom Scheidt et al., 2020; Chicco et al., 2021).

The MAPE is calculated as the mean of the percent deviation from predicted loads $P_t^{predict}$ from real observations P_t^{real} over all timesteps T, multiplied by 100:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{P_t^{predict} - P_t^{real}}{P_t^{real}} \right| \times 100 \tag{4.4}$$

The RMSE is calculated as the root of the mean squared deviation between $P_t^{predict}$ and P_t^{real}:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} |P_t^{predict} - P_t^{real}|^2} \tag{4.5}$$

The MAE is the mean of the absolute errors:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |P_t^{predict} - P_t^{real}| \tag{4.6}$$

R^2 is calculated by dividing the squared error between $P_t^{predict}$ and P_t^{real} and the squared error between the average load P^{mean} and actual values P_t^{real}:

$$\text{R}^2 = 1 - \frac{\sum_{t=1}^{T} (P_t^{predict} - P_t^{real})^2}{\sum_{t=1}^{T} (P^{mean} - P_t^{real})^2} \tag{4.7}$$

Higher values of R^2 indicate a higher forecast quality; values can range from -∞ to 1. When a model yields negative R^2 values, it indicates that its predictions for the target variable are less accurate than simply using the mean as a forecast (Chicco

et al., 2021).

4.3.8 Applicability

Most forecasting literature solely focuses on comparing and improving methods based on widely proliferated metrics, such as the MAPE or RMSE (vom Scheidt et al., 2020). However, a critical evaluation of how well the presented methods perform in actual use cases is often missing. Hence, our study compares the performance of the presented forecasting methods for the whole energy community load in an actual use case: reducing the peak aggregated energy community load by scheduling day-ahead charging and discharging of a BESS based on the day-ahead load forecast (Reihani et al., 2016). Reducing the peak load of the energy community is important to save the underlying distribution grid from degradation, to avoid costly reinforcement measures and to reduce possible peak power grid charges (Uddin et al., 2018).

We formulate the underlying optimization based on Reihani et al. (2016), simultaneously targeting peak shaving and load smoothing. For that, we minimize the squared sum of the forecasted load $P(t)^{predict}$ and the BESS power $P(t)^{BESS}$ multiplied by the time resolution Δt, over the forecasting horizon N, as depicted in Equation 4.8. The BESS operations are accommodated with a few constraints: the maximum power P^{max} shall never be exceeded (Constraint 4.8b). The State of Charge SOC at time t is defined by the previous charging operations $P(t)^{\mathrm{BESS}}$ multiplied with the time resolution Δt and divided by the maximum BESS capacity E^{tot} (Constraint 4.8c). Furthermore, the SOC has to be kept within 0 and 1 (Constraint 4.8d). For our study, we simply set the maximum BESS capacity E^{tot} at the peak load of the previous year $P^{max,y-1}$ times the time resolution Δt, which is in our case one hour. The maximum BESS charging and discharging power P^{max} is set at half the capacity E^{tot}, divided by the time resolution Δt. Finally, in Constraint 4.8e, we limit the amount of allowed full cycles to one equivalent full cycle per day d (equaling to one full charging and one full discharging cycle). We note that investigating different BESS sizes and peak shaving strategies might bring additional insights, but this would go beyond the scope of this study.

$$\min \quad \left(\sum_{t=1}^{T} \Delta t [P(t)^{\text{predict}} + P(t)^{\text{BESS}}]^2 \right) \tag{4.8a}$$

$$\text{s.t.} \quad P^{\max} \geq |P_t^{\text{BESS}}|, \quad \forall t \in T, \tag{4.8b}$$

$$SOC_t = SOC_{t-1} + \frac{\Delta t \cdot P(t)^{\text{BESS}}}{E^{\text{tot}}}, \quad \forall t \in T, \tag{4.8c}$$

$$0 \leq SOC_t \leq 1, \quad \forall t \in T, \tag{4.8d}$$

$$\sum_{t=1}^{T} |P(t)^{\text{BESS}}| \cdot \Delta t \leq E^{\text{tot}} \cdot \Delta t \cdot D \cdot 2 \tag{4.8e}$$

For every forecasting method, the previously depicted peak shaving and smoothing optimization is conducted using the Mixed-Integer Linear Programming solver Gurobi (Gurobi Optimization, LLC, 2023). Then, the suggested BESS charging operations P^{BESS} are applied to the actual observed loads P^{real}. Subsequently, based on the respective forecasting methods, we can compare the achieved peak reductions through a day-ahead scheduling of BESS operations. Thereby, we can evaluate the actual applicability of the presented methods for an energy community peak shaving task.

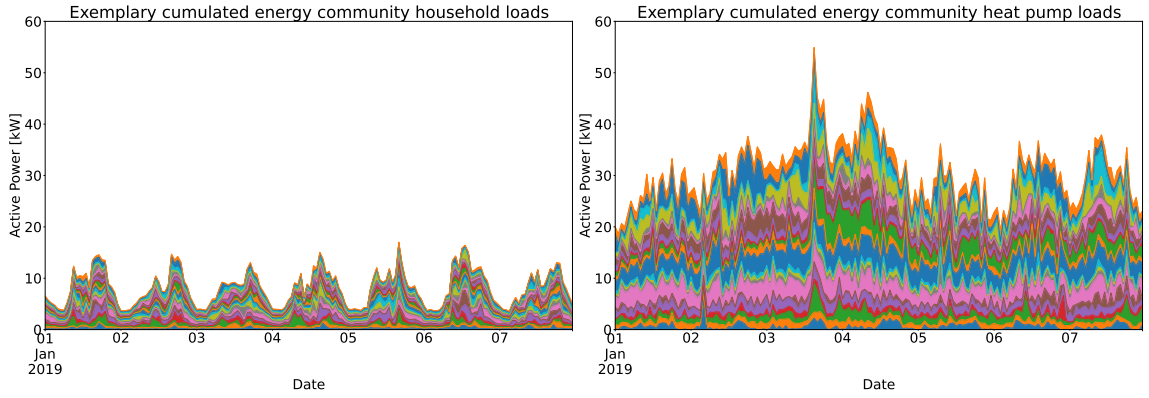
4.4 Case study

In this section, the underlying dataset and the results of our hyperparameter tuning process are presented.

4.4.1 Data

The previously presented methodology is applied to a high-quality dataset of household loads in an energy community in Hamelin, Germany (Schlemminger et al., 2022). Initially, the dataset includes active and reactive power, voltage and current measurements of 38 households equipped with water-to-water heat pumps and an additional heating rod as backup heater. The dataset includes separate measurements for the households and heat pumps in 10 seconds to 60 minutes resolution from mid-2018 to the end of 2020. The heat pumps from the dataset are both responsible for covering heating and hot water demand. For our study, we use the

hourly resolution of the active power and 21 out of the 38 households that do not have missing data. We see the agglomeration of the 21 households as an exemplary, small energy community, which can be found in a comparable size in existing studies (Van Der Stelt et al., 2018; Dong et al., 2020). We note that we are solely focusing on forecasting the aggregated active power. However, phase imbalance or voltage issues might arise through the installation of heat pumps (Navarro-Espinosa and Mancarella, 2014) in single- or three-phase configurations, and these are interesting directions for future work.

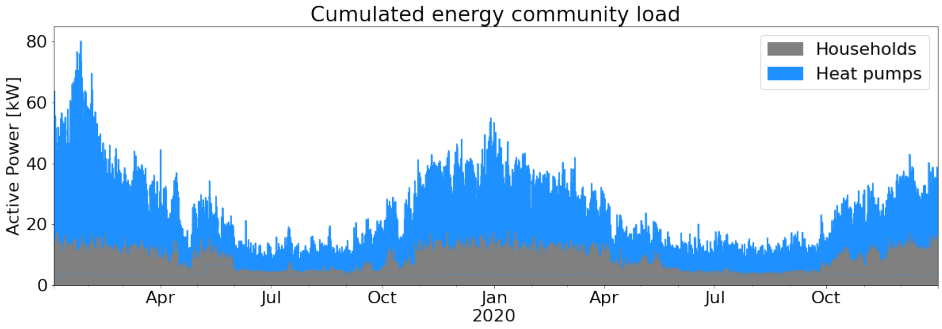


(a) (b)

Figure 4.2.: Exemplary load profiles of energy community in Hamelin, Germany

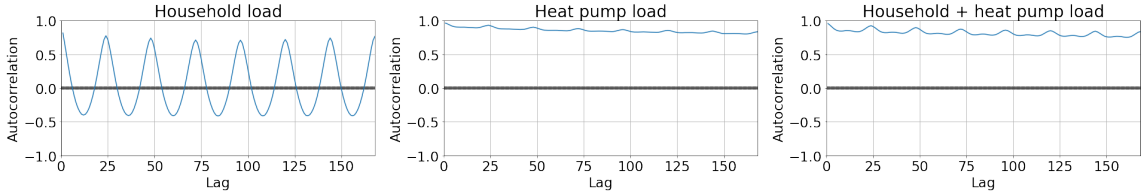
We depict an exemplary weekly load profile of the household and heat pump loads in Figure 4.2. We can observe that the underlying household loads follow a completely different pattern than the heat pump loads. While the household loads follow a daily pattern, with load peaks in the morning and evening, and load valleys in the night, the heat pump loads are rather on a constant high level over days, mainly caused by low temperatures. Furthermore, the heat pumps are for the exemplary illustrated winter weeks up to 8 times higher than the household loads, underlining the additional stress caused by heat pump installations on distribution grids (Protopapadaki and Saelens, 2017). Overall, through the installation of heat pumps the peak load in our dataset is raised from 20.1kW to 80.1kW, which represents a fourfold increase.

In Figure 4.3, the yearly energy community load is illustrated, showing again the heavy impact of heat pump installations on the load curve, with distinct new peak loads during winter months. Also, the autocorrelation profile depicted in Figure



(a)

Figure 4.3.: Yearly aggregated energy community load.



(a)

Figure 4.4.: Autocorrelation of household and heat pump loads.

4.4 shows the differences between heat pump and common household loads. While household loads are strongly correlated with the same daily hours, heat pump loads strongly correlate with loads in the previous hours. Potential autocorrelation patterns of future energy communities, including both household and heat pump loads, rather resemble the heat pump load autocorrelation structure.

The distinct profile of heat pump loads and their difference to the regular household loads underlines our question if separately forecasting aggregated household and heat pump loads before summing them might yield an advantage over directly forecasting the whole energy community loads, due to the different distributions and properties of the load curves.

Of the presented dataset, we use observations between the beginning of 2019 till the end of 2020 for our study. The data of 2019 is used for hyperparameter selection, with the first 6 months being used for training and the last six months for testing. We split our data half-half during the hyperparameter tuning process, since we want to cover different seasonalities. Then, the data in 2020 is used for the actual benchmarking of

the methods, with the previously determined parameters. We use a sliding window forecast, that trains the models at the beginning of each month based on the last twelve months. For forecasting the day-ahead hourly loads, features from the past two days are used, as previously described.

In the following, we present the results of the feature selection process and the subsequent hyperparameter selection, as well as the particular structure of the utilized models.

4.4.2 Feature selection

After conducting our feature selection process for the household, heat pump and combined dataset separately, we obtain the resulting feature sets in Table 4.1. We conduct separate feature selection processes to ensure an unbiased evaluation process in light of the comparison between individually aggregated or directly combined energy community load forecasting.

Table 4.1.: Resulting feature sets for different load models. Features that were excluded by the feature selection process, but were kept in the dataset to maintain temporal relationships, are marked in brackets (further details in the text).

	HH	HP	Comb
Apparent Temperature		✓	✓
Apparent Temperature: Rolling Average 24 Hours	✓	✓	✓
Apparent Temperature: Rolling Average 48 Hours	✓	✓	✓
Past loads (48 hours)	✓	✓	✓
Probability of Precipitation		✓	
Relative Humidity	✓	✓	✓
Temperature		✓	✓
Wind speed	✓	✓	✓
Cosine of hour	✓	(✓)	✓
Sine of hour	(✓)	(✓)	(✓)
Average load at same hour last week	✓	✓	✓
Dropped features (e.g. wind direction, ...)			

We find differences in relevant features between the different types of loads. For instance, the cosine of the hour would usually not be included in the heat pump load dataset, which is instead more focused on temperature features such as the apparent temperature or the probability of precipitation, which would not be included in the household and directly combined model. We note that we include the sine and cosine

of the hour nonetheless (marked in brackets in Table 4.1), to maintain the temporal relationship between current and past loads for the tree-based models, as explained in Section 4.3.1.

4.4.3 Hyperparameters and model structure

Based on the previously described Bayesian hyperparameter optimization and selected feature sets, we investigate for each model optimal parameters.

We use the first six months of 2019 for training during the hyperparameter selection process, and evaluate based on the last six months. Through our Bayesian hyperparameter optimization, we obtain different hyperparameters for heat pump, household and directly combined loads, as depicted in Table 6.2. The XGBoost parameter search space is based on Chung et al. (2022) and Wang et al. (2020b), the Random Forest (RF) search space is based on Walther et al. (2019) and Chung et al. (2022), the LSTM parameters are based on He and Tsang (2021) and Semmelmann et al. (2022), and the Transformer search space is based on Hertel et al. (2023).

The LSTM neural network is built with one bi-directional LSTM layer, two dense layers, from which the second dense layer has half the neurons of the first ones, and one dropout layer, before one final dense layer with neurons in the amount of the prediction horizon (in our case 24 hours) (Semmelmann et al., 2022).

Model	Parameter	Values	HH	HP	Comb
XGBoost	max_depth	(3,10)	3	5	6
	subsample	(0.4,1.0)	0.569	0.467	0.711
	min_child_weight	(2,6)	2.172	2.059	4.693
	colsample_bytree	(0.6,1.0)	0.633	0.887	0.744
	n_estimators	(10,200)	29	21	19
RF	max_depth	(1,500)	326	32	368
	learning_rate	(0.01,0.2)	0.035	0.079	0.01
	min_samples_split	(2,10)	2	6	10
	min_samples_leaf	(1,10)	5	4	1
	n_estimators	(10,200)	20	176	162
LSTM	batch_size	[32,64,128,256,512,1024]	256	128	64
	lstm_neurons	[32,64,128,256]	256	256	128
	lstm_first_layer	[32,64,128,256]	256	256	32
	dropout	(0.3,0.7)	0.3	0.7	0.3
	activation_function	[tanh, relu, sigmoid]	sigmoid	sigmoid	relu
	optimizer	[adam, adagrad, rmsprop]	sigmoid	sigmoid	sigmoid
	learning_rate	(0.0001,1)	0.0001	0.0001	0.0001
Transformer	num_layers	(1,6)	1	1	1
	d_model	[16,32,64,128,256,512,1024]	128	1024	128
	num_heads	[1,2,4,8]	8	1	2
	batch_size	[16,32,64,128,256,512]	16	64	16
	learning_rate	(0.0001,0.1)	0.0008	0.0001	0.0014

Table 4.2.: Results of the Bayesian hyperparameter tuning process for each model.

	MAPE	MAE	RMSE	R^2
Random Forest	**12.84**	**959.35**	1362.61	**0.74**
Random Forest CEEMDAN	15.51	1093.85	1481.39	0.69
XGB	13.11	969.79	1362.03	**0.74**
XGB CEEMDAN	14.53	1036.19	1421.26	0.71
LSTM	14.08	1041.63	1447.21	0.70
LSTM CEEMDAN	17.12	1161.97	1522.47	0.67
Transformer	13.19	965.39	**1352.30**	**0.74**
Transformer CEEMDAN	15.60	1049.55	1373.21	0.73

Table 4.3.: Household load forecasting results for evaluation data set (2020).

4.5 Results

In this section, we present the results of our study. First, reached metrics for household and heat pump load forecasting are presented. Second, the results of different aggregation strategies are depicted. Finally, we show the results for the application of the forecasted loads on the peak shaving case study.

4.5.1 Forecasting results

Table 4.3 displays the results for the forecasting of household loads. Depending on the metrics the best results are achieved by the Transformer (RMSE, R^2) and Random Forest (MAPE, MAE) model. Although the Transformer model is amongst the best models, we do not see a remarkably better performance of it over tree-based models such as the Random Forest. In addition, the CEEMDAN extension deteriorates the models rather than improving them.

Table 4.4 shows the forecasting results for heat pump-only loads, which differ strongly from the household-only results. The Transformer models significantly outperform the alternative models. Based on the MAPE and MAE, the Transformer model yields the best results, while based on the RMSE and R^2 metric, the Transformer-CEEMDAN model yields the best results. The CEEMDAN extension improves the forecast quality for the neural network-based methods while significantly deteriorating the tree-based methods. We note that we have compared the variance of the forecasting results over multiple runs of the underlying methods after obtaining the initial results, to analyze the uncertainty connected with them, as

	MAPE	MAE	RMSE	R^2
Random Forest	52.20	2054.40	2756.48	0.88
Random Forest CEEMDAN	97.54	2985.40	3897.60	0.76
XGB	49.17	2064.44	2817.45	0.88
XGB CEEMDAN	70.06	2617.10	3458.96	0.81
LSTM	70.51	2240.22	2908.37	0.87
LSTM CEEMDAN	49.32	1860.33	2482.39	0.90
Transformer	**33.03**	**1602.49**	2280.83	**0.92**
Transformer CEEMDAN	47.93	1690.77	**2226.14**	**0.92**

Table 4.4.: Heat pump load forecasting results for evaluation data set (2020).

depicted in the Appendix. Although we observe a higher standard deviation of the RMSE of the neural network-based methods, the order of the results remains the same.

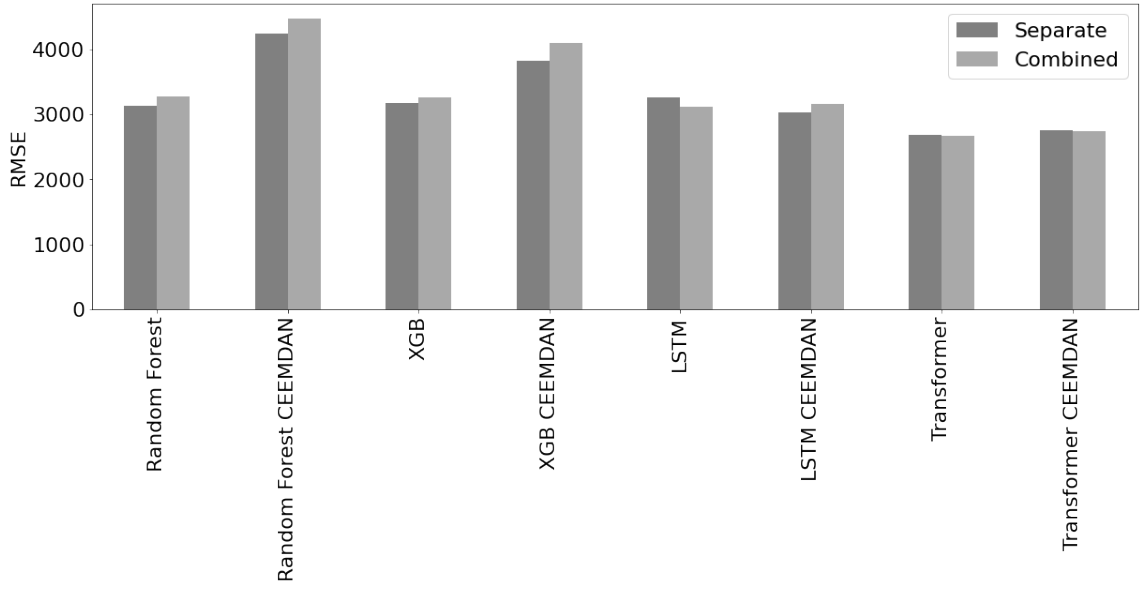
4.5.2 Aggregation level

Table 4.5 presents the results over the aggregation levels and methods. Overall, the results underline the superiority of the Transformer models: the "Transformer-CEEMDAN: combined" model achieves the best result for two of four metrics (RMSE, R^2), the "Transformer: separate" model achieves the best results for three of four metrics (MAPE, MAE, R^2). All Transformer models reach the highest observed R^2 score of 0.9. While the tree-based methods yielded comparable results for the household-only case, they are significantly worse when heat pump loads are added. This has implications for energy community and grid operators: forecasting models that have achieved good results in the past might not be the most suitable ones in a future with significant heat pump loads.

We compare the effects of the aggregation level in Figure 4.5. Although separately forecasting heat pump and household loads and then aggregating the forecast brings improvements for some methods, especially the best-performing methods only exhibit negligible performance differences, or even perform better when using the directly combined aggregation level (Transformer-CEEMDAN).

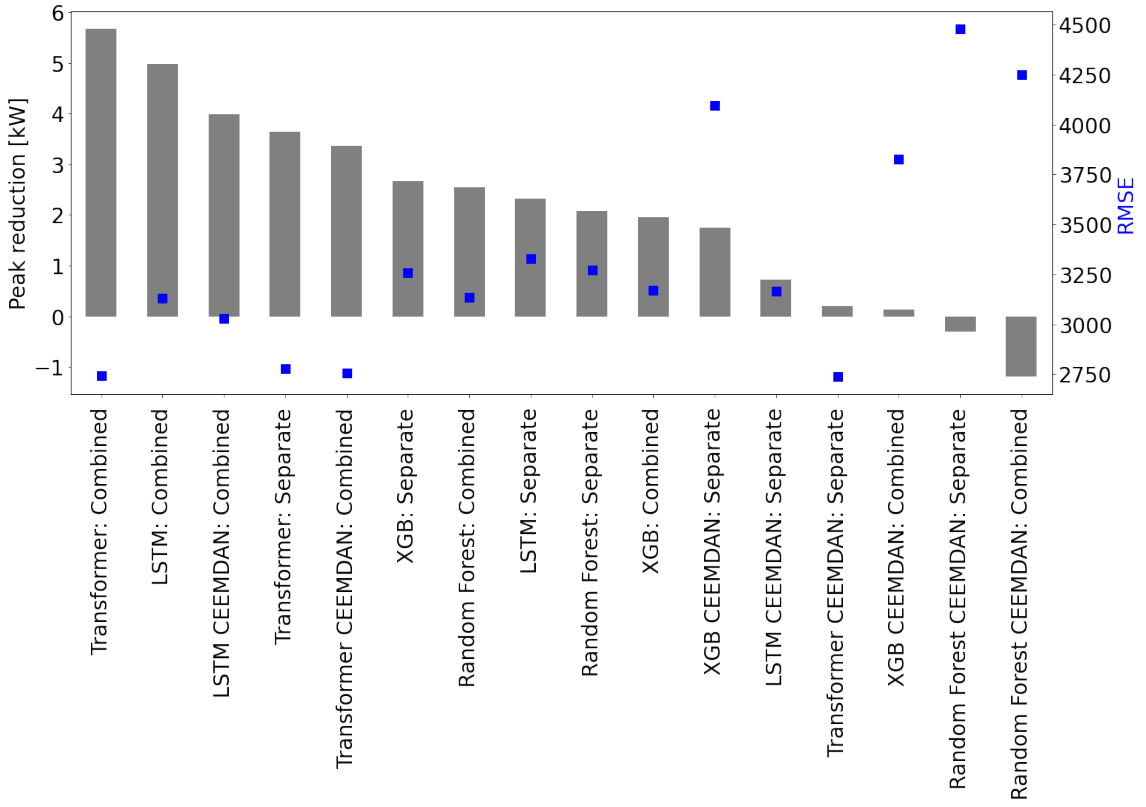
	MAPE	MAE	RMSE	R^2
Random Forest: Separate	16.79	2377.94	3137.00	0.87
Random Forest: Combined	17.24	2466.57	3269.30	0.86
Random Forest CEEMDAN: Separate	25.40	3274.45	4247.91	0.77
Random Forest CEEMDAN: Combined	22.94	3277.46	4478.73	0.74
XGB: Separate	16.48	2375.88	3168.89	0.87
XGB: Combined	16.58	2432.93	3256.33	0.86
XGB CEEMDAN: Separate	20.80	2913.93	3827.78	0.81
XGB CEEMDAN: Combined	21.31	3056.01	4097.77	0.78
LSTM: Separate	18.76	2525.47	3265.96	0.86
LSTM: Combined	17.11	2509.65	3330.41	0.86
LSTM CEEMDAN: Separate	16.47	2314.98	3029.79	0.88
LSTM CEEMDAN: Combined	17.66	2426.39	3167.11	0.87
Transformer: Separate	**13.43**	**2014.11**	2743.59	**0.90**
Transformer: Combined	13.76	2062.61	2776.97	**0.90**
Transformer CEEMDAN: Separate	16.16	2150.27	2754.22	**0.90**
Transformer CEEMDAN: Combined	14.95	2089.65	**2736.40**	**0.90**

Table 4.5.: Forecasting results for the whole energy community, including household and heat pump loads, for evaluation data set (2020).



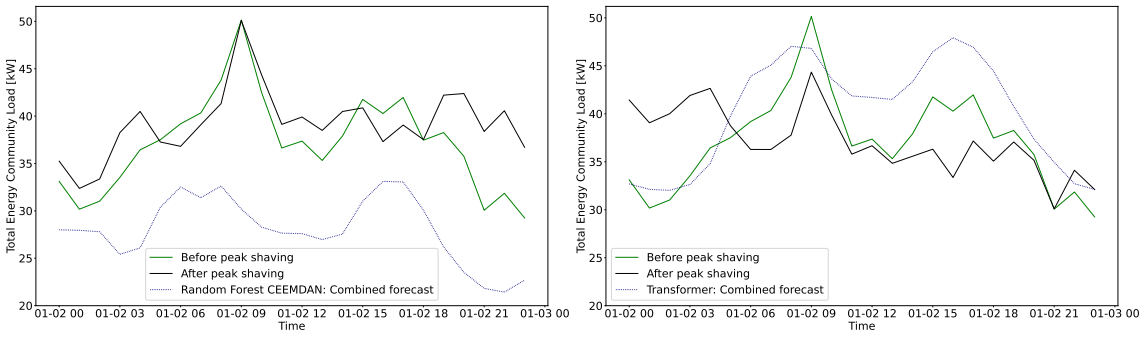
(a)

Figure 4.5.: RMSE per method and aggregation level.



(a)

Figure 4.6.: Achieved peak reduction based on day-ahead scheduling of BESS.



(a) (b)

Figure 4.7.: Peak shaving results on the day with the highest peak load.

4.5.3 Applicability

In the following, we investigate the applicability of the presented forecasts for day-ahead scheduling of a BESS sized at the hourly peak energy consumption of the year

before (80.1 kWh). The overall yearly peak reduction, based on the optimization model detailed in Section 4.3.8, is presented in Figure 4.6. We can again observe a strong performance of the Transformer-based methods, yielding solid peak reductions. The highest peak reduction of 5.7kW is achieved with the Transformer model and the "combined" aggregation level, which represents 57% of the theoretical optimal peak reduction based on a perfect foresight forecast.

On the other hand, models with weak forecasting performance, such as the Random Forest CEEMDAN model, yield negligible or even negative peak reductions through the day-ahead scheduling, underlining the importance of high-quality forecasts. We compare peak shaving on the day with the highest peak load with the method that yields the best peak reduction results (Transformer Combined) and the worst (Random Forest CEEMDAN Combined) by considering the load curves depicted in Figure 4.7. The Random Forest CEEMDAN method forecasts the highest peak in the afternoon, while the actual peak takes place in the morning. In addition, the load is consistently underestimated. Consequently, the peak load is not sufficiently reduced. Although the Transformer model also forecasts an afternoon peak, the overall load level and timing of peaks is forecasted better. Consequently, the BESS operation reduces the peak load level. We also observe that the three methods achieving the highest peak reduction follow the "combined" aggregation level, thereby delivering another indication that obtaining separate heat pump measurements does not necessarily lead to an operationally relevant improvement of the forecasting quality.

Although we see a tendency that models with good overall forecasting quality yield a reasonable decision basis for day-ahead BESS peak shaving scheduling, the "LSTM-CEEMDAN: Separate" or "Transformer-CEEMDAN: Separate" forecast with individually predicted household and heat pump loads raise awareness for potential problems when relying too much on forecast results. Even though both methods exhibit a solid forecasting performance, they lead to only limited peak reductions since they fail to correctly forecast the peak shape on the day with the worst peak load.

4.6 Discussion

The results of our study have implications for energy community and grid researchers and operators. We show that through the installation of heat pumps in

energy communities, the autocorrelation patterns and peak load magnitudes significantly change. Based on that, the choice of adequate forecasting methods should be reviewed and re-evaluated. While for traditional energy communities, tree-based forecasting models, such as XGBoost or Random Forests, are delivering reasonable forecasting results, it is not the case anymore when heat pumps are installed. Then, in our case study, Transformer-based models are significantly outperforming all other investigated models.

Thereby, we also contribute to the general discussion about Transformer models in load forecasting: Zeng et al. (2023) find that linear models outperform multivariate Transformer models for long-term forecasting. However, with a global training strategy, Transformer models outperform linear models and other baselines for short-term and long-term forecasting (Hertel et al., 2023; Nie et al., 2023; Emami et al., 2023). On very aggregated load time series, Transformer models can also outperform several baselines significantly (Hertel et al., 2022). In our experiments with traditional household loads, we see no large difference in forecasting quality between the Transformer model and tree-based models. However, when investigating heat pump loads and total energy community loads including heat pumps, the Transformer models considerably outperform the other models.

We can see a comparable pattern for the CEEMDAN technique, which decomposes the load time series in different Intrinsic Mode Functions that are separately forecasted and later aggregated. While the method does not improve the traditional energy community forecast, it constitutes one of the best methods for forecasting loads of heat pumps and energy communities with them.

We can transfer these results also to our application case, in which peak loads are reduced through an external BESS, based on the forecast of the depicted methods. The highest peak reductions are consistently reached through the Transformer method, which also achieved good forecasting results. However, we note that we only analyzed a limited case study with a given storage size and limited load data. Hence, future research should also critically evaluate the applicability of load forecasts on energy community- and grid operation-related, actual tasks. Although the forecasting methods that achieve good forecasting metrics also tend to show good results in the actual peak reduction task, we can also see discrepancies between forecast quality and actual effectivity, for instance through failing predicting the peak

load. Since most forecasting literature is focused on evaluating common metrics like the RMSE, without considering an actual use case, we call for a more task-centric forecast evaluation and the consideration of alternative metrics that might be more aligned with the task at hand.

Our result that separately predicting aggregated household and heat pump loads and then summing the forecasts up does not bring a meaningful advantage over forecasting directly the load of the whole energy community indicates that efforts to gather separate heat pump data might not be worthwhile. Instead, energy community and grid operators should focus on gathering solid load measurements at the transformer level, which can be used for forecasting models and operational decisions built upon them. We note that our study is focused on the low-voltage level. It has to be investigated if our findings hold true for the medium- and high-voltage levels.

Our study is based on the energy community household and heat pump load data from Schlemminger et al. (2022). The water-to-water heat pumps from the dataset are operated based on desired household temperature levels, neglecting potential price-based demand response signals (Klaiber et al., 2015). Through an increasing level of households with dynamic household prices, the load forecasting uncertainty could also rise, which should be considered in future studies (Klaiber et al., 2015). In addition, new heat pump technologies, tariff structures and regulation can lead to concept drifts that make an adjustment of forecasting models necessary (Gama et al., 2014). For instance, the German government has announced a new set of rules for controllable consumer devices – which will be implemented from 1st January 2024 on (§14a EnWG) (Bundesnetzagentur, 2023) – allowing grid operators to reduce the electricity consumption of heat pumps and electric vehicle chargers down to 4.2 kW during overloading events. Applying these new rules could lead to considerable changes of heat pump load profiles and an increase in forecasting uncertainty. Also, we note that the most proliferated type of heat pumps in Germany are air-to-water ones (Bundesverband Wärmepumpe, 2023), which can exhibit slightly different load profiles than the water-to-water heat pumps from in the underlying data set. However, due to comparable heat pump coefficients of performance, we argue that the heating demand of energy communities with water-to-water and air-to-water heat pumps – and respective forecasting outcomes – should be comparable (Çakır et al., 2013).

Overall, we understand our work as a first step towards the discussion of forecasting energy community loads with heat pumps, which will get increasingly important over coming years, given the increasing number of heat pump installations (Bundesverband Wärmepumpe, 2023). Because of the observed changes in load curves, novel forecasting methods should be discussed and applied. However, we note that we have only explored a limited amount of models, given the high number of novel models published in recent years (Wang et al., 2022c). Hence, we publish large parts of our study open-source, including the feature engineering and selection steps, the underlying final data set, the evaluation pipeline, the peak shaving application, the best-performing methods and our resulting forecast. Thereby, we simplify benchmarking novel methods against our results and contribute to open-source load forecasting research[3].

We note that our study has a couple of limitations. The peak shaving application of our forecasted loads is based on a retrospective simulation, which might neglect practical factors. We aim to empirically validate our results in further studies. Also, we focus on forecasting aggregated energy community loads. In further studies, it might be interesting to analyze the effects of forecasting individual household loads before aggregation. Furthermore, the underlying dataset includes perfect foresight weather data, which might lead to slightly overestimating the forecasting quality. We argue that this does not interfere with the general direction of our results, given the overall good level of weather data forecasts and that all models are based on the same data, hence a fair comparison is given.

4.7 Conclusion

Our study investigates the impact of the installation of heat pumps in energy communities on day-ahead load forecasting with a state-of-the-art forecasting pipeline. The installation of heat pumps leads to remarkable changes in autocorrelation patterns and peak loads of the energy community. This has implications for the overall load forecasting process. In particular, we find that:

[3]Our best performing forecasting models, all our feature-engineered and preprocessed data, our benchmarking pipeline and all our results are published open-source at `https://github.com/leloq/load-forecasting-with-heatpumps`. We encourage fellow researchers to benchmark novel forecasting methods against our results.

- The best-performing forecasting methods change after the installation of heat pumps. While for traditional energy communities, also tree-based models such as Random Forests or XGBoost deliver a reasonable forecasting quality, after installing heat pumps, Transformer-based methods outperform them significantly.

- The day-ahead energy community load forecasting quality cannot be notably increased by obtaining separate measurements of heat pump loads, which would constitute an additional effort for energy community or distribution grid operators.

- Transformer-based models are also delivering the best performance in a real-world peak reduction BESS use case for the investigated energy community with heat pumps. However, we see a discrepancy between forecasting metrics and actual results in the application task for some models.

Our findings have practical implications for operators of distribution grids and energy communities, making a re-evaluation of applied forecasting methods necessary and advocating against potentially expensive efforts to obtain separate heat pump load measurements.

We note that our study has some limitations: it is limited to a selected energy community, the practical application is based on a retrospective simulation and has not been empirically validated, the underlying data only includes water-to-water heat pumps, the forecast is based on perfect foresight weather data and only a selection of forecasting methods were applied.

Hence, we encourage researchers to use our dataset, results and evaluation pipeline, which we publish open-source, to benchmark novel methods against them to advance accurate forecasting techniques for loads of energy communities with heat pumps and to apply our methodology on alternative datasets. We also call for a more task-centric evaluation of forecasting methods, which might include the introduction of novel metrics that are more aligned with the application area of the produced forecasts. Also, future studies should investigate the impact of heat pump installations and aggregation levels on the forecast quality of the medium- and high-voltage grid, and empirically validate our results for energy communities and underlying low-voltage

grids. Further studies should also consider the effect of severe weather events on the forecasting quality of energy communities with heat pumps.

CHAPTER 5

A NOVEL LSTM-XGB FORECASTING MODEL BASED ON SMART METER DATA

This chapter introduces a novel hybrid forecasting algorithm that combines an LSTM model for predicting day-ahead load patterns with an XGBoost-based approach for correcting peak load forecasts. Additionally, the impact of aggregators' access to smart meter load measurements on forecast accuracy is examined. Forecasting peak loads is important from both an economic perspective due to potential peak demand charges and a technical perspective, considering the risk of transformer or line overloading. The investigation into the role of smart meter data further contributes to understanding the data retrieval considerations of aggregators.

This chapter comprises the following article: L. Semmelmann, S. Henni and C. Weinhardt. *Load forecasting for energy communities: A novel LSTM-XGBoost hybrid model based on smart meter data*, Energy Informatics, 2022.

5.1 Introduction

Day-ahead load forecasting is an essential task for grid operators and utilities in modern smart power systems to optimize balancing groups and to match upcoming demand and supply. Currently, standard load profiles, which are provided by the German Federal Association of the Energy and Water Industry in every year, are widely used by grid operators and modelers to approximate energy consumption (Peters et al., 2020). However, sector-coupled smart grids require improved forecasting methods, since new large consumers, such as heat pumps and electric vehicles, add significant loads to residential households. In addition, intermittent renewable generation, especially from photovoltaic, changes traditional load patterns. In smart

grids, more accurate forecasts could enable an improved management of emerging flexibility potentials, e.g., from battery storage, electric vehicles and heat pumps. The emergence of smart meters creates further possibilities in the field of day-ahead forecasting through the availability of high-resolution load data on household level. The authors of Zufferey et al. (2016) show with smart meter data from over 10,000 households in Basel, Switzerland, that a higher number of smart meter load profiles increases the general prediction accuracy significantly.

Improving day-ahead load forecasts also plays a vital role for (smart) energy communities. Energy communities are an emerging concept in research and practice, where local communities are collectively managing and optimizing their electricity production and consumption, e.g., through peer-to-peer trading or the joint utilization of storage systems (Shrestha et al., 2019; Henni et al., 2021). The importance of energy communities has been recognized by the European Union who plans to promote and strengthen decentral structures and has introduced the concept of "Citizen Energy Communities" in the 2019 *Directive on common rules for the internal market for electricity* (Golla et al., 2020; European Parliament and Council of the European Union, 2019). A central task in these energy communities will be the planning and management of flexibility potentials and electricity production. By improving community load forecasts, energy management can be improved, costs can be lowered and CO_2 emissions reduced (Wen et al., 2019; Grundmeier et al., 2014). While (day-ahead) load forecasting plays an important role on all levels of future smart grids, we specifically focus on energy communities in this work. A special feature of energy communities is their level of aggregation within a smart grid. In literature, energy communities typically consist of usually in between 2 to 500 households: in Coignard et al. (2021), communities between 2-95 households are analyzed, in Reijnders et al. (2020), 47 Dutch households are regarded, while Schlund et al. (2018) focus on 500 distributed households within a network section. This makes (day-ahead) load forecasting of energy communities based on smart meter data a different task than in individual households or larger grid sections. In individual households, smart meter data is either available or not, and load profiles may differ significantly from one household to another. In energy communities, there is already some level of aggregation which means that standard load profiles could be applied here as a (naive) forecast. However, the level of aggregation is much lower than in the case of

grid-level forecasts which can contain 10,000s of households (Zufferey et al., 2016). (Day-ahead) load forecasting in energy communities therefore deserves special attention, since the question arises whether smart meter data can be utilized strategically (e.g., by only installing smart meters in selected households) to improve load forecasts. This work thus aims at investigating the potential to improve day-ahead load forecasting of smart energy communities.

Recent research works like Wang et al. (2019a) have identified bi-directional Bi-directional long Short-Term Memory recurrent neural networks (Bi-LSTM) as suitable method to achieve high load forecasting accuracy. Although Bi-LSTM-based forecasts often enable high prediction accuracy in general, the forecasting of peak load hours and peak load quantities remains an important issue, as shown in Sarduy et al. (2016) and Liu and Brown (2019). Previous works in the field consider the forecasting of peak loads and peak load hours as part of the overall forecasting process, instead of separating the forecasting of the general load pattern (e.g., through an LSTM) from the explicit forecasting of peak loads. Forecasting peak loads is especially important for grid operators that have to prevent possible congestion situations in the grid or at transformer stations (Kucevic et al., 2021b). Only a fraction of existing works in the load foreacasting field incorporates smart meter data into the (LSTM-based) forecasting process. Furthermore, selection criteria for smart metered-households are rarely discussed (Haben et al., 2021; Kong et al., 2017; Ghiani et al., 2019).

In this work, we therefore contribute to the field of community load forecasting through two extensions of previous works. First, we demonstrate the improvements that can be achieved by incorporating smart meter data into day-ahead community load forecasts. We use the concept of feature permutation importance to identify the most important features for the training of a LSTM. This information could potentially be used to install smart meter infrastructure selectively by targeting the most relevant households for the community forecast. Second, we tackle the shortcomings regarding the incorporation of accurate peak load forecasting in previous works by proposing a hybrid bi-directional LSTM-XGBoost forecasting model. In the hybrid model, we deploy a LSTM which is suitable to accurately predict the general trend of aggregated community load. We then separately forecast peak load time and quantity with an XGBoost model using on smart meter data. Lastly, we

combine the peak load forecast with the LSTM-based general forecast to obtain a holistic community load forecast. Also, cyclical type-of-day features, such as the *sin* and *cos* transformation of the hour, are engineered to further improve the forecast quality without requiring additional data as demonstrated in Haben and Giasemidis (2016).

We therefore aim to investigate (i) if smart meter data can improve existing LSTM load forecasting models of energy communities and (ii) whether the problem of insufficient peak forecasts can be tackled with a novel hybrid model. The contributions of this work are thus threefold:

1. A bi-directional LSTM-based model for the forecast of the aggregated load of an energy community using individual and aggregated smart meter data as input.

2. The identification of the most important forecast input features in terms of type-of-day data as well as smart meter data of individual households using feature permutation.

3. A novel hybrid LSTM-XGBoost approach is proposed to incorporate accurate peak load forecasting and to improve overall accuracy of existing day-ahead aggregated load forecasting methods.

The remainder of this study is structured as follows. The first section covers the theoretical background of LSTM-based day-ahead load forecasting and XGBoost. The second section describes the methodology of this study and additional feature engineering steps that were undertaken. The third section describes the underlying dataset and the setup of the case study in which we demonstrate the developed methodology. The fourth section gives an overview of the results, whereas the fifth section discusses the findings of the case study. The final section summarizes the results and gives an outlook on further research directions in the field.

5.2 Theoretical Background

Day-ahead load forecasting has been a relevant topic in research for years. A traditional approach is the Autoregressive moving average (ARIMA) method, mostly combined with other methods like the lifting scheme (Lee and Ko, 2011), generalised

autoregressive conditional heteroscedasticity (Hor et al., 2006) or artificial neural networks (Dube et al., 2017). More recent works have shown the good applicability and performance of LSTMs for day-ahead forecasting problems (Kong et al., 2017). LSTMs, which were first introduced by Hochreiter et. al. in Hochreiter and Schmidhuber (1997), are based on Recurrent neural networks (RNNs). RNNs are sequence-based networks that can establish temporal correlations between previous and current information. This makes RNNs suitable for load forecasting problems, since upcoming loads often depend on daily patterns and routines as well as past load data. In Bouktif et al. (2018), France's metropolitan electricity loads are forecasted with a combined model of LSTMs and genetic algorithms for feature selection and hyperparameter tuning. The forecasting error, compared with an ExtraTree model, can be reduced by over 20%. In Jiao et al. (2018), LSTMs are used to forecast the electricity consumption of 48 non-residential consumers. By using LSTMs, a Mean absolute percentage error (MAPE) in the amount of 22.45% is reached. In comparison, with the traditional ARIMA method only a MAPE of 35.87% is achieved. As stated in Bouktif et al. (2020), it is important to find the right combination of LSTM hyperparameters in order to achieve accurate load forecasting results.

Load forecasting in energy communities is a special form of day-ahead load forecasting due to the level of load aggregation. For instance, the authors of Coignard et al. (2021) evaluate energy community load forecasts from 2 to 95 households. Furthermore, in Coignard et al. (2021), the importance of peak-load hour forecasts is emphasized in energy communities since, through accurate forecasts, the scheduling of battery storage systems and flexible loads can be optimized for high self-sufficiency rates.

Another recent development in machine learning is the so-called Extreme gradient boosting (XGBoost), which was introduced by Chen and Guestrin (2016). XGBoost is an efficient implementation of gradient boosting that is based on parallel tree learning and efficient proposal calculation and caching for tree learning. The XGBoost algorithm has found a wide variety of use cases, also in the context of energy systems research. In Zheng and Wu (2019), the framework is used for short-term wind power forecasting. In Wang et al. (2017), next month's electricity consumption is forecasted through a hybrid wavelet transform and XGBoost model. The first works have also combined XGBoost with day-ahead load forecasting models. For instance,

in Wang et al. (2021), an adaptive decomposition method is used together with an XGBoost-based regression model to forecast loads of industrial customers in China and Ireland. The authors of Li et al. (2019) separately forecast day-ahead loads through an LSTM neural network and XGBoost. Subsequently, an error-reciprocal method is used to combine the forecasts. However, both methods are used for a general load forecast instead of focusing the XGBoost forecast on peak loads. Previous works like Shwartz-Ziv and Armon (2022) have shown that XGBoost outperforms neural networks for regression and classification tasks on tabular data.

Several studies have shown that LSTM models accurately capture temporal dependencies but often underestimate peak values (Karimian et al., 2019; Feng et al., 2020). Hence, this study combines the LSTM day-ahead forecast, which generally depicts the temporal structure of the load, with a XGBoost forecast of peak load times and quantities. To our knowledge, no studies have pursued this approach so far.

In machine learning, feature importance measures help to better understand relevant inputs. A commonly used method for feature importance analysis is the permutation importance measure, which was introduced by Altmann et al. (2010). In this method, the decrease of prediction accuracy is measured after permuting input features. Thereby, a permutation importance score can be calculated for every feature to assess its importance for the model.

Building on these previous findings, we first develop a LSTM-based day-ahead forecast model and identify the most important input features in terms of easy-to-observe and smart-meter data using permutation importance. We then expand previous models by introducing an XGBoost model for forecasting both peak load time and quantity and combine the two approaches into one holistic hybrid model to improve overall accuracy of day-ahead aggregated load forecasts of energy communities.

5.3 Methodology

In this section, we describe our methodology for smart meter data-based LSTM forecasting of day-ahead aggregated community loads. An overview of the research framework of this study is depicted in Figure 9.1. In the following, we describe each component of the framework in detail.

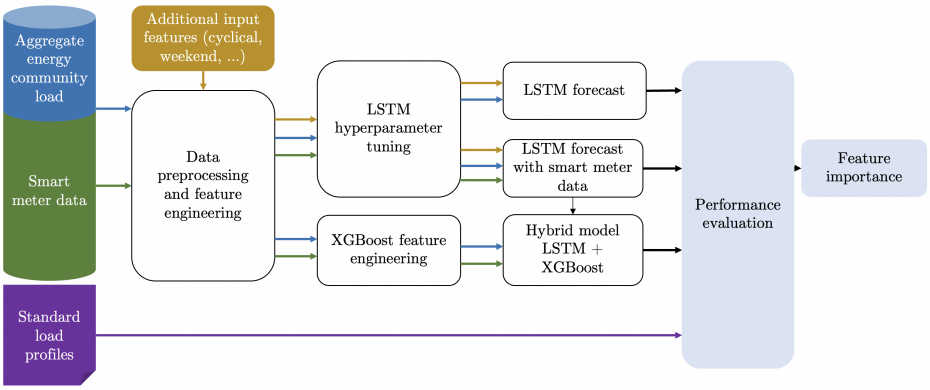


Figure 5.1.: Proposed research methodology.

Input data and type-of-day features. In a first step, the underlying smart meter data is preprocessed to create additional input features and to create the aggregate load of all smart meters , which serves as target variable. The aggregate load at time t can be calculated by summing up every load $\mathrm{P_{n,t}}$ of all smart metered households: $P_{agg,t} = \sum_{\mathrm{n=1}}^{\mathrm{N}} \mathrm{P_{n,t}}$.

As shown in Kanda and Veguillas (2019), adding additional type-of-day features to the underlying dataset can improve the general forecasting accuracy. Type-of-day features in this work include variables for the weekday, hour and month. To achieve periodicity for type-of-day variables, sinusoidal transformation is used as described in Haben and Giasemidis (2016). Also, a binary variable for weekends is added.

Data preprocessing. For the use of LSTM neural networks, the input data has to be preprocessed first. Every input feature I can be seen as a sequence of data points for the past timesteps, as stated in Equation 5.1:

$$\boldsymbol{I} = \{i_{t-K}, \ldots, i_{t-2}, i_{t-1}\} \tag{5.1}$$

In our case, represents the amount of timesteps per day in the underlying dataset. Due to the sensitivity of LSTMs to the data scale, all input vectors are normalized to the range of (0,1) by min-max-normalization. The input matrix for the forecast of any day d in the dataset consists of all input features I:

$$\boldsymbol{X_d} = \{\boldsymbol{I_1}, \boldsymbol{I_2}, ...\} \tag{5.2}$$

LSTM model. LSTMs are a special form of Recurrent neural networks (RNN), which solves the problem of exploding and vanishing gradients by adding a memory cell and gate (Wang et al., 2019a). Thereby, long-distance relationships between elements in sequence data can be processed. To create these temporal relationships, the LSTM defines and maintains a memory cell state over its life cycle. Three different types of timing modules exist in LSTMs: an input gate, a forget gate and an output gate. In turn, every timing module maintains its own memory cell and has its own task. The input gate is used to process incoming information, the forget gate decides about information retention of the historical cell state and the output gate processes outgoing information. The decision about information affecting the cell's state can be made selectively by using sigmoid activation functions. The output of the gates lies between 0 and 1. Thereby, a decision is made about the amount of information that is passed through the respective structure. A recent advance of LSTMs are Bi-directional long Short-Term Memory recurrent neural networks (Bi-LSTM), which can process both past and future information. In contrast, traditional LSTMs can only work with one-way transmission of information. Several works have shown that Bi-LSTM neural networks outperform traditional LSTMs in load forecasting problems (Wang et al., 2019a; Atef and Eltawil, 2020). Hence they are preferred over traditional LSTMs in this work. The unfolded structure of a Bi-LSTM is depicted in Figure 5.2.

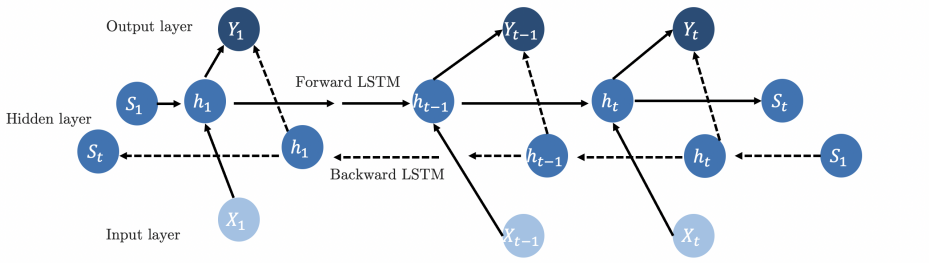


Figure 5.2.: Structure of Bi-LSTM network that both processes past and future information

The bi-directional LSTM layer in this study is followed by a dense layer, another bi-directional LSTM layer, two dense layers and a dropout layer to prevent overfitting (Tang et al., 2019).

LSTM hyperparameter tuning. To achieve a good combination of computational effort and accuracy, a randomized grid search is conducted for hyperparameter

tuning based on Wang et al. (2019b). The parameters listed in Table 6.2 represent the parameter search space, 100 runs are conducted with new random combinations of hyperparameters. The parameters for the search space itself are defined based on existing studies that use LSTM neural networks for load forecasting (Kong et al., 2017; Muzaffar and Afshari, 2019; Zheng et al., 2017; Bouktif et al., 2018; Jiao et al., 2018; Bouktif et al., 2020; Jahangir et al., 2020).

Hyperparameter	Value
Past input timesteps as multiple of	$1, 2, 3$
Batch size	$32, 64$
Epochs	$50, 75, 100$
Steps per epoch	$75, 100, 125, 150, 200$
Learning rate	$0.1, 0.01, 0.005, 0.001$
Number of units in LSTM-layer	$20, 70, 130, 200, 260$
Number of units in dense layers	$20, 70, 130, 200, 260$

Table 5.1.: In each run, a random combination of hyperparameters is tested.

Feature importance. Since this paper also aims to improve the general understanding of LSTM neural networks for energy community forecasting, the importance of the respective input features is investigated. Therefore, the measure importance Permutation importance (PIMP) is used, which was introduced by Altmann et al. (2010). The permutation feature importance metric is deployed in many load forecasting studies and is model-agnostic (Huang et al., 2016; Lahouar and Slama, 2015). To evaluate the importance of a certain feature I through permutation importance, its values are randomly shuffled to create a permuted input vector I_ξ. Now, the decrease in prediction accuracy in terms of $MAPE_{I_\xi}$ is compared to the MAPE of the unpermuted baseline model, as stated in Equation 5.3:

$$PIMP_I = MAPE_{I,\xi} - MAPE \tag{5.3}$$

A higher $PIMP_I$ means the model gets worse through a randomization of feature I, which indicates a higher feature importance.

XGB feature engineering. Previous studies on LSTM-based aggregated day-ahead load forecasting have shown improvements over alternative methods. However they are less well suited to predict varying peak load times and (extreme) peak quantities, as a time series forecast will always try to predict an expected value

rather than extreme events. To improve the accuracy of peak load prediction within our day-ahead aggregated load forecast, we therefore rely on a classification approach that specifically predicts peaks. We divide the task of peak load forecasting into two sub-tasks: predicting the *time* and *quantity* of the next day's peak load. Therefore, two XGBoost models are separately trained to forecast peak load quantities and times. For the model input, the whole data set of smart meter loads is reduced to daily load indicators. Each day d is depicted as vector of consecutive timesteps t, thus $d = [t_1, \ldots, t_K]$.

The two target variables and are calculated for every day d. In Equation 5.4, the peak load is obtained by getting the highest load $P_{t,d,agg}$ on day d:

$$P_{max,d,agg} = Max(P_{1,d,agg}, \ldots, P_{K,d,agg}) \tag{5.4}$$

In Equation 5.5, the peak time is obtained by getting the time step of the previously determined $P_{max,d,agg}$:

$$t_{P_{max,d,agg}} = t(P_{max,d,agg}) \tag{5.5}$$

Then, for every day d a range of statistical measures is calculated, as noted in Table 5.2, based on the previous day $d-1$ or up to 21 previous days $d-1, \ldots, d-21$. In detail, maximum loads, minimum loads, mean loads, median loads and load standard deviations are regarded. The subscript n denotes input features that are derived for each individual household in the respective community, whereas the subscript agg denotes that the input features are derived based on the aggregated energy community load. For the peak time $t_{P_{max}}$ forecasting model, also the peak times of the 20 smart metered households with the largest annual energy consumption, N_{large}, are regarded for the past 21 days. Only the 20 largest households are regarded due to computational limitations.

Feature	$t_{P_{max}}$ input data	P_{max} input data
$P_{max,d-1,n}$ $\forall$ n $\in$ N	Yes	Yes
$P_{min,d-1,n}$ $\forall$ n $\in$ N	Yes	Yes
$P_{mean,d-1,n}$ $\forall$ n $\in$ N	Yes	Yes
$P_{median,d-1,n}$ $\forall$ n $\in$ N	Yes	Yes
$P_{\sigma,d-1,n}$ $\forall$ n $\in$ N	Yes	Yes
$t_{P_{max,d-1,agg}}$, ..., $t_{P_{max,d-21,agg}}$	Yes	No
$t_{P_{max,d-1,n}}$, ..., $t_{P_{max,d-21,n}}$ $\forall$ n $\in N_{large}$	Yes	No
$P_{max,d-1,agg}$, ..., $P_{max,d-21,agg}$	No	Yes
$P_{max,d-1,agg}$	No	Yes
$t_{P_{max,d-1,agg}}$	Yes	No

Table 5.2.: Features for XGBoost datasets

XGBoost model. XGBoost was introduced by Chen and Guestrin (2016). The approach builds upon gradient tree boosting algorithms, which are extended by a second-order Taylor expansion for a faster optimization process and to avoid overfitting. Previous works have shown that the XGBoost algorithm can be well applied for load forecasting tasks. For instance, the authors of Wang et al. (2021) apply XGBoost to the load forcasting of industrial customers in Ireland and China.

XGBoost is based on an ensemble of Classification and Regression Tree (CART), which are used as weak learners. Weak learners are usually performing slightly better than random guesses in classification and prediction tasks and are modified over the iterations of the optimization process to form a well-performing ensemble model. The prediction for sample i is defined by Equation 5.6,

$$\widehat{y}_i = \sum_{m=1}^{M} f_m(i) \tag{5.6}$$

where M is the number of Classification and Regression Trees, and $f_m(i)$ is the forecasted value for the sample i in tree m. The underlying objective function is introduced in Equation 5.7:

$$\text{Obj} = \sum_{i \in I_j} l(y_i, \widehat{y}_i) + \sum_{m=1}^{M} \Omega(f_m) \tag{5.7}$$

where I_j is the set of all samples in leaf j and l is the second-order loss function that measures the difference between predicted value $\widehat{y}_i$ and actual value y_i. The

regularization term , as defined in Equation 5.8, consists of the number of leaf nodes T. The score of leaf j is measured by w_j. γ and β are parameters of the tree:

$$\Omega\left(f_m\right) = \gamma T + \frac{1}{2}\beta \sum_{j=1}^{T} w_j^2 \tag{5.8}$$

The structure of the Classification and Regression Trees and exact split points are determined by the quadratic objective function, which is simplified through the aforementioned second-order Taylor expansion, as noted in Equation 5.9:

$$\text{Obj } = \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \beta \right) w_j^2 \right] + \gamma T \tag{5.9}$$

where g_i is the first derivative of the loss function and h_i is the second derivative. The quadratic equation 5.9 is solved to obtain the leaf node score :

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \beta} \tag{5.10}$$

As a scoring function , Equation 5.11 is introduced to evaluate the quality of the tree structure q:

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \beta} + \gamma T \tag{5.11}$$

Finally, to determine the tree structure and splitting decisions , a greedy algorithm is used that starts with one leaf and then iteratively adds branches, as noted in Equation 5.12:

$$\mathcal{L}_{split} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \beta} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \beta} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \beta} \right] - \gamma \tag{5.12}$$

where I_L are sample sets of left nodes and I_R are sample sets of right nodes. Given that $I = I_L \cup I_R$, the loss reduction after a split is denoted by $\mathcal{L}_{split}$. Through Equation 5.12, possible split candidates are evaluated. For a more detailed explanation of the XGBoost algorithm, we refer to Chen and Guestrin (2016).

Based on the previously introduced approach, two separate XGBoost models are

trained to forecast and . Since forecasting is a classification problem, the Receiver Operating Characteristic Curve (ROC AUC) is used as optimization metric. For the forecasting model of , the Mean squared error (MSE) is used as optimization metric, since this is a regression task.

The parameters of the XGBoost model are also determined through a hyperparameter search, based on parameters from Zheng et al. (2017); Wang et al. (2021); Li et al. (2019). The parameter search space is described in Table 5.3. Parameters are separately determined for the peak time and peak load model. In total, 1000 runs are conducted per model.

Hyperparameter	Space	Distribution
N estimators	[40,1000]	Randint
Max depth	[1, 100]	Randint
Learning rate	[0.01, 0.59]	Uniform
Subsample	[0.3, 0.6]	Uniform
Colsample bytree	[0.5, 0.4]	Uniform
Min child weight	[0.05, 0.1, 0.02,1, 2, 3, 4]	None
Gamma	[0,0.5,2,10]	None

Table 5.3.: Hyperparameter search space.

Hybrid LSTM-XGB model. After forecasting and with the XGBoost model, the results have to be incorporated into the LSTM forecast, which is a vector of forecasted loads : $\{\hat{\mathrm{P}}_1, \ldots, \hat{\mathrm{P}}_\mathrm{k}, \ldots, \hat{\mathrm{P}}_\mathrm{K}\}$. For readability, we simplify the outputs of the XGBoost prediction as $t_{XGB} = t_{Pmax,d}$ and $P_{XGB} = Pmax, d$.

The most straightforward approach would be to simply replace the value of the original LSTM load forecast, $\hat{\hat{P}}_k$ at time step $k = t_{XGB}$ with the predicted peak load quantity $\hat{P}_{XGB}$. However, this bears the risk that in case the peak load time has not been predicted correctly, the prediction will extremely overestimate the true load. We therefore scale down the predicted peak load by a parameter $\lambda \in [0,1]$. In our case, we set $\lambda = \frac{1}{2}$ and calculate the new peak value according to Equation 5.13.

$$\hat{\mathrm{P}}_{\mathrm{t_{XGB}}} = \hat{\mathrm{P}}_\mathrm{t} + \frac{1}{2}(\hat{P}_{XGB} - \hat{\mathrm{P}}_\mathrm{t}) \tag{5.13}$$

Since load peaks are usually patterns of subsequent, elevated loads, in Equation 5.14 also the previous load and subsequent load are adapted by a quarter of the difference between the XGB and LSTM-based peak load forecast:

$$\hat{\mathrm{P}}_{\mathrm{t_{XGB}-1}} = \hat{\mathrm{P}}_{\mathrm{t-1}} + \frac{1}{4}(\hat{P}_{XGB} - \hat{\mathrm{P}}_{\mathrm{t}}) \qquad \hat{\mathrm{P}}_{\mathrm{t_{XGB}+1}} = \hat{\mathrm{P}}_{\mathrm{t+1}} + \frac{1}{4}(\hat{P}_{XGB} - \hat{\mathrm{P}}_{\mathrm{t}}) \tag{5.14}$$

Thereafter, the adjusted values are inserted into the forecasting vector:

$$\{\hat{\mathrm{P}}_1, \ldots, \hat{\mathrm{P}}_{\mathrm{t_{XGB}-1}}, \hat{\mathrm{P}}_{t_{XGB}}, \hat{\mathrm{P}}_{\mathrm{tXGB+1}}, \ldots, \hat{\mathrm{P}}_{\mathrm{k}}, \ldots, \hat{\mathrm{P}}_{\mathrm{K}}\} \tag{5.15}$$

Performance evaluation. Finally, the forecasting performance is evaluated by the most commonly used metric in day-ahead forecasting, the Mean absolute percentage error (MAPE). The MAPE divides the sum of percentual deviations from the forecasted loads P_{ft} by the actual loads P_{rt} with the number of time steps , as described in Equation 7.2:

$$\mathrm{MAPE} = \frac{1}{K} \sum_{t=1}^{K} \left| \frac{P_{ft} - P_{rt}}{P_{rt}} \right| \times 100 \tag{5.16}$$

As a second metric, the Root-mean-squared error is used, which is the root of the mean squared error from P_{ft} and P_{rt}, as denoted in Equation 5.3:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} |P_{ft} - P_{rt}|^2} \tag{5.17}$$

In this work, the MAPE is calculated for all forecasted day-ahead loads as well as only for the highest forecasted load, averaged over all days in the test data set. For the general load forecast, also the RMSE metric is regarded. Through this, we can assess both overall load forecast quality and the peak load forecasting capabilities of our model.

In order to achieve more stable and unbiased results, the dataset is further split with a twelvefold-cross-validation, where every split represents 30 days (Burman, 1989). To achieve comparable results within splits and even-sized train-validation-test sets, the dataset is shifted for 30 days in every iteration.

In the following, we apply the developed methodology to a case study in order to demonstrate the achievable improvements in energy community load forecasting through our developed model.

5.4 Case study

In this section, the setup of our study is described. In particular, the underlying dataset is described, the results of the LSTM and XGBoost hyperparameter tuning are presented and the four forecast scenarios are introduced.

Dataset. The introduced method is evaluated based on a dataset of German smart meter household data from 2019 published by Beyertt et al. (2020). The dataset includes 200 households that agreed on the publication of their loads, and 70 households participated in a behavioral experiment. The data of the remaining 130 households is used in this study. The households from the study are distributed all over Germany, which prevents us from adding geographically dependent weather features to the data set. The calculated aggregated load of all 130 households represents the load of a hypothetical energy community. In Table 5.4, the dataset is described. In Figure 5.3, an exemplary load of the energy community is depicted. We can observe a repeating pattern of load peaks in the morning and evening and load valleys in the night. The households in the dataset are relatively small with a mean annual household consumption of 779kWh.

The number of households in the energy community constructed in this paper lies within the range of community sizes from existing studies. In Coignard et al. (2021), the communities are randomly sampled with 5 to 95 households with 4MWh annual consumption each, resulting in an aggregated load between 20MWh to 380MWh. In a case study from Heeten, Netherlands an energy community of 47 households is depicted, with a calculated energy usage of 164.500kWh per year (Reijnders et al., 2020). In Schlund et al. (2018), different configurations of up to 500 distributed households are regarded.

The dataset is split in twelve parts for the twelve-fold cross-validation. The first 252 days (36 weeks) of data serve as training data, the following 83 days (11.86 weeks) for validation and the remaining 30 days (4 weeks) as test data, representing approximately one month each. After every iteration, the dataset is shifted by 30 days. Therefore, our train-validate-test split is 70%, 23% and 7%.

LSTM model. The proposed LSTM is set up based on best practices from existing research (Kong et al., 2017; Muzaffar and Afshari, 2019; Zheng et al., 2017;

Parameter	Value
Time covered	01.01.2019 - 31.12.2019
Time resolution	15 min
Smart-metered households (N)	130
Minimum annual household consumption	116.65kWh
Maximum annual household consumption	2011.03kWh
Mean annual household consumption	779.91kWh
Aggregated annual load of energy community	101.4 MWh
Mean load of energy community	11.57 kW
Minimum load of energy community	4.92 kW
Maximum load of energy community	31.57 kW

Table 5.4.: Descriptive statistics about the underlying dataset.

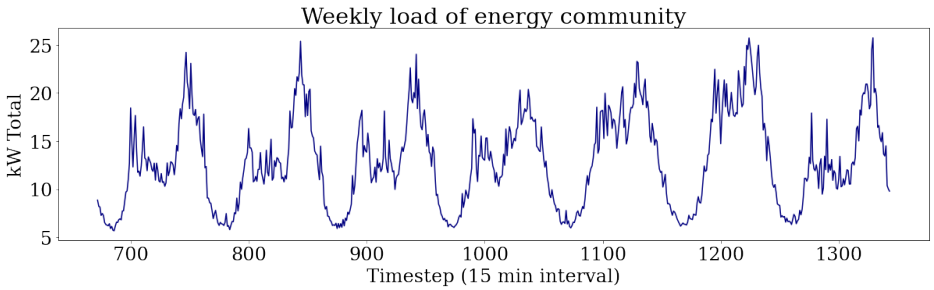


Figure 5.3.: Aggregated energy community load from 07.01.2019 to 14.01.2019.

Bouktif et al., 2018; Jiao et al., 2018; Bouktif et al., 2020; Jahangir et al., 2020). Several optimizers are compared (SGD, Adagrad, RMSProp, Adam). Due to slightly better results, the Adam optimizer is used. For improved computational efficiency, training is stopped early when no further improvements in valuation loss can be observed. The final LSTM parameters obtained from the hyperparameter search are listed in Table 5.5.

The models are trained and evaluated on a Google virtual machine with 8 virtual CPUs and 64 GB RAM. The LSTM neural networks are realized with the help of the TensorFlow toolkit (Abadi et al., 2016).

XGBoost. To find the optimal parameters for the XGBoost models for peak time and peak load forecasting, a hyperparameter search has been conducted. The resulting parameters are listed in Table 5.6.

Parameter	Value
Optimizer	*Adam*
Time steps per day	96
Past input timesteps as multiple of	1
Batch size	64
Epochs	75
Steps per epoch	200
Learning rate	0.001
Number of units in LSTM-layer	200
Number of units in dense layers	130

Table 5.5.: Final LSTM parameters.

Parameter	Value peak time model	Value peak load model
N estimators	87	929
Max depth	41	96
Learning rate	0.18	0.43
Subsample	0.88	0.46
Colsample bytree	0.75	0.88
Min child weight	4	0.05
Gamma	10	2

Table 5.6.: Final XGBoost parameters for peak time and peak load model.

Scenarios. In this work, four different scenarios are compared. Standard load profiles (SLP) for the year 2019 are used as baseline case, obtained from Stromnetz Berlin GmbH (2019). The standard load profiles are scaled proportionally to the aggregated energy community load (Meier, 2000). In a second scenario, the LSTM is used to forecast day-ahead energy community loads, with the only input features being day-before aggregated energy community load and type-of-day features as inputs, such as the *sin* and *cos* of the hour, weekday or month. The second scenario is in the following, denoted as LSTM. In the third scenario, we add the smart metered loads of the last day of each household of the 130 consumers (LSTM SM). Finally, in the fourth scenario, we combine the results of the third scenario with the XGB peak load finetuning (LSTM SM XGB). All four scenarios and the respective input datasets are summarized in Table A.2.

Scenario	Standard Load Profiles	LSTM	Smart Meter Data	XGB Finetuning
SLP	Yes	No	No	No
LSTM	No	Yes	No	No
LSTM SM	No	Yes	Yes	No
LSTM SM XGB	No	Yes	Yes	Yes

Table 5.7.: Summary of datasets and included data.

5.5 Results

In this section, we describe and compare the results of the four introduced scenarios. We also evaluate the standalone performance of the XGBoost model and present the results of the permutation feature importance analysis.

In Figure 5.4, the day-ahead forecast for October 17 2019, a weekday, is displayed for the standard load profiles (SLP), the general LSTM model (LSTM) and the LSTM model with smart meter data (LSTM SM). We can observe that both the LSTM and LSTM SM manage to forecast the general load pattern quite well, whereas the SLP overestimates the actual load profile on this certain day. When we also take the day-ahead forecasts of other days into account, we can see that the SLP follows a rather generic pattern, that only manages to match the daily load irregularly. We also note that the LSTM SM forecasts the day-ahead loads slightly better than the LSTM.

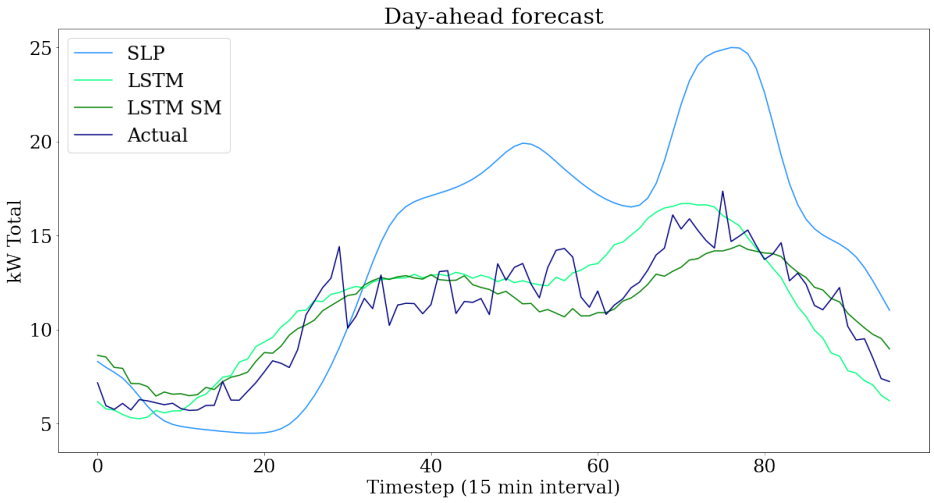


Figure 5.4.: Day-ahead forecasts for 2019-10-16

Before its integration into the LSTM model, the peak load and peak time forecasting performances of the XGBoost model are compared to a forecast based on historical values. The XGBoost model is compared with a day-ahead forecast based on the peak load and peak time of the same day in the week before. For the evaluation, a twelvefold cross-validation is conducted in the same way as described in the previous chapter. For the peak load forecast, the averaged XGBoost MAPE over the twelvefold cross-validation (M = 7.75, SD = 1.35) compared to the averaged MAPEs of the forecast through week-before peak loads (M = 9.65, SD = 1.45) demonstrated a significant improvement, t(20) = 3.2, p = .005. For the peak time forecast evaluation, the amount of correctly forecasted peak times is compared first. Again, comparing the averaged matches of the XGBoost peak time forecast (M = 2.82, SD = 1.54) with the matches of the forecast with week-before values (M = 1.55, SD = 1.37) yields a significant improvement, t(20) = -2.05, p = .05. As final metric, it is counted how often the forecasted peak time is amongst the five highest day-ahead load time steps. Comparing the averaged top five occurrences through the XGBoost model (M = 19.27, SD = 3.26) and the top five occurrences through the week-before forecast (M = 16.64, SD = 3.32), a slightly significant improvement can be observed once again, t(20) = 1.87, p=.07.

After evaluating the standalone performance of the XGBoost model, the forecast of the hybrid LSTM-XGBoost model is depicted in Figure 5.5. For this exemplary day it can be seen how incorporating the XGBoost-based peak load and peak time forecast can improve the overall forecast quality.

The results of the twelvefold cross-validation of the four scenarios are depicted in Table A.2 and Table 5.9. We can observe that, on average, the LSTM SM XGB outperforms all other models regarding overall MAPE. Compared with the LSTM SM, an average improvement of 0.14 percentage points is achieved. Within the test period between the 28.10.-26.11., the LSTM SM XGB model manages to improve the accuracy by 0.4 percentage points compared to the LSTM SM model. Another remarkable observation is that adding individual smart meter data as additional input data significantly improves the model. Compared to the simple LSTM model, LSTM SM reaches a MAPE of 16.95 compared to 21.64 without smart meter data, an improvement of 4.69 percentage points. Only in the second evaluated period did the LSTM model perform better than the LSTM SM. All models consistently

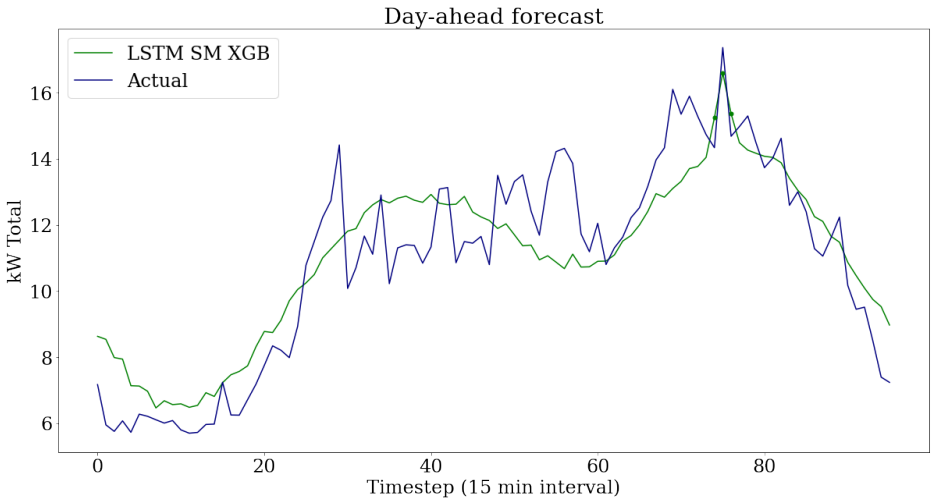


Figure 5.5.: Day-ahead forecast for 2019-10-16

Evaluated Period	MAPE (overall) [%] / RMSE (overall) SLP	LSTM	LSTM SM	LSTM SM XGB
01.01 - 30.01.	24.85 / 3.52	19.63 / 3.70	18.25 / 3.27	**17.98 / 3.23**
31.01. - 01.03.	25.64 / 3.87	**19.42 / 3.20**	25.35 / 3.87	25.04 / 3.83
02.03. - 31.03.	24.05 / 3.94	17.82 / 2.79	11.18 / 2.14	**11.04 / 2.11**
01.04. - 30.04.	24.95 / 3.97	25.63 / 3.00	19.40 / 2.83	**19.13 / 2.79**
01.05. - 30.05.	28.03 / 4.71	24.81 / 2.84	**23.57 / 2.79**	23.61 / 2.79
31.05. - 29.06.	27.53 / 4.78	28.50 / 3.33	10.71 / 1.60	**10.70 / 1.59**
30.06. - 29.07.	27.76 / 4.81	16.69 / 1.96	**11.00 / 1.41**	11.15 / 1.43
30.07. - 28.08.	27.01 / 4.87	23.68 / 2.60	20.41 / 2.35	**20.39 / 2.35**
29.08. - 27.09.	28.64 / 4.90	21.61 / 2.68	**17.36 / 2.06**	17.40 / 2.06
28.09. - 27.10.	27.19 / 4.46	16.89 / 2.19	11.29 / 1.73	**11.20 / 1.71**
28.10. - 26.11.	21.50 / 3.21	17.38 / 2.52	15.39 / 2.74	**14.99 / 2.67**
27.11. - 26.12.	26.20 / 3.71	27.63 / 4.13	19.52 / 3.36	**19.16 / 3.30**
Average	26.11 / 4.23	21.64 / 2.91	16.95 / 2.51	**16.81 / 2.49**

Table 5.8.: MAPE and RMSE results by periods, for overall forecasting accuracy

outperform the SLP. The outperformance of the LSTM SM XGB model is confirmed by the RMSE metric.

Furthermore, we evaluate the MAPE of forecasted peaks. Again, the LSTM SM XGB outperforms all other models. In comparison with the LSTM SM, an improvement of 3.55 percentage points is reached on average. In 9 out of 12 months, the LSTM SM XGB outperforms the other models in terms of overall MAPE. In 8 out of

Evaluated Period	MAPE (forecasted peak) [%] SLP	LSTM	LSTM SM	LSTM SM XGB
01.01 - 30.01.	33.71	22.86	14.64	**9.22**
31.01. - 01.03.	33.91	20.04	10.36	**10.2**
02.03. - 31.03.	38.99	14.75	17.02	**12.27**
01.04. - 30.04.	43.61	13.19	13.62	**11.45**
01.05. - 30.05.	68.92	25.7	26.99	**23.24**
31.05. - 29.06.	69.35	37.44	**9.61**	10.11
30.06. - 29.07.	71.47	19.48	**13.80**	19.04
30.07. - 28.08.	82.33	36.48	31.64	**27.42**
29.08. - 27.09.	72.80	36.32	22.98	**18.99**
28.09. - 27.10.	60.94	18.05	12.61	**9.72**
28.10. - 26.11.	33.02	12.97	20.54	**10.09**
27.11. - 26.12.	30.87	**11.44**	22.03	11.52
Average	53.33	22.39	17.99	**14.44**

Table 5.9.: MAPE results by periods and accuracy of forecasted peaks

12 periods, the LSTM SM XGB forecasted peak MAPE outperforms the other models. Once again, we can observe that adding smart meter data (LSTM SM) improves the forecast accuracy from a MAPE of 22.39 to a MAPE of 17.99, which reflects an improvement of 4.4 percentage points. Most notably is the improvement in peak forecast accuracy compared to the SLP, with an improvement of 38.89 percentage points between SLP and LSTM SM XGB.

As the addition of individual smart meter data significantly improved the overall community forecast performance, we are interested in finding out which features, and especially which households' smart meter data, is important to improve forecast quality. This information could be used to identify characteristics of households in which it is particularly helpful for forecasting tasks to install smart meters.

The Permutation importance (PIMP) for the LSTM SM are depicted in Figure 5.6. We can observe that the aggregated energy community load (sum) is by far the most important feature. Further important features are the sin and cos transformed hour and day, as well as the binary variable for weekends. Also, the loads of selected customers are important input features for the LSTM. While the feature importances of the households seem relatively low in comparison to the sum and the cyclical features, we know from the results in Table 8 and 9 that the addition of smart meter data leads to significant improvements and therefore even though seemingly small,

these feature importances should not be neglected. Most of the households with a high feature importance are also households with relatively high annual electricity consumption. For instance, household 147 is the fourth largest household amongst the 130 smart metered households with an annual electricity consumption of 1,700 kWh. Household 177 is the 8th largest household with an annual consumption of 1,448kWh, household 181 is 11th with 1,352kWh annual consumption. However, there are also several households with high feature importances that do not belong to the largest households.

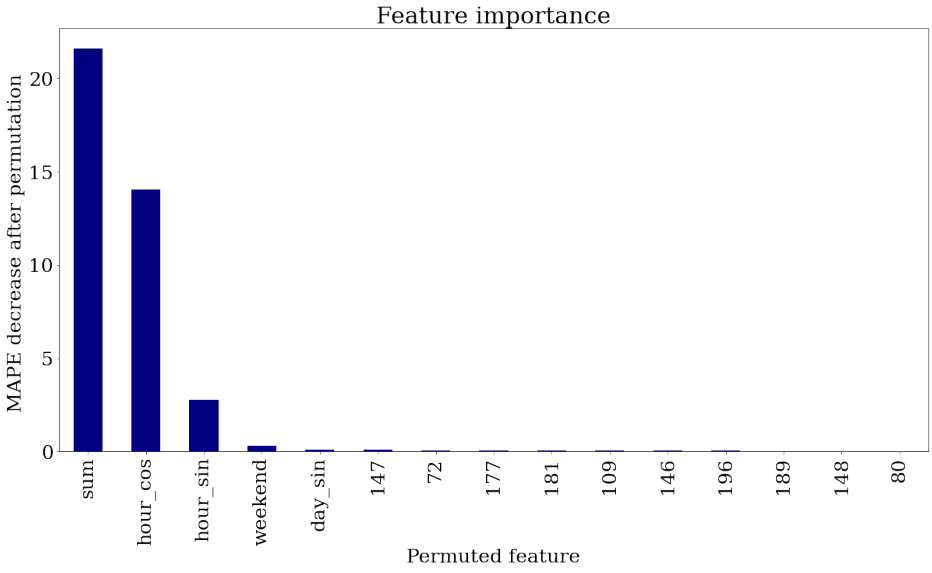


Figure 5.6.: Average feature importances for smart meter-based LSTM (LSTM SM)

5.6 Discussion

In this section, we discuss the results presented and their implications for day-ahead load forecasting in energy communities. The study was conducted with load data from a limited number of German households. Hence, it has to be investigated if the results of this study still prove valid in communities with a higher number of smart metered households, as well as data from other countries and differing community configurations. Also, we were not able to include weather data as an input feature due to the geographic distribution of the households from the underlying dataset. This leaves opportunities for further research. In the following, we discuss two aspects of our work in particular.

First, we observe that the addition of smart meter data in energy communities can significantly improve the day-ahead forecasting accuracy of energy communities in our case study. This confirms the results of Zufferey et al. (2016), where also a higher accuracy in aggregated load prediction was reached by increasing the number of smart meters. Hence, we suggest to consider the installation and implementation of smart meters in the planning process for energy communities. Our results indicate that selected households contribute more to improving forecasting quality than others. For instance, households with a larger annual consumption seem to have a larger impact on the forecast than smaller households. Still, this does not hold true for all households with a high feature importance. Thus, further research has to focus on identifying characteristics of households that improve the forecasting quality. With this information, grid operators and energy community managers could selectively install smart meters to optimize their day-ahead forecasting model.

Our feature importance analysis showed that the most important factor for forecasting day-ahead loads of energy communities is the past aggregated energy community load itself. It has to be noted that engineered type-of-day features, such as the *sin* and *cos* transformation of the hour, are by far the second most important input features. Hence, we strongly propose that coming works in the field of load forecasting also include *sin* and *cos* transformed type-of-day features.

Second, we introduce a novel hybrid LSTM-XGBoost model that enables improved peak load forecasts by separately forecasting the general load pattern and peak loads. To our knowledge, we are the first ones to propose peak load time and quantity forecasting through a dedicated XGBoost model and to combine an LSTM and XGB forecast into a holistic model. By using the hybrid LSTM-XGBoost model, we can improve the overall model performance and peak forecasting performance in our study. In addition, we propose that further research also evaluates the performance of a hybrid peak load forecasting XGBoost model in combination with other recent proposed algorithms like temporal attention-based convolutional networks Tang et al. (2022) or federated learning Fekri et al. (2022).

5.7 Conclusion

In this paper, we propose a framework for smart meter-based day-ahead forecasting in energy communities with bi-directional LSTM neural networks and a combined

LSTM-XGBoost model. Furthermore, we contribute to the general understanding of important input features in smart meter-based energy community load forecasting. We can draw three main conclusions.

First, our results confirm that the LSTM-based models achieve significantly higher accuracy than forecasting based on standard load profiles. In addition, using smart meter data as additional input further improves the accuracy of forecasting in our case study.

Second, the novel hybrid LSTM-XGBoost manages to further increase the forecasting accuracy of smart meter-based models, especially in terms of peak load forecasting.

Third, the most important features for the forecast of the aggregated energy community load are, in our case study, the past aggregated load itself, transformed hour and day data, a binary weekend variable, and past loads of selected households. We see a tendency that the past loads of households with higher annual consumption may be more important features, but this needs to be confirmed and further investigated in future research.

This paper gives scope for further research in the field of energy community load forecasting. Future work should further confirm and deepen the assessment of the hybrid LSTM-XGBoost model and its viability in cases without smart meter data or in combination with alternative forecasting algorithms. Furthermore, adding weather data to the forecasting process could be an interesting addition to this study.

CHAPTER 6

PRIVACY-PRESERVING PEAK TIME FORECASTING

This chapter addresses the challenge of forecasting peak times for aggregated electricity loads from a privacy-preserving perspective. Predicting peak load times is critical for the operational efficiency of aggregators and grid operators. To tackle this, the concept of Learning-to-Rank is introduced and applied to the task of peak time forecasting. The results demonstrate that this approach delivers comparable accuracy in predicting peak times while relying solely on the ranks of load data rather than precise measurements. This method directly addresses data privacy concerns, offering households the potential to share ranked load information with third parties instead of disclosing their exact load measurements, thereby enhancing data privacy.

This chapter comprises the following article: L. Semmelmann, O. Resch, S. Henni and C. Weinhardt. *Privacy-preserving peak time forecasting with Learning to Rank XGBoost and extensive feature engineering*, IET Smart Grid, 2024.

6.1 Introduction

Ensuring a balance between power supply and demand is essential for electrical grids' stable and efficient operation. In this context, forecasting future electrical loads plays an integral role (Bunn and Farmer, 1985). Load forecasting is performed for various planning horizons, from long-term over medium-term to short-term load forecasting with annual, monthly, or daily planning horizons, respectively (Srinivasan and Lee, 1995). It can additionally be classified by the aggregation level considered (Groß et al., 2021).

One essential discipline of load forecasting is peak load forecasting. Peaks are

the occurrence of the maximum load in a specific timeframe (e.g., a day) and can be characterized by two dimensions: peak time and peak load quantity. Peak time describes the timestep where the maximum load occurs, while peak load quantity describes the maximum load measured in the respective timestep (Dai et al., 2021). As peaks constitute the maximum strain on the grid, predicting the maximum load and especially the timing of peaks is crucial for grid stability. Yet, most approaches in the discipline of electrical peak demand forecasting are either concerned with only predicting the peak load quantity (Lee and Cho, 2022) or the peak time forecast is inferred from an overall load forecast by considering the time when the predicted load curve is at its maximum (Goia et al., 2010; Haida and Muto, 1994). As a high general prediction accuracy does not necessarily imply a good prediction quality concerning peak loads and peak times (Semmelmann et al., 2022), further research specifically about peak times is necessary.

Another increasingly important aspect of load forecasting is the privacy preservation of load data. By sharing exact load data, new potential data security vulnerabilities are created. For instance, in case of a data leak, the load data of an industrial firm could allow conclusions about the current economic situation of the respective firm. Substantial industrial customers see data security as a significant barrier to participation in load shifting programs - and thereby a barrier to sharing load data (Olsthoorn et al., 2015). Hence, to motivate industrial customers to share their load data, participate in grid provider programs, and reduce potential vulnerabilities, it is essential to work on privacy-preserving methods to work with load data.

This paper combines a focus on peak time forecasting with a privacy-preserving Learning to Rank model in the context of the BigDEAL Challenge 2022. In this international competition organized by Dr. Tao Hong, Duke Energy Distinguished Professor at UNC Charlotte and Director of the Big Data Energy Analytics Laboratory (BigDEAL), 78 international teams competed along different tracks related to peak load forecasting. The challenge consisted of a qualification and a final round. In the final round, the three tracks of the challenge were targeted at forecasting peak load quantities, peak times, and the shape of the load in a five-hour timeframe around the peaks. The findings presented in this paper stem from the approach followed by Team SGEM KIT in the final round of the BigDEAL Challenge for the track peak time forecasting.

We propose a novel approach to forecasting peak times with a Learning to Rank extreme gradient boosting (XGBoost) model, also used for the peak time forecasting track of the BigDEAL Challenge. We compare our results with a naive day-before benchmark forecast and a general state-of-the-art XGBoost-based load forecasting model that delivers load forecasts for every time step. The latter model achieved the fourth rank in the peak load forecasting track of the final round of the BigDEAL challenge and can hence be seen as a relevant benchmark. The Learning to Rank model only requires ranks of loads, instead of their actual magnitude, as input. Thus, it can be used in application areas where protecting actual load data is highly relevant, thereby, for instance, potentially encouraging customers to share their data with grid operators.

This paper shows that the Learning to Rank XGBoost model yields comparable forecasting accuracy as a well-performing baseline XGBoost model. Furthermore, we conduct extensive feature engineering. We investigate which features are most important for the general XGBoost and Learning to Rank models. Additionally, we use Bayesian Hyperparameter Optimization to find optimal hyperparameter combinations.

In conclusion, we aim to make the following contributions:

- An extensive feature engineering process is described, including the implementation of rolling averages and type-of-day features, thereby significantly improving the peak time forecasting accuracy.
- The Learning to Rank XGBoost algorithm is used for peak time forecasting, working only with ranks of loads instead of absolute loads as target feature, thereby offering potential privacy-preserving properties. The model is compared to a conventional XGBoost load forecasting model, from which peak time forecasts are inferred.
- The XGBoost-based models are optimized with a state-of-the-art Bayesian hyperparameter optimization, enabling further increases in prediction accuracy.

The remainder of this paper is structured as follows. In the second chapter following the introduction, we set our study in the context of related work. In the third chapter, we describe our general methodology. We depict our feature engineering

process, the utilized XGBoost models as well as the regarded metrics. In the fourth chapter, we present the BigDEAL case study and the underlying data. Subsequently, we describe the case study results and the achieved forecasting accuracy according to the previously introduced metrics in chapters five and six. Finally, in chapter seven, we discuss our results and give an outlook to further research questions.

6.2 Related Work

In this section, we give an overview of related peak load and peak time forecasting studies, with a special focus on works that cover privacy-preserving features.

Most load forecasting-related studies focus on an overall load forecast, most often through neural network-based methods, Support Vector Machines (SVM), or Auto-Regressive Integrated Moving Average (ARIMA) (Kong et al., 2017; Nti et al., 2020; Tang et al., 2022; Wang et al., 2022b). The first advances in load forecasting were, amongst others, made with ARIMA-based models. In Juberias et al. (1999), an hourly short-term electrical ARIMA load forecasting model is introduced. Lee et al. improve the ARIMA model by using a lifting scheme wavelet transformation to enhance the forecasting accuracy in Lee and Ko (2011). Another way of enhancing the ARIMA load forecasting is suggested by Nie et al. (2012), employing a hybrid ARIMA and SVM model, where the ARIMA model forecasts the linear basic load component and the SVM is used for non-linear components.

In contrast, recent load forecasting studies mostly focus on neural network-based methods. In Kong et al. (2017), the authors use Long Short-Term Memory (LSTM) recurrent neural networks to forecast single-residential household loads. Also, Lin et al. (2022) use a dual-stage model, based on an LSTM and an attention-based encoder, for a probabilistic load forecasting model. The temporal attention mechanism, in combination with a Convolutional Neural Network, is employed by Tang et al. (2022) as well. Another type of neural network is suggested by Wang et al. (2022b), who use a transformer model based on an encoder-decoder architecture for a multi-energy load forecasting problem.

For the planning and operation of modern power systems and distribution grids, especially the forecasting and subsequent reduction of peak loads are essential (Kucevic et al., 2021a,b). Besides the fact that the previously mentioned studies are focused on an overall load forecast, several studies emphasize the tendency of neural

network-based approaches to underestimate peak loads (Syed et al., 2021; Semmelmann et al., 2022; Amjady, 2001). Hence, Syed et al. (2021) adapt the LSTM cost function to penalize underestimation of the load. The authors of Semmelmann et al. (2022) pursue another approach by combining LSTMs with a dedicated peak time and peak load XGBoost forecast. Thereby, the overall load and peak load forecasting accuracy are improved. Also Dai et al. (2022) develop a hybrid LSTM-XGBoost load forecasting model. The authors state that the XGBoost forecast could be further improved by employing a Bayesian hyperparameter optimization method to find more suitable parameters. Our study implements the suggested Bayesian hyperparameter search approach for improving forecasting accuracy.

Haida and Muto (1994) were amongst the first to focus on peak load forecasting specifically. The authors combine a transformation technique to consider seasonal load changes and annual load growth with a multivariate regression analysis. Thereby, the authors reduce forecasting errors in transitional seasons such as spring and fall. In Amjady (2001), the importance of peak load forecasting for dispatching centers in power networks is highlighted. The authors focus on peak load forecasts with a dedicated ARIMA model alongside the overall hourly load forecast. In Saini (2008), the peak loads for up to 7 days ahead were forecasted with a feed-forward neural network, combined with a Principal Component Analysis for factor extraction. Besides peak loads itself, also forecasting the time of peak loads plays an essential role. The authors of Haq and Ni (2019) show that significant improvements in overall load forecasting accuracy can be reached by focusing on peak time forecasting. Within the study, the load demand time series is decomposed into low-frequency components. Then, a peak load binary variable is derived from the value at risk concept, to improve forecasting accuracy during peak times. Finally, a deep belief network is trained to forecast future loads. We remark that there are - to the extent of our knowledge - no studies solely focusing on predicting peak times.

Another research stream in smart grid research deals with implementing privacy-preserving methods. In Chin et al. (2019), the authors discuss two potential privacy protection schemes for short-term load forecasting: model-distribution predictive control (MDPC) and load-level, in combination with support vector regressions. The study concludes that the MDPC has a slight negative impact on forecasting accuracy for smaller aggregations of loads, which diminishes with higher aggregation

levels. On the other hand, the load-leveling approach improves load forecasting accuracy. Another privacy-preserving load forecasting approach is presented in Hou et al. (2020), where load forecasting models of residential customers are trained on distributed smart meters and handled locally. Only the forecasting outputs are reported to the cloud through fog nodes. The authors of Dong and Liu (2022) are suggesting a privacy-preserving model for electricity theft detection by adding Gaussian noise to the consumption data of customers before applying a Convolutional Neural Network, aiming to achieve a balance of customer privacy and model accuracy.

A further important concept in privacy-preserving forecasting research is differential privacy (Dwork, 2006). In the context of differential privacy, noise is added to the input data until it can no longer be used to confidently predict which individual delivered the underlying data. In Li et al. (2020b), differential privacy is granted by adaptively controlling the gradients of training data, combined with a framework to allocate privacy budgets. Le et al. introduce a novel, privacy-preserving adaption of the XGBoost framework for federated learning (Le et al., 2021). The authors use a secure matrix multiplication method and a noise perturbation approach in a separate model. In a comparable approach, in Fernández et al. (2022), a combination of federated learning and differential privacy is utilized for short-term load forecasting. One outcome of the study is that increasing the number of participating consumers leads to enhanced forecasting results and potentially too high computational costs, especially for complex neural network architectures. Also, in Venkataramanan et al. (2022), a federated learning model for privacy-preserving forecasting of distributed energy resources, such as solar PV, EV storage or flexible loads, is developed. The authors validate their study with 1,000 IoT nodes and show that the approach can be used for grid services like predicting curtailment events or load swings. A recent advance of federated learning models for privacy-preserving forecasting has been made in He et al. (2023), where a hierarchically federated model exploits all underlying datasets while enabling information exchange of users with similar load patterns. The state-of-the-art model enables a significant improvement in forecasting accuracy over benchmark models while maintaining a high fault tolerance. For a better balance between privacy and data quality, in Tran et al. (2022) a two-step model is suggested. In the first step, a distributed perturbation method is applied on the underlying high-frequency load data. In the second step, through a private noise dis-

tribution protocol, noise elements are distributed over the smart meters of individual customers. In a case study, the authors show the utility of the data is maintained while preserving the privacy of users. Also the authors of Eibl et al. (2018) investigate the impact of differential-privacy on forecasting quality, underlining that for some methods the introduction of differential-privacy leads to significantly worse forecasts.

We can observe that many past privacy-preserving methods focus on adding noise to the underlying data, e.g., the hourly loads, sometimes at the cost of worse forecasting accuracy. Another possible approach to change the underlying data could be the transformation of actual loads to less sensitive ranks of loads, which can then be used to forecast peak times, e.g., through the Learning to Rank method introduced by Chapelle and Chang (2011). The Learning to Rank method has been applied with the XGBoost method (Vidović et al., 2021), but not in the context of peak time forecasting. The transformation of loads to ranks could yield one big advantage over more sophisticated methods: it is likely to be easier for end-users to implement and comprehend, which, as various studies have shown, is essential for the adoption of novel technologies and smart grid applications (Syed et al., 2020; Rajapaksha and Bergmeir, 2022; Bharadiya, 2023; Herm et al., 2021).

Our study fills two essential gaps in research. In the context of the BigDEAL Challenge, we primarily focus on the peak time forecasting task, which has only been discussed to a small extent in prior research. Furthermore, we provide a detailed examination of suitable features for the peak time forecasting problem. Second, we employ the Learning-to-Rank XGBoost algorithm to forecast peak times. We are analyzing its performance compared to benchmark approaches, which use actual loads instead of ranks of loads. Thereby, we make an essential contribution to privacy-preserving peak time forecasting.

6.3 Methodology

This section aims to provide a comprehensive overview of the methodology used in the BigDEAL Challenge to forecast peak load times and quantities. First, we give an overview of our general forecasting framework. The following subsections describe the respective steps, beginning with our feature engineering approach. Then, we describe the models we utilized and our Bayesian Hyperparameter Optimization.

Finally, we introduce the metrics used to evaluate the performance of the previously engineered models.

6.3.1 Forecasting framework

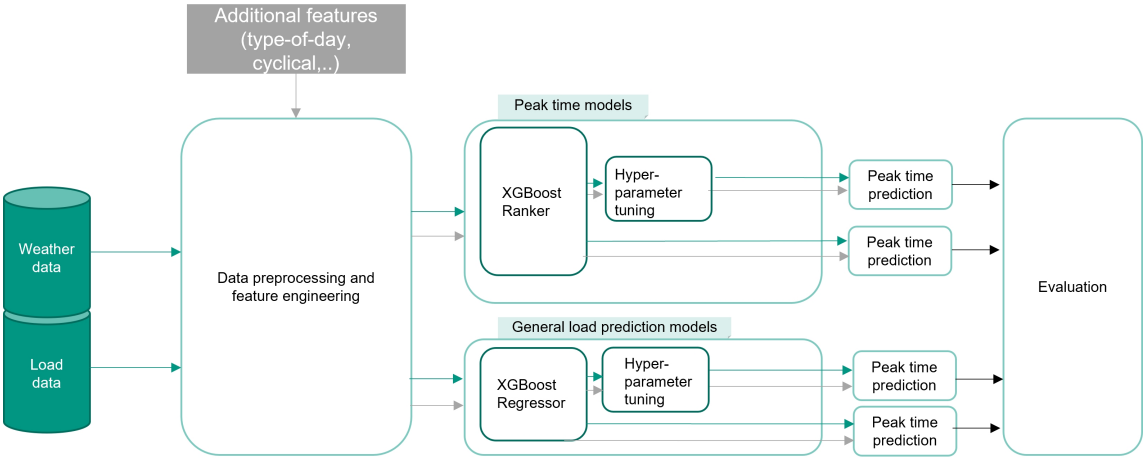


Figure 6.1.: Graphical overview of the research approach

The final stage of the BigDEAL challenge consisted of several rounds, each including respective historical load and temperature observations. For the to-be-forecasted time horizon, only temperature data was given. In the first step, we enrich the given dataset through extension with additional features, which are explained in detail in the following section. The colour of the arrows indicates which data is used as input for the respective steps. All the next steps use fully feature-engineered data. We distinguish between two model types: a dedicated peak time model and a general load forecasting model. The latter serves as a solid, well-performing benchmark and reflects the predominant approach in the existing literature to infer the peak load and peak time forecast from the whole daily forecast (Dai et al., 2021; Goia et al., 2010; Haida and Muto, 1994). Both model types are described in Section 6.3.3. For the peak time models, we differentiate between an XGBoost model with standard parameters and a model with tuned hyperparameters. The overall hyperparameter tuning approach is introduced in Section 6.3.4. The forecasts are then evaluated according to the metrics defined in Section 6.3.5.

Overall, the utilized forecasting framework can be structured as depicted in Figure 6.1.

6.3.2 Feature Engineering

Feature engineering describes the process of creating representations of the raw data that can improve the models' effectiveness. For a high prediction quality, adequate feature engineering is essential, with the effect of feature selection surpassing that of selecting different models in many cases (Kuhn and Johnson, 2019). Below, the different feature engineering techniques used to transform the input data are described:

Type-Of-Day Features. Type-of-day features are variables that are created by categorizing dates, for instance, in groups of working days and non-working days. Past studies have shown that type-of-day features can improve the overall forecasting accuracy when added to the feature set Kanda and Veguillas (2019). Hence, we added binary variables for determining whether the day is a weekday, holiday, preceded, or followed by a holiday, respectively. In addition, the weekday, as well as month and day of the month, are provided to the model as input after a sine and cosine transformation.

Sine and Cosine Transformation of Cyclical Features. Past studies, such as Semmelmann et al. (2022); Gürses-Tran et al. (2022) have shown that the sine and cosine transformation of cyclical features, such as the hour or weekday, result in high feature importance and are therefore essential features for electrical load forecasting. The advantage of the sine and cosine transformation lies in a better representation of the cyclical variables, for example, allowing the model to learn that 11 pm is closer to 2 am than 8 am. Hence, we implement sine and cosine features for hour, weekday, and month. For the hour, we additionally implement a 2x and 4x sine and cosine, which repeats two and four times, respectively, per day. We apply the sine and cosine transformation by calculating the number of past time steps since the beginning of the respective seasonal period, e.g., a day or month, scaled to a range between 0 and 2π.

Relative changes of features. For every continuous feature, relative changes up to the past 24 hours in one-hour increments are calculated and used as additional input features. This is based on the hypothesis that the rate of change in those features might affect the resulting load pattern, especially if those changes happen suddenly. The formula for the calculation of relative change $G_{\tau,T}$ of feature τ and the time frame tf in every time step t is calculated in Equation 6.1:

$$G_{\tau,tf} = \frac{\tau_t - \tau_{t-tf}}{\tau_{t-tf}} \tag{6.1}$$

Rolling Averages. Since the selected model predicts the respective timesteps independently of each other, one potential drawback could be its difficulty in considering dependencies over multiple timesteps. Furthermore, as described in Mayrink and Hippert (2016), it is also essential to include lagged temperature features in the load forecasting model due to the thermal inertia of buildings. Therefore, rolling averages are calculated for all temperature features over different time frames ranging from one hour to 192 hours. For up to ten hours, this is done for every interval length. Above, only the rolling average for 12, 15, 18, 24, 36, 48, 96, and 192, respectively, are calculated. The calculation of the rolling average $RA_{\tau,tf,t}$ for temperature feature τ and time frame tf at time step t is described in Equation 6.2.

$$RA_{\tau,tf,t} = \frac{1}{tf} \sum_{n=1}^{tf} \tau_{t-n} \tag{6.2}$$

Rolling average features have shown a high feature importance in previous research where ensemble-based prediction models were used on time series data, for example, in Natras et al. (2022). We, therefore, calculate a "DiffToRollingAverage" (dRA) data point for every rolling average to detect deviations of the underlying temperature feature from the rolling average, as described in Formula 6.3:

$$dRA_{\tau,tf,t} = RA_{\tau,tf,t} - \tau_t \tag{6.3}$$

6.3.3 Prediction Models

As mentioned in section 6.3.1, two different XGBoost-based models are trained to predict the load in every time step of the test set. The first model is a general load prediction model that forecasts the load of every hour in the respective period. From the forecasted load patterns of the general model, the peak times are inferred to serve as a benchmark against which the second model can compete. The second model is a dedicated peak prediction model, which is based on a novel approach to forecast peak times by employing a Learning to Rank XGBoost model. First, the XGBoost algorithm is presented in general since both used models are based thereon. The

following two sections describe both peak load forecasting models in greater detail.

Extreme Gradient Boosting (XGBoost)

XGBoost was introduced by Chen and Guestrin (2016) and has been proven as a highly efficient and accurate model for regression and classification tasks. In the load forecasting context, approaches based on XGBoost have often outperformed other models, as shown in Liao et al. (2019); Abbasi et al. (2019). The model is based on an ensemble of classification and regression tree weak learners. Furthermore, the quadratic objective function is simplified through a second-order Taylor expansion, which yields enhanced runtimes and limits overfitting.

General XGBoost Load Prediction Model The general load prediction model consists of an XGBoostRegressor that predicts the load for each timestep of the test set individually. Then, for every day d in the test set, the timestep t of the highest load P_{max} is taken as peak time prediction $t_{d,P_{max}}$, as depicted in Equation 6.4

$$t_{d,P_{max}} = max(P_{d,t}, ..., P_{d,T}) \tag{6.4}$$

Hereafter, this model is referred to as XGBP (XGBoost Pattern). We are also considering a hyperparameter-optimized version, which is called XGBPH, in the following. The XGBPH model also served as a model for the peak load and shape prediction tracks of the BigDEAL challenge.

Learning to Rank XGBoost Peak Time Model The idea of the proposed Learning to Rank XGBoost peak prediction model is the essential characteristic of a peak, which is that it is the highest load in the considered timeframe, such as a day. In other words, if the loads of a day were ranked by descending load, the peak would always have rank one. Thus, we propose a model that learns to rank the timesteps of a day. Since the Learning to Rank model works with day-wise ranks as a target variable instead of loads, it requires less sensitive data than traditional approaches. The initial idea of the Learning to Rank model was first described by Chapelle and Chang (2011).

The Learning to Rank model requires a transformation of the target variable from load to rank. Every load P on a day d can be mapped to a rank r, which ranges from 1 to 24:

$$\left\{\left(\mathbf{P}_d^t, r_d^t\right)\right\} \tag{6.5}$$

A rank r of 1 represents the peak load, while rank 24 represents the lowest load occurring on a certain day d.

In Figure 6.2, the load is transformed into day-wise rankings of the respective timesteps by descending load for an exemplary time series consisting of two days with three timesteps per day.

Timestamp	Load (kW)
2018-01-01 12:00am	12
2018-01-01 08:00am	50
2018-01-01 04:00pm	10
2018-01-02 12:00am	20
2018-01-02 08:00am	40
2018-01-02 04:00pm	25

Timestamp	Rank
2018-01-01 12:00am	2
2018-01-01 08:00am	1
2018-01-01 04:00pm	3
2018-01-02 12:00am	3
2018-01-02 08:00am	1
2018-01-02 04:00pm	2

Figure 6.2.: Visualization of the target transformation

The previous target variable, "Load," is discarded and not used by the model. The peak time model thus learns with a different target variable than the general load prediction model. The input features remain unchanged. For the prediction of ranks, for every possible rank R a score is calculated based on comparisons of the timesteps in each day. For a detailed description of the score calculation methodology, we refer to Chapelle and Chang (2011). The scores can, in turn, be sorted and turned into rankings for each day.

The prediction of our proposed models is thus day-wise rankings for all timesteps in the test set. In the final transformation step, the timestamp associated with rank one is selected as the peak time for each day. As stated in the overview, we do not only consider one single Learning to Rank peak time model but different variations of it. This leads to two dedicated peak time models that are investigated: XGBR (XGBoost Ranker), a plain XGBoost Ranker without hyperparameter tuning, and XGBRH (Learning to Rank XGBoost hyperparameter tuned), where hyperparameter tuning using the methodology defined in the next section is applied to improve the model.

6.3.4 Bayesian Hyperparameter Tuning

Hyperparameter tuning is a crucial task in machine learning and describes the practice of optimizing the parameters of the selected model in order to obtain a higher prediction quality. It has been shown that in several cases, baseline models could better be improved by hyperparameter adjustments of existing models than by inventing new models (Bardenet et al., 2013). There are several approaches to hyperparameter tuning, including using a Bayesian optimization, as proposed in Snoek et al. (2012), which belongs to the class of automated hyperparameter tuning. In automated hyperparameter tuning, the model is considered to be a black box function that, given validation data, returns a score, which generally is the chosen error metric to be optimized. The goal of the optimization is to find hyperparameters that minimize this error. Contrary to the Bayesian optimization method, traditional optimization approaches are not suited for this kind of optimization problem. Bayesian optimization leverages Bayes theorem for selecting parameters to be evaluated in the true objective function by using a probability model of the objective function, which is, in turn, based on sample data from previous iterations. For a more comprehensive introduction to the general principles of Bayesian optimization for hyperparameter tuning, we refer to Wu et al. (2019).

Bayesian optimization leads to a significant improvement of hyperparameters with only a few iterations (Wu et al., 2019), making it both effective and time efficient. Time efficiency played a particular role in selecting this approach as the time for obtaining a prediction and, thus, hyperparameter tuning was limited in the competition the approach was developed for. Furthermore, several studies in the load forecasting field have shown that more accurate predictions can be reached by using Bayesian optimized model parameters Trierweiler Ribeiro et al. (2020); Jin et al. (2021); Munem et al. (2020).

Our implementation of the Bayesian optimization hyperparameter selection model is depicted in Figure 6.3. The approach is based on two major parts: an optimizer and a hyperparameter evaluation function that is minimized throughout the iterations.

The optimizer describes the Bayesian optimization model with its parameters and the search space. For the optimization model itself, we use the Bayesian optimization

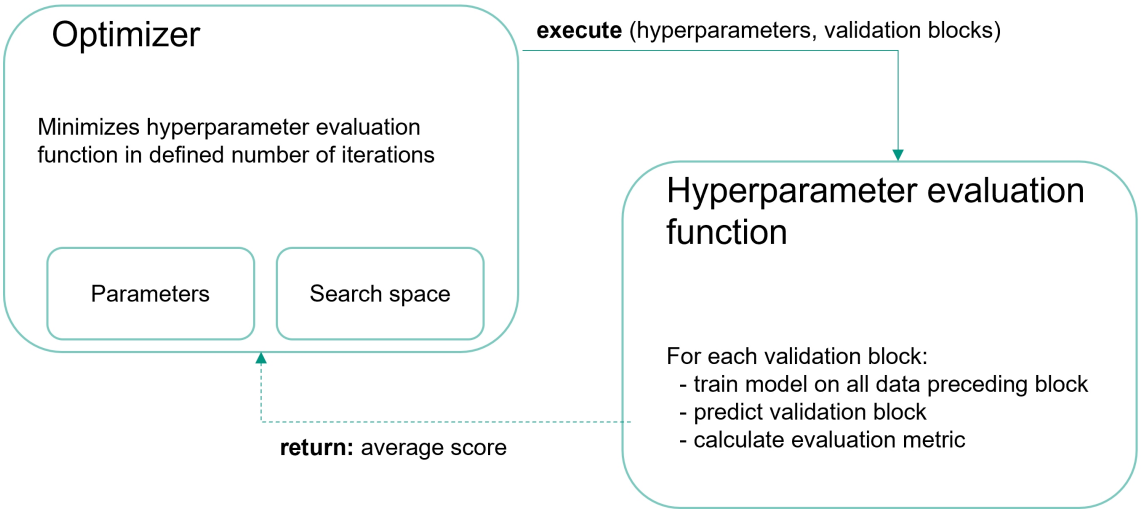


Figure 6.3.: Structure of hyperparameter optimization

package PyPi (2022). Table 6.1 depicts the chosen parameters of the optimization model. The model is initialized with ten random points and runs for 100 iterations. The parameters *alpha*, which is used for the internal Gaussian process, and *kappa*, which controls the relation between exploitation and exploration, are set as specified in the table. For reproducibility, a random seed is used.

Parameter	Value
init_points	5
niter	100
gp_alpha	10^{-10}
kappa	1.5
seed	112

Table 6.1.: Bayesian optimization parameters

In addition to selecting a model, a search space needs to be set. It defines the values the parameters selected by the optimization model can take. As all of our models are XGBoost-based, they have a shared search space. It is presented in Table 6.2, where for each hyperparameter that is optimized, the upper and lower boundaries are defined. The parameter *max_depth* is discreet and thus rounded before use. We derive the parameters from past XGBoost-based load forecasting studies (Semmelmann et al., 2022; Wang et al., 2021; Massaoudi et al., 2021).

As depicted in Figure 6.3, in each iteration, the optimizer calls the hyperparameter evaluation function with a set of hyperparameters and validation blocks. The

Hyperparameter	Lower bound	Upper bound
max_depth	3	10
learning_rate	0.01	1.0
subsample	0.5	1.0
min_child_weight	0.5	5.0
colsample_bytree	0.5	1.0

Table 6.2.: Hyperparameter search space

hyperparameters are selected by the optimization function and thus differ in each iteration. The validation blocks constitute the time frames used by the hyperparameter evaluation function to calculate the model's evaluation metric. We use the whole year 2017 as a validation block to get parameters for an overall robust model.

In the second part, the hyperparameter evaluation function is used to determine the performance of the model for the hyperparameters of the current iteration. For each validation block, the respective model is trained on all available data with time stamps preceding the validation block. Subsequently, the model's performance is evaluated on the validation block. This is performed for each validation block individually, leading to three-fold cross-validation. The function returns the average score over the three blocks. The evaluation metric employed varies depending on which model type is evaluated. For general prediction models, the mean absolute percentage error for the true and predicted loads for each time step is used. For the dedicated peak time prediction models, the score is based on a daily MAE-based metric, defined in section 6.3.5. In the last step, the hyperparameters are selected based on the best scores. For every different Local Distribution Company (LDC), a dedicated hyperparameter tuning run has been conducted.

6.3.5 Metrics

In the following, we present the used metrics for the evaluation of our peak time forecasting models.

Accuracy. In many studies, e.g., J. Liu and L. E. Brown (2019), peak time forecasting is considered to be a classification task. The reasoning for this is that a peak time forecast is only of use if it predicts the exact time of the peak load event. Thus, accuracy, which is in many cases used to evaluate binary classifications, is a popular metric to evaluate peak time forecasts. The accuracy metric $\mathbb{P}$ is defined as

in Equation 6.6:

$$\mathbb{P}(\text{Actual} = \text{Predicted}) \;=\; \frac{Tp + Tn}{N} \tag{6.6}$$

with Tp and Tn being the amount of correctly predicted positive and negative labels, respectively. N constitutes the total amount of predictions. Hence, accuracy measures the share of correct predictions.

Mean Absolute Error (MAE). As a second error metric, we calculate the Mean Absolute Error (MAE), which punishes wrong predictions linearly to the distance to the true prediction of the respective day. It is calculated as the mean of day-wise the absolute deviations of the predicted from the true peak time, as depicted in Equation 6.7:

$$MAE = \frac{1}{D}\sum_{i=1}^{D} |t_{d,P_{max},pred} - t_{d,P_{max}}| \tag{6.7}$$

With D being the considered amount of days and $t_{d,P_{max},pred}$ and $t_{d,P_{max}}$ being, respectively, the predicted and the actual peak time of day i.

BigDEAL Peaktime Metric (BDPM). We also evaluate a dedicated metric, which was introduced in the context of the BigDEAL challenge, called BigDEAL Peaktime Metric (BDPM). The BPDM is a modified version of a cumulative absolute error that punishes higher deviations more strongly by introducing a punishment factor. At the same time, it is capped for deviations greater than five hours. In its original form, the error is cumulated over the whole considered timeframe. To ensure comparability over test blocks with differing lengths, we decided to norm this error by the number of days considered. Our normed version of the BDPM is defined in Equation 6.8 and 6.9:

$$BDPM = \frac{1}{D}\sum_{i=1}^{D} f(|t_{d,P_{max},pred} - t_{d,P_{max}}|) \tag{6.8}$$

with:

$$f(r) = \begin{cases} r & \text{for } r \leq 1 \\ 2 * r & \text{for } 2 \leq r \leq 4 \\ 10 & \text{for } r \geq 5 \end{cases} \tag{6.9}$$

6.4 Case Study

The previously described methodology was applied on the data set provided by the initiators of the BigDEAL Challenge 2022. The data set comprises historical load data of three U.S. neighboring local distribution companies (LDCs), along with temperature data from six weather stations in the same region. Initially, data from 2015 to 2017 was provided in hourly resolution. Then, during the final round, data for 2018 was provided in six subsequent iterations, which serve as test blocks for the model evaluation and as the basis for the hyperparameter optimization.

6.4.1 Exploratory Data Analysis

The given data set is structured as an hourly time series. Each row has a timestamp containing the year, date and hour. The feature variables include the weather data columns T1 up to T6 and the timestamps. When analyzing the weather data, it becomes visible that the pairwise correlation between the columns is extremely high and in no case smaller than 0.95, which fits the assumption that the weather stations are located close together.

For the load of the LDCs, three different target variables, LDC1, LDC2 and LDC3, are given. No unit of measurement is provided for the target variables. The respective loads of the LDCs are forecasted separately. In Figure 6.4, the distribution of the values of the target variables is depicted. We can observe that the loads of the different LDCs vary significantly in magnitude, with LDC3 having by far the largest loads. Despite these differences in magnitudes, the LDC columns are highly correlated. The lowest pairwise correlation observed is 0.91. As there still are variations between the different LDC load profiles, predicting and evaluating multiple LDCs can be considered some form of additional cross-validation of the model.

Model	Feature Engineering	Regression-based	Rank-based	Hyperparameter
Baseline (Day-before peak time)	No	No	No	No
XGBP*	No	Yes	No	No
XGBP	Yes	Yes	No	No
XGBPH	Yes	Yes	No	Yes
XGBR	Yes	No	Yes	No
XGBRH	Yes	No	Yes	Yes

Table 6.3.: Model variations investigated in this study

Apart from the timestamps, weather features are the only feature variables that are initially provided. Figure 6.5 depicts the relationship between the observed average LDC loads and the average temperature features. The relationship between temperature and load appears to be nonlinear, with both high and low temperatures being associated with a high load. This indicates the use of electricity for both heating when cold temperatures occur and cooling when the temperature is high. We can also underline that observation from a statistical point of view: the overall Pearson correlation between the average temperature measurements and average LDC load measurements is quite low at 0.063. However, when we only regard all observations during temperature measurements below 60, the correlation is strongly negative at -0.87: the lower the temperature, the higher the loads. When we only regard the remaining observations at temperature measurements above 60, the correlation amounts to 0.87.

On the described data set, feature engineering following the methodology introduced in Section 6.3.2 was performed. This increased the number of input features to over 350. 16 of those features are related to the timestamp, i.e., cyclical features, while the remaining features constitute various transformations of the respective temperature features.

6.4.2 Train-Test Splits

The train-test splits in this work are based on the iterations of the final round of the BigDEAL Challenge 2022. As training data, all observations preceding the first date of the respective test set are used. Notably, the time frames of the test sets vary in length, requiring forecasts ranging up to three months ahead. We consider 6 different test sets, four of which have a length of two months, while the others have a length of one month and three months, respectively. All test sets are in the year

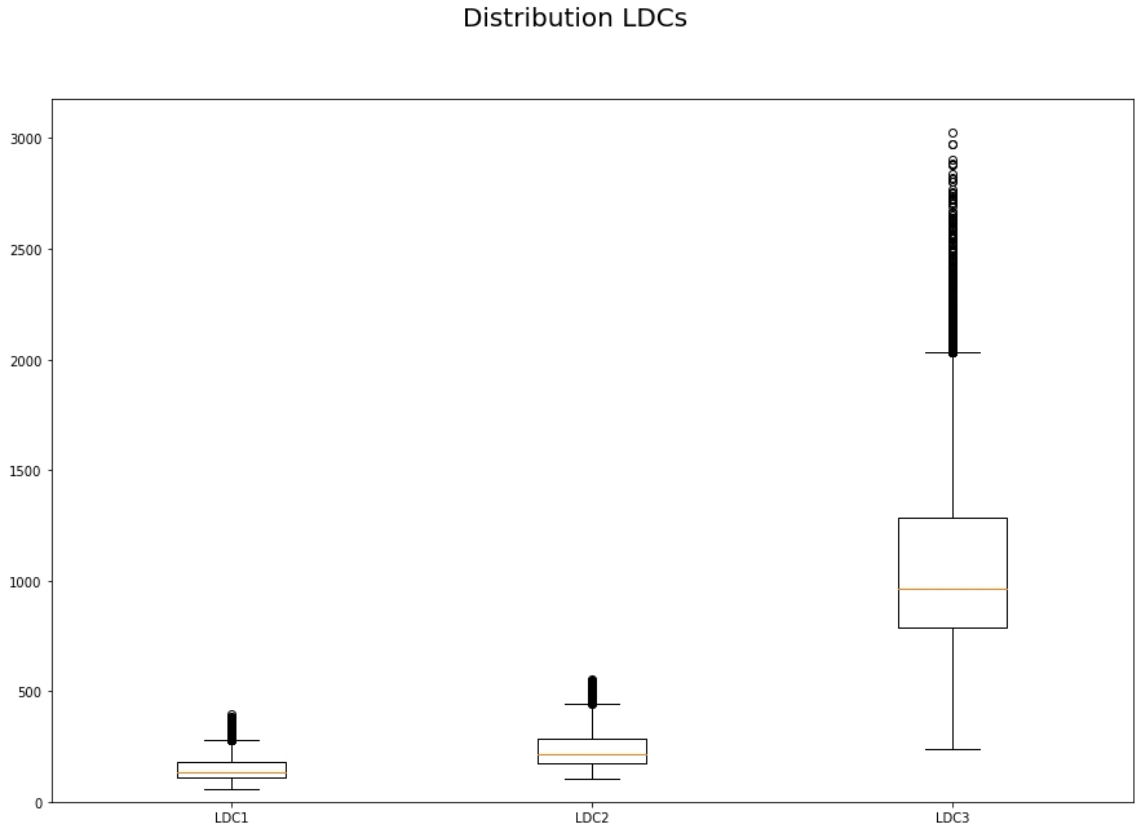


Figure 6.4.: Distribution of Local Distribution Company (LDC) loads

2018. The exact time frames are depicted in Table 6.4.

Test set No.	Start date	Length (Months)
1	2018/01/01	2
2	2018/03/01	3
3	2018/06/01	2
4	2018/08/01	1
5	2018/09/01	2
6	2018/11/01	2

Table 6.4.: Description of test data

6.4.3 Model variations

In the following, we compare different variations of the XGBoost model, as described in our methodology. We compare our results with a naïve benchmark model that simply takes the day-before peak time as forecast, as in Bracale et al. (2017). In Table 6.3, the setup of the different models is shown. First, we introduce a base-

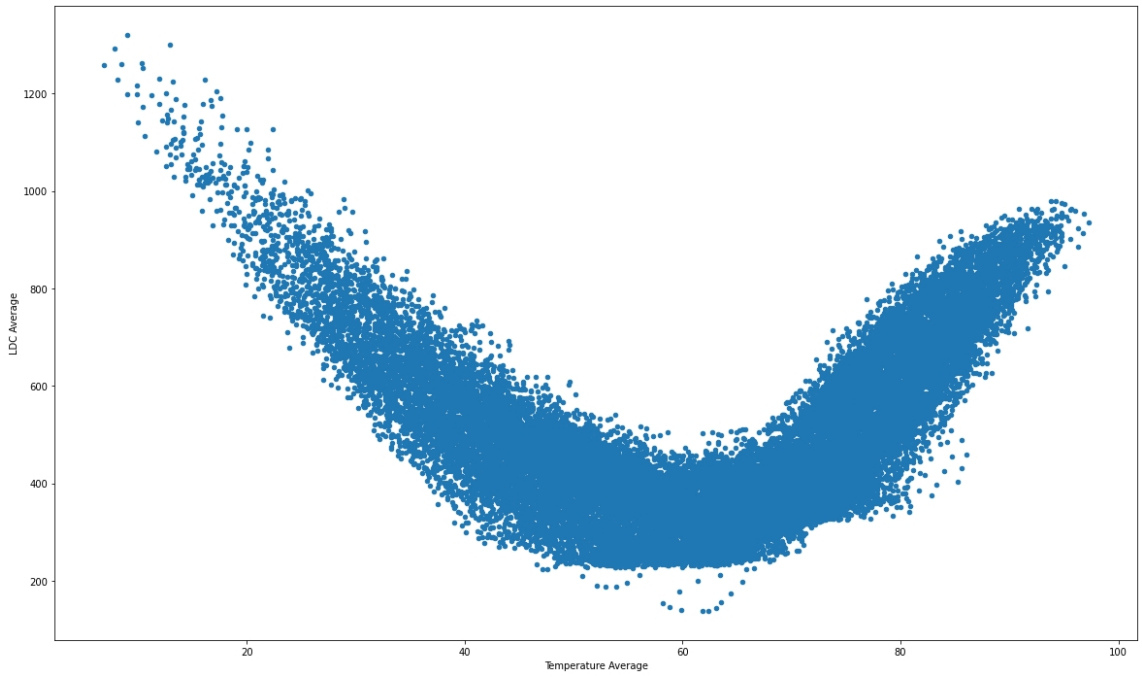


Figure 6.5.: Relation between average temperature measurements and LDC loads

line, regression-based XGBoost model called $XGBP*$, which is only trained with given data, without any feature engineering or hyperparameter tuning. Second, we introduce the baseline XGBoost model $XGBP$, which is trained with enriched feature-engineered data, but only with standard parameters instead of Bayesian-optimized parameters per test. Third, we add to the $XGBP$ model the Bayesian-hyperparameter optimization, which yields model $XGBPH$. The first three models all deliver a general load forecast for every time step in the respective test set, from which the peak time forecast is inferred, as described in the Methodology. The $XGBPH$ model served as a model for the daily peak load forecast and the overall load forecast during the BigDeal Peak Time challenge. For the daily peak load forecast, the model achieved an overall fourth rank amongst all competitors.

The two last models, $XGBR$ and $XGBRH$ are based on the previously introduced Learning-to-Rank XGBoost algorithm, which uses as input ranks of loads instead of absolute loads and which yields a forecast of daily ranks of loads. The $XGBRH$ utilizes Bayesian-optimized hyperparameters.

6.4.4 Bayesian-optimized hyperparameters

As described in our methodology, we conduct a Bayesian hyperparameter optimization, based on test data of 2017, for every LDC. In Table 6.5, we show the resulting hyperparameters for the $XGBPH$ and $XGBRH$ models for each LDC. We can observe that within the respective models, hyperparameters tend to go in the same direction. However, comparing the general load XGBoost model $XGBPH$ and the Learning to Rank $XGBRH$ model, we see significant differences. The max depth is around 5 for every general LDC model, while it is between 8 and 9 for the ranked models. The learning rate, as well as the subsample, are lower for the ranked models.

Hyperparameter	XGBPH	XGBRH
max_depth	5/5/5	8/9/8
learning_rate	0.15/0.11/0.07	0.019/0.03/0.046
subsample	0.97/0.96/0.95	0.53/0.63/0.54
min_child_weight	3.58/3.49/3.64	1.63/4.81/4.36
colsample_bytree	0.77/0.72/0.80	0.52/0.82/0.62

Table 6.5.: Hyperparameter overview

Test Set	Baseline	XGBP*	XGBP	XGBPH	XGBR	XGBRH
1	27 / 34 / 36	36 / 42 / 37	36 / 51 / 47	46 / **69** / **68**	41 / 59 / 54	**58** / **69** / 61
2	31 / 47 / 36	38 / 29 / 35	44 / 53 / **58**	52 / **66** / 54	**58** / 51 / 55	57 / 54 / 55
3	44 / 45 / 32	41 / 52 / 34	46 / 56 / 34	52 / 67 / 61	**59** / 69 / 67	**59** / **77** / **69**
4	41 / 52 / 32	32 / 32 / 35	39 / 58 / 58	**55** / 55 / **71**	52 / 63 / 68	48 / **74** / 58
5	44 / 43 / 51	39 / 39 / 48	41 / 54 / 39	49 / **54** / 49	**52** / 46 / **52**	48 / 46 / 51
6	30 / 31 / 36	30 / 31 / 30	36 / 33 / 49	41 / **48** / **56**	38 / 39 / 46	**43** / 46 / 53
Average	36 / 42 / 37	36 / 38 / 37	41 / 51 / 48	49 / 60 / **60**	50 / 55 / 57	**52** / **61** / 58
Std	7.01 / 7.30 / 6.40	3.90 / 7.93 / **5.56**	**3.34** / 8.27 / 8.92	4.60 / **7.86** / 7.73	7.93 / 10.23 / 6.30	6.09 / 12.48 / 5.96

Table 6.6.: Accuracy in percent

Test Set	Baseline	XGBP*	XGBP	XGBPH	XGBR	XGBRH
1	4.85 / 4.86 / 5.29	2.23 / 1.49 / 1.85	1.88 / 1.46 / 1.31	1.29 / 1.46 / **0.90**	1.69 / 0.78 / 1.22	**1.14** / **0.75** / 0.93
2	4.32 / 3.54 / 3.64	2.07 / 2.37 / 1.26	1.25 / 1.36 / 0.97	1.26 / **1.02** / **0.72**	**1.13** / 1.60 / 1.22	1.20 / 1.70 / 0.95
3	1.02 / 0.95 / 1.02	0.79 / 0.74 / 0.87	0.80 / 0.54 / 0.79	0.65 / 0.41 / 0.49	0.54 / 0.46 / 0.41	**0.51** / **0.36** / **0.36**
4	1.13 / 0.68 / 1.12	1.16 / 0.84 / 1.03	0.84 / 0.45 / 0.55	**0.68** / 0.55 / **0.35**	0.81 / 0.48 / 0.42	0.74 / **0.35** / 0.65
5	2.02 / 1.98 / 2.14	1.57 / 1.79 / 1.85	1.75 / **0.56** / 1.67	**1.21** / 0.67 / **1.59**	1.62 / 1.36 / 1.80	1.31 / 1.16 / 1.61
6	5.54 / 4.44 / 5.26	2.18 / 2.22 / 2.16	2.39 / 1.54 / 1.69	2.08 / **1.59** / **1.08**	**1.86** / 1.86 / 1.39	2.02 / 1.61 / 1.26
Average	3.10 / 2.72 / 3.07	1.67 / 1.52 / 1.50	1.49 / 1.09 / 1.16	1.16 / **0.95** / **0.86**	1.27 / 1.09 / 1.08	**1.15** / 0.99 / 0.97
Std	1.76 / 1.63 / 1.78	0.54 / 0.71 / 0.47	0.57 / 0.45 / 0.43	0.41 / **0.48** / **0.41**	0.48 / 0.55 / 0.51	**0.30** / 0.54 / **0.41**

Table 6.7.: MAE per test set

Test Set	Baseline	XGBP*	XGBP	XGBPH	XGBR	XGBRH
1	4.19 / 4.36 / 4.25	2.24 / 1.53 / 1.93	1.86 / 1.46 / 1.47	1.41 / 1.47 / **0.98**	1.78 / 0.90 / 1.32	**1.17** / **0.81** / 1.00
2	4.11 / 3.33 / 3.52	2.58 / 2.58 / 1.74	1.56 / 1.37 / 1.25	1.43 / **0.97** / **0.88**	**1.27** / 1.53 / 1.32	1.36 / 1.61 / 1.14
3	1.72 / 1.55 / 1.78	1.13 / 1.16 / 1.18	1.22 / 0.74 / 0.95	0.98 / **0.54** / 0.59	0.82 / 0.64 / 0.54	**0.67** / 0.56 / **0.43**
4	2.03 / 1.03 / 1.84	1.97 / 1.06 / 1.52	1.26 / **0.52** / 0.74	**1.06** / 0.71 / **0.48**	0.96 / 0.71 / 0.58	1.16 / **0.52** / 1.00
5	2.74 / 2.18 / 2.26	2.07 / 1.78 / 2.00	1.90 / **0.72** / 1.80	**1.52** / 0.92 / **1.64**	1.87 / 1.46 / 1.85	1.61 / 1.33 / 1.69
6	4.77 / 4.27 / 4.60	2.48 / 2.69 / 2.38	2.14 / 1.86 / 1.64	2.07 / 1.87 / 1.08	**1.80** / 2.11 / 1.56	1.90 / **1.80** / **1.36**
Average	3.26 / 2.79 / 2.95	2.07 / 1.80 / 1.79	1.65 / 1.11 / 1.31	1.41 / **1.08** / **0.94**	1.42 / 1.22 / 1.20	**1.31** / 1.11 / 1.10
Std	1.16 / 1.29 / 0.18	0.47 / 0.64 / **0.37**	**0.34** / 0.48 / 0.37	0.36 / 0.46 / 0.37	0.52 / **0.38** / 0.48	0.39 / 0.50 / 0.38

Table 6.8.: BDPM per test set

6.5 Results

We evaluate the different model variations, based on the six test blocks, for all three LDCs. For every scenario, we evaluate the peak time forecasting performance according to the Accuracy, the Mean Absolute Error (MAE), and BigDeal Peak Time Metric (BPDM).

In Figure 6.6, the actual load of LDC3 for an exemplary day in the first test set is depicted alongside the forecasts of the three general load prediction models $XGBP*$, $XGBP$ and $XGBPH$. We can observe that all three models roughly match the shape of the daily load curve, with two peaks, one in the morning and one in the evening. The higher peak lies in the evening at 23:00. The plot shows a tendency that is later also confirmed in absolute results: $XGBP*$, the model without an enriched, feature-engineered data set, has difficulties in forecasting accurate absolute load values, whereas the two remaining models match the load pattern better. From each general load forecast, the highest forecasted load is inferred as peak time forecast. Here, only the hyperparameter optimized model $XGBPH$ manages to forecast the actual peak at 23:00 accurately.

In contrast to the general load prediction models, the ranked models do not deliver hourly load forecasts, from which the point of time of the highest load is inferred as peak time. Hence, they are first plotted in Figure 6.7, where for every day in the first test period from January to February 2018, the respective peak times and peak time forecasts are plotted for LDC3. The plot follows a certain colour scheme: real peak times are plotted in green, Learning to Rank-based forecasts are plotted in tones of blue, and general load forecasts are plotted in tones of red and velvet. The hyperparameter-tuned models $XGBPH$ and $XGBRH$ are plotted with higher opacity. First, we can observe that, in general, all forecasting models deliver a solid

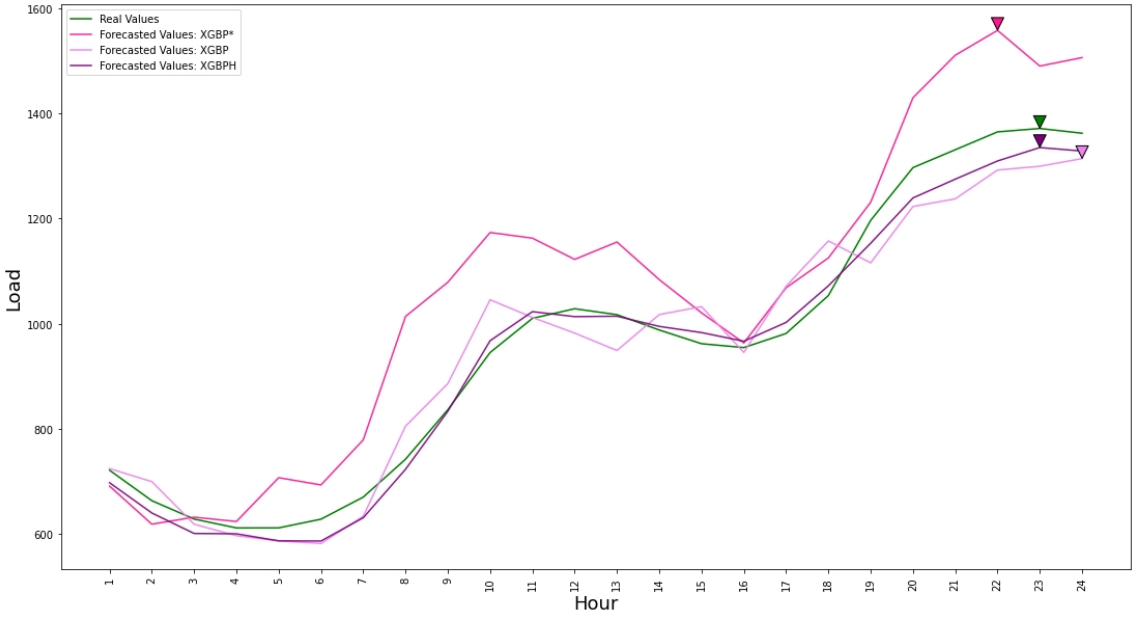


Figure 6.6.: Exemplary load forecast

performance, mostly forecasting the peak time at least at hours around the peak, if not predicting it correctly. We assume that one reason for this is the high aggregation level of local distribution companies, the high data quality, and the fact that the training data covers multiple years. We can also observe the tendency of peaks either occurring in the morning hours around 9:00 or in the evening hours around 21:00. We note that all XGBoost-based forecasting models manage quite well to forecast the peak times, even when there is a switch from periods of morning peaks to evening peaks. If we base our forecast on a recency-based model that always uses the day-before or week-before peak time, this would lead to significant losses in the MAE metric and accuracy if the switches switch from morning peaks to evening peaks. We also note that the Bayesian hyperparameter-optimized models consistently predict the peak time more accurately than the models with standard parameters. The worst performing XGBoost-based model is the one without a feature-engineered data set and hyperparameter tuning, $XGBP*$.

In Table 6.6, the accuracies for the different models are depicted, respectively for each LDC. First, we can observe that all models with the feature-engineered data outperform the model with the base data set, $XGBP*$, by far. Second, both Bayesian-optimized yield the best accuracies on average.

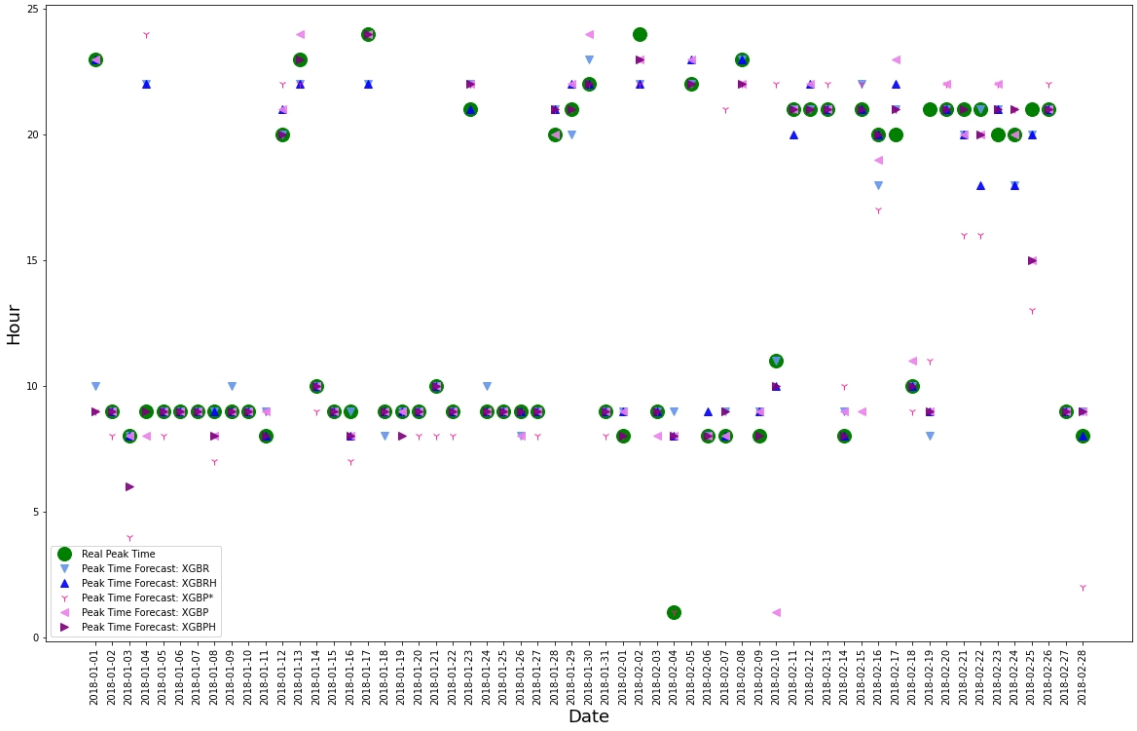


Figure 6.7.: Exemplary peak time forecasts

The same picture occurs when analyzing the resulting MAEs in Table 6.7. On average, the Hyperparameter-optimized models outperform the models without hyperparameter optimization. For LDC1, the *XGBRH* models yield the best MAEs on average; for LDC2 and LDC3, the *XGBPH* yields the best results. All MAE values for the models based on the feature-engineered data set are around 1, which can be interpreted as a mean deviation of the forecasted peak time from the true peak time of one hour.

Similarly, for the BDPM in Table 6.8, the Bayesian hyperparameter-tuned models mostly outperform the standard models, and *XGBPH* delivers the best results for LDC2 and LDC3, while the *XGBRH* model delivers the best BDPM results for LDC1. The average monthly BDPM values reached through the *XGBPH* model are significantly better than the ones reached with the $XGBP*$ model (P=*0.016*). Whereas the average monthly BDPM results reached by the *XGBRH* model are not significantly different than the ones achieved through the *XGBRH* model (P=*0.90*).

This observation underlines two integral findings in our study. First, the peak time forecasting quality is significantly increased by our feature engineering process and

the Bayesian hyperparameter optimization. Overall, our $XGBPH$ and $XGBRH$ models have achieved an exceptional level of peak time forecasting quality. With a Mean Absolute Error of just 1 hour, our performance is by far superior to the baseline case, where the day before peak time is used for the forecast, resulting in a Mean Absolute Error of approximately 3 hours. Second, transforming the actual load values to ranks of loads and employing a Learning to Rank XGBoost model does not significantly lower the peak time forecasting quality compared to the well-performing, regular XGBoost model with feature engineering and hyperparameter optimization, from which peak times are inferred. Thereby, we show that a more privacy-preserving peak time forecasting approach does not necessarily negatively influence the overall forecasting quality. Nonetheless, we note that the information on the ranks of loads still contains some information about the load data providers and could be used to identify them. However, industrial customers could be more open to sharing ranks of loads instead of actual load values with the grid operator since they do not contain information about machine utilization and company activity.

6.6 Feature Importances

In the previous section, we show that our feature engineering process reaches significant accuracy improvements. Hence, we are interested in investigating the average overall feature importances for the $XGBPH$ and $XGBRH$ models through the XGBoost feature weights. The XGBoost feature importance weight can be explained as the number of times that a certain feature is used in the trees of the model. The weight is then calculated as a share of the sum of all feature weights (Ma et al., 2020).

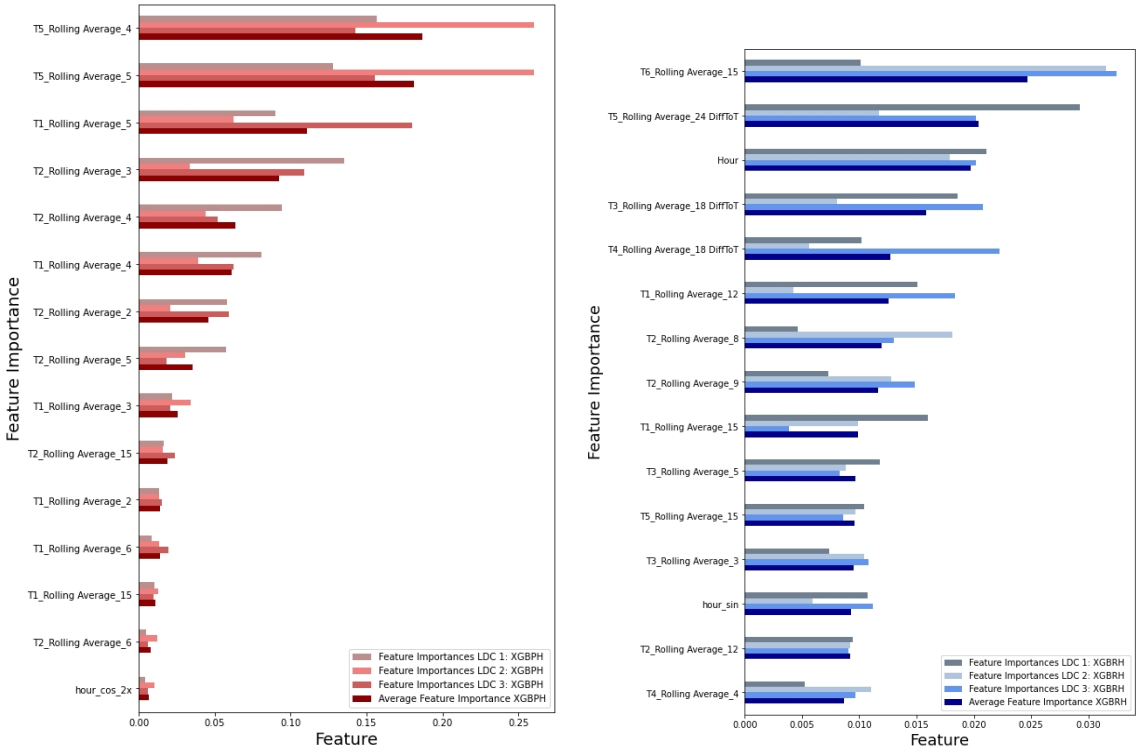
In Figure 6.8a, the averaged feature importance weights are depicted for the $XGBPH$ model. One striking observation is that out of the 15 most important features, 14 features are rolling averages of temperatures instead of the temperature measurements themselves. As mentioned before, we assume that the reason for the high importance of temperature rolling averages is the thermal inertia of buildings. Moreover, we observe that the most important rolling temperature features are those which cover time spans of three to five hours. We can also see that some temperature measurements are relatively more important for the models of the respective LDCs. For instance, $T5$ temperature measurements seem to be more relevant for

LDC2, while $T1$ and $T2$ measurements are relatively more important for LDC1. This might be connected to the distance of the temperature measurement stations to the respective LDCs: the closer the measurement stations are to the LDCs, the more relevant the measurements are likely to be. We see this point especially relevant in light of our previous observation of the strong negative correlation of loads and temperature for lower temperatures and the strong positive correlation for higher temperatures. In addition, it underlines the relevance of considering appropriate temperature measurements in peak time and peak load forecasting tasks, in settings where the temperature is connected to the respective loads. The 15th most important feature is the 2x cosine of the hour, which supports the claim of Semmelmann et al. (2022); Gürses-Tran et al. (2022) that it is reasonable to calculate the sine and cosine of cyclical features.

For comparison, we depict the feature importance weights of the $XGBRH$ model in Figure 6.8b. Again, we can see the high feature importance of rolling average-related features. However, for the Learning to Rank-based models, the important rolling average features cover longer time spans of up to 15 hours. Furthermore, we can see the high feature importance of the "Difference of Temperature to Rolling Average" features, especially for the 24-hour rolling average. High values of this feature indicate temperature peaks, which could lead to its high importance in the Ranked XGBoost model. Furthermore, we can observe relatively lower feature importances per feature, indicating that a wider array of features is used in the Learning to Ranked XGBoost trees.

6.7 Conclusion

This paper offers a novel privacy-shielding approach to the peak time forecasting problem for local distribution companies by leveraging the Learning to Rank XGBoost algorithm. The Learning to Rank model is based on ranks of loads instead of absolute magnitudes of loads, requiring less confidential data. To analyze the accuracy of our approach, we conducted a case study in the context of the BigDEAL load forecasting challenge, where the peak times of three LDCs had to be forecasted. Furthermore, we conducted extensive feature engineering and selected model parameters through a Bayesian hyperparameter optimization. Finally, we analyze the importance of the respective engineered features.



(a) XGBPH models (b) XGBRH models

Figure 6.8.: Comparison of averaged XGBoost feature importances for XGBPH and XGBRH models

We show that the hyperparameter-tuned Learning to Rank XGBoost model delivers the highest average accuracy for two LDCs and the highest MAE and Big Deal Peak Time Metric for one LDC. For the remaining cases, the hyperparameter-tuned general load prediction model, which serves as a baseline in this work and achieved the fourth rank for the peak load forecasting track of the BigDEAL challenge, delivers the best results. Furthermore, we show that all XGBoost-based models significantly outperform a day-before recency-based benchmark model, thereby highlighting the value of XGBoost models for peak time forecasting. Also, we show a strong increase in forecasting accuracy by adding additional features, such as rolling averages of temperature measurements. Future works in this field should apply the rank-based forecasting approach to other methods, such as neural networks, and compare the results with the XGBoost-based Learning to Rank model introduced in this study. Leveraging neural network models, such as recursive neural networks or convolutional neural networks, might show superior performance.

CHAPTER 7

GENERATING SYNTHETIC LOAD PROFILES OF RESIDENTIAL HEAT PUMPS

This chapter addresses the challenge posed by the limited availability of historical heat pump load data by introducing a novel method based on k-means clustering to generate synthetic load profiles. By providing households, policymakers, and aggregators with the ability to create synthetic heat pump load profiles, this approach facilitates assessments of the feasibility and economic potential of dynamic tariffs.

This chapter comprises the following article: L. Semmelmann, P. Jaquart, and C. Weinhardt. *Generating synthetic load profiles of residential heat pumps: a k-means clustering approach*, Energy Informatics, 2023.

7.1 Introduction

Germany plans to install up to 500,000 heat pumps annually from 2024 onwards. In doing so, the government aims to transition the heating sector to carbon-free, electricity-based heating (DeutscheWelle, 2023). While the installation of heat pumps comes with several benefits, such as the high energy efficiency, the reduced reliance on gas, and the absence of direct carbon emissions, the adoption of the technology might also entail certain pitfalls. For instance, Protopapadaki and Saelens (2017) show that heat pump penetrations of more than 20-30% could cause severe issues in distribution grids. Hence, it is essential to analyze the impact of increasing heat pump loads on our energy system and to develop heat pump operation strategies. However, the absence of widely available heat pump load data has led to limitations in existing heat pump-related studies: often, heat pump load profiles are derived by simulation models like TRNSYS (Maranghi et al., 2023; Roccatello et al.,

2023). While this approach is feasible for small-scale, building-level investigations, it is not viable for large-scale studies with a large number of heat pump loads. In Gunkel et al. (2023), the authors study the impact of heat pumps on national peak load hours. The authors use electricity consumption data of 720,000 households to discover that heat pump installations lead to 14% more peak load hours in Denmark than electric vehicles. Albeit the study yields relevant insights on the impact of heat pumps on the national electricity demand and peak loads, the underlying data is not available as open-source data set and hence cannot be used for further studies. Schlemminger et al. (2022) have published the first high-quality and high-resolution data set of heat pump and household loads for 38 households from Hamelin, Germany, which has since served as the data basis for multiple studies in the field. For instance, Yang et al. present a model to manage and coordinate loads in order to reduce distribution grid operation costs. They use the data from Schlemminger et al. to model daily load peaks Yang et al. (2023). Zhu et al. (2023) use the combined household and heat pump data from the data set to model a household in Hamburg and present a carbon reduction- and savings-aware operation mechanism for a combined PV-BES-EV system (photovoltaics - battery storage - electric vehicle). Their study highlights the problems of using a limited heat pump load data set, as the utilization of the Hamelin-based data set to model a Hamburg-based household increases the degree of simulation inaccuracy.

To overcome spatial and temporal restrictions of open-source data sets, researchers have worked on methods to generate synthetic load profiles, especially for residential customers (Pinceti et al., 2019; El Kababji and Srikantha, 2020). Further studies deal with the synthetic generation of industrial and commercial heat load profiles. Jesper et al. (2021), for instance, apply a k-means clustering method on 797 annual gas load profiles to create synthetic industrial and commercial heat load profiles. However, to the extent of our knowledge, there exist no studies and open-source tools for the creation of synthetic heat pump load profiles. In this study, we aim to fill this research gap by introducing a k-means-based clustering model to create synthetic regional heat load profiles based on regional weather data and the data set of Schlemminger et al. (2022).

The contributions of this work are summarized in the following. This paper aims to present the first model for synthetic heat pump load profile generation, applying a

k-means clustering approach. We validate our model based on metrics from existing literature (Li et al., 2020a). We also contribute a novel method to determine the optimal number of clusters, by finding a trade-off between load profile diversity and accuracy. Furthermore, we publish our model and a web-based heat pump load profile generator open-source to enable research models in the field of energy informatics to integrate heat pump energy consumption, thereby helping to overcome the lack of publicly available heat pump load profiles [4].

The remainder of the paper is structured as follows. First, related work is presented and the implications for this study are discussed. Second, the methodology of our approach and the respective evaluation metrics are introduced. Third, we present the case study on which our methodology is applied. Fourth, we evaluate our results according to the introduced metrics and discuss the optimal selection of clusters. Finally, we summarize our work in the conclusion and give an outlook on further research questions.

7.2 Related Work

Most synthetic load profile generation studies are focused on household load profiles. Pillai et al. use artificial neural networks to create normalized residential load profiles based on weather data (Pillai et al., 2014). The authors show the opportunities of synthetic load profile creation, especially for simulations in regions without adequate data, which would otherwise have to rely on inaccurate methods, such as working with a constant load assumption. While Pillai et al. focus on standardized load profiles for whole regions based on temperature profiles, Fischer et al. (2015) introduce a stochastic model to create synthetic residential household load profiles with high resolution, implementing socio-economic features as well as seasonal effects. In a later work, the model is extended with space heating and hot water load profiles (Fischer et al., 2016). In their study, the importance of diverse load profiles is especially addressed. In a further study, Li et al. use an iterative process based on geographic locations and load compositions to create bus-level load time series (Li et al., 2020a). The authors also discuss the validation of their results in detail, which serves as the basis for the evaluation of our case study. In another relevant

[4] Web-based heat pump load profile generator available at `https://heatpump.ninja`. Source-code available at `https://github.com/leloq/synthetic-heat-pump-load-profile-generator`.

study, Jesper et al. use a k-means clustering approach to create synthetic industrial and commercial heat load profiles based on 797 annual natural gas profiles. In their study, the correlation between heat loads and ambient temperature is used to create synthetic heat profiles. Another relevant contribution in the field is made by Ruhnau et al. (2019b), introducing the "When2Heat" data set, which includes national heat pump load profiles and coefficients of performance (COP) for 16 cold-temperature climate countries in the European Union. The authors underline the need for open energy data for electricity market simulations. However, the data set is targeted at nationwide studies, thereby being less appropriate for simulations on the household level and in smaller grid-level aggregations.

While the depicted studies made valuable contributions to the field of synthetic load profile generation, to the extent of our knowledge, no past works are focused on the synthetic generation of household-level heat pump load profiles for varying geographies. Hence, we introduce our k-means-based model in the following section.

7.3 Methodology

In this section, we describe our methodology to create synthetic heat pump load profiles based on the underlying data set. We employ a k-means clustering approach based on the overall methodology introduced by Jesper et al. (2021) to create synthetic heat profiles. First, we describe the overall functionality of the k-means algorithm. Then, we depict the necessary steps to use the k-means algorithm to obtain synthetic load profiles. Finally, we describe suitable synthetic heat profile validation metrics from the existing literature.

K-means algorithm: The k-means algorithm was introduced by MacQueen (1967). The clustering algorithm is highly computationally efficient and easily implementable. Hence, many studies in the field of energy informatics and other domains rely on the k-means algorithm (Jesper et al., 2021; Panapakidis and Christoforidis, 2017; Azad et al., 2014; Jessen et al., 2022). The algorithm iteratively partitions a data set into K clusters, with the aim of minimizing the sum of squared Euclidean distances from every observation to chosen cluster centroids μ_i. Every cluster centroid has the same dimension as the observations. In this study, we aim to cluster temperature profiles day-wise with an hourly resolution. Hence, every cluster centroid μ_i and observation x_j is represented as a 24-dimensional vector, each dimension

representing one hour. After the initiation of cluster centroids and the respective assignment of observations to a specific cluster by minimization of the Euclidean distance, each cluster centroid is iteratively redefined as μ_{i*}, by calculating the mean of all observations M_I that are assigned to cluster centroid μ_i:

$$\mu_{i*} = \frac{1}{M_I} \sum_{m=1}^{M} x_m \tag{7.1}$$

Overall, the k-means algorithm is based on the following steps:

1. *Initialization:* we randomly choose K cluster centroids.
2. *Assignment of observations:* every observation x_j is assigned to its nearest cluster centroid μ_i, based on the Euclidean distance.
3. *Update of centroids:* new cluster centroids μ_{i*} are calculated based on Equation 7.1.
4. *Termination:* the algorithm is terminated when there are no further changes in partitions. Otherwise, the algorithm is repeated again from step 2 onwards.

Synthetic heat pump profile clustering model: We use the k-means algorithm to cluster daily temperature profiles in K clusters, based on the previously introduced procedure. Thereby, for every day d and every household h, the respective heat pump load profile $P_{d,h}$ belongs to the respective cluster $P_{d,h} \rightarrow k$. Therewith, every cluster k has per household its own set of associated heat pump load profiles: $k_h : \{P_{d,h}, ...\}$. As there are multiple possible temperature measurements, we calculate the correlation between all measurements and the target heat pump load to select the most representative temperature profile as the basis for the clustering process.

We can use the fact that every temperature cluster has its own set of associated daily heat pump load profiles to create synthetic profiles (e.g., for new geographic locations) with the following steps:

1. *K-means initialization:* We apply the k-means algorithm on our underlying data set and create k clusters and the associated set of heat pump load profiles.

2. *Temperature processing:* We transform the temperature profile of the desired geographic location and time horizon into 24-dimensional vectors.

3. *Clustering:* We map the daily temperature profiles of the target geography to the previously determined clusters.

4. *Heat pump profile creation:* For every day, we randomly draw a heat pump load profile from the set associated with the respective cluster.

Generally, we can apply this procedure for every household h in the set of all underlying households H, to create a synthetic household load profile. Thereby, we can implement different usage patterns and sizes of households and heat pumps. Through the random selection process in step 4, every synthetically generated heat pump load profile based on the same household is unique. To create a synthetic heat pump data set of N households, we can randomly draw N households from our underlying set of households H.

Validation metrics: Our synthetically generated load profiles can be validated from two perspectives. First, they should follow the distribution of the underlying data set (Snoke et al., 2018). Second, they should vary over different iterations, exhibiting a desired degree of diversity. For the first point, we are evaluating key characteristics of the synthetic data for a given test period and validation metrics suggested by Li et al. (2020a). In detail, we regard load factors over time, which depict the ratio of mean loads and peak loads, as well as load distribution curves, which depicts the percentage of loads in relation to the mean load. For a detailed introduction of the metrics we refer to Li et al. (2020a). Furthermore, we compare the deviation between weekly real and synthetic and heat pump electricity consumption over the regarded test period, and the correlation of synthetic and real profiles, as in Fischer et al. (2016).

Besides the quality of the synthetically generated load profiles, we want to evaluate the diversity of the generated load profiles. Fischer et al. (2016) underline the importance of diversity in synthetic load profiles to avoid aggregations of peak loads. To analyze the diversity of our generated profiles, we compare the mean-variance of L synthetically generated heat pump load profiles. Then, for every time step t in the test period, the variance over all synthetically generated profiles σ_L^2 is calculated.

Finally, we calculate the mean-variance MV over all time steps and synthetic load profiles:

$$\mathrm{MV} = \frac{1}{T} \sum_{t=1}^{T} \sigma_L^2 \tag{7.2}$$

For robust results, we compare the aggregation of synthetic and real loads for all underlying households H in the respective test period, $P_{agg,gen}$ and $P_{agg,real}$. To analyze the diversity of generated load profiles, we create a high number of L and evaluate the MV of $P_{agg,gen}$.

Cluster number selection: Previous studies, such as Jesper et al. (2021), use the elbow method to find the optimal number of K clusters that yield additional input and low distortion within clusters. However, the elbow method is solely focused on the temperature clustering itself. In this study, we adopt a broader view when selecting the optimal number of clusters: an increase in the number of clusters reduces the number of associated heat pump profiles per cluster and thereby makes the clustering model more deterministic and less diverse. When the number of clusters K equals the number of days in the data set D, every synthetically generated heat pump load profile for a day d and household h is similar. Hence, we argue that the accuracy of synthetic profiles should be compared with the diversity reached to find the optimal cluster number K.

7.4 Case Study

We apply the methodology introduced previously on the data set of Schlemminger et al. (2022). The data set consists of household and heat pump load measurements for 38 single-family homes in Hamelin, Germany, for the period between May 2018 and the end of 2020. The households have an average annual household load of 2829 kWh and a heat pump load of 4993 kWh. To the best of our knowledge, the data set is the first of its kind, providing high-quality and high temporal resolution heat pump load profiles. The households in the data set are equipped with water-to-water heat pumps and an additional 6kWh heating rod as backup. Furthermore, the houses are equipped with a 300-liter storage tank. In addition, the households have solar thermal systems installed, which mainly take over the production of hot water during summer. Although this alters the heat pump load profile in summer months

compared to heat pump households without additional solar thermal systems, we argue that this constitutes an acceptable pitfall, as most critical load peaks occur in winter months, when the solar thermal systems remain inactive. We also note that the main type of installed heat pumps in Germany are air-to-water heat pumps Bundesverband Wärmepumpe (2023), which can exhibit different load profiles and react differently to cold temperatures. However, we argue that due to comparable coefficients of performance over different heat pump types, our results can also indicate various other heat pump types, especially in high-load winter weeks (Çakır et al., 2013).

We utilize 21 out of the 38 household load profiles that have no missing data in the period between January 2019 and December 2020 to train our clustering algorithm. We then use the period from May 2018 to the end of 2018 for model testing by creating synthetic heat pump loads and comparing them to the aggregated real loads.

The data set of Schlemminger et al. (2022) includes various temperature features. To ensure that there is a sufficient connection between temperature and heat pump loads, we conduct an initial correlation analysis between the aggregated heat pump load and the temperature features. We observe that the temperature and apparent temperature have the highest correlation with the aggregated heat pump load, as depicted in Figure 7.1. Hence, we base our clustering model on the apparent temperature.

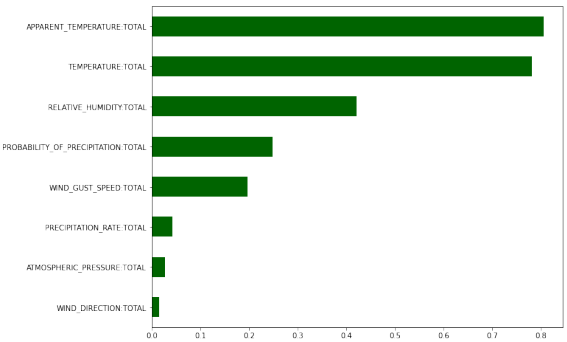
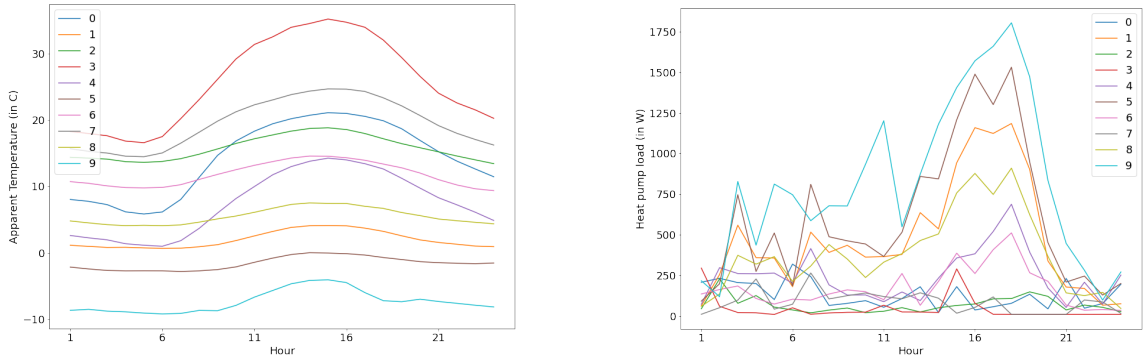


Figure 7.1.: Correlation of temperature features with aggregated heat pump load



(a) Temperature centroids

(b) Mean heat pump load profiles

Figure 7.2.: Clustering results with $K = 10$

7.5 Evaluation

In this chapter we evaluate our synthetically generated heat pump load profiles. First, we depict the general outcome of the clustering algorithm. Second, we compare the aggregated synthetic heat pump profiles for the test period with their actual values. For this period, we also calculate the previously introduced validation metrics. Third, we compare the interplay of accuracy and diversity over a range of possible K clusters to indicate the optimal number of clusters. Finally, we create synthetic data for exemplary cities in Germany to show the generalizability of our approach. In showing the internal validity of the synthetically generated load profiles, rather than benchmarking our model to other possible models, we follow the overall approach of other notable works in the field of synthetic load profile generation (Li et al., 2020a; Jesper et al., 2021).

Clustering: Figure 7.2 shows the result of the clustering process, based on $K = 10$ clusters. We observe ten different centroids of the K-means model based on varying temperature profiles. Therewith associated, we illustrate the corresponding mean heat pump load profiles. For instance, centroid 9 exhibits constantly negative temperatures. The corresponding load profile of the centroid also exhibits the highest heat pump loads with especially high peaks in the afternoon, which would also expected rationally. Overall, we observe that most mean heat pump load profiles have afternoon peaks. Centroid 9, having the highest temperatures, has mean heat pump loads close to 0W.

Validation: To validate the results of our clustering process, we create synthetic

heat pump load profiles for all considered households in 2018 and compare them with the real heat pump load profiles that were unseen in the training process of the k-means algorithm. Figure 7.4 presents the results of this comparison. The synthetic heat pump profiles match the real profile, especially showing low loads during summer time and increasing heat pump loads during colder winter time. This is confirmed in Figure 7.4, where weekly synthetic versus real weekly heat pump energy consumption is displayed. After repeating the synthetic creation of all considered households 50 times and comparing the heat pump energy consumption over the whole testing period with the actual consumption, we find a relatively low error of 2.4% We conclude that during the test period and for the Hamelin data set, the synthetic heat pump load profiles match the shape of the real observed profiles.

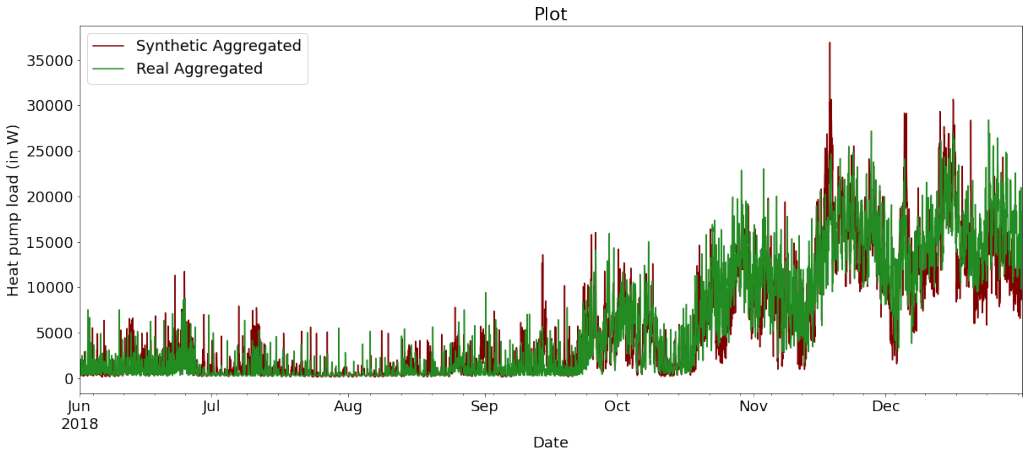


Figure 7.3.: Synthetic versus real heat pump load profile for all considered households

We further validate our synthetic heat pump generation process according to two metrics from Li et al. (2020a), namely load factors and load distribution curves. In Figure 7.5a, we depict the load factors over time, comparing the real and synthetic heat pump load profiles. The load factor depicts the ratio of monthly average loads and peak loads. We can observe that the overall shape of the synthetic data represents the real load profile well. Higher load factors in winter months are especially depicted correspondingly. In Figure 7.5, we depict the load distribution curves of the real and synthetic data, which show the percentage of load being at different values, relative to its mean values. Both curves follow the same pattern. Setting our results side by side with the results of Li et al. (2020a), the deviation of load factors and load distribution curves of synthetic and real profiles is comparable, although the

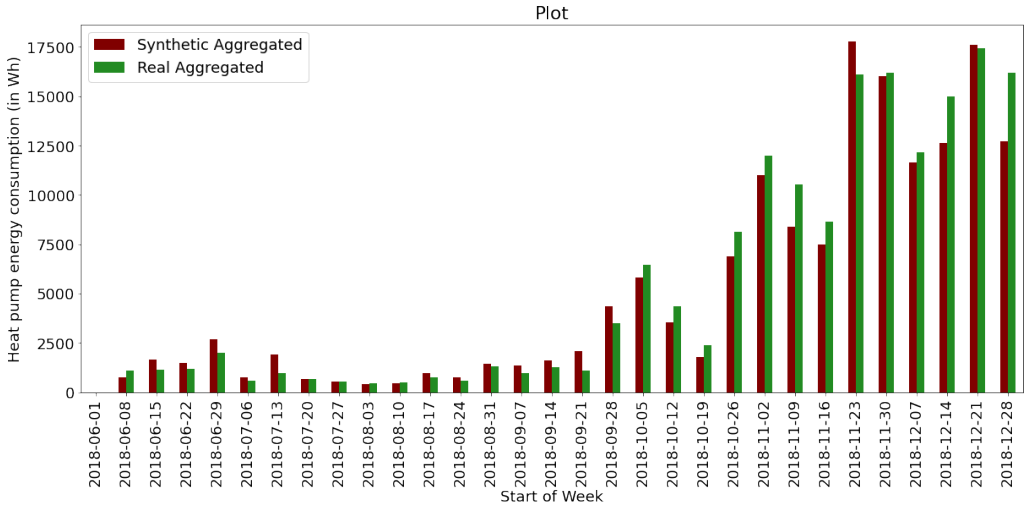


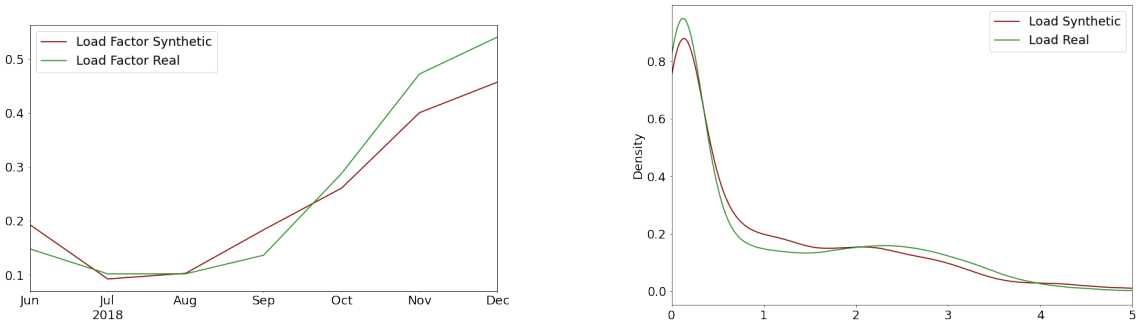
Figure 7.4.: Synthetic versus real weekly heat pump energy consumption for all considered households

overall structure of the metrics varies significantly since Li et al. generate bus-level electricity load profiles of power grids.

As another validation metric, we regard the Pearson correlation coefficient. For our test case, the Pearson correlation between real and synthetic profiles lies at 0.88. For comparison, in Fischer et al. (2016), where synthetic energy demand profiles are created with a stochastic bottom-up method, the correlation lies only slightly higher at 0.92.

Overall, comparing the various depicted validation metrics, we conclude that our synthetically generated heat pump load profiles match the distribution of the underlying data well and can be used for generating synthetic heat pump load time series.

Accuracy vs diversity: To find the optimal amount of clusters, we focus on a comparison of accuracy, in terms of the error of the heat pump energy consumption during our testing period and the diversity of the generated load profiles, represented by the previously introduced mean-variance (MV). To reduce the impact of stochastic effects, we calculate the average of both metrics for 50 generated synthetic heat pump profiles during our testing period for all possible numbers of clusters. Then, we scale down the metrics to a range from 0 to 1 to ensure comparability. Figure 7.6a, depicts the results of this analysis. We find that a low number of clusters leads to high annual consumption errors, providing inaccurate load profiles. On the other



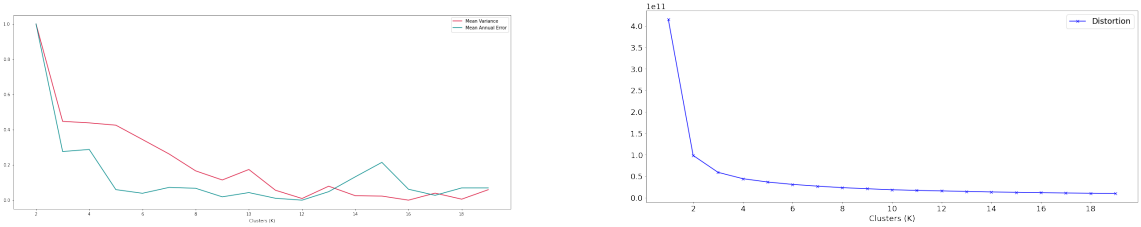
(a) Load factors

(b) Load distribution curves

Figure 7.5.: Validation of synthetically generated heat pump load profiles

hand, a small number of clusters comes with a high degree of variance and, therefore, connected diversity in load profiles. We suggest working with up to 10 clusters for a balanced trade-off between accuracy and diversity. In comparison, using the elbow method, we would work with two clusters, which correspond to the recommended trade-off between distortion and the number of clusters, as depicted in Figure 7.6b. However, as shown in the previous graph, this would lead to a relatively high annual error of the associated synthetic heat pump load profile generation process. Hence, we recommend regarding the trade-off between mean-variance and the error of the generated profiles instead of the clustering-focused elbow method.

We note that our approach is subject to stochastic effects and the underlying data set, although we aimed to reduce the stochastic effects by regarding the average results of multiple runs. We recommend a case-specific selection of the number of clusters K, while in our case study, using up to 10 clusters yields a good trade-off between mean-variance and accuracy.



(a) Scaled mean-variance versus heat pump energy consumption error

(b) Distortion (elbow method)

Figure 7.6.: Cluster amount selection

Transferability: We create synthetic load profiles for two German municipalities, Kuehnhaide and Koeln-Stammheim, for the year 2019, according to our

previously introduced model. Kuehnhaide, with a mean temperature of 7.6°C in 2019, belongs to the coldest German regions, whereas Koeln-Stammheim, with a mean temperature of 11°C, belongs to the warmest regions in Germany. Figure 7.7 presents the transferability of our model and shows that the synthetic aggregated load profile in Kuehnhaide exhibits significantly higher peaks than the profile of Koeln-Stammheim, which would also be expected rationally. Furthermore, the overall base heat pump load level in Kuehnhaide is constantly higher than in Koeln-Stammheim. We interpret this observation as indication that we can use our approach for the synthetic generation of representative heat pump load profiles in other locations with comparable temperature profiles and building structures as in Hamelin, e.g. in Germany, Austria or Switzerland. However, we also observe that days with particular high heat pump loads in Kuehnhaide follow a more similar load pattern than in Koeln-Stammheim. This might indicate that load profiles on these days are drawn from a low temperature cluster with fewer observations. Hence, the publication of further open-source heat pump load data sets, especially from regions with colder temperatures, could contribute to the overall quality of synthetic heat pump load profile generation approaches.

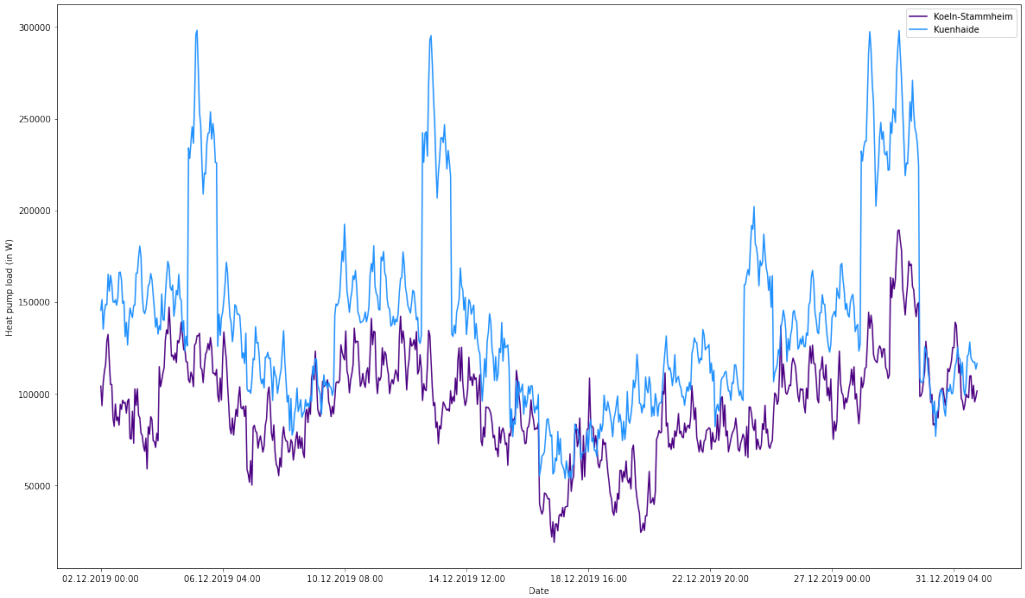


Figure 7.7.: Aggregated synthetic heat pump load profiles in Kuehnhaide and Koeln-Stammheim

7.6 Conclusion

This work presents a k-means-based model to generate synthetic heat pump load profiles. We show that the synthetically generated data follows the structure of the real data according to load factors, load distribution curves and the Pearson correlation coefficient, which underlines the applicability of the proposed heat pump load profile generator for power system simulations. We suggest choosing the number of clusters for the model by comparing the accuracy of generated profiles and their diversity, expressed by the mean-variance of generated load profiles. Future research can work on alternative synthetic load profile generation methods, such as Generative Adversarial Networks and benchmark them with the presented method. Furthermore, future studies may also apply the presented methodology on data sets of other heat pump types and use time series transformation techniques to increase the diversity of the synthetically generated heat pump load profiles.

Part IV.

Behavioral uncertainty

INTRODUCTION TO PART IV

A significant barrier to the adoption of dynamic tariffs is the loss aversion exhibited by many consumers (Nicolson et al., 2018). Dynamic tariffs inherently carry a degree of risk: households that consume electricity primarily during periods of high prices or when wholesale electricity prices unexpectedly spike may incur financial losses. However, this risk can be mitigated by effectively managing household flexibility potentials.

In this part, the economic implications of dynamic tariffs for households are analyzed, and an approach to estimate potential savings based on household characteristics and preferences is proposed. The economic impact of the price risk associated with dynamic tariffs being transferred to an aggregator is examined. These contributions provide a foundation for developing price guarantee mechanisms, which could reduce household price risks and promote the broader adoption of dynamic tariffs.

CHAPTER 8

PRICE GUARANTEES FOR HOUSEHOLDS WITH DEMAND-SIDE FLEXIBILITY

This chapter introduces a novel concept aimed at overcoming households' loss aversion toward dynamic tariffs: household-specific electricity price guarantees. Under this approach, aggregators offer households equipped with heat pumps a guaranteed electricity rate tailored to their specific endowments (e.g., heat pump size, photovoltaic systems, battery energy storage systems, thermal storage), building characteristics (e.g., insulation quality, level of modernization), and preferences (e.g., thermostat setpoint flexibility, lower nighttime temperatures). In exchange, aggregators gain the right to manage the household's flexibility potential.

The implementation of household-level price guarantees involves several key components. First, a sophisticated formulation of home energy management optimization systems is developed, accounting for flexibility endowments, building characteristics, thermal inertia, and heating behavior under dynamic tariffs. Second, a Monte Carlo simulation is performed using representative household samples across multiple years, incorporating day-ahead electricity prices and weather data from four German cities. Lastly, household-specific price guarantees are predicted using a quantile regression approach and evaluated from the perspective of aggregators, providing insights into this concept's feasibility and economic implications.

This chapter comprises the unpublished article: L. Semmelmann, S. Kimbrough, and P. Staudt. *Price guarantees for households with demand-side flexibility potential and thermal building inertia under dynamic electricity tariffs*, Working Paper, 2025.

8.1 Introduction

8.1.1 Background and aim

The sharp increase in intermittent renewable energy generation in the power system of many countries requires an expansion of demand-side flexibility potential (McPherson and Stoll, 2020). Besides the ramp-up of battery storage capacity, household demand response is an integral tool to increase flexibility (Aghaei and Alizadeh, 2013).

One widely discussed policy option to encourage demand response is the introduction of dynamic tariffs for electricity consumers to signal temporal scarcity and provoke a response from flexible loads. While there are various possible implementations of dynamic tariffs (e.g., time-of-use, critical peak pricing, real-time prices, etc.), their underlying rationale is the same: Customers pay time-variable prices for their consumption that are aligned with power system scarcity signals (Freier and von Loessl, 2022). The operation of home energy management systems in the context of dynamic electricity tariffs has been widely studied from an operational research perspective, with a focus on optimization-based approaches to improve energy scheduling and cost management (Althaher et al., 2015; Hubert and Grijalva, 2012).

Despite their potential benefits, the adoption of dynamic electricity tariffs remains limited in major power systems. A recent survey in Germany revealed that only 7% of households have subscribed to such tariffs (Verbraucherzentrale Bundesverband e.V., 2024). While in theory, dynamic tariffs enable an alignment of consumer demand and market signals (Guo and Weeks, 2022), consumers are reluctant in their adoption (Nicolson et al., 2017). A UK study concluded that two-thirds of households would not switch to a time-of-use dynamic tariff (Nicolson et al., 2017). Loss aversion was one of the most prominent drivers of this reluctance.

We argue that the limited adoption of dynamic tariffs is fundamentally rooted in uncertainty. For households, uncertainty about future electricity prices and potential savings contributes to reluctance, driven by behavioral factors such as loss aversion. This uncertainty could be further exacerbated by the growing adoption of electricity-based heating systems, such as heat pumps. These systems introduce ad-

ditional complexities for home energy management, requiring models to account for temperature-dependent heat pump efficiencies and the thermal inertia of buildings, which is influenced by insulation properties (Fitzpatrick et al., 2020; Sperber et al., 2020).

While households may be hesitant to adopt dynamic tariffs due to risk aversion and uncertainty, aggregators offer a viable mechanism to address these challenges (Burger et al., 2017). Aggregators are entities that pool and manage decentralized flexibility resources, acting as intermediaries between end-users and electricity markets to optimize market participation (Burger et al., 2017). While significant progress has been made in developing algorithms and operational strategies for aggregators to manage decentralized fleets of households equipped with battery storage (Angeli et al., 2023) or heat pumps (Kircher and Zhang, 2021; Zhang et al., 2019b), a unified and comprehensive formulation of the operational problem for households under dynamic tariffs remains an open challenge. Existing studies do not fully integrate the interplay of PV and battery energy storage (BESS) systems (Stute and Klobasa, 2024), heat pumps and insulation-dependent thermal inertia of buildings (Sperber et al., 2020), and the uncertainties associated with occupant heating behavior and comfort preferences (Baeten et al., 2017). This gap in the literature creates uncertainty for aggregators, complicating their ability to accurately evaluate and harness the flexibility potential of residential energy systems.

Our study seeks to address the barriers to dynamic tariff adoption by mitigating uncertainties for both households and aggregators. For households, we propose the introduction of price guarantees, which serve as fixed cost per unit of electricity consumed under dynamic tariffs. In other domains, guarantee schemes have been established as a proven tool for mitigating risks (So and Song, 1998; Urban, 2009; Li et al., 2024). The suggested guarantees, offered by aggregators, provide households with a predictable cost framework while transferring the associated risks of dynamic tariffs and volatile electricity prices to the aggregator. In exchange, the aggregator gains operational control over the household's flexibility potential. To address the uncertainties faced by aggregators, we develop a comprehensive formulation of household flexibility potential, integrating factors such as energy storage, thermal inertia, and occupant behavior. This formulation enables a more accurate estimation of the value derived from managing residential flexibility. The proposed

system is then utilized to design household-level price guarantees while quantifying the associated risks and profit potentials for aggregators.

The control of household flexibility potentials through aggregators consistently reduced electricity costs by an average of 7.36% (2.5 ct/kWh). Notably, 78.4% of households benefited from guarantees below the competitive retail benchmark price. Aggregators also experienced improved financial outcomes at the household level, albeit with a moderately wider confidence interval and associated uncertainty. Additionally, the study provides insights into the key factors influencing guarantee levels, emphasizing the critical role of modeling building thermal dynamics based on insulation levels to enhance household flexibility potential. These findings support the development of decision support systems which are capable of offering real-time price guarantees tailored to individual household characteristics.

8.1.2 Proposed approach

Our methodology for formulating individual, household-level price guarantees is based on a three-stage process. First, given a certain household setup and corresponding spot market prices, we formulate a model to calculate the achievable electricity price guarantees. A household setup consists of endowment (e.g., size of heat pumps, PV, BESS, etc.), heating requirements, building insulation, household load profiles, and local weather profiles. Second, we simulate randomly drawn household setups and corresponding possible guarantee prices. The first two stages are simulated under perfect foresight, i.e., we know exact household load curves, price profiles, weather measurements, heating setpoints, and PV generation for the investigated years. In the third stage, we implement a model to suggest actual price guarantees. To this end, we train a forecasting model based on input features that are observable characteristics (e.g., household endowment) and aggregate price information (mean annual day-ahead market price, i.e., a yearly power future price). The model cannot access actual market price curves, weather profiles, PV generation, and heating setpoints. The decision problem directly models the uncertainty utilities and aggregators face when offering individual price guarantees in exchange for household flexibility potential. It captures real-world complexities such as variable consumption patterns, fluctuating market prices, and differing household endowments, align-

ing closely with the practical challenges of implementing such guarantees. We then go on to evaluate the associated risks for aggregators and evaluate the features that influence the level of potential price guarantees.

8.1.3 Related work

The literature has widely discussed operating household flexibility potential with dynamic tariffs to achieve cost savings. The same is true for providing guarantees under uncertainty. This subsection summarizes past studies' main findings and illustrates the corresponding research gap.

Operating household electricity consumption under dynamic tariffs: A high number of studies investigate households subscribing to dynamic tariffs. These studies can be divided into two main groups. First, various studies focus on the overall scheduling problem of flexibility potential, such as battery storage, electric vehicles, and heat pumps (Zhou et al., 2020; Pallante et al., 2020; Pena-Bello et al., 2017). Second, other studies investigate the policy implications of dynamic tariff adoption on a system or grid level (Stute and Kühnbach, 2023).

The literature on load scheduling under dynamic tariffs aims at optimizing the dispatch of flexibility potential for household cost minimization (Zhou et al., 2020; Pallante et al., 2020; Pena-Bello et al., 2017). For instance, battery storage systems can be scheduled with genetic algorithms to shift loads to lower price periods (Pena-Bello et al., 2017), electric vehicle charging events can be scheduled to accommodate user convenience and dynamic price variations (Zhou et al., 2020) or heat pumps can be operated to shift the heating demand based on temperature constraints and day-ahead prices (Pallante et al., 2020; Nolting and Praktiknjo, 2019). Some studies combine different flexibility potentials, e.g., by developing control algorithms for an all-electric dwelling with a heat pump, PV, thermal storage, and an electric vehicle (Pallonetto et al., 2016). While these studies contribute to the overall understanding of the operation of home energy management systems under dynamic pricing, they are mostly tailored to specific use-case studies, such as an Italian office building (Pallante et al., 2020) or a residential building in Ireland (Pallonetto et al., 2016). However, to provide a price guarantee for households with flexibility potential, the formulation of an optimization model of a home energy management system is re-

quired, given varying building structures, heating patterns, and installed flexibility potential. Studies considering multiple sources of flexibility (e.g., electric vehicles, PV, and BESS installations) frequently apply a simplified heating model that only considers a singular building type and heating profile (Stute and Klobasa, 2024; Aniello and Bertsch, 2023).

However, the load-shifting potential of buildings heavily varies across insulation standards (Sperber et al., 2020). Yet, reduced-order thermal response models (also called RC models) are described as viable ways to model the electrical heating demand of households given their building structure (Sperber et al., 2020).

One noteworthy study from Winzer et al. (2024) contributes to the field by proposing and evaluating profile contracts as a mechanism to balance flexibility incentives with price risk hedging for electricity consumers. Profile contracts are real-time tariffs with a hedging component, where customers agree on a fixed price for a predefined consumption profile. At the same time, deviations are settled at spot market prices, allowing for both price stability and dynamic flexibility incentives. While this study acknowledges the need to mitigate price risks and encourages demand-side flexibility, our work extends this by focusing on the integration of household-specific factors, such as thermal building inertia and individualized energy management, to design tailored price guarantees. Unlike profile contracts, which rely on predefined consumption profiles, we propose a framework that dynamically adjusts to individual household behaviors and external uncertainties, enabling a more granular and operationally flexible approach to balancing aggregator risks and consumer incentives.

Most of the aforementioned studies assume that households eventually *will* switch to dynamic tariffs. However, it is unclear if that is the case, given the reluctance of customers to expose themselves to price risks (Nicolson et al., 2017). Hence, we explore in the following section the existing literature on the formulation of guarantees, which are used in other disciplines to transfer and mitigate risks.

Formulating Guarantees: Guarantees are employed across various industries through distinct mechanisms, each aimed at addressing specific risks and enhancing market competitiveness. For example, service time guarantees can provide a competitive advantage by differentiating a firm from its competitors (So and Song, 1998), credit guarantees enable capital-constrained entities to secure loans by transferring risk to a guarantor (Li et al., 2023c; Wang et al., 2022a), and performance guaran-

tees for hybrid power plants facilitate financing by reducing uncertainty about future power output and revenues (Ackermann et al., 2022). In these contexts, the guarantor enhances the attractiveness or accessibility of its products or services, while the counterparty benefits from risk mitigation. Typically, the risk is transferred to an entity that possesses the expertise to manage it. For instance, in China, regulators require operators of peer-to-peer lending platforms to transfer credit risks to specialized guarantee providers (Wang et al., 2022a). A similar risk transfer underpins the approach proposed in our study. While households may lack the expertise or risk tolerance to handle the volatility of dynamic tariffs, aggregators and utilities are well-versed in managing and hedging such risks (Bruninx et al., 2019). Although guarantees benefit all involved parties, the underlying risk remains and must be carefully modeled and analyzed. For instance, guaranteed service times may still be exceeded (So and Song, 1998), borrowers may default on loans (Li et al., 2023c; Wang et al., 2022a), and the guaranteed output of a power plant may fall short of expectations (Ackermann et al., 2022). Therefore, robust risk modeling and analysis are essential to supporting informed decision-making. While such models have been developed and discussed in other fields, the specific problem of formulating guaranteed constant electricity prices in exchange for operating household flexibility potential has not yet been addressed. This problem is particularly challenging because it combines the complexity of optimization with various interconnected constraints and behavioral variability, while also incorporating unique operational restrictions like thermal inertia and the dynamic potential of household flexibility. These elements require a novel approach to balancing risk and resource allocation in a way that ensures both system reliability (e.g., thermal comfort of inhabitants) and economic viability (e.g., limited aggregator risk) under uncertainty. This study seeks to fill this research gap.

Based on existing guarantee-formulation literature, we devise a sector-agnostic scheme in Figure 8.1, which is described in the following. It later serves as the foundation for our guarantee model (So and Song, 1998; Li et al., 2023c; Wang et al., 2022a; Ackermann et al., 2022).

First, a common denominator of guarantee literature is that **case-specific information** about the counterparty and the desired guarantee is ascertained and shared with the guarantee-giving entity. This can include system characteristics like the

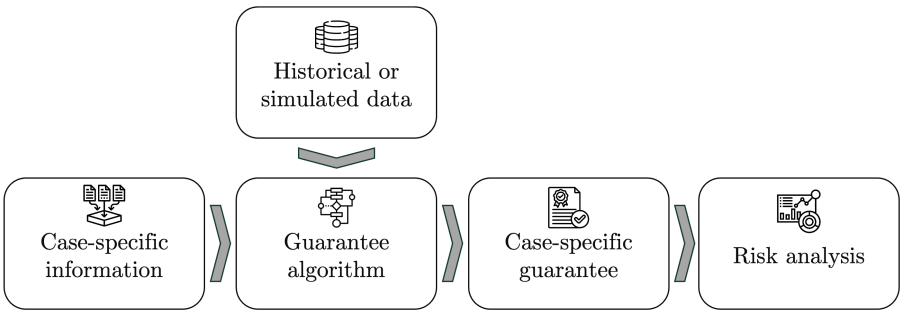


Figure 8.1.: Sector-agnostic guarantee-giving scheme

location or desired output of a power plant (Ackermann et al., 2022), fixed values for costs (Li et al., 2023c; Luo and Wu, 2018), or known probability functions for uncertain inputs like demand (Li et al., 2023c). Then, a connection between the case-specific information and the output guarantee is established based on **historical or simulated data**. For instance, the underlying data set can be built up by simulating a high number of scenarios, consisting of possible parameter combinations and uncertain input parameters (Ackermann et al., 2022; Luo and Wu, 2018). Alternatively, an empirical data set exists upon which the guarantees can be formulated (Consiglio et al., 2008). We note that there are also studies that are not based on historical or simulated data, which then directly go on with a theoretical guarantee formulation (Urban, 2009). However, for high-order problems with time-dependent systems and input parameters (like in our case), this is hardly possible. The actual relationship between case-specific information, potential historical or simulated data points, and viable guarantees is established with a **guarantee algorithm**. The selected algorithm depends on the context of the guarantee. When there is an interaction between the guarantee-giving entity and counterparty, the decision process is frequently modelled with a Stackelberg game (Li et al., 2023c; Wang et al., 2022a). When there is no interaction and the guarantee should be conditioned upon the previously simulated scenario data, a guarantee curve can be obtained with a linear optimization (Ackermann et al., 2022). A guarantee curve illustrates the relationship between a guaranteed outcome given a varying input condition (i.e., guaranteed power output of a PV plant given varying irradiance levels). In other cases, an analytical guarantee formulation can be obtained (Urban, 2009). Then, a **case-specific guarantee** is the output of the guarantee algorithm. This can either

be a fixed-guarantee for given, case-specific input parameters (Urban, 2009), or a curve that describes varying guarantees conditioned on exogenous input parameters (Ackermann et al., 2022; Li et al., 2023c). Eventually, the risk and competitiveness of the formulated guarantees is evaluated in a **risk analysis** step, which mostly focuses on describing the distribution of the results of the guarantees. This may include evaluating the mean of the distribution (Ackermann et al., 2022), confidence interval width (Van Lint and van Zuylen, 2005), specific percentiles or Value-at-Risk measures (Vehviläinen and Keppo, 2003; Ackermann et al., 2022), or the variance of the results (Luo and Wu, 2018).

8.1.4 Contributions

Based on the state-of-the-art research and the resulting research gaps, we make the following contributions:

- **General contribution:** We address the gap between the theoretical flexibility potential of residential energy systems and the observed reluctance of households to adopt dynamic tariffs in practice. Our approach introduces a comprehensive modeling framework for home energy management systems, explicitly incorporating building thermal inertia and behavioral constraints. To mitigate the uncertainty that hinders adoption, we propose a novel mechanism of household-level electricity price guarantees. These guarantees provide cost predictability for households while enabling aggregators to harness residential flexibility potential, effectively lowering barriers to dynamic tariff adoption and fostering greater alignment between consumer behavior and market signals.
- **Specific contribution (1):** We formulate a comprehensive cost minimization framework for dynamic tariffs, integrating diverse flexibility potentials. These include PV generation, battery storage, thermal storage, and the often-overlooked impact of building-level insulation and modernization status on thermal inertia. Additionally, we account for behavioral constraints such as occupant-defined setpoint profiles. This holistic approach combines technical and behavioral dimensions to more accurately capture the real-world flexibility potential of households, providing a practical basis for effective demand-side management strategies.

- **Specific contribution (2):** We quantify the value generated for aggregators through the acquisition and management of household flexibility potentials, providing a detailed assessment of the economic benefits derived from the suggested guarantees.
- **Specific contribution (3):** We predict and interpret household-level electricity price guarantees using a quantile regression approach and evaluate their reliability to ensure effective mitigation of uncertainty for aggregators.

8.1.5 Article organization

This article is structured as follows: Section 8.2 presents the proposed guarantee prediction methodology, which adapts a sector-agnostic guarantee-giving scheme to the context of electricity price guarantees. We detail our home energy management system optimization model, which integrates household flexibility potential and operational constraints into the decision-making process. In addition, we describe the Monte Carlo simulation process, the utilized quantile regression method, and how we evaluate the associated risks of given guarantees. Section 8.3 describes our model validation. In particular, we introduce the building typology that serves as input for our thermal building models, heat pump and thermal storage sizing decisions, empirical setpoint data, and real-world distributions of PV and BESS installations, as well as price and weather data. Section 8.4 presents the results of the model validation. We start with a validation of the modeled behavior of home energy management systems and benchmark it against actual empirical data of the German building stock. We then evaluate the calculated price guarantees. Finally, we investigate the factors that influence the level of price guarantees. In Section 9.7.3, we discuss the implications of our study and conclude in Section 8.6.

8.2 Guarantee Model

We adapt the previously introduced sector-agnostic framework for guarantee algorithms (Figure 8.1) to the specific context of this study. In particular, we design an algorithm that proposes household-specific electricity price guarantees in exchange for granting the operator the right to manage the household's flexibility potential in accordance with the household's stated preferences.

The steps of the proposed method are illustrated in Figure 8.2. At the core of this process is the observation of **household appliance endowments and preferences**, which ultimately shape the flexibility potential of each household (Stute and Klobasa, 2024; Sperber et al., 2020). Household endowments encompass the building type, modernization measures (e.g., improved insulation), PV systems, BESS, heat pump installations, and thermal storage. Preferences determine whether and to what extent the household permits the operator to adjust its heating setpoint profile (e.g., lowering the indoor temperature during high-price periods). The operation of household flexibility potential is subsequently modeled using a mixed-integer linear program, which captures the household-level **home energy management optimization** process.

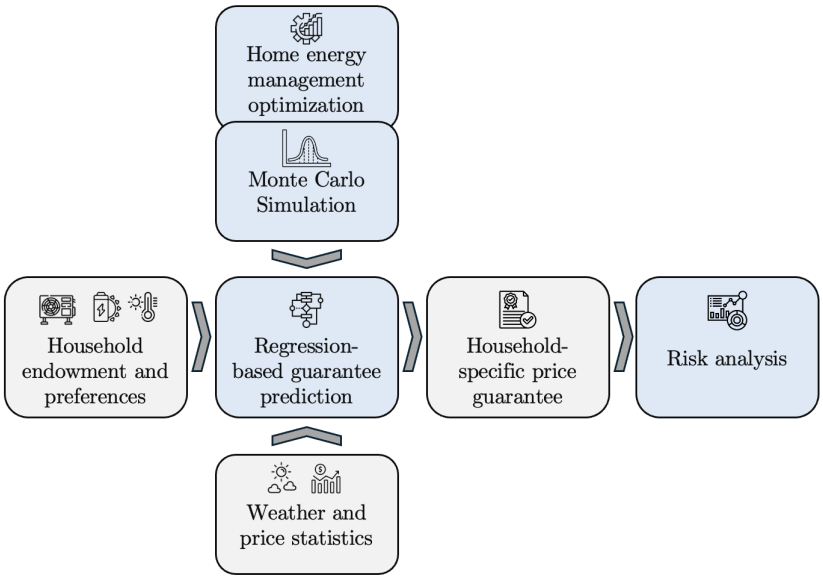


Figure 8.2.: Electricity price guarantee methodology

We divide the investigation period into training and testing years. From the training years, exact setpoint profiles, day-ahead prices (which serve as input for the utilization of flexibility potential), and weather data (outside temperature and solar irradiance) are known. Then, yearly household electricity costs under dynamic pricing can be calculated, given any combination of endowment and preferences. The resulting total household electricity costs can be divided by the total consumption to calculate a price per unit of consumed electricity, which could have been guaranteed to a household. To build up a historical dataset that can serve as input for the guarantee algorithm, a **Monte Carlo Simulation** of potential endowment, prefer-

ence, locations (which influence weather and irradiance), and historical price years is conducted. We resort to a stochastic Monte Carlo approach since calculating all possible parameter combinations would be computationally infeasible (Ackermann et al., 2022). In the resulting data set, the actual weather and price time series are reduced to singular data points (average electricity price, average temperature), based on the assumption that aggregators can hedge future electricity prices (Bruninx et al., 2019) and make assumptions about future weather data. The generated data sets depict the relationship between household characteristics, spot market prices, weather, and potential price guarantees.

Then, the generated data set is used to suggest price guarantees for the testing years based on randomly drawn combinations of endowment and preferences and only the average spot market and weather data points for future years (given the assumption that the operator can hedge these). The underlying temperature setpoint profiles used in the test period have not been seen in the training period. A **quantile regression model** (Somers and Whittaker, 2007) is employed to predict suitable price guarantees given the household characteristics. We employ a quantile regression model due to its ability to extrapolate, the ability to consider varying risk preferences and the interpretability of input factors (Somers and Whittaker, 2007).

Eventually, the suggested **household-specific price guarantees** are calculated for the test period. The price guarantee suggestion can be seen as the real-world decision process of an aggregator under uncertainty (since exact future day-ahead prices, weather data, and setpoint profiles cannot be known). We then employ the home energy management optimization model for these exact years, using perfect knowledge of prices, weather, and setpoints. This allows us to conduct a **risk analysis** of the provided guarantees. We then compare the risk profiles of the outspoken price guarantees with those of a scenario where households remain subscribed to constant flat-rate electricity prices, and aggregators have no control over household flexibility. This comparison highlights the relative trade-offs in financial risk and operational benefits, providing aggregators with insights into whether transitioning households to dynamic tariffs with price guarantees is a more viable strategy than maintaining the status quo. Finally, the impact and hence importance of the **input variables** on the proposed guarantee is analysed in order to investigate the main factors determining household-level electricity price guarantees. This can include

household characteristics (e.g., building insulation, thermal and electrical storage capacity), preferences (e.g., allowed setpoint deviations), weather characteristics and raw market characteristics (e.g., mean annual electricity price).

8.2.1 Home energy management system optimization model

The following section presents the mathematical formulation of the underlying home energy management system optimization model, building on state-of-the-art research in cost minimization modeling (Stute and Klobasa, 2024; Semmelmann et al., 2024), building thermal dynamics (Sperber et al., 2020), heat pump operations (Verhelst et al., 2012; Emhofer et al., 2022), thermal storage systems (Fischer et al., 2017b), and BESS integration (Semmelmann et al., 2024). The objective of the energy management system is to minimize household costs, considering time-varying spot market prices, household load, PV generation, and flexibility potential. Household load comprises the common load of appliances and the electricity demand induced by the heat pump. The flexibility potential is determined by a potential BESS installation, the operational flexibility of the heat pump and a backup heater (which is used when the heat pump's thermal power is not sufficient), building thermal inertia, and the possible presence of thermal storage, as well as some potentially granted thermostat setpoint flexibility.

Objective function: The objective function of a given household seeks to minimize its yearly electricity costs $C^{tot.}$ (Equation 8.21). These yearly costs are computed as the sum of the household's electricity consumption $E_t^{tot.}$ multiplied by the spot market prices p_t^{spot}, minus the PV generation fed back to the grid $E_t^{PV,ext.}$ multiplied by the feed-in tariff p_t^{FIT} at each time step t:

$$\min C^{Tot.} = \sum_t \left(E_t^{Tot.} \cdot p_t^{spot} - E_t^{PV,Ext.} \cdot p_t^{FIT} \right) \tag{8.1}$$

The optimization problem is subject to the following constraints that govern the physical and comfort constraints of the household:

$$E_t^{tot.} = E_t^{HH} + E_t^{HP} + E_t^{HR} + \frac{1}{\eta^{BESS}} \cdot E_t^{BESS,Ch.} - \eta^{BESS} \cdot E_t^{BESS,Dch.} - E_t^{PV,Int.}, \tag{8.2}$$

$$E_t^{PV,Tot.} = E_t^{PV,Int.} + E_t^{PV,Ext.}, \tag{8.3}$$

$$E_t^{tot.} \cdot E_t^{PV,Ext.} = 0, \tag{8.4}$$

$$E_t^{HH},\ E_t^{HP},\ E_t^{BESS,Ch.},\ E_t^{BESS,Dch.},\ E_t^{PV,Int.},\ E_t^{PV,Ext.} \geq 0. \tag{8.5}$$

In Constraint 8.2, the total energy consumption is expressed as the sum of the household load E_t^{HH}, the heat pump consumption E_t^{HP}, the backup heater consumption E_t^{HR}, and the BESS charging demand $E_t^{BESS,Ch.}$, minus the PV generation consumed internally ($E_t^{PV,Int.}$) and the discharged energy from the BESS ($E_t^{BESS,Dch.}$). The BESS inverter efficiency is represented by η^{BESS}.

In Constraint 8.3, the total production of the PV installation is divided into consumed generation and the remainder that is fed into the grid, accounting for potential profits from feed-in tariffs. The mutual exclusivity constraint in Constraint 8.4 ensures that PV generation is only fed to the grid once the household's demand has been satisfied. Additionally, the non-negativity of all energy components is enforced in Constraint 8.5.

The optimization problem could be altered to account for alternative regulatory regimes, for instance, by compensating the PV feed-in with spot market prices or by allowing BESS to discharge into the grid. However, for the sake of conciseness, we orient our formulation around the regulatory environment typically used in feed-in tariff environments (Semmelmann et al., 2024).

Battery operation: The household BESS is operated under the following constraints:

$$E_t^{BESS,Ch.},\ E_t^{BESS,Dch.} \leq P^{BESS,Max.} \cdot \Delta t, \tag{8.6}$$

$$E_{t+1}^{BESS} = E_t^{BESS} + E_t^{BESS,Ch.} - E_t^{BESS,Dch.}, \tag{8.7}$$

$$0 \leq E_t^{BESS} \leq E^{BESS,Max.}, \tag{8.8}$$

$$N^{cycles} \geq \frac{1}{2 \cdot E^{BESS,Max.}} \sum_{t=0}^{T} \left(E_t^{BESS,Ch.} + E_t^{BESS,Dch.} \right), \tag{8.9}$$

In Constraint 8.6, the energy charged to and from the BESS $E_t^{BESS,Ch.}$ and $E_t^{BESS,Dch.}$ is constrained by the maximum BESS power $P^{BESS,Max.}$ multiplied by the duration of a time step Δt. In Constraint 8.7, the energy stored in the battery E_{t+1}^{BESS} is updated as a result of charging and discharging operations and the previous state of charge. The stored energy is further restricted by the BESS energy capacity $E^{BESS,Max.}$, as described in Constraint 8.8. Notably, in cases where a household does not possess a BESS, $E^{BESS,Max.} = 0$, thereby prohibiting any charging or discharging operations. Finally, in Constraint 8.9, the yearly full equivalent cycles of the BESS are limited to prevent excessive usage and the resulting degradation (Semmelmann et al., 2024).

Heat pump operation and thermal constraints: The operation of the household's heat pump is designed to maintain the thermal comfort of the inhabitants, defined by the manually selected thermostat set point T_t^{set} at each time step. We model the impact of heat pump-generated thermal energy on indoor temperature while considering outdoor temperatures, irradiance, and the building structure. To this end, we implement a 1R1C reduced-order thermal model, also referred to as an RC model (Sperber et al., 2020; Wang et al., 2019c). These RC models simplify the thermal dynamics of buildings by using analogies of resistors (R) and capacitors (C) to represent heat transfer and the thermal storage capacity of building materials (Wang et al., 2019c). The specific RC values depend on the building's thermal properties, such as insulation, construction materials, window size, and window type.

The equations governing the indoor and outdoor temperatures in relation to external heat inputs and the building's structure based on the discrete formulation of the 1R1C differential equation from Zhang et al. (2024) are as follows:

$$T_t = T_{t-1} + \frac{1}{R^{ia} * C^i}(T_t^{out} - T_t) + \frac{Q_{t-1}^c + Q_{t-1}^i}{C^i} \tag{8.10}$$

In this equation, T_t^{in} represents the indoor air temperature at time t, T_t^{out} is the outdoor temperature, Q_t^c refers to the thermal energy provided by the controllable equipment (e.g., heat pump, backup heater, thermal storage), Q_t^i is the thermal energy from irradiance, R_{ia} is the thermal resistance between the building interior and envelope, and C_i represents the thermal capacitance of the indoor air.

The thermal input from irradiance, Q_t^i, is defined as:

$$Q_t^i = P_s \dot{A}_i \tag{8.11}$$

where P_s is the solar irradiation and A_i is the effective window area for absorption of solar heat gains on internal air (Sperber et al., 2020).

We couple the thermal energy generated by the heat pump and the electrical energy E_t^{HP} required for that in Equation 8.12, based on the Coefficient of Performance (COP) of the heat pump:

$$Q_t^{HP} = COP_t \cdot E_t^{HP}, \quad E_t^{HP} \leq E^{HP,Max.} \tag{8.12}$$

The COP is determined by the outdoor temperature T_t^{out} and is derived from Verhelst et al. (2012), based on an air-to-water heat pump system connected to a residential floor heating system:

$$COP(T_t^{\text{out}}) = c_0 + c_1 T_t^{\text{out}} + c_2 T_{w,s} + c_3 (T_t^{\text{out}})^2 + c_4 T_{w,s}^2 + c_5 T_t^{\text{out}} T_{w,s} \tag{8.13}$$

The parameters c_0, c_1, c_2, c_3, c_4 and c_5 are based on Verhelst et al. (2012); the water supply temperature $T_{w,s}$ is set at 45°C based on field measurements (Emhofer et al., 2022).

The total heating demand Q_t^c can be satisfied by the heat pump thermal load Q_t^{HP}, the load discharged from the thermal storage $Q_t^{St.,Dch.}$ and the backup heater load Q_t^{HR}. We model optional thermal storage installations (in the form of a buffer tank) by setting up an energy balance of the storage like in Fischer et al. (2017b), neglecting storage losses to the room and assuming homogeneous temperature distribution

within the tank:

$$Q_t^c = Q_t^{HP} + Q_t^{HR} + \eta^{St.} \cdot Q_t^{St.,Dch.} - \frac{1}{\eta^{St.}} \cdot Q_t^{St.,Ch.} \tag{8.14}$$

In this equation, the thermal demand Q_t is met by the heat pump generation Q_t^{HP} and the thermal energy discharged from the storage ($Q_t^{St.,Dch.}$), minus the energy charged into the storage ($Q_t^{St.,Ch.}$). The efficiency losses of the storage are represented by $\eta^{St.}$. The thermal load of the backup heating element, Q_t^{HR}, is equal to E_t^{HR}, as resistance heating operates with a $COP = 1$. We note that we only consider thermal demand for heating, which represents the largest thermal consumption in households (Berger and Worlitschek, 2018) and ignore running hot water demand (as in Le Dréau and Heiselberg (2016); Hedegaard et al. (2017)), which we leave for future research.

We limit the thermal charging and discharging rate to a maximum rate $Q^{Ch./Dch.,Max.}$, to model realistic and feasible operation speeds (Finck et al., 2018):

$$Q_t^{St.,Dch.},\ Q_t^{St.,Ch.} \leq Q^{Ch./Dch.,Max.} \tag{8.15}$$

To model the thermal energy stored in the tank, we track its current charge $Q_t^{St.}$ over time:

$$Q_{t+1}^{St.} = Q_t^{St.} + Q_t^{St.,Dch.} - Q_t^{St.,Ch.} \tag{8.16}$$

The maximum energy stored in the thermal storage is limited by the size of the storage $Q^{St.,Max.}$:

$$0 \leq Q_t^{St.} \leq Q^{St.,Max.} \tag{8.17}$$

Analogous to the BESS, $Q^{St.,Max.} = 0$ implies that no buffer tank is installed. Both the maximum thermal energy stored in the buffer tank and the maximum discharging power are derived from an experimental evaluation (Finck et al., 2018).

The indoor temperature T_t^{in} at time step t may deviate from the setpoint temperature T_t^{set}. This deviation is denoted as ΔT_t, which is defined as:

$$T_t^{\text{in}} - T_t^{\text{set}} = \Delta T_t \tag{8.18}$$

Finally, we denote deviations from the actual indoor temperature and the setpoint temperature as ΔT_t:

$$\Delta T_t = |T_t^{\text{in}} - T_t^{\text{set}}| \tag{8.19}$$

We use these deviations to calculate discomfort costs, which are then integrated into the household optimization problem based on Baeten et al. (2017). We implement discomfort costs in the optimization problem to incentivize an operation according to the setpoint profile set by the household. The discomfort cost $C_t^{\text{Discomfort}}$ is determined by the amount of deviation beyond a permissible flexibility range T^{Flex}. If the deviation ΔT_t exceeds the allowed flexibility T^{Flex}, a discomfort cost is applied. Otherwise, the cost is zero:

$$C_t^{\text{Discomfort}} = \begin{cases} p^{\text{Discomfort}}(\Delta T_t - T^{\text{Flex}}), & \text{if } \Delta T_t > T^{\text{Flex}} \\ 0, & \text{otherwise} \end{cases} \tag{8.20}$$

The discomfort costs $p^{\text{Discomfort}}$ are set relatively high ($=100\frac{EUR}{K}$ to prioritize thermal comfort over cost savings.

The originally introduced objective function is complemented with the discomfort costs to achieve cost savings while maintaining thermal comfort for inhabitants (Baeten et al., 2017):

$$\min C^{Tot.} = \sum_t \left(E_t^{Tot.} \cdot p_t^{spot} - E_t^{PV,Ext.} \cdot p_t^{FIT} + C_t^{\text{Discomfort}} \right) \tag{8.21}$$

8.2.2 Monte Carlo Simulation

We conduct a Monte Carlo simulation to generate possible cost outcomes (and hence cost guarantees) for a diverse dataset of household, market, and weather parameters, which we subsequently use to train an algorithm for predicting cost guarantees at the individual household level. The Monte Carlo method is a widely utilized approach for estimating result distributions based on random input parameters and uncertain variables (Hammersley, 2013; da Silva Pereira et al., 2014). We employ a Monte Carlo simulation because an exhaustive simulation of all potential parameter combinations

is too computationally expensive[5].

We simulate multiple iterations of price guarantee calculations, as described in the mathematical formulation in Section 8.2.1, using randomly drawn input parameters detailed in Section 8.3. The generated dataset serves as input for price guarantee forecasting in Section 8.2.3, where the resulting distributions are analyzed to evaluate risks from the perspective of the aggregator. The sample size and the convergence of the Monte Carlo simulation are evaluated based on the Central Limit Theorem (Yang, 2011), which ensures the stability and reliability of the Monte Carlo simulation by approximating the sampling distribution of the mean as normal with increasing sample size. Convergence is validated by monitoring the width of the 99% confidence interval for the target variable, which stabilizes when additional iterations no longer significantly impact the confidence interval.

8.2.3 Quantile regression-based guarantee prediction and benchmark methods

After having generated the underlying sample of price guarantees given different household characteristics and price and weather years, we employ a quantile regression to build a price guarantee prediction model (Somers and Whittaker, 2007). Quantile regression is a statistical method used to estimate specific percentiles (quantiles) of a target variable based on a set of independent variables. Unlike linear regression, which predicts the average relationship between variables, quantile regression focuses on different points of the distribution, such as the median or upper/lower percentiles. This makes it particularly useful for the prediction of the desired price guarantees, as it allows aggregators to tailor guarantees to different levels of risk tolerance by focusing on specific segments of the price distribution.

The quantile loss function, ρ_τ, for a given quantile $\tau \in (0, 1)$ is defined as follows:

$$\rho_\tau(y) = y(\tau - I_{y<0}(y)),$$

[5]For instance, assuming 12 different building types, 3 modernization options, 100 setpoint profiles, 4 PV sizes, 4 BESS sizes, 4 cities, and 4 price years, the total number of combinations would be $12 \cdot 3 \cdot 100 \cdot 4 \cdot 4 \cdot 4 \cdot 4 = 921,600$. If each simulation iteration required 1 minute of computation time, the process would take approximately 921,600 minutes, or around 639 days. Thus, a Monte Carlo-based approach is adopted to ensure computational feasibility.

where $I_{y<0}$ is the indicator function. By minimizing the expected loss $\mathbb{E}_{\rho_\tau}(Y - a)$ with respect to a, the solution corresponds to the τ-th quantile of Y.

The presented quantile regression method can be used to balance the competitiveness of the offered guarantees and the risk appetite of the aggregator. Lower quantiles (e.g., $\tau = 0.1$) represent guarantees offering lower prices to households, which are riskier for the aggregator as they increase the potential for financial loss in unfavorable market conditions. Conversely, higher quantiles (e.g., $\tau = 0.9$) correspond to higher guaranteed prices, which are less risky for the aggregator but may be less appealing to households.

The resulting regression model enables aggregators to forecast the guaranteed electricity price p_i^g for a household i as:

$$p_i^g = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i,$$

where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients of the predictors $X_{i1}, X_{i2}, \ldots, X_{ip}$ (e.g., size of the BESS, size of the thermal storage, average price of the investigated year), and ϵ_i represents the residual error.

We set $\tau = 0.5$ for a balanced aggregator risk profile and conduct an additional sensitivity analysis for alternative quantiles.

In our approach to calculating price guarantees, we use the quantile regression to forecast relative changes as percentage deviations from the average electricity price rather than directly predicting the absolute guarantees. This method has proven more effective in preliminary experiments, as it focuses on the underlying relationships between the guarantees and the input features, rather than being influenced by the absolute magnitude of electricity prices. By transforming the predicted percentage deviations back into absolute guarantees using the average electricity price, we ensure that the model remains adaptable to varying market conditions. This approach reduces the risk of overfitting to specific price levels or anomalies in the training data, thereby enhancing the robustness of the model, especially in years characterized by high price volatility. It also improves the model's generalizability, as percentage deviations inherently normalize the data.

We evaluate the performance of the quantile regression-based household-level guarantees by benchmarking them against a baseline scenario. In this baseline, house-

holds are offered a uniform retail rate by their utility or aggregator, and operational control remains with the local home energy management system rather than being transferred to the aggregator. The uniform retail rate is determined as the volume-weighted wholesale market price, calculated using representative standard load profiles for the respective years (Katz et al., 2016; Bundesverband der Energie- und Wasserwirtschaft e.V., 2024). This baseline provides a realistic reference point, as uniform retail rates are widely used in existing energy market structures, allowing for meaningful comparisons with the proposed guarantee model. Moreover, by keeping operational control local, the baseline highlights the added value of transferring control to the aggregator.

8.2.4 Risk evaluation and feature importance

We can evaluate the quality of the predicted guarantees by comparing them to the actual achievable profit guarantee (p_i^{g*}) in the test years. The calculation of actual achievable guarantees is also subject to information not available to the prediction algorithm, e.g., the realization of thermostat setpoint profiles, detailed weather profiles, and actual day-ahead price curves.

When assuming that a household i has been granted guarantee p_i^g, we compare the realized total household costs $C^{Tot.}$ with the guaranteed costs considering the guaranteed rate. The resulting difference is denoted as "household result" in the remainder of the study. The following metrics are applied to analyze the risk of the given price guarantees based on the distribution of household results:

- **Realized average electricity cost per unit:** This metric evaluates the yearly electricity costs per unit of consumed electricity for households under different flexibility management scenarios. It captures the economic impact of aggregator-controlled flexibility, where aggregators optimize household energy use to respond to dynamic electricity prices, compared to household-controlled flexibility, which typically focuses on maximizing PV self-consumption. The metric highlights the cost savings achieved through optimized energy management and market price responsiveness.

- **Mean household result:** The mean household result is calculated for each test year by comparing the pre-determined guarantees granted to households

with the actual costs that accrued during the year. Based on (Ackermann et al., 2022), we interpret a mean result of 0 as ideal, as it reflects a perfect balance where the guarantee-giving entity neither incurs excessive losses nor gains disproportionate profits. This ensures that the guarantees align closely with the actual household costs, maintaining fairness and financial sustainability while reinforcing the reliability of the prediction algorithm. We note that a household result of zero does not necessarily mean that there are no profits for the aggregator, since he could add a profit margin on this value.

- **Width of Distribution:** We evaluate the accuracy of the household result distributions by calculating the width of the 5%-95% confidence interval. The width of a distribution is a key metric for evaluating variability and reliability (Van Lint and van Zuylen, 2005). In our case, narrower distributions of household results, reflected by a smaller width of the confidence interval, would be a sign of improved predictability and consistency in the outcomes of the guarantee framework. This makes width a particularly suitable metric, as it directly captures the variability in household results. By minimizing the width, the aggregator can ensure a more uniform and predictable financial exposure.

We evaluate the factors influencing potential price guarantees by examining the effect size, direction, and significance of the coefficients in an ordinary least squares linear regression, using the relative difference between the guarantees and the average electricity price for the respective years as the target variable.

8.3 Model validation

In this section, we describe the parameters and distributions we use in the Monte Carlo simulation. We base our simulations on empirical price, weather and building data from Germany.

8.3.1 Building models

We model the building thermodynamics of the investigated households using reduced-order 1R1C thermal response models, which consolidate the thermal properties of buildings into a single thermal resistance (R) and capacitance (C) network. This approach provides a computationally efficient and accurate representation of

thermal dynamics (Sperber et al., 2020). The input for these models is derived from representative German building typologies, encompassing various construction periods and insulation states. These typologies were developed as part of a large-scale statistical analysis of the European (and specifically German) building stock (Ballarini et al., 2014).

The dataset includes 12 distinct building types (as shown in Table 8.1) and three modernization variants: original condition, conventional renovation with moderate insulation, and deep renovation with high insulation. We utilize the 1R1C values computed in Sperber et al. (2020), along with the distribution of building sizes and living areas, which are subsequently used to determine appropriate heat pump sizes as follows.

Code of building type	Construction period	Heated living area in m^2	Building stock (in thousands)	A_i	C_i	R_{ia}
SFH A	< 1859	199	330	1.45	3.17	4.76
SFH B	1860 - 1918	129	966	1.12	2.74	6.08
SFH C	1919 - 1948	275	1,131	2.89	3.82	3.35
SFH D	1949 - 1957	101	859	0.92	2.10	7.51
SFH E	1958 - 1968	110	1,509	1.36	2.72	5.57
SFH F	1969 - 1978	158	1,507	1.71	3.33	5.27
SFH G	1979 - 1983	169	704	1.35	2.67	6.22
SFH H	1984 - 1994	137	1,160	1.75	2.88	6.19
SFH I	1995 - 2001	111	1,035	2.00	2.09	7.31
SFH J	2002 - 2009	133	907	5.11	8.00	7.35
SFH K	2010 - 2015	160	494	5.61	8.00	6.59
SFH L	> 2016	160	258	5.72	8.00	6.71

Table 8.1.: Summary of single-family house (SFH) typologies for the building stock at the end of 2018

8.3.2 Heat pumps and thermal storage

We derive the size of the heat pump from the living area of the investigated building types presented in Table 8.1. We follow an established sizing rule (Fraga et al., 2018), which recommends a nominal heat pump power of 69.8 W/m^2 for non-retrofitted buildings, across all building types, including retrofitted ones. This conservative approach ensures sufficient capacity for peak heating demands and simplifies the simulation process. Additionally, we assume that every household is equipped with a 6kW backup heater (Schlemminger et al., 2022). The heat pump COP is calculated based on field measurements from an air-to-water heat pump with a water supply temperature of 45°C, fitted to a polynomial function of the outside temperature (Emhofer et al., 2022). Detailed parameters for this calculation are provided in the Supplementary Material (see Table A.1).

We also model the potential inclusion of a thermal storage installation. We assume that a $0.5m^3$ (500l) thermal storage water tank provides 17.8kWh of thermal flexibility potential, which can be fully discharged in one hour (Finck et al., 2018). Currently, there is no empirical data on the prevalence and distribution of thermal storage in heat pump-equipped households. Therefore, we assume that 50% of such households are equipped with additional heat storage, with half of these having a small 500-liter tank and the other half a large 1,000-liter tank. This assumption aligns with product variants offered by major heat pump and water storage manufacturers (Vaillant Group, 2024).

8.3.3 Thermostat setpoints

The thermostat settings of the investigated households represent a source of flexibility but also uncertainty for aggregators. To address this, we utilize 1,000 empirical yearly setpoint profiles from Luo, Na and Hong, Tianzhen (2022), measured in 2017. We use 80% of these profiles during the Monte Carlo simulation, reserving the remaining 20% for the test period. This approach ensures that the setpoint profiles of the households in the test period are unseen during training, thereby reflecting the decision-making challenges and uncertainties that aggregators would encounter in practice. It should be noted that the setpoint dataset from Luo, Na and Hong, Tianzhen (2022) is based on a US sample, which may exhibit heating patterns differing from those typical in other countries. However, there is no comparable dataset available.

Additionally, we model the option for households to adopt a nighttime setback feature, in which the thermostat temperature is reduced to 15.6°C between 10 PM and 6 AM, as described in Moon and Han (2011). This feature is incorporated to evaluate whether offering such a contract option affects the possible price guarantees. We assume that 75% of households opt for unobstructed thermostat operation, corresponding to a “heating-as-desired" scenario. This reflects findings from studies such as Sachs et al. (2012), which indicate that a significant majority of users prioritize maintaining thermal comfort over energy-saving behaviors, with many overriding energy-saving thermostat settings or avoiding setback options entirely. Furthermore, we assume that all households may offer 0°C, 1°C, or 2°C setpoint flexibility as a contract detail (i.e., allowing the operator to deviate from the desired setpoint),

with these values being equally probable for the Monte Carlo simulation. This contracted flexibility introduces the potential for gaming (e.g., offering a 2°C flexibility potential and subsequently increasing the daily setpoint profile by 2°C). Designing corresponding incentive systems to ensure the revelation of correct underlying preferences remains a subject of future research.

8.3.4 PV and BESS

We derive the distribution of PV installations and BESS sizes from empirical data of a German governmental registry of energy-related installations (Marktstammdatenregister, 2024), which we parsed and processed using the *open-mastr* package (Kotthoff et al., 2024). In our analysis, we focus on PV systems with a nominal power of 10kW or below and BESS systems with a capacity of 10kWh or less, which we consider representative of typical household sizes (Hartner et al., 2017). This excludes large-scale and industrial storage installations, which are beyond the scope of this study. As of December 12th, 2024, there were 1,672,942 BESS installations and 3,210,749 PV systems recorded within the set range.

We assume all simulated households to be equipped with a PV system, as we consider this a viable future scenario for those households that might provide demand-side flexibility given the rapidly increasing number of PV installations (Destatis, 2024a). Since larger PV systems are typically associated with larger batter storage capacities, we do not assume independent distributions for PV sizes and BESS capacities in our Monte Carlo simulation. Instead, we define four PV size buckets (0.1–2.5kW, 2.5–5kW, 5–7.5kW, 7.5–10kW) and five BESS capacity buckets (0kWh [no storage], 0.1–2.5kWh, 2.5–5kWh, 5–7.5kWh, 7.5–10kWh). The resulting distributions are shown in Figure A.1 in the Supplementary Material.

An analysis of this data shows that larger PV installations tend to correlate with larger storage systems, while smaller PV systems are predominantly paired with smaller or no BESS. In our Monte Carlo simulation, the PV size for each household is first drawn from the defined PV size buckets based on their assigned probabilities. Subsequently, the BESS size is drawn from the corresponding BESS capacity bucket associated with the selected PV size bucket. This conditional sampling approach reflects the observed real-world relationship between PV and BESS sizes, providing

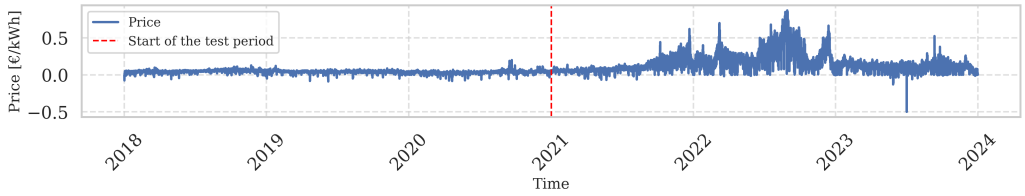


Figure 8.3.: Mean daily day-ahead spot market prices for the DE/LU market zone

a more realistic representation of household configurations in the simulation.

Based on the drawn PV sizes, hourly PV generation time series are created following the widely-proliferated method from Pfenninger and Staffell (2016), incorporating variations in cloudiness, irradiance, and other weather factors, while assuming a 10% system loss, 35° tilt, and 180° azimuth, though real-world panels vary in azimuth, tilt, and efficiency. However, for the sake of conciseness, we consider one uniform system. The allowed BESS Equivalent Full Cycles are (N^{cycles}) set at 365, which represents one cycle per day. The maximum discharge power $P^{BESS,Max}$ is calculated by multiplying the sampled BESS size by 0.41, based on German field measurements (Semmelmann et al., 2024).

8.3.5 Day-ahead prices and grid fees

The dynamic tariffs in our study are based on the day-ahead market price of the German market zone in the European electricity market, obtained from Bundesnetzagentur (2024c). We chose the German-Luxembourg market zone as it is the largest and most liquid market zone within the European market. We depict the observed prices in Figure 8.3, showing high price fluctuations from 2021 onwards. The years 2021, 2022, and 2023 are then used as test periods to evaluate the model's performance even under challenging market conditions characterized by high price volatility. esting the model in such an environment is important to assess its robustness and adaptability, ensuring that it can effectively handle extreme fluctuations and provide reliable outcomes even under highly dynamic and uncertain market scenarios.

In addition, we consider grid charges, as well as taxes and levies from Bundesverband der Energie- und Wasserwirtschaft (BDEW) (2024) in our study, to model household electricity costs realistically. We use yearly average electricity prices as input for the quantile regression, as described in Section 8.2.3. The resulting grid

charges, taxes, levies, average market prices and volume-weighted average prices are shown in Table Table A.2 in the Supplementary Material.

8.3.6 Weather data

The underlying thermal building model is influenced by two weather variables: the solar irradiance and outside temperature (Sperber et al., 2020). We obtain weather data for four German cities, Munich, Cologne, Berlin, and Potsdam (Yang et al., 2021; Visual Crossing, 2024), to model the impact of different weather patterns and associated temperature differences.

Figure 8.4.: Mean weekly temperatures and solar irradiation (GHI) in the four investigated cities

Figure 8.4 shows the temperature and Global Horizontal Irradiance (GHI) over the target cities and investigated years. We observe the same pattern over all cities (lower temperatures and irradiance in winter, higher values in the summer), with some regional differences. For instance, Munich consistently exhibits lower irradiance, while temperatures tend to be higher in Cologne.

The RC values in Sperber et al. (2020) are fitted based on Southern Vertical Irradiance instead of the more common Global Horizontal Irradiance. Hence, we transform the GHI values in an additional step to Southern Vertical Irradiance with the Python package *pvlib* (Holmgren et al., 2018). We depict the resulting change of the irradiance curve in the Supplementary Material in Figure A.2.

8.4 Results

The results are divided into three parts: First, we validate the outcomes of our home energy management optimization and thermal models by comparing the simulated household electricity and thermal energy demand to official German statistics. Second, we analyze the economic value of aggregators taking over the flexibility po-

tential of households. Third, we examine the resulting price guarantees and the associated risk for aggregators. And fourth, we investigate the most important factors that influence the guaranteed prices.

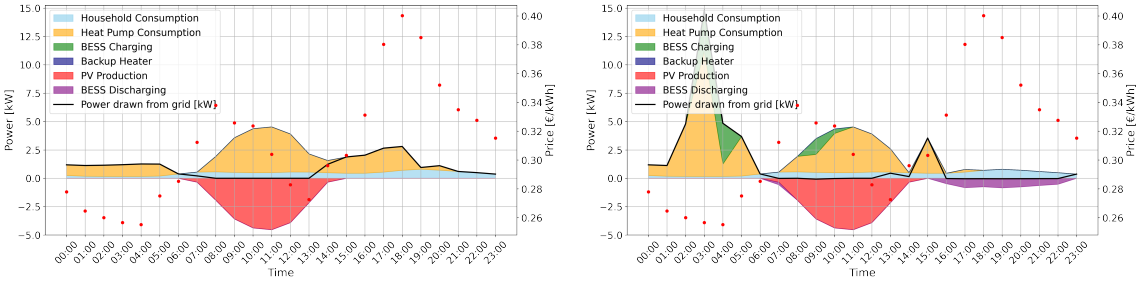
8.4.1 Home energy management system validation

We optimize the household operation based on the model from Section 8.2. While with fixed guaranteed prices, households would not have the incentive to shift demand, the aggregator would be exposed to dynamic price risk. To manage the risk, the aggregator actively utilizes household flexibility, which may involve optimizing battery and thermal storage operations or adjusting building (pre-)heating schedules. These actions are performed within the boundaries of user-defined comfort constraints.

In Figures 8.5, 8.6 and 8.7, we compare the way the home energy management system operates based on different flexibility provision on the same day for the same household (a single family home (SFH) A building that has been deeply renovated, an 8.75kW PV installation, 8.75Wh BESS, 1,000-liter thermal storage buffer tank, 1°C of granted setpoint flexibility, in Berlin on the 20th January, 2022).

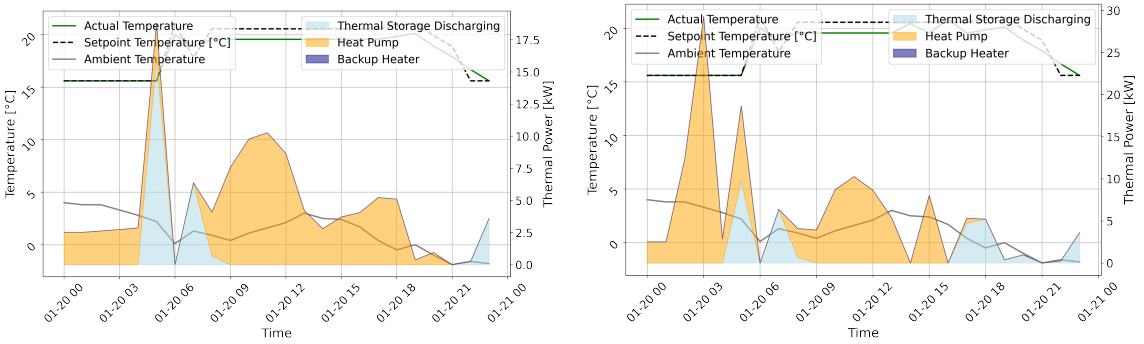
Figure 8.5a illustrates the electrical energy balance of the household, comparing demand components (heat pump, backup heater, household consumption, BESS charging) and supply components (PV production, BESS discharging) alongside the day-ahead market prices for the given day. The resulting power drawn from the grid is depicted as a solid black line. In the case of the home energy management system controlling its flexibility potential on its own, the household primarily focuses on maximizing self-consumption during the PV generation peak around noon. However, during the evening price peak, the household remains unaffected by price fluctuations, leading to most of the power being drawn from the grid. Conversely, in the aggregator flexibility scenario, the flexibility potential is utilized to minimize power draw from the grid during the evening price peak (Figure 8.5b). In this case, heat pump loads are shifted forward temporally, and the BESS is discharged during peak price times, having been charged during the low-price period at night.

We present the outcomes of the thermal building model in Figure 8.6b, which shows the desired setpoint temperature (dotted grey line), the actual indoor tem-



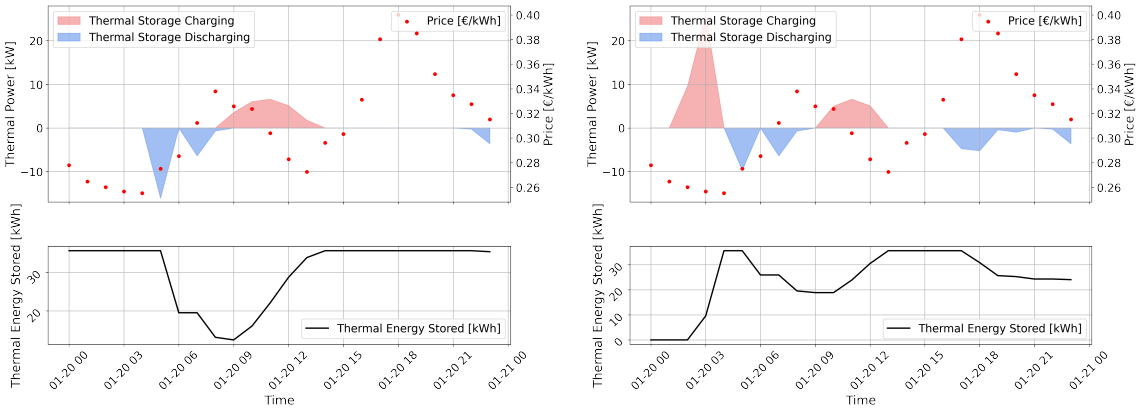
(a) Household controls flexibility (b) Aggregator controls flexibility

Figure 8.5.: Electrical energy balance.



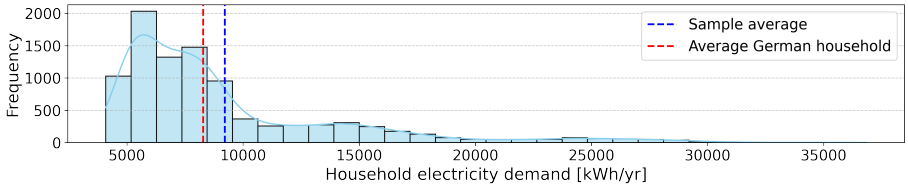
(a) Household controls flexibility (b) Aggregator controls flexibility

Figure 8.6.: Thermal model

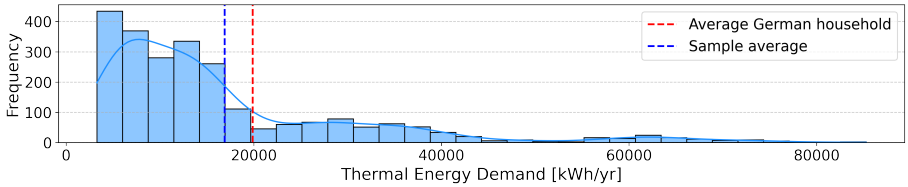


(a) Household controls flexibility (b) Aggregator controls flexibility

Figure 8.7.: Thermal storage operation



(a) Household electricity consumption



(b) Thermal energy Demand

Figure 8.8.: Histogram of yearly household electricity consumption and thermal demand of the whole sample

perature (green line), the ambient temperature (solid grey line), and the thermal energy provided by the heat pump and the discharge of the thermal storage. In both scenarios, we observe that the heat pump operates in the morning to accommodate a setpoint temperature increase. Later in the day, the heat pump activates earlier than in the case where the aggregator has no control over the household flexibility, and the thermal storage is discharged to mitigate the evening price peak (Figure 8.6b). The granted 1°C setpoint flexibility is employed to pre-heat the building during low-price periods, allowing the indoor temperature to decline during peak-price hours, effectively using the building as thermal storage.

Figure 8.7b also displays the operation of the thermal storage. We can observe that in the dynamic case, the thermal storage is charged during the night with lower prices.

We proceed to evaluate the household electricity and thermal energy demand across the sample obtained after 9,404 iterations of the Monte Carlo simulation as depicted in Figure 8.8. We have analyzed the convergence of the Monte Carlo simulation according to the Central Limit Theorem, as described in Section 8.2.2. After 2,805 runs, convergence based on a 99% confidence interval has been reached, as depicted in Figure A.3 in the Supplementary Material. These results are compared with German average values. We assume a benchmark average household electricity consumption of 3,383 kWh, based on official German statistics from Destatis (2024c)

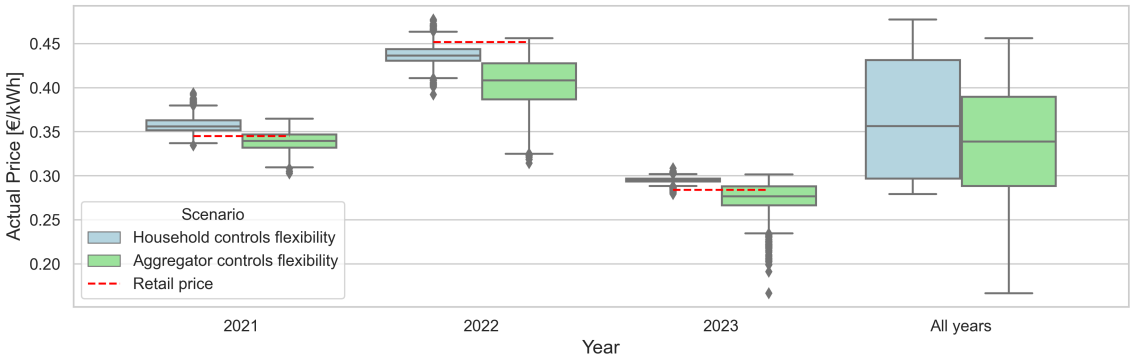
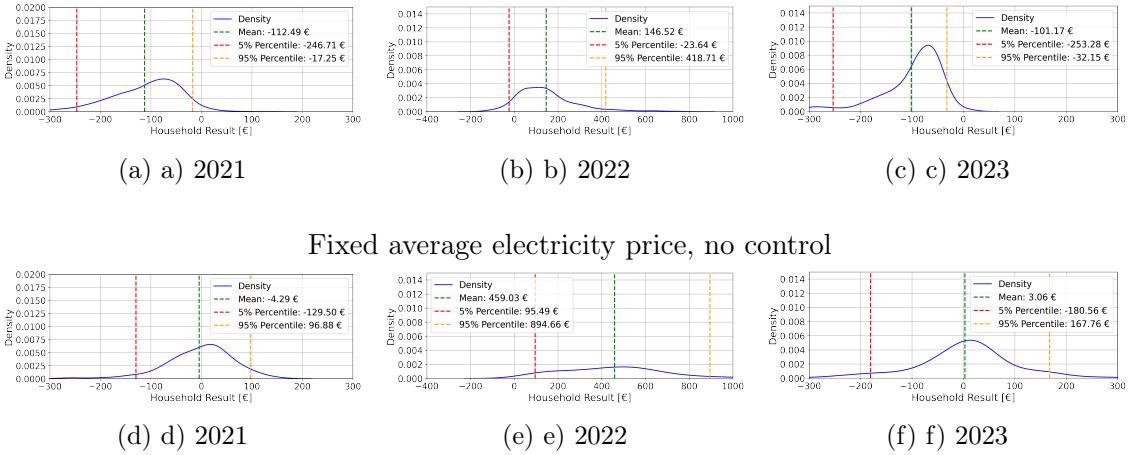


Figure 8.9.: Costs of households per unit of consumed electricity

and an average annual heat pump electrical consumption of 34.5 kWh/m^2, based on manufacturer estimates (Bosch Home Comfort, 2024). For thermal energy demand, we assume an average of 130 W/m^2, as reported in Destatis (2024b), which is then multiplied by the average area of the investigated households (153 m^2). The average household electricity consumption in the sample is close to the calculated German average value. Also the average thermal energy demand aligns closely with the German average. For both household electricity consumption and thermal energy demand, we observe long-tailed distributions with significant outliers. These outliers are predominantly associated with poorly insulated buildings, leading to disproportionately high heat demand, consistent with findings from existing research (Alabid et al., 2022).

The analysis conducted in this section evaluates the accuracy of the simulation framework and modelling of the home energy management system by comparing the resulting electricity demand to national statistics, ensuring the model's reliability, which is required for the design of robust price guarantees.



(a) a) 2021 (b) b) 2022 (c) c) 2023

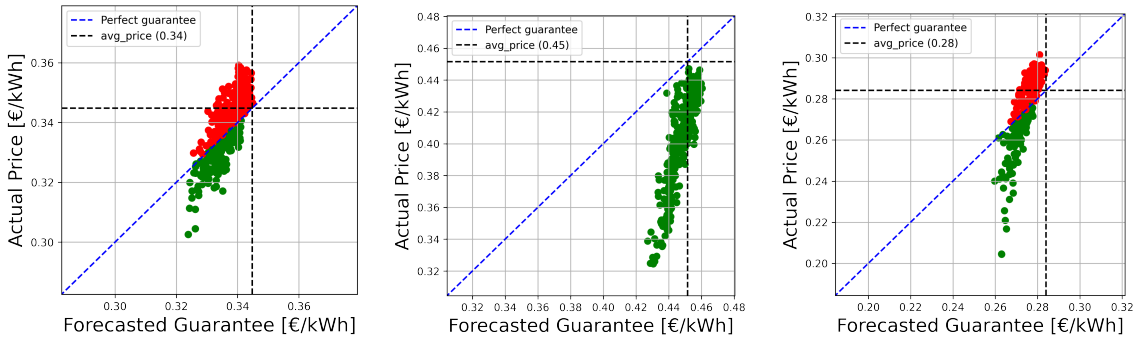
Fixed average electricity price, no control

(d) d) 2021 (e) e) 2022 (f) f) 2023

Individual guarantees, with control

Figure 8.11.: Distributions of household results from the aggregator's perspective

8.4.2 Value of aggregator control over household flexibility



(a) 2021 (b) 2022 (c) 2023

Figure 8.10.: Comparison of guaranteed and realized prices per household

In this section, we evaluate the benefits of the aggregator managing household flexibility by comparing the average electricity prices achieved under two scenarios: one where the home energy management system operates based on constant electricity prices, and another where the aggregator controls flexibility using dynamic market prices. We note that the "aggregator controls flexibility" scenario is equivalent to one where the household opts for a dynamic tariff-oriented control on its own. However, considering the aversion of households towards dynamic tariffs, we denote this case in the following consistently as "aggregator controls flexibility" case.

In Figure 8.9 (and in Table A.5 in the Supplementary Material), the distribution

of realized yearly electricity prices over all households in the sample is depicted, demonstrating the clear economic benefits of the aggregator managing household flexibility. Across all years, the aggregator's control consistently leads to lower mean average electricity prices compared to a scenario where the home energy management system relies on constant prices and is optimized to increase self-generated PV power consumption. The most significant relative reduction occurs in 2022, with an 8.48% (3.4ct) lower mean electricity price, reflecting the value of adapting to dynamic market prices during a volatile period.

While the absolute reductions in prices vary by year, they consistently translate into meaningful percentage decreases, emphasizing the aggregator's ability to leverage flexibility to optimize costs. On average, over all years analyzed, the aggregator reduces electricity prices by 7.36% (2.5ct), underlining the consistent value of dynamic control across varying market conditions.

In addition, we observe that under aggregator control, the majority of households achieved electricity prices below the volume-weighted average electricity price, which can be interpreted as a competitive retail rate. Specifically, 78.40% of households realized a lower price per unit of electricity consumed through dynamic control, compared to only 32.67% in the baseline scenario.

While we show the economic benefit of dynamic tariff-oriented operation of household flexibility potentials, the observation has been made ex-post. However, since we want to reduce the price risk for households while controlling the risk for aggregators, we aim to suggest household-specific cost guarantees even before we know the exact realizations of prices, weather and behaviour, which we proceed within the following section.

8.4.3 Price guarantee risk evaluation

To overcome the loss aversion of households, we suggest that aggregators offer guarantees. This section is dedicated to evaluating the associated risks of these guarantees for aggregators.

We use the previously described Monte Carlo sample to train our quantile regression-based guarantee prediction model, which is then used to forecast guarantees on a household basis for the previously unseen years 2021, 2022 and 2023.

We calculate household results based on the approach described in Section 8.2.4. We assume that a certain electricity price has been granted to the household (either as a usual flat household tariff or in the context of a guaranteed lower price in exchange for the right to control the household's flexibility potential). Then, we calculate costs (or profits) that occur on a household basis for the aggregator. We analyze the distributions of household results (the difference between costs that were guaranteed and the ones that actually accrued) to evaluate the associated risks.

Figure 8.10 illustrates the household-level guarantees for each year of the study. In an ideal scenario with perfect predictions, these guarantees would align precisely with the actual achievable electricity prices, as represented by the blue angle bisector. However, due to the inherent uncertainties in thermostat setpoints, weather, irradiance, PV output, and realized electricity prices, deviations between the predicted guarantees and the actual prices are evident. This discrepancy is especially noticeable in 2022, where the forecasted guarantees consistently exceeded the actual household electricity costs. The primary reason for this is the price dynamics shown in Figure 8.3. During 2022, there were significant price spikes in the summer months. However, the guarantee prediction algorithm is based on the average annual electricity price. Since the simulated households had PV production during the summer, the effect of these price spikes was alleviated, leading to an overestimation of the guarantees for that year.

Overall, the quantile regression demonstrates an ability to capture the directional relationship between individual household-level guarantees and realized costs. Households with lower guarantees generally exhibit lower realized prices, while those with higher guarantees tend to align with higher realized costs. This finding is significant, as it indicates that even under the inherent uncertainty of weather patterns, price dynamics, and individual heating behavior, it is possible to formulate meaningful guarantees using only household characteristics and aggregated price and weather statistics. Furthermore, as already highlighted in Figure 8.9, most households achieve actual electricity costs below the volume-weighted average electricity price, commonly viewed as a competitive retail rate. This reinforces the earlier observation that these households would have realized financial benefits by adopting a dynamic tariff or guarantee structure.

However, while the majority of households benefit from guarantees and the dy-

namic control of their flexibility potential, some households experience guarantees and realized prices that exceed the competitive retail rate. This can occur, for example, when a household's heating demand often aligns with price peaks or when the building's limited thermal inertia restricts its thermal flexibility. The way the operator manages these households impacts the financial outcome of the portfolio. For example, offering these households a rate lower than the retail price could result in financial losses, while rejecting them (or offering them rates higher than the retail price) could improve the performance.

In Figure 8.11 and Table 8.2, the resulting distributions of profit (or loss) per household from the aggregator's perspective are depicted for the benchmark- and guarantee-case. A desirable distribution of household results is characterized by a mean close to zero, indicating that the aggregator has accurately predicted guarantee values and has a balanced portfolio (Ackermann et al., 2022). A mean household result close to zero ensures that aggregators neither incur excessive losses nor achieve disproportionate gains. A mean close to zero also suggests that the guarantees align well with actual household costs, reflecting the algorithm's capacity to account for diverse variables such as weather fluctuations, household consumption patterns, and market price dynamics. Additionally, we assess the width of the distribution.

In two of the three years analyzed, the mean household results are significantly closer to zero when the aggregator provides individual guarantees and manages the flexibility potential. This demonstrates that offering guarantees leads to a more accurate alignment with actual household costs while maintaining a balanced portfolio, where losses from some households are offset by profits from others. However, 2022 stands out as an outlier again due to summer price spikes and overly conservative price predictions, resulting in strongly positive household results for the aggregator in that year.

In all three years, household-level guarantees resulted in higher mean outcomes for the aggregator compared to scenarios without control over flexibility potential. This suggests that guarantee contracts can provide mutual benefits: households gain access to lower prices than the retail rate without compromising comfort, while aggregators enhance their financial performance. However, this improvement comes with a trade-off in the form of a wider confidence interval, reflecting greater variability in household results. For example, in 2022, the baseline case showed a 90% confidence

interval ranging from -23.64€ to 418.71€, whereas for the individual guarantees, the interval extended from 95.49€ to 894.66€.

Case	Target Year	Mean	Interval Width
Fixed average electricity price, no control	2021	-112.49	229.46
Individual guarantees, with control	2021	-4.29	226.37
Fixed average electricity price, no control	2022	146.52	442.36
Individual guarantees, with control	2022	459.03	799.16
Fixed average electricity price, no control	2023	-101.17	221.13
Individual guarantees, with control	2023	3.06	348.33

Table 8.2.: Results per case and target year

8.4.4 Feature importance of input factors

While the economic value of aggregators managing household flexibility and its financial impact on the aggregator's portfolio has been previously discussed, we now focus on the input factors influencing guarantee levels. This is done through a linear regression analysis that links household characteristics and endowments to the difference between the guarantees and the average electricity price for the respective years[6]. Interpreting the regression results provides valuable insights into the key drivers of guarantee levels and associated savings. From an operational standpoint, these insights could inform the development of a decision support system for aggregators, enabling real-time guarantee suggestions tailored to customers.

The R-squared value of 0.484 indicates that the model explains 48.4% of the variance in the relative difference between the guarantees and the average electricity price, suggesting moderate predictive power. While the model captures key drivers, half of the variance remains unexplained. This intuitively makes sense, given that the model has no information about actual household behavior, and exact weather and price curves in the respective years.

We note that all variables but the option to opt for a nighttime setback are significant on a 0.1% level, indicating that each contributes meaningfully to explaining the variance, and their inclusion strengthens the reliability of the model. This finding

[6]As detailed in Section 8.2.3, the model targets the relative differences between guarantees and the yearly average electricity price to avoid overfitting to the price variable, thereby improving its generalizability.

Dep. Variable:	Relative Difference Guaranteed to Avg. Price [%]	**R-squared:**	0.484
Model:	OLS	**Adj. R-squared:**	0.483
F-statistic:	551.0	**Prob (F-statistic):**	0.00
BIC:	2.917e+04	**AIC:**	2.911e+04

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	30.9576	1.014	30.532	0.000***	28.970	32.945
building_type	-0.1691	0.022	-7.537	0.000***	-0.213	-0.125
Modernization Status	-1.0983	0.094	-11.716	0.000***	-1.282	-0.915
E_s_max	-0.1721	0.005	-32.879	0.000***	-0.182	-0.162
E_bess_max	-1.1545	0.025	-46.742	0.000***	-1.203	-1.106
sampled_pv_size	0.1585	0.028	5.723	0.000***	0.104	0.213
weather_avg	-2.5502	0.088	-29.011	0.000***	-2.722	-2.378
flexibility	-0.6179	0.094	-6.556	0.000***	-0.803	-0.433
setpoint_option_nighttime_setback	0.3495	0.178	1.958	0.050*	-0.000	0.699

Table 8.3.: Summary of the OLS regression results

has operational implications for the aggregator, as it guides the selection of relevant customer information to request before offering household-level guarantees, ensuring a more focused and effective decision-making process.

The coefficients of the independent variables in the regression illustrate how a one-unit increase in a given characteristic translates into a percentage point change in the relative difference between the guaranteed price and the average electricity price. For example, an increase in modernization status or newer building types is associated with lower guarantee values (a greater negative relative difference to the average price)[7]. This underscores the significant role of thermal building dynamics in determining savings potential. Modernized buildings, with improved insulation and thermal properties, function as more effective thermal mass storage, enabling greater savings and lower guarantee levels.

The coefficient for PV size demonstrates a counterintuitive effect. This arises from the focus on modeling the relative difference in guarantees rather than absolute costs. While PV installations reduce overall energy consumption and household electricity costs, they do not enhance flexibility potential. Instead, they primarily reduce grid electricity usage during high feed-in periods, which often align with low

[7] We have opted to include modernization status as a continuous variable in the regression model to simplify the analysis and allow for an approximate interpretation of its influence on the relative difference between the guaranteed and average price. However, this approach assumes a linear relationship between modernization status and the target variable, which may not fully capture the distinct effects of each discrete category (1, 2, 3).

spot market prices. A similar rationale applies to the effect of the nighttime setpoint setback option (which is only significant on a 5% level). Although this option can lower heating demand and reduce electricity bills, it does not add to the household's flexibility potential and thus has limited impact on the guarantees.

8.5 Discussion

We see uncertainty as a major obstacle to the widespread adoption of dynamic tariffs: households may be discouraged by the risk of unpredictable price fluctuations, while aggregators may avoid offering guarantees due to the difficulty of accurately modeling household flexibility across electrical, thermal, and behavioral dimensions. Our study addresses these challenges by demonstrating that the market-oriented management of household flexibility, such as through an aggregator, can generate substantial savings for households. At the same time, aggregators gain from improved financial performance, reflected in favorable mean household results and a manageable, quantifiable distribution of outcomes.

Our findings align with previous studies that highlight the economic value of household flexibility potential, but we take a different approach regarding how this flexibility should be utilized in the market. While earlier research often assumes households will eventually adopt dynamic tariffs independently, we advocate for transferring the management of this flexibility to aggregators in exchange for fixed price guarantees. In this context, household flexibility potential can be viewed as a tradeable good. Our study provides valuable insights into how this tradeable good can be priced under uncertainty.

We observed that the mean household result for the aggregator was consistently higher than the baseline status quo and close to zero in two of the three years analyzed. The first observation emphasizes the financial incentives for aggregators to offer guarantees, while the second highlights the portfolio-balancing effects of aggregator control: losses incurred from some households are offset by profits from others. On an individual household level, however, such losses could discourage customers from adopting dynamic tariffs. From a market perspective, both households adopting dynamic tariffs and those opting for guarantees contribute positively by responding to market scarcity signals. For example, they help balance supply and demand by shifting energy usage away from peak demand periods to times of high

renewable energy generation and lower day-ahead prices.

While our study focuses on the operation and pricing of household flexibility potentials, further research into their integration with operational constraints and decision-making under uncertainty faced by aggregators represents a valuable avenue for future investigation. For example, one promising area is the development of pricing strategies for guarantees that account for an aggregator's existing renewable energy portfolio. This would involve understanding how the variability and predictability of renewable generation influence the pricing and risk associated with household-level guarantees. Another relevant area is the selection and optimization of long-term electricity procurement strategies, such as future contracts, in the context of a managed (already existing) portfolio. Aggregators must balance the cost and availability of these contracts against the variability of household flexibility and market conditions.

Our study aims to realistically model a distribution of households, for instance, by considering real-world distributions of buildings, PV and BESS sizes, price data, and weather data. However, we note that the generalizability of the achieved guarantees is limited. Only four cities were considered, and a uniform heat pump type and COP were assumed. Further, the PV systems are modeled with the same tilt, azimuth, and system efficiency. Nonetheless, we see the selected input parameters as a valid overall representation of households with different heating profiles, building types, and flexibility endowment. Future research could enhance the granularity of the model by incorporating additional flexibility sources, such as electric vehicles, employing more advanced thermal models (e.g., 2R2C models), and accounting for variations in PV system configurations. These extensions would provide an even more comprehensive representation of household flexibility potential and their impact on price guarantee schemes.

The methodology can be easily applied to alternative markets and regulatory scenarios. For instance, the feed-in tariffs considered in the home energy optimization problem could be replaced with the net metering scheme from the US (Schelly et al., 2017), and the distributions for the Monte Carlo simulation can be replaced with local values, weather, and price curves. Additionally, a comparative analysis of price guarantees across different international markets could provide valuable insights for future research. Such a study would explore the applicability and effectiveness of

guarantees within diverse international power systems, shedding light on how market-specific factors influence their design and implementation.

We note that for the Monte Carlo simulation, the home energy management model is optimized under perfect foresight, which we see as a fair assumption, given the theoretical character of our study. However, the practical implementation for aggregators would require an operation under the uncertainty of future setpoint profiles and household behavior. Previous literature has shown that this can be achieved well with Model Predictive Control algorithms, which can yield 63-98% of profits realized under perfect foresight (Junker et al., 2020). However, we see an investigation of the implementation of price guarantees in the field as an important area for further research. We also acknowledge that our study does not explore consumer attitudes toward these novel price guarantees. A behavioral investigation into household perceptions and acceptance of such guarantees in future research would significantly enhance the understanding and practical applicability of this concept.

8.6 Conclusion

Our study proposes a framework for enabling the adoption of dynamic electricity tariffs among households by introducing price guarantees tailored to their flexibility potential, including electrical storage systems, thermal storage systems and the thermal mass of buildings. Our methodology is designed to transfer price risks from households to aggregators, thereby addressing consumer reluctance to engage with dynamic pricing schemes due to perceived financial uncertainty.

We perform a three-stage process that formulates household-specific price guarantees based on a combination of deterministic and stochastic simulations. These guarantees account for household characteristics such as building insulation, energy storage capacities, and thermal flexibility. Through quantile regression, we demonstrated how aggregators can predict price guarantees while managing uncertainties in market prices, weather conditions, and household behavior.

Our results indicate that the proposed price guarantees can reduce households' risk while enabling aggregators to optimize the use of household flexibility potential, such as batteries and thermal storage. By leveraging this potential, aggregators can reduce exposure to peak electricity prices and operate more efficiently within the constraints of volatile energy markets. Furthermore, we identified key factors

influencing price guarantees, including building type, modernization status, and the availability of thermal and electrical storage systems.

This work contributes to the broader discussion on demand response by providing a practical solution to bridge the gap between household reluctance and the system-level benefits of dynamic pricing. The methodology can be adapted to any region by adapting the formulation of the regulation (i.e., grid charges or feed-in tariffs) and the underlying distributions within the Monte Carlo simulation. Future research could explore the behavioral response of households to price guarantees and the real-world implementation challenges faced by aggregators.

Part V.

Technological uncertainty

INTRODUCTION TO PART V

When an increasing number of households adopt dynamic pricing, this ultimately alters load patterns. As illustrated in Part I, this can have a beneficial impact on the power system as a whole by shifting loads aligned to the system's signals. However, this can also have counterintuitive effects on a local, technical level: when a high number of flexible devices react on the same price signal, this might lead to issues on the grid level.

In this part of my thesis, a multi-step simulation study alleviates the uncertainty about the technical distribution grid-level impact of a high number of households subscribed to dynamic tariffs.

CHAPTER 9

THE IMPACT OF DYNAMIC TARIFF ADOPTION ON DISTRIBUTION GRIDS

In this chapter, I analyze the impact of increasing shares of dynamic tariff adoption on distribution grid reinforcement costs. This chapter tackles the uncertainty around the technical implications of dynamic tariffs on physical power flows in the distribution grid, potentially leading to transformer or line overloading. Based on empirical data sources for household and heat pump loads, electric vehicle charging events and PV and BESS sizes, the operation of an increasing share of households subscribed to dynamic tariffs is modelled. Then, the resulting loads are applied to different German distribution grid topologies, in which power flow simulations are then conducted. Based on resulting power flows, necessary reinforcement measures can be identified. This simulation serves as the basis for the analysis of potential alternative grid charge and feed-in remuneration policies.

This chapter comprises the article: L. Semmelmann, K. Kaiser, A. Heider, K. Kircher, G. Hug, C. Weinhardt. *Analyzing the Impact of Dynamic Tariff Adoption and Regulatory Options on Distribution Grids with an Open-Source Framework*, forthcoming in Proceedings of the Sixteenth ACM International Conference on Future Energy Systems, 2025.

9.1 Introduction

The introduction of dynamic electricity tariffs is an essential step to reduce peak demand in modern power systems and, thereby, the need to run costly peak power plants (Faruqui et al., 2010). Various types of dynamic tariffs are discussed in the literature, ranging from time-of-use pricing to critical peak pricing, real-time pricing,

and other variants. Real-time pricing is considered to be the most direct implementation of dynamic tariffs since it offers customers prices directly linked to the wholesale electricity market (Faruqui et al., 2010). Dynamic tariffs are also increasingly promoted by regulators internationally. In a 2019 directive, the European Union required its member states to enable customers to adapt their consumption to market signals (European Parliament and the Council of the European Union, 2019). As a translation into national law, Germany requires utilities with more than 100,000 customers to offer dynamic tariffs from 2025 onwards (Bundesministerium für Wirtschaft und Klimaschutz, 2023). Dynamic tariffs offered by German utilities today are mostly based on the wholesale day-ahead market price (Tibber, 2023).

Although the roll-out of dynamic tariffs, such as real-time pricing, is already underway, its effects on the power system are still not entirely clear (Dallinger and Wietschel, 2012; Kühnbach et al., 2021). Studies have shown that when multiple devices are controlled by an automated optimization algorithm that reacts to a given price signal, it might cause new, unprecedented demand peaks. This phenomenon, the avalanche effect, has been discussed for electric vehicles (EVs) (Dallinger and Wietschel, 2012; Kühnbach et al., 2021) and heat pumps (HPs) (Patteeuw et al., 2016). However, previous studies on avalanche effects have neglected three important aspects. First, the impact on the distribution grid is mostly modeled by analyzing aggregated loads, thereby neglecting potential issues arising at specific locations within the distribution grid, which could lead to costly reinforcement measures. Second, previous research often considers only standard load profiles, ignoring the interplay of HPs, battery energy storage systems (BESSs), and solar photovoltaic (PV) installations. Third, although studies discuss potential regulatory options to reduce avalanche effects, a thorough comparison is still missing.

This paper addresses these research gaps by analyzing the impact of increasing real-time tariff adoption under different regulatory options. We do this in the scope of a potential 100% electrified future with EVs, HPs, PV, and BESSs installed in every household. We leverage empirical data on HP and household load, battery size, and EV usage to create realistic household profiles. Then, we formulate an optimization problem for the operation of the flexible devices at the household level, exploring different policy options concerning grid charges and PV feed-in compensation. Finally, we connect the individual household loads within various distribution

grid topologies to calculate the necessary grid reinforcement costs. We conduct our analysis for different types of days, such as days with the highest household or HP load, or the highest PV feed-in.

In summary, our study has three main contributions. First, the impact of increasing adoption of real-time pricing for residential households on distribution grid reinforcement costs, given different day types and grid topologies, is investigated. Second, the effectiveness of regulatory options for grid charges and PV feed-in remuneration are investigated in light of an increasing share of dynamic tariff adoption. Third, we provide the underlying models and empirical data open-source to enable other researchers to integrate new data sources or test other regulatory options [8].

The remainder of this work is structured as follows. In Section 9.2, we give an overview of existing studies in the field of dynamic tariff adoption and its impact on distribution grids and highlight the resulting research gaps. In Section 9.3, we describe our overall methodology and simulation framework. Section 9.4 presents the mathematical formulation of our underlying optimization problem on a household level. Section 9.5 introduces various policy options for grid charges and PV remuneration. In Section 9.6, we give an overview of the empirical data used to carry out the evaluations presented in Section 9.7. We conclude in Section 9.8.

9.2 Background

The introduction of dynamic pricing is a promising demand-side management measure that has the potential to flatten load curves, reduce electricity bills, and lower CO_2 emissions (Dutta and Mitra, 2017). Various forms of dynamic pricing have been suggested. In time-of-use pricing, the electricity price varies between on-peak and off-peak hours and is predetermined for a given period, e.g., per season. In critical peak pricing, a particularly high price is charged when the system-wide load reaches its peak, resulting in a stronger price signal during these times than time-of-use tariffs. The most direct form of dynamic pricing, however, is real-time pricing, where prices change regularly, reflecting wholesale market prices. Although considered the most effective dynamic pricing scheme, real-time pricing requires advanced technology for scheduling distributed devices and communication of price signals, as well as

[8] Our open-source framework and underlying data is published open-source on Github, see: `https://github.com/leloq/dynamic-tariffs-in-distribution-grids`.

advanced metering infrastructure (Dutta and Mitra, 2017; Faruqui et al., 2010).

However, it has been shown in literature that the required technology to adapt the operation of devices to real-time tariffs with the objective to realize energy cost savings is already available, e.g. for HPs (Schibuola et al., 2015; Finck et al., 2020; Fischer et al., 2017a), EVs (Aljohani et al., 2021; Erdinc et al., 2014) and batteries installed in combination with a PV system (Yang and Fang, 2017; Bedi et al., 2018). Furthermore, smart metering equipment can provide high-resolution electricity consumption measurements (Faruqui et al., 2010). Proposed operation methods range from online stochastic optimization methods (Bedi et al., 2018) to model predictive control (Fischer et al., 2017a) and deep reinforcement learning (Zhang et al., 2019a).

Studies have investigated how an increased share of households operated under dynamic tariffs influences the overall power system (Roozbehani et al., 2012; Stute and Klobasa, 2024). It was concluded that an increasing share of load subjected to real-time pricing increases consumers' price elasticity, potentially leading to increased volatility in the overall system. The authors call for assessing the trade-off between economic efficiency and stability of the system. Several studies have investigated this trade-off for specific cases. In Faria and Vale (2011), optimal real-time pricing curves from a utility perspective have been determined for a 33-bus distribution network with the goal of achieving peak demand reductions. In Savolainen and Svento (2012), the potential of real-time tariffs to reduce the required peak generation capacities in the Norwegian power system was demonstrated. In Arlt et al. (2024), potential cost savings for households with automated heating, ventilation, and air conditioning systems in Texas have been quantified. Switching to real-time pricing led to 30% reductions in peak system load and reduced grid investments. Although these studies underline the relevance and potential benefits of real-time pricing tariffs, they focus on limited grid topologies and certain controllable assets, neglecting the possible effects of their interplay.

A potential issue of high shares of households subscribed to real-time pricing schemes is the aforementioned avalanche effect, which can lead to new peak loads and additional stress on the grid (Dallinger and Wietschel, 2012). In Müller et al. (2022), a study that simulates the impact of different EV charging strategies in German distribution grids shows that 71% of investigated grids would have to be reinforced to handle new load peaks caused by simultaneous charging of EVs if their charging is

optimized for real-time tariffs. While the authors emphasize that signals sent from the wholesale market do not necessarily reflect the local grid situation and acknowledge potential needs for grid reinforcement, they do not discuss potential options to mitigate these effects. Important regulatory options influencing household flexibility utilization are grid charges (Hanny et al., 2022) and feed-in remuneration (Ossenbrink, 2017). A first discussion on policy design in light of the increasing adoption of dynamic tariffs was presented by Stute and Klobasa (2024). However, the work is focused on a few low-voltage grid topologies and lacks a detailed discussion of potential grid charges.

We close this research gap by analyzing a future 100% electrified grid scenario in light of potential avalanche effects connected to increasing real-time pricing adoption rates of households. Based on that, we analyze and discuss the effects of regulatory options on grid reinforcement costs on the low- and medium-voltage grid level. We publish our work open-source with an easy-to-use workflow, enabling an analysis of alternative regulation and flexibility patterns in future studies.

9.3 Methodology

This section describes our overall methodology, which consists of four steps: household preprocessing, household optimization, grid power flow, and reinforcement simulation, as depicted in Figure 9.1. Given different regulatory contexts, we aim to simulate the impact of increasing shares of households with dynamic tariffs on grid reinforcement costs. We conduct our study in a future scenario with 100% electrified heating, EVs, distributed PV generation, and home batteries. This scenario resembles the case investigated by Müller et al. (2022). Given the current political development that many European countries, for instance, Germany (Bundesverband Wärmepumpe, 2023), incentivize the installation of large numbers of HPs over the coming years and the fact that already a large fleet of PV and battery storage systems are installed in households (Peper et al., 2022), we consider this as a realistic scenario. However, our flexible open-source model allows the analysis of different scenarios, for instance, with lower shares of EVs, PV, or battery systems, thereby reflecting alternative scenarios in future studies.

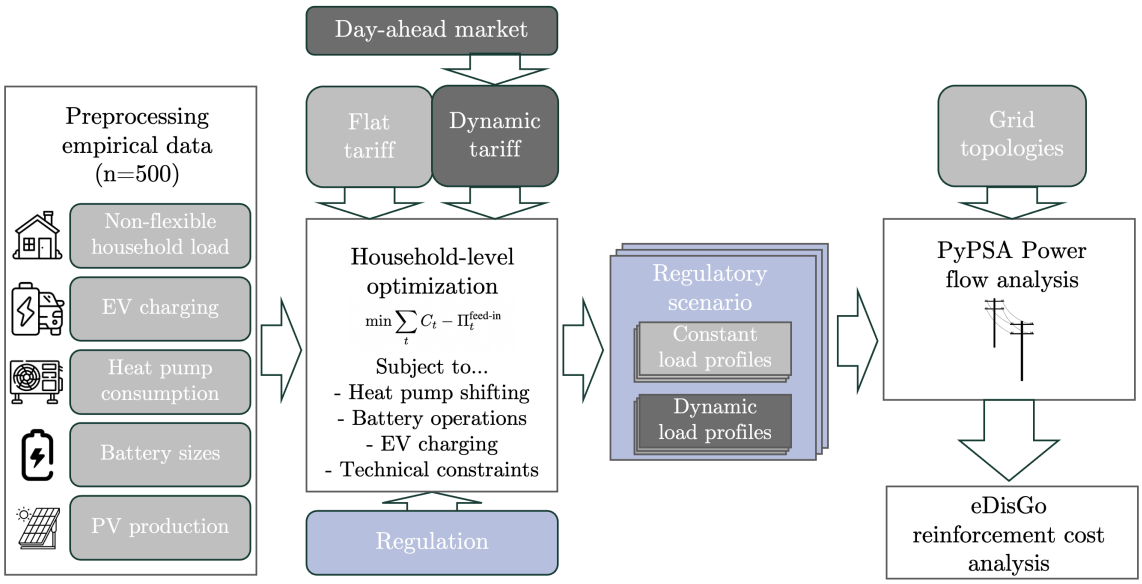


Figure 9.1.: Overview of simulation steps and overall methodology

9.3.1 Household preprocessing

In this step, we generate N residential profiles consisting of regular household loads, HP loads, PV generation, and EV profiles. Furthermore, sizing decisions for the BESS are made. The regular household and HP load profiles are based on a German dataset (Schlemminger et al., 2022), extended using the methodology described in Semmelmann et al. (2023a). For the EV load, we use empirically observed EV usage data from Norway (Sørensen et al., 2021a,b), considering charging events and plug-in and plug-out times. The household BESS are sized according to an empirical sample from 947 German households (Semmelmann et al., 2024). We match households with larger HP and household consumption with larger batteries to make the scenario more realistic. We generate the PV output power profiles based on the widely proliferated approach introduced by Pfenninger and Staffell (2016). Overall, our household preprocessing is designed to constitute a realistic sample of households based on empirical data while still exhibiting variance between households through randomly drawing HP, household, and EV profiles as well as sizes of PV and BESS installations. Section 9.6 describes our data source in further detail.

9.3.2 Household optimization

We study two different cases with regard to the operation of flexible devices in households. The first case assumes that EV and HP consumption are equal to the empirical profiles, neglecting their flexibility potential. Only the BESS is operated to increase the self-consumption of the households, thereby leading to reduced electricity drawn from the grid, as illustrated in Müller et al. (2022). Since a constant electricity tariff is present in this case, charging and discharging decisions solely depend on whether there is surplus PV generation.

In the second case, by contrast, the operation of EVs and HPs is flexible, i.e., consumption can be shifted to reduce costs further and exploit cost benefits when subscribing to pricing schemes other than a constant volumetric charge (Fischer et al., 2017a; Ortega-Vazquez, 2014). We model the HP flexibility potential based on two real-world projects (Lechwerke AG, 2021; Kaiser et al., 2023). We ensure that the daily HP consumption from the empirical observations is satisfied. However, we assume that based on Lechwerke AG (2021), the HP can be blocked during high price periods three times per day and at most for two consecutive hours without losses of comfort for inhabitants. After being unblocked, the HP has to be operated for at least two hours to maintain thermal comfort. Regarding EV charging, the energy demand, arrival time, and departure time are given for each charging session. Charging can be shifted within the specified time window of the parking event, considering also the power constraints of the underlying technical equipment. The exact mathematical formulation of the optimization model on a household level is described in Section 9.4.

Given the large share of additional demand constituted by HPs and EVs, we restrict the household flexibility potential to these loads, assuming the remaining household load to be inflexible in all scenarios (Haakana et al., 2018). We note that in future studies, also the flexibility of additional loads from appliances like dishwashers could be considered (Ampimah et al., 2018).

We assume that each household has a home energy management system, which schedules the flexible loads based on an optimization problem. We analyze the impact of an increasing share of households subscribing to dynamic tariffs by gradually increasing the number of households that use a dynamic price signal as input to

the optimization. This dynamic household tariff is based on the day-ahead price (Golmohamadi et al., 2021) as this pricing scheme is already implemented in practice (Tibber, 2023). Furthermore, the optimization is subject to varying regulatory scenarios, i.e., different combinations of PV feed-in remuneration (e.g., flat feed-in tariff, market-price oriented) and grid charge designs (e.g., volumetric, peak demand, etc.) (Aniello and Bertsch, 2023), as described in Section 9.5. The resulting load profiles are then used as input to the power flow computations.

9.3.3 Grid power flow and reinforcement simulation

Every investigated low-voltage grid topology has M nodes where households are connected. Due to the computational complexity of the previously introduced household-level optimization problem, only N households are simulated. To scale to a number of M households in every investigated grid topology, M households are randomly sampled with replacement from the set of N simulated households. For comparability purposes, we keep the same households connected to the same nodes for each investigated regulatory framework and dynamic tariff penetration rate.

For each scenario and investigated grid topology, the AC power flow computation is carried out in PyPSA (Brown et al., 2017; Heider et al., 2023; Semmelmann et al., 2023c). PyPSA is an open-source tool that uses power supply and demand time series to calculate voltages and power flows on lines and transformers. PyPSA uses the Newton-Raphson algorithm to numerically solve the nonlinear algebraic equations that govern AC power flow.

In a consecutive step, required grid reinforcement measures and their costs are calculated through the open-source framework eDisGo (Heider et al., 2023). The resulting branch currents and bus voltages from the PyPSA simulation are compared to predefined limits, i.e. 100 % for component loading and $\pm 10\,\%$[9] deviation from the nominal voltage. If any of these bounds are violated, eDisGo iteratively installs new components or splits feeders until all grid issues are resolved. As a last step, the total cost of grid reinforcement is calculated using standard costs for the installed components. We run five simulations per grid topology to account for statistical deviations in the placement of different household loads.

[9] There are additional voltage-level-specific bounds for the voltage drops and rises. These are further detailed in Reiner Lemoine Institut gGmbH (2021).

9.3.4 Assumptions

The modeling framework described above entails five key assumptions, which we explain below.

Modeling households as price takers: We assume the households to be price takers, i.e., their consumption does not impact wholesale market electricity prices. This is a strong assumption, especially in the case that 100% of households are subscribed to a real-time price. However, given that households do not directly participate in wholesale electricity markets, this is a reasonable assumption if the real-time prices incorporate the potential shift of the loads or if the inflexible load is still the dominant part of the load. Furthermore, our flexible open-source model could be integrated with a power system model that evaluates the impact of household decisions on wholesale price curves.

Same operation strategies for every household: Although we model a wide variety of households following distinct empirical electricity consumption patterns, the optimization strategy that schedules flexible load is the same for each household. We note that in practice, there might be a variety of home energy management systems following different strategies, which would ultimately lead to different household load profiles. However, we argue that our proposed optimization algorithm yields a good approximation for the best possible scheduling of flexible devices, following the formulations in Müller et al. (2022); Mulleriyawage and Shen (2020); Aniello and Bertsch (2023), and hence delivers a realistic approximation of household behavior under real-time pricing.

Focus on a 100% electrified household scenario: While we analyze the impact of varying HP, EV, PV, and BESS adoption rates on peak loads in a sensitivity analysis, our primary focus is on a fully electrified scenario where 100% of households are equipped with HPs, EVs, PV, and BESS installations. Given the rapidly increasing installation numbers of these technologies (International Energy Agency (IEA), 2024a; Kühnbach et al., 2021; Semmelmann et al., 2024), we see this as a potential future scenario. In addition, several related studies investigate a comparable scenario with full or high electrification of heating and transportation (Rinaldi et al., 2021; Müller et al., 2022; Ruhnau et al., 2019a). Moreover, we focus on residential households, neglecting potential dynamic tariffs and regulatory perspec-

tives for industrial and commercial customers. We model industrial and commercial customers in the distribution grids with constant consumption not influenced by dynamic tariffs. Our framework theoretically could also implement these customer types' dynamic behavior and scheduling. However, this would go beyond the scope of this study.

Perfect foresight: We assume perfect foresight of prices and energy demand in the household optimization model. Given our grid and policy analysis focus, we argue that this is a fair assumption, which is also common in related studies (Kühnbach et al., 2021; Aniello and Bertsch, 2023). In reality, the scheduling of devices and the utilization of flexibility potentials would need to be implemented through dedicated control algorithms, which do not have the luxury of perfect foresight (Fischer et al., 2017a; Hubert and Grijalva, 2012).

Time-series-based grid reinforcement analysis: We use time-series-based simulations to estimate the grid reinforcement costs. Traditionally, estimating grid reinforcement needs follows a more conservative approach, namely calculating only worst-case snapshots (see e.g. Gupta et al. (2021)). For these, values for the simultaneity of different load and feed-in technologies are assumed, and calculations are run for worst-case situations with high load and high feed-in. Compared to a time-series-based analysis, the worst-case simulations result in higher costs and more robust grids since they indirectly assume that all peaks occur simultaneously even though, in reality, the peaks likely occur at different times and, therefore, may require less reinforcement. In our investigations, we only assess strictly necessary grid reinforcement and investigate the influence of dynamic tariffs on these requirements. Furthermore, we assume that with increasing observability in the grids, time-series-based reinforcement calculations will become more realistic, and future grid planning will move more in this direction. Moreover, policymakers have begun advocating for time-series-based grid reinforcement analysis (Energy Systems Integration Group, 2024). Therefore, time-series-based grid reinforcement is used in the following investigations. Although conducting a worst-case snapshot reinforcement analysis lies beyond the scope of this study, we consider it a valuable avenue for future research.

9.4 Mathematical formulation

This section presents the mathematical formulation of the optimization problem for scheduling HPs, EV charging, and battery storage systems in the investigated households. This optimization problem is separately applied to every household within the set of all simulated N households. We base our optimization model and the necessary constraints on several studies covering operation strategies in households with batteries, PV installations, EVs, and HPs, namely Mulleriyawage and Shen (2020); Aniello and Bertsch (2023); Fischer et al. (2017b). From a household perspective, the energy costs C_t minus the feed-in revenues $\Pi_t^{\text{feed-in}}$ are to be minimized over the simulation period with T time steps, resulting in the following objective function:

$$\min \sum_{t=1}^{T} \left(C_t - \Pi_t^{\text{feed-in}} \right) \tag{9.1}$$

The components in this expression can be computed by

$$C_t = E_t^{\text{total}} \cdot p_t^{\text{tariff}}, \qquad \forall t \tag{9.2}$$

$$\Pi_t^{\text{feed-in}} = E_t^{\text{PV,feed-in}} \cdot p_t^{\text{feed-in}}, \qquad \forall t \tag{9.3}$$

$$E_t^{\text{PV}} = E_t^{\text{PV,feed-in}} + E_t^{\text{PV,Internal}}, \qquad \forall t \tag{9.4}$$

$$E_t^{\text{total}} = E_t^{\text{HH}} + E_t^{\text{EV}} + E_t^{\text{HP}} + E_t^{\text{BESS,Charge}} - E_t^{\text{BESS,Discharge}} - E_t^{\text{PV,Internal}}, \qquad \forall t \tag{9.5}$$

$$E_t^{\text{total}} \cdot E_t^{\text{PV,feed-in}} = 0, \qquad \forall t \tag{9.6}$$

$$E_t^{\text{total}}, E_t^{\text{PV,feed-in}}, E_t^{\text{PV,Internal}}, E_t^{\text{EV}}, E_t^{\text{HP}}, E_t^{\text{BESS,Charge}}, E_t^{\text{BESS,Discharge}} \geq 0, \qquad \forall t \tag{9.7}$$

where the costs C_t are calculated by multiplying the respective net household consumption E_t^{total} with the electricity tariff p_t^{tariff} at time t, while the feed-in revenues $\Pi_t^{\text{feed-in}}$ are the product of the feed-in energy $E_t^{\text{PV,feed-in}}$ and the feed-in remuneration $p_t^{\text{feed-in}}$ at time t. The values for p_t^{tariff} and $p_t^{\text{feed-in}}$ are time-variable or constant, depending on the respective policies in place. In (9.4), the PV production E_t^{PV} is split up into two components, namely $E_t^{\text{PV,Internal}}$ directly consumed by the household and

$E_t^{\text{PV,feed-in}}$ fed into the grid. Constraint (9.5) defines the overall net energy demand of the household E_t^{total} as the sum of the non-flexible component E_t^{HH}, the consumption for EV charging E_t^{EV}, the HP consumption E_t^{HP}, and the energy charged to the BESS $E_t^{\text{BESS,Charge}}$, minus the energy discharged from the BESS $E_t^{\text{BESS,Discharge}}$ and the internally used PV production $E_t^{\text{PV,Internal}}$. Constraint (9.6) prevents simultaneous PV feed-in and power draw from the grid, and (9.7) enforces positive energy values.

BESS: The necessary constraints for the BESS are given as follows:

$$E_t^{\text{BESS,Charge}} \leq P^{\text{BESS,Max}} \cdot \Delta t, \qquad \forall t \tag{9.8}$$

$$E_t^{\text{BESS,Discharge}} \leq P^{\text{BESS,Max}} \cdot \Delta t, \qquad \forall t \tag{9.9}$$

$$E_t^{\text{BESS,Charge}} \leq E_t^{\text{PV,Internal}}, \qquad \forall t \tag{9.10}$$

$$E_t^{\text{BESS,Charge}} \cdot E_t^{\text{BESS,Discharge}} = 0, \qquad \forall t \tag{9.11}$$

$$E_t^{\text{BESS,SOE}} = E_{t-1}^{\text{BESS,SOE}} + \eta \cdot E_t^{\text{BESS,Charge}} - \frac{1}{\eta} \cdot E_t^{\text{BESS,Discharge}}, \qquad \forall t \tag{9.12}$$

$$E_0^{\text{BESS,SOE}} = E^{\text{BESS,min}} \tag{9.13}$$

$$E^{\text{BESS,min}} \leq E_t^{\text{BESS,SOE}} \leq E^{\text{BESS,max}}, \qquad \forall t \tag{9.14}$$

The maximum charging/discharging rate $P^{\text{BESS,Max}}$ limits the energy that can be charged and discharged during one time step with duration Δt, as captured in (9.8) and (9.9). We note that in practice, the rated power could vary for the charging and discharging case. However, we choose the same value for simplicity. Furthermore, the battery charge $E_t^{\text{BESS,Charge}}$ is bounded by the internally used PV production $E_t^{\text{PV,Internal}}$, as defined in (9.10). Thereby, charging the BESS from the grid is prohibited, as is the case in Germany (Bundesministerium der Justiz und für Verbraucherschutz, 2023), as well as indicated in related literature analyzing dynamic tariff impacts (Parra and Patel, 2016). In (9.11), we enforce mutual exclusivity of charging and discharging operations. The state of energy of the battery $E_t^{\text{BESS,SOE}}$ is tracked in (9.12) by integrating the charging/discharging decisions and accounting for the BESS efficiency η. Constraint (9.13) specifies that the initial state of energy of the battery $E_0^{\text{BESS,SOE}}$ at $t = 0$ is set to the minimum allowed charge level $E^{\text{BESS,min}}$. Finally, (9.14) keeps the BESS state of energy within the capacity limits

$E^{\text{BESS,min}}$ and $E^{\text{BESS,max}}$.

Heat pump: We model the flexibility potential of the HP by assuming that it can be blocked for some hours of the day with a negligible loss of comfort for the occupants (Lechwerke AG, 2021; Kaiser et al., 2023; Fischer et al., 2017b; Semmelmann et al., 2023c). Hence, the optimized HP consumption time series per household corresponds to the empirical observations with parts of it shifted within the day. This results in reduced computational complexity compared to detailed thermal HP models such as implemented in Crawley et al. (2001); Pergantis et al. (2024) with a loss of accuracy acceptable for the purpose of our studies. The constraints regarding the blocking decisions are defined as follows:

$$E_t^{\text{HP}} \geq \left(1 - \chi_t^{\text{HP,Block}}\right) \cdot E_t^{\text{HP,Empirical}}, \qquad \forall t \tag{9.15}$$

$$E_t^{\text{HP}} \leq \left(1 - \chi_t^{\text{HP,Block}}\right) \cdot P^{\text{HP,Max}} \cdot \Delta t, \qquad \forall t \tag{9.16}$$

$$\chi_t^{\text{HP,Switch}} \geq \chi_{t-1}^{\text{HP,Block}} - \chi_t^{\text{HP,Block}}, \qquad t \in \{2, \ldots, T\} \tag{9.17}$$

$$\sum_{i=t}^{t+23} \chi_i^{\text{HP,Switch}} \leq 3, \qquad t \in \{1, \ldots, T-23\} \tag{9.18}$$

$$\chi_t^{\text{HP,Block}} + \chi_{t+1}^{\text{HP,Block}} \leq 2 \cdot \left(1 - \chi_t^{\text{HP,Switch}}\right), \qquad t \in \{1, \ldots, T-1\} \tag{9.19}$$

$$\chi_t^{\text{HP,Block}} + \chi_{t+1}^{\text{HP,Block}} + \chi_{t+2}^{\text{HP,Block}} \leq 2, \qquad t \in \{1, \ldots, T-2\} \tag{9.20}$$

where $\chi_t^{\text{HP,Block}}$ is a binary variable representing the blocking decision in each time step t. When $\chi_t^{\text{HP,Block}}$ equals 1, the HP is blocked and the HP energy consumption E_t^{HP} is zero, as captured in (9.15) and (9.16). When $\chi_t^{\text{HP,Block}}$ equals 0, the HP may operate, and the consumption is restricted to lie between the empirically observed consumption $E_t^{\text{HP,Empirical}}$ and the maximum consumption, which results from the maximum power $P^{\text{HP,Max}}$. Constraints (9.17) and (9.18) limit the number of blocking events per 24-hour window to three, where the binary variable $\chi_t^{\text{HP,Switch}}$ denotes the end of a blocking event. Finally, based on the real-world implementations Lechwerke AG (2021) and Kaiser et al. (2023), (9.19) ensures that the HP remains unblocked for at least two hours after a blocking event, and (9.20) ensures that the duration of a blocking event does not exceed two hours.

Additionally, we enforce that the total HP energy consumption within each six-

hour interval remains the same as in the empirical observations:

$$\sum_{t=d\cdot 24+\kappa}^{d\cdot 24+\kappa+5} E_t^{\mathrm{HP}} \geq \sum_{t=d\cdot 24+\kappa}^{d\cdot 24+\kappa+5} E_t^{\mathrm{HP,Empirical}}, \quad \forall d, \forall \kappa \tag{9.21}$$

where $d \in \{0, \ldots, \frac{T}{24} - 1\}$ denotes the day and $\kappa \in \{0, 6, 12, 18\}$ the hour of the day at which the considered six-hour interval starts. Thereby, we ensure that the heating demand is covered in a similar time frame as in the empirical observations so that there is no significant loss in occupants' comfort.

Electric vehicle: For each charging session c in the empirical data, the arrival time, departure time, and the charged energy $E_t^{\mathrm{EV,Empirical}}$ are known. Based on this information, $t_c^{\mathrm{EV,a}}$ is defined as the time step in which the EV arrives, and $t_c^{\mathrm{EV,d}}$ denotes the time step in which the EV departs. EV charging can be shifted within the given time window but must meet the empirical demand:

$$\sum_{t=t_c^{\mathrm{EV,a}}}^{t_c^{\mathrm{EV,d}}} E_t^{\mathrm{EV}} = \sum_{t=t_c^{\mathrm{EV,a}}}^{t_c^{\mathrm{EV,d}}} E_t^{\mathrm{EV,Empirical}}, \quad \forall c \tag{9.22}$$

Additionally, charging is constrained by the maximum charging power $P_t^{\mathrm{EV,Max}}$:

$$E_t^{\mathrm{EV}} \leq P^{\mathrm{EV,Max}} \cdot \rho_t^{\mathrm{EV}} \cdot \Delta t, \quad \forall t \tag{9.23}$$

where ρ_t^{EV} denotes the share of time for which the EV is plugged in during time step t. Bidirectional charging, e.g., to feed electricity back to the grid or to cover the demand of other devices, is currently not considered but may be a valuable extension in future work.

9.5 Regulatory options

Previous studies have identified grid tariffs and feed-in remuneration rates as key policy items that shape consumption patterns and scheduling decisions of flexibility potentials (Aniello and Bertsch, 2023; Kaschub et al., 2016; Fett et al., 2019).

9.5.1 Grid tariffs

Volumetric charges (status quo): Volumetric grid charges, where consumers pay a fixed charge per unit of energy consumed, are the most proliferated form of grid tariffs (Hoarau and Perez, 2019; Hanny et al., 2022). To include these charges in our model, we compute the electricity tariff p_t^{tariff} in (9.2) as the sum of a wholesale electricity price $p_t^{\text{wholesale}}$ (which can be either constant or time-varying) and a fixed grid charge p^{grid} per unit of energy consumed:

$$p_t^{\text{tariff}} = p_t^{\text{wholesale}} + p^{\text{grid}} \tag{9.24}$$

While constant volumetric charges are applied in many countries (e.g., Germany, Great Britain, or Australia), it has been shown that they can lead to an unfair allocation of costs, especially when there is a high share of PV-BESS installations in the system (Dehler et al., 2017; Schittekatte et al., 2018). Furthermore, volumetric charges tend to limit the potential of flexible devices as they do not reflect the current state of the electricity system (Bergaentzlé et al., 2019).

Peak demand charges: In this setting, consumers are charged based on their highest consumption within the billing period (Stokke et al., 2010). Thereby, households are incentivized to flatten their load curve and avoid high peak loads. From a modeling perspective, this can be implemented by setting p^{grid} in (9.24) to zero and adding the following term to the objective function (9.1):

$$\max_t \left(E_t^{\text{total}}\right) \cdot p^{\text{grid,demand}} \tag{9.25}$$

where the highest occurring energy consumption E_t^{total} is multiplied by a given demand charge $p^{\text{grid,demand}}$. Demand charges potentially yield a better alignment of price signals and system costs and could improve fairness (Hledik, 2014).

Segmented tariff: Another approach to incentivize households to flatten their load profile is a segmented tariff (Li et al., 2023b), also referred to as increasing block rate tariff (Bloch et al., 2019). In the scope of this grid tariff design, consumers pay a different volumetric charge depending on the consumption level in the given time

step. The costs C_t in the objective function (9.1) are computed as:

$$C_t = E_t^{\text{total}} \cdot p_t^{\text{wholesale}} + \sum_{s=1}^{S} E_t^s \cdot p^{\text{grid},s}, \quad \forall t \tag{9.26}$$

where S is the number of different tariff segments, and E_t^s and $p^{\text{grid},s}$ denote the energy and volumetric charge corresponding to segment s. The energy assigned to each segment cannot exceed the predetermined value $E^{s,\text{Max}}$, i.e.,

$$0 \leq E_t^s \leq E^{s,\text{Max}}, \quad \forall s, \forall t \tag{9.27}$$

and the sum over all segments must be equal to the household's net consumption:

$$\sum_{s=1}^{S} E_t^s = E_t^{\text{total}}, \quad \forall t \tag{9.28}$$

Tariff values are increasing with increasing consumption, i.e., $p^{\text{grid},1} \leq p^{\text{grid},2} \leq \cdots \leq p^{\text{grid},s}$, and consumers can reduce their costs by flattening their load profile. A potential benefit compared to peak demand charges is that there is always an incentive to keep consumption below the defined thresholds, while for demand charges, the level to which consumption is reduced may be determined by a few time steps with unavoidable high consumption.

Rotating tariffs: Finally, we introduce a novel "rotating tariff", which aims to reduce the unintended avalanche effects by assigning different time-variable volumetric grid charges to different households. In this system, some households face higher grid charges when wholesale electricity prices are low, potentially discouraging them from shifting their load to those hours. For the implementation of rotating grid tariffs, households are assigned to V different tariff groups, meaning each household n is assigned to a tariff group $v^n \in \{1, \ldots, V\}$. The grid charge $p_t^{\text{grid,rotating},n}$ for household n at time step t is determined by a time-dependent binary parameter ζ_t^v that indicates whether a higher grid charge applies to tariff group v:

$$p_t^{\text{grid,rotating},n} = \begin{cases} V \cdot p^{\text{grid}} & \text{if } \zeta_t^{v^n} = 1 \\ \frac{p^{\text{grid}}}{V} & \text{if } \zeta_t^{v^n} = 0 \end{cases} \tag{9.29}$$

Thus, when $\zeta_t^{v^n}$ is 1, the grid charge is V times higher than the baseline volumetric charge p^{grid}. For the remaining time steps, the grid charge p^{grid} is divided by V.

The point of time of high grid charges varies depending on which tariff group the respective household is mapped to. We analyze a simple rotating grid charge with two tariff groups ($V = 2$) that have alternating high grid charges. The advantage of this approach lies in its relatively simple setup. However, the optimal implementation of such a tariff design, its impact on fairness, and its cost-reflectivity have to be discussed in future studies.

9.5.2 Feed-in remuneration

In addition to the different grid charges, we investigate two different regulatory options for the remuneration of PV production fed into the grid. Feed-in remuneration design is seen as a major driver of investments in renewable energy infrastructure and the way the flexibility potentials of households are used (Couture and Gagnon, 2010; Ossenbrink, 2017).

Constant feed-in tariffs (status-quo): Many countries (e.g., Germany, Japan, UK) have introduced feed-in tariffs that compensate every unit of PV production fed into the grid with a constant rate $p^{\text{feed-in}}$ (Fett et al., 2019; Aniello and Bertsch, 2023; Dijkgraaf et al., 2018). While constant feed-in tariffs have been identified as an effective policy tool to increase PV penetration, they also proved to be expensive and not cost-effective (Poponi et al., 2021). Furthermore, it has been shown that constant feed-in tariffs lead to a feed-in profile that is not necessarily aligned with the power system's needs (Klein et al., 2019).

Dynamic feed-in tariffs: A potential alternative and widely discussed measure to align PV feed-in with market signals is the implementation of a dynamic feed-in remuneration. In this case, the compensation of excess feed-in is based on the wholesale market price (Aniello and Bertsch, 2023; Klein et al., 2019). An exemplary implementation of dynamic feed-in tariffs can be found in the state of Victoria, Australia, where retailers can offer their customers a time-varying remuneration for electricity exported to the grid Essential Services Commission (2024). We model this approach in our study by setting the feed-in compensation to the respective market price $p_t^{\text{feed-in}} = p_t^{\text{wholesale}}$ for each t, based on Aniello and Bertsch (2023).

The scenarios derived from the potential policy compositions consisting of the described grid charges and feed-in remuneration are summarized in Table A.2 in the Appendix. We investigate the interplay of these scenarios and varying dynamic pricing adoption rates.

9.6 Data

This section describes the empirical data sources used in our study. By leveraging empirical data instead of purely simulated profiles, we aim to create a realistic representation of residential loads in distribution grids (Aniello and Bertsch, 2023; Müller et al., 2022). We put together a dataset of 500 single-family households in Hamelin, Germany, for the year 2019. The dataset includes time series for the inflexible load, HP consumption, EV charging, and PV generation of every household in hourly resolution. Furthermore, the dataset includes a specific BESS capacity and power for every household. The various data sources are introduced in the following.

Heat pump and inflexible household load: Our study is based on household and HP load profiles from Schlemminger et al. (2022). The dataset consists of 38 single-family households in Hamelin, Germany, and includes hourly energy consumption data from mid-2018 to the end of 2020. The buildings from the dataset have water-water HPs with 7.4-11.3 kW thermal power and a 6kW backup heating rod. We follow the approach from Semmelmann et al. (2023a) to increase the HP load profile variance, leveraging a k-means clustering and random sampling method. The method clusters temperature measurements of the empirical dataset from Schlemminger et al. (2022) and maps the observed HP load profiles on these days to the clusters. Then, based on the temperature measurements observed in Hamelin in 2019, daily HP and household load profiles are randomly drawn from the clusters. For a more detailed description of the method, we refer to Semmelmann et al. (2023a).

Electric vehicles: The EV charging time series are based on the empirical dataset in Sørensen et al. (2021b). The data was collected from a housing cooperative in Norway between December 2018 and January 2020 and includes the plug-in time, plug-out time, and the charged energy for a total of 6878 charging sessions of 97 EV users. We only consider the charging sessions of users with a private charging point, as this reflects the setting of single-family homes as investigated in our study, and there are significant differences between the charging habits of EV users with a

private charging point and EV users with shared charging points (Sørensen et al., 2021a). As not all users in the dataset are registered from the start of the specified time period, we only consider users for which the first charging session is no later than March 2019. With all of these exclusions, 13 EV charging time series with 2359 charging sessions remain. For each of the 500 households, one of these 13 EV charging time series is randomly assigned, and the time series is randomly shifted by $k \in [-4, 4]$ weeks to increase variance between households while keeping the usage patterns on different day types intact.

Figure A.1 in the Appendix depicts the households' resulting weekly aggregated energy consumption. It shows that the inflexible household load and EV load are rather constant throughout the year, while there are HP-induced peak loads during the winter months.

BESS sizing: For the sizing of household battery storage systems, we draw from an empirical sample of systems installed in Germany, adhering to the size distribution. Within the representative sample, the systems either have a capacity of 2.5kWh (6.7%), 5kWh (37.2%), 7.5kWh (31.5%), or 10kWh (24.6%). The average power-to-energy ratio, which sets the BESS power rating in relation to energy capacity, is 0.41. We assign larger batteries to households with a larger yearly household and HP energy consumption.

PV generation: We create an hourly PV generation profile for a 1kW installation in Hamelin in 2019, based on the open-source tool renewables.ninja (Pfenninger and Staffell, 2016). We use the default settings of renewables.ninja, assuming a 35° tilt of the panels, a south orientation (180° azimuth), and a 10% system loss. Furthermore, we assume that households with larger PV installations tend to have larger battery storage systems. Therefore, we assume that the individual household's PV peak power takes the same value as the previously assigned BESS capacity, i.e., for a household with a 10kWh BESS, the PV peak power is assumed to be 10kW. The PV profile of each household results from scaling the PV profile obtained from Pfenninger and Staffell (2016) with the household's PV peak power. We note that in reality, installed PV panels would likely exhibit more diverse tilt and orientation angles, leading to slightly different generation profiles.

Wholesale market prices: Our study compares two different retail pricing schemes for household electricity consumption: a flat and a dynamic tariff. Both

are based on the hourly day-ahead price in Germany for 2019 (Bundesnetzagentur for Electricity, Gas, Telecommunications, Post and Railway, 2024), but shifted to the positive range. We set the flat tariff to the mean of the observed, shifted series. For the dynamic tariff and feed-in remuneration, we directly use the shifted series. We note that this is a rather simplistic implementation of dynamic prices. However, we argue that our approach reflects the structure of more sophisticated dynamic tariffs and is easy to comprehend. Furthermore, our open-source model enables a simple adaption of tariffs and market prices in future studies.

Grid charges and feed-in tariffs: The constant volumetric grid charges are set to 0.0722 EUR per kWh, based on actual values of the local distribution grid operator in Hamelin (Avacon Netz GmbH, 2019; Bundesnetzagentur, 2019). The demand charge is set at 67.94 EUR per kW per year, derived from the demand charge of the grid reserve capacity for the same year (Avacon Netz GmbH, 2019). Weekly or monthly peak demand charges could also be considered in further studies. For the segmented tariff, we consider three segments. The price values are defined as 0.0361, 0.0722, and 0.1444 EUR per kWh, which corresponds to the constant volumetric charge scaled by a factor of 0.5, 1, and 2, respectively. Similar to Bloch et al. (2019), the corresponding consumption limits are set to $E^{1,\mathrm{Max}} = E^{2,\mathrm{Max}} = 2\,\mathrm{kWh}$, while there is no upper limit on E_t^3. We set the fixed PV feed-in remuneration to 0.1187 EUR, as was the case in Germany for PV installations below 10kWp installed in January 2019 (Bundesnetzagentur, 2024b).

Grid topologies: We use representative German distribution grid topologies based on geo-referenced load and generation data (Amme et al., 2018). The data comprises medium voltage (MV) grids with underlying low voltage (LV) grids for the whole of Germany. These German grids were then clustered with a k-medoids approach into ten representative grids based on their installed capacities of PV, wind, HPs, and electromobility (Reiner Lemoine Institut, 2024). In total, we select ten representative grids based on the clustering process. These grids vary in their urban settings (urban, suburban, or rural). Figure A.5 in the Appendix depicts an overview of the ten analyzed grid topologies. We note that the grids vary in installed PV and wind generation, as well as residential demand. The grid topologies include highly spatially resolved data on lines, transformers, and connected load, generation, and storage units. One exemplary grid is displayed in Figure A.4 of the Appendix.

9.7 Results and discussion

In this section, we present the results of the household-level optimization for the different regulatory settings. In the second step, we randomly assign the 500 household profiles to the investigated grid topologies. Finally, we evaluate the associated reinforcement costs per scenario and discuss the implications of our results.

9.7.1 Household-level optimization

When a household switches to a dynamic retail tariff, the individual flexibility decisions and resulting load profiles change: the EV is charged during the night with low prices, HP operation during high-price hours is avoided, and battery charging decisions are altered.

Figure 9.2 displays the resulting net load profiles for all scenarios, aggregated over all 500 households. We differentiate between 100% of households with constant retail tariffs and 100% of households with dynamic retail tariffs and show the results on different specific days. These days correspond to the days with the highest aggregated inflexible household load (3rd March), EV load (18th November), HP load (1st December), and PV feed-in (13th May), respectively. While we run the household-level optimization for the whole year, we focus on investigating the reinforcement costs on these particular days to reduce the computational complexity of the simulation. Figure A.6 and Table A.3 in the Appendix depict the loads on these days without any load shifting or BESS usage.

Figure 9.2 shows that, especially on the days with the highest inflexible load, EV demand, and HP consumption, the switch from constant tariffs (top row of plots) to dynamic tariffs (middle row) significantly alters the aggregated consumption profile. In the dynamic tariff scenario, households react to the price signals from the wholesale market, for instance, by shifting EV charging to low-price morning hours. The scenarios with the volumetric grid charges consistently lead to the highest peak loads, while the alternative grid charges reduce them. This preliminary result is in line with previous studies that mention avalanche effects as a result of a high degree of dynamic tariff adoption (Dallinger and Wietschel, 2012; Kühnbach et al., 2021). We note that even on the peak inflexible load day, the aggregated consumption profile is shifted according to the price signals since a considerable HP demand is

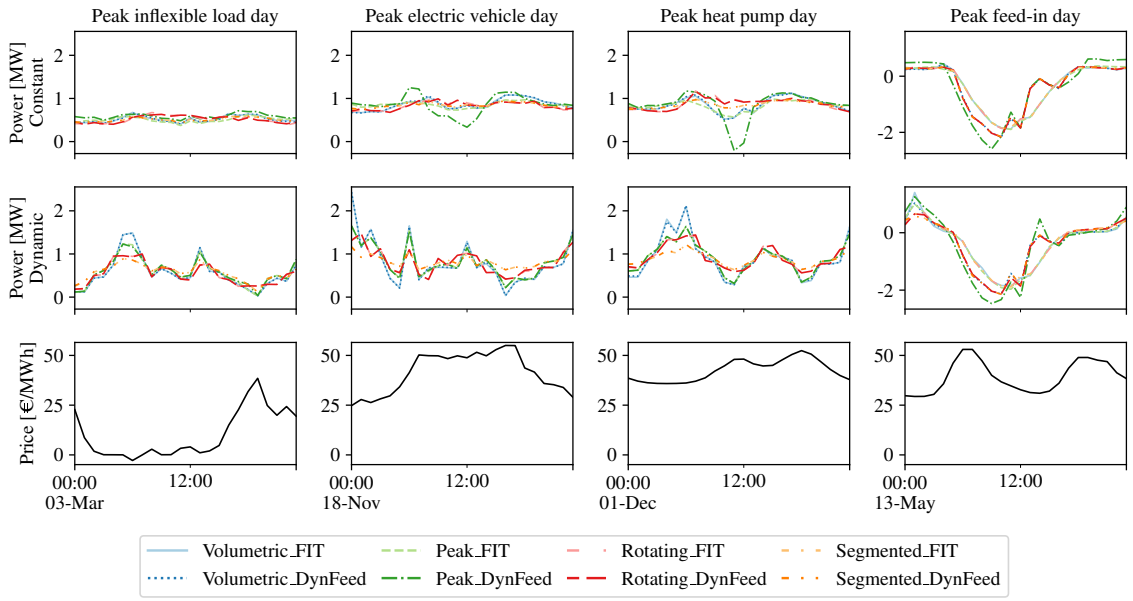


Figure 9.2.: Aggregated household load for the simulated households

present (as depicted in Figure A.6 in the Appendix)[10].

The load profiles on the peak feed-in day are largely unaffected by the level of dynamic tariff adoption but vary significantly across regulatory scenarios. Notably, scenarios with dynamic feed-in tariffs cause the peak feed-in to shift to earlier hours, coinciding with periods of higher spot market prices. This effect is particularly pronounced under dynamic feed-in tariffs with peak grid charges.

While our study focuses on a potential future scenario with 100% PV, BESS, EV, and HP adoption, we conduct a peak load sensitivity analysis for varying adoption rates in the Appendix (Figures A.2 and A.3). We note that also for lower adoption rates (e.g., only 25% of households are equipped with PV, BESS, EV, and HP installations), the conclusions remain the same: when the households switch to dynamic tariffs under the current regulatory scenario with volumetric grid charges, peak loads substantially increase. On the other hand, an exemplary alternative regulatory scenario with segmented grid charges leads to reduced dynamic tariff-induced peak load increases. We note that, in general, peak loads increase with higher PV, BESS, EV,

[10] On the HP type of day, the scenario with peak demand charges and dynamic feed-in remuneration led to a noticeable load valley at 12:00. The reason for this lies in the peak wholesale price, which then leads to an HP blocking decision over all households, enabling well-remunerated feed-in of excess PV production.

and HP adoption independent from the regulatory scenario, which is in line with the results of previous research (Muratori, 2018; Say and John, 2021; Love et al., 2017).

9.7.2 Impact on grid reinforcement costs

After the initial evaluation of aggregated load profiles, we now randomly assign the individual household profiles to the distribution grid nodes. Thereby, we simulate the impact of an increasing share of households subscribing to dynamic tariffs on grid reinforcement costs and evaluate alternative regulatory settings. Figure 9.3 shows the reinforcement costs that result from the computations for the investigated types of days and a combination of all days[11].

Status-quo regulation (volumetric charges and constant feed-in tariffs)

The results of our grid reinforcement cost analysis confirm the ones from previous studies Müller et al. (2022); Stute and Klobasa (2024): under the currently widely proliferated constant volumetric grid charges, high additional grid reinforcement costs can be expected when a large share of households switch to dynamic tariffs. These reinforcement needs are triggered by transformer and line overloading and voltage issues. Some of the investigated topologies already require reinforcement measures starting at a 25% dynamic tariff adoption rate onwards. At the same time, at 100% adoption, all grids except one have to be substantially reinforced to endure simultaneous household reactions on common price signals. An additional analysis in the Appendix (Figure A.8) shows that the previously calculated peak loads correlate with the resulting grid reinforcement costs.

Our analysis of different types of days enables a more granular analysis of the causes of the high reinforcement costs under constant volumetric grid charges, fixed feed-in tariffs, and high dynamic tariff penetration. On the day with the highest aggregated household load, even in a 100% dynamic tariff scenario, only minor reinforcement costs occur, while most grids do not even have to be reinforced. We note that the EV and HP consumption and PV feed-in on this particular day are lower compared to the other days. On the other hand, on the day with the highest

[11] In the Appendix, the detailed reinforcement costs per type day and regulatory setting are depicted in Tables A.6 to A.10, as well as an overview of occurring grid issues (i.e., voltage violations and overloading events) in Figure A.7. In addition, for the *Volumetric FIT* scenario, also a detailed breakdown of reinforcement costs per grid and penetration rate is provided in Table A.5.

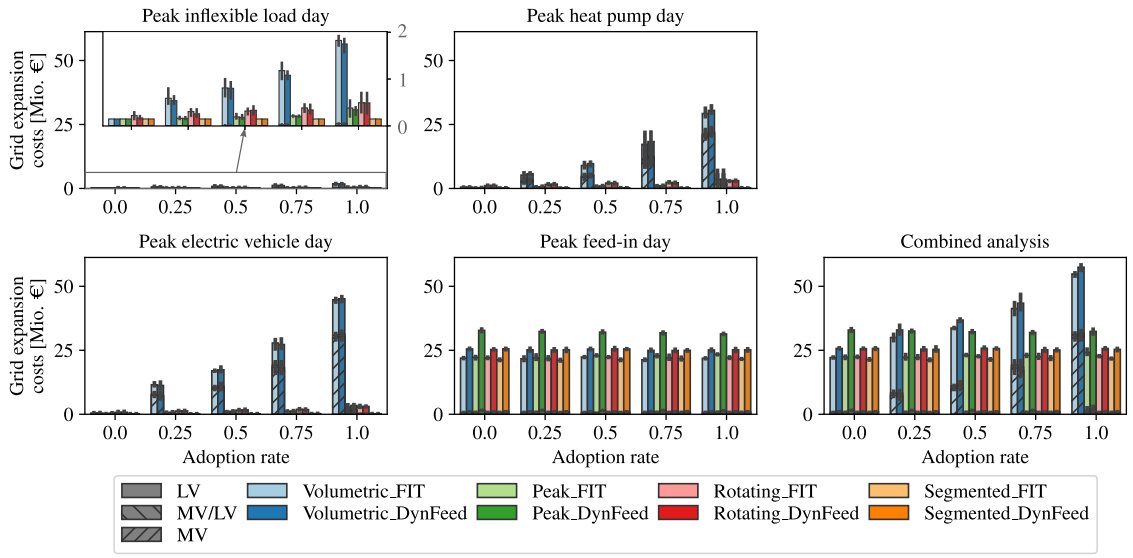


Figure 9.3.: Grid reinforcement costs across all investigated grid topologies and policy options

aggregated EV or HP consumption, the higher the share of households with dynamic tariffs, the higher the reinforcement costs and the number of grid topologies that must be reinforced. These results suggest that automated load-shifting from EVs, and to a lesser extent HPs, are the primary drivers of avalanche effects under dynamic tariffs. Most of these reinforcement needs occur in the medium-voltage grid.

On the day of the feed-in peak, there are consistent reinforcement needs independent of the dynamic tariff adoption rate. This is caused by simultaneous feed-in, which also cannot be alleviated by demand shifting from HPs or EVs. As a result, overvoltage issues in the investigated distribution grids arise (which Figure A.7 illustrates in further detail in the Appendix).

Susceptibility of grid topologies

When analyzing grid reinforcement costs under the scenario with volumetric grid charges and constant feed-in tariffs (the current regulation in Germany), we find that the impact of rising dynamic tariff adoption rates varies significantly between grids (see Table A.5 in the Appendix). For instance, the "rural-wind-1" or "suburban-balanced" grids show no increase in reinforcement costs due to dynamic tariffs. In contrast, the "rural-pv" grid experiences a substantial increase, with costs rising

from €1.3 million at 0% adoption to €10.4 million at 100% adoption—an overall increase of €9.1 million. Both grids that do not require reinforcement are relatively small, featuring short medium-voltage (MV) lines and a low number of transformers (see grid characteristics in Table A.4 in the Appendix). In contrast, the "rural-pv" grid has a relatively high number of transformers and the second-highest line length on the MV level. In addition, the grid has the lowest mean nominal transformer capacity.

A similar trend is observed when examining the correlation between reinforcement costs and grid characteristics across all topologies (see Fig. A.9 in the Appendix). The length of medium-voltage (MV) lines show the highest positive correlation, while the mean nominal power of transformers exhibits a negative correlation. It is important to note that these findings represent statistical associations rather than causal relationships. Additionally, the analysis is based on a limited sample of grids, which may affect the generalizability of the results.

Impact of alternative regulatory options

In addition to the prevalent regulatory framework featuring volumetric grid charges and constant feed-in tariffs, we have analyzed the impact of alternative tariff schemes on grid reinforcement costs. As previously discussed and shown in Figure 9.2, alternative grid charges, such as peak charges, rotating tariffs, and segmented tariffs, effectively reduced peak loads on the inflexible load, EV and HP peak days. Similarly, Figure 9.3 demonstrates that these alternative grid charge designs can mitigate the adverse effects of a high share of households adopting dynamic tariffs on grid reinforcement costs. For example, in the combined analysis, aggregated reinforcement costs across all grids in the Volumetric FIT scenario rise significantly, from €22.2 million[12] to €54.8 million—an increase of €32.6 million. In contrast, under the Segmented FIT regulation, the costs increase only marginally, from €21.4 million to €21.8 million, representing a modest rise of €0.4 million (see Table A.6 in the Appendix). The alternative grid charge designs create a monetary incentive to flatten the demand curve and distribute it over the day, reducing the number of undervoltage and overloading issues significantly, thereby reducing the associated

[12] We note that also initial grid reinforcement costs can vary across the different scenarios investigated, as regulatory options influence how household flexibility potentials are utilized, even in the absence of households subscribing to dynamic tariffs.

grid reinforcement costs.

Although alternative grid charge designs mitigate the adverse effects of increasing dynamic tariff adoption rates on grid reinforcement costs, a similar impact cannot be observed for the introduction of dynamic feed-in remuneration rates, despite their prominence in policy discussions (Aniello and Bertsch, 2023; Klein et al., 2019). On the contrary, the introduction of dynamic feed-in remuneration rates is associated with slightly higher reinforcement costs across all levels of dynamic tariff adoption. The underlying reason is illustrated in Figure 9.2: under dynamic feed-in remuneration, households collectively shift their feed-in to hours with higher prices, ultimately leading to increased grid reinforcement costs. However, this effect on reinforcement costs is relatively minor compared to the steep cost increases observed under volumetric grid charges. This is primarily because, for the households studied, excess PV production on sunny days significantly surpasses their flexibility potential, given the limitations of small battery capacities and low heating demand during the summer months. Apart from the volumetric case, the reinforcement requirements for the feed-in day exceed the measures on the other types of days in most cases. Or, in other words: when we consider alternative grid charge designs, the grid reinforcement measures required to accommodate 100% of households having PV installations are mostly sufficient to also deal with dynamic tariff-induced grid stress.

9.7.3 Discussion

Our work underlines an important result from past studies (Dallinger and Wietschel, 2012; Kühnbach et al., 2021; Stute and Klobasa, 2024): developments that are beneficial on a transmission system level, e.g., the widespread adoption of dynamic electricity tariffs, can lead to issues on the distribution grid level. Simultaneous reactions to the considered wholesale market prices trigger these issues. Our findings indicate that replacing volumetric grid charges with alternative designs that mitigate the simultaneous operation of flexible devices, such as HPs and EVs, prevents spot market-induced price signals from generating new peak loads and associated issues at the distribution grid level.

The investigated alternative grid charge designs include peak demand charges, segmented tariffs, and rotating tariffs. All three options demonstrate reduced grid

reinforcement costs with increased dynamic tariff adoption rates, compared to the volumetric grid charge design. However, their practical applicability requires critical examination. Peak demand charges, while effective in the idealized conditions of our simulation, may face challenges in real-world implementation due to the need for accurate predictions of future household demand. Reliable scheduling of flexibility potentials to reduce peak loads is essential for these charges to have a tangible impact. Without robust home energy management systems capable of reliably shifting flexibility to reduce peak loads, the effectiveness of this policy would remain uncertain in practice. Segmented tariffs achieved the lowest grid reinforcement costs at 100% dynamic tariff adoption (see Table A.6). Unlike peak demand charges, segmented tariffs do not target individual peak demands but incentivize a general flattening of the household load curve. However, the design of segmented tariffs by grid operators must carefully balance grid reinforcement cost reductions with network fee revenues. Striking this balance to achieve a Pareto improvement for grid operators presents a valuable area for further research. Rotating grid charges also succeeded in avoiding sharp increases in reinforcement costs at higher dynamic tariff adoption rates. Similar to segmented tariffs, rotating tariffs are not aimed specifically at peak load reduction but employ time-variable charges that rotate among households. While these tariffs could discourage some households from shifting load to low-price periods, they may conflict with the "non-discriminatory tariff" requirement imposed on distribution system operators (DSOs) in many countries (Stute and Klobasa, 2024). However, in certain jurisdictions, such as Finland, DSOs have greater flexibility in tariff design, potentially enabling the inclusion of rotating tariffs (Wang et al., 2023). The detailed design of rotating tariffs—such as the number of tariff groups and the differentiation between high and low tariff levels—remains an open question for future research. Furthermore, to enable dynamic tariff adoption and the suggested alternative grid charges, a high penetration rate of smart meters would be necessary, which is not yet the case in many European countries Zhou and Brown (2017).

Our study also examined an alternative policy design for feed-in tariffs, specifically dynamic, market-oriented remuneration. However, due to the significant excess feed-in from the investigated households and the misalignment between wholesale market price valleys (where feed-in is discouraged) and local peak PV production, no positive

effects on grid reinforcement costs were observed. On the contrary, dynamic feed-in tariffs resulted in higher grid reinforcement costs across all scenarios.

The chosen time-series-based approach likely underestimates the necessary reinforcement costs compared to the currently used worst-case estimations. However, it is necessary to analyze detailed time series data to accurately capture changes in simultaneity and peak loads between the different policy options. Furthermore, when more households are equipped with smart meters, the grid operators can likely access data and measurements across the grids. Grid planning might move from the worst-case approach towards using real-world data on grid utilization in this case, which aligns more with our chosen approach.

We also note that our study is focused on a perfect foresight scenario, where household loads are determined by an optimization problem. Translated into a real-world setting, a household energy management system would have to make real-time decisions under uncertainty. However, we argue that also in a real-world scenario, the results should go in the same direction since previous studies have shown that through economic model predictive controllers, 63-98% of the perfect foresight savings can be reached (Junker et al., 2020). Different providers may use different load and price forecasts, particularly in a system with a high share of volatile renewable generation. This can lead to more diverse decisions and load profiles and potentially lower load peaks. In addition, we note that the results are purely simulation-based and may neglect individual behavior and preferences. However, a German empirical study on the interaction of households with their PV/BESS energy management system showed that the observed behavior in the field resembles the outcomes of simulation-based studies (Semmelmann et al., 2024). To better understand the impact of household behavior and preferences on the resulting decisions of dynamic tariff-operated home energy management systems, future studies should also empirically investigate this.

Our study is constrained to a limited number of empirical load profiles and only considers one year of day-ahead market prices. Furthermore, we acknowledge that due to limited data availability, our analysis relied on datasets from two different geographies (Germany and Norway), which may introduce minor modeling inaccuracies. For instance, the colder climate in Norway could influence factors such as personal mobility patterns and EV battery efficiency. However, our open-source

framework facilitates the analysis of alternative datasets and more price years in future research.

Furthermore, our study focuses on energy systems with a regulatory framework that promotes the self-consumption of households' PV generation. In this regulatory setting, PV production fed to the grid is remunerated with a fixed feed-in tariff. Through the consumption of energy stored in the battery, volumetric grid charges can be avoided (Fett et al., 2019). In contrast, several states like the United States (Gregoire-Zawilski and Siddiki, 2023), Spain or Ecuador (Ordóñez et al., 2022) have net metering or net billing policies in place, where feed-in is compensated by credits, which reduce the electricity bills. While we have not investigated the interplay of an increasing dynamic tariff adoption, policy options, and net metering policies for the sake of conciseness, we see it as a worthwhile avenue for further research.

9.8 Conclusion

Our study investigates the impact of an increasing share of households with dynamic tariffs on grid reinforcement costs on the low- and medium-voltage levels. We provide a four-step open-source simulation framework that includes data preprocessing of empirically observed individual household loads, household-level optimization of shiftable loads like HPs and EVs, as well as the operation of battery storage, a PyPSA-based power flow analysis and a subsequent eDisGO-based calculation of grid reinforcement costs. We compare the status-quo policy design with volumetric grid charges and fixed feed-in tariffs with alternative policy designs. We find that the currently widely proliferated volumetric grid charges lead to simultaneous reactions of households to wholesale market price signals, leading to avalanche effects, which in turn lead to substantial reinforcement costs. We show that alternative grid charges, such as segmented or rotating ones, potentially alleviate the negative effects of high penetration of dynamic tariffs. We publish our code, data, and model as an open-source framework. Thereby, we enable researchers, policymakers, and practitioners to evaluate the impact of novel policies and technologies - in combination with an increasing share of dynamic tariffs - on the distribution grid.

Future valuable research avenues include investigating varying household settings (e.g., varying shares of PV, EV, and HP adoption), fairness issues (e.g., energy equity of grid charge options), novel technological options (e.g., vehicle-to-grid), and alter-

native regulatory options. In addition, future studies could elaborate on particular modeling aspects of the study at hand, for instance, by implementing sophisticated thermal building models or by incorporating operation under uncertainty on a household level (e.g., through a model predictive control approach).

Part VI.

Finale

CHAPTER 10

CONTRIBUTIONS AND IMPLICATIONS

Dynamic tariffs are a potential key tool for policymakers to ensure a cheaper, greener, and more efficient electricity supply by aligning market signals with customer electricity demand. This chapter evaluates the answers to the overarching research questions posed in Chapter 1 and distills the contributions of this dissertation.

The first research question (*"What are the empirically observed effects of self-consumption-promoting regulation on the operation of battery energy storage systems and their economic impact on the power system?"*) examines the system benefits of household battery storage systems in practice. The observed German systems, operated under a self-consumption optimization goal with static electricity tariffs, were evaluated based on the market outcomes of their charging and discharging operations. The findings reveal that average market profits were minimal (5€ per household per year), with some households even incurring losses, which must then be socialized among consumers.

In contrast, the findings of Research Question 2 (*"How do dynamic regulatory approaches and alternative influencing factors enhance the value of BESS for both the energy system and their owners?"*) demonstrate that alternative regulatory frameworks, such as enabling households to adjust their charging and discharging operations based on day-ahead prices or to operate fully aligned with them, significantly improve theoretical profits from the battery operating on the day-ahead market (up to 225€ per household per year).

The empirical findings for Research Question 1 highlight the significant impact of policies on the practical utilization of flexibility potentials in the market. Although households *could* have adjusted their consumption profiles and home energy

management systems to prioritize increased feed-in of their PV production during periods of high system demand, the actual operation of their systems aligns with the economic incentives established by self-consumption-oriented policies, regardless of market signals. Although also some previous studies have questioned the benefits of self-consumption-promoting policies (Green and Staffell, 2017; Aniello and Bertsch, 2023), the empirical analysis based on a field study with 947 households makes a significant contribution by providing the first empirical evidence of this policy's potential negative system effects.

The findings in the scope of the first two research questions underscore the critical role of home energy management systems in leveraging flexibility potentials and dynamic tariffs. As a result, Part II focused on mitigating the operational uncertainty of home energy management systems, particularly with the integration of emerging technologies such as heat pumps, while addressing data retrieval challenges for operators. The findings of Research Question 3 (*"To what extent do the same forecasting methods used for day-ahead predictions of traditional household loads also perform effectively for heat pump loads?"*) highlight the need for methodological advancements. In response to this question, the forecasting accuracy of advanced methods such as LSTM, XGBoost, and Transformer models are benchmarked using metrics like MAE, MAPE, and RMSE, both before and after heat pump installation in an energy community. For traditional household loads, computationally efficient methods like XGBoost provide reasonable accuracy; however, achieving high-quality forecasts requires more computationally expensive Transformer models after heat pumps are installed. In the day-ahead energy community load forecast - after heat pump installations - Transformer models achieve a MAPE of 13.43% with an R^2 of 0.9. In comparison, random forest and XGBoost models perform worse, with MAPEs of 16.79% and 16.48%, respectively, and an R of 0.87 for both. While several existing studies already evaluate day-ahead load forecasting methods for energy communities (e.g., Coignard et al. (2021)), the findings within the scope of Research Question 3 contribute a specific analysis of the impact of heat pumps and incorporate the novel consideration of Transformer models.

In Research Question 4, titled *"How can smart meter data and hybrid LSTM-XGBoost models improve day-ahead aggregated load forecasts for energy communities, particularly in addressing insufficient peak load predictions?"*, the focus lies on

evaluating the impact of separate access to household smart meter data and the application of a novel hybrid LSTM-XGBoost algorithm on the quality of day-ahead load forecasting for aggregators. Through a case study involving German households, it is found that access to historical disaggregated smart meter load measurements significantly enhances forecast accuracy, as assessed by MAPE and RMSE. Moreover, the proposed hybrid model, which incorporates a separate XGBoost forecast for peak load amounts, improves the overall forecast results and the MAPE of the forecasted peak loads. Compared to a standard LSTM model using aggregated load data, the LSTM-XGBoost model incorporating additional smart meter data improves load forecast accuracy, reducing the MAPE from 21.64% to 16.81% and the RMSE from 2.91 to 2.49. Research Question 5 (*"How can smart meter data and hybrid LSTM-XGBoost models improve day-ahead aggregated load forecasts for energy communities, particularly in addressing insufficient peak load predictions?"*), addresses the question of what data is necessary for accurate peak time forecasts. In a case study of US local distribution companies, the Learning-to-Rank method is applied, which generates peak time forecasts based on the rank of historical loads rather than absolute values. When measured in terms of Accuracy and MAE, the Learning-to-Rank algorithm performs comparably to peak time predictions derived from regular forecasts without significant differences in performance. While Research Questions 3–5 are not explicitly focused on dynamic tariffs or the operations of aggregators and utilities offering them, they provide valuable insights into practical data retrieval considerations. These questions lie at the intersection of technology and behavior. Aggregators that require less data for reliable operation may be more attractive to households, potentially encouraging the adoption of dynamic tariffs. Additionally, operating with less granular data—such as not having access to separate heat pump load measurements -as highlighted in Research Question 3- may offer advantages from a privacy perspective. Also, in Research Question 5 (*"How can synthetic heat pump load profiles be generated using a k-means clustering- based approach?"*), a method for generating synthetic heat pump load data without relying on historical customer measurements is presented. This approach helps address the uncertainty surrounding novel heat pump loads, which is further exacerbated by the limited availability of open-source historical load profiles. While the underlying k-means clustering algorithm has already been used to generate synthetic gas

load profiles in other studies (Jesper et al., 2021), the proposed model within the scope of Research Question 5 is the first to apply it to the creation of synthetic heat pump load profiles. The synthetically generated profiles show a relatively low overall error of 2.4% compared to actual measurements over the test period, highlighting the model's validity. Further evaluation based on load factors and load distribution curves confirms the method's quality and practical applicability.

Part III explores the limited adoption of dynamic tariffs, where loss aversion and uncertainty about price risks act as key barriers. The economic implications of dynamic tariffs for households are addressed in Research Question 7, titled *"How can the operation of household home energy management systems under dynamic tariffs be modeled by incorporating flexibility potentials like battery storage, heat storage, heat pumps, and building thermal inertia, alongside behavioral constraints such as thermostat setpoint profiles?"*. In response, a linear programming model that simulates the operation of heat pump-equipped households under dynamic tariffs is developed, incorporating the thermal inertia of buildings and other flexibility potentials. The thermal mass of buildings—considering different building types and modernization measures—is identified as a significant yet often overlooked flexibility potential in the existing literature.

Building on this formulation, the concept of household-specific electricity price guarantees is introduced. These guarantees are offered by aggregators to households in exchange for the right to operate their flexibility potentials, provided these operations remain within the household's comfort boundaries. Following this, in Research Question 8 (*"What is the economic value of managing household flexibility potentials for aggregators, and how can it be quantified under the proposed price guarantee mechanism?"*), the economic value of aggregator control over household flexibility potentials is investigated. The findings from a German case study reveal that aggregator-managed household flexibility consistently reduced electricity costs by an average of 7.36% (2.5 ct/kWh) over the test years 2021–2023. Furthermore, 78.4% of households would benefit from price guarantees lower than the competitive retail price. Finally, in Research Question 9 (*"How accurate are quantile regression-based predictions of household-level electricity price guarantees?"*), the proposed decision support system for aggregators, which recommends household-specific electricity price guarantees, is evaluated. These guarantees are determined based on

household characteristics and preferences while accounting for uncertainty in future prices and weather conditions. The proposed guarantee prediction model achieves an R-squared value of 0.484, indicating that it explains 48.4% of the variance in the difference between the guaranteed and actual household electricity costs, demonstrating moderate predictive power. Across the three test years, the findings show that the aggregator achieves improved financial portfolio results when implementing the suggested price guarantee mechanism. However, this improvement comes at the cost of a wider confidence interval for household costs, which introduces higher uncertainty for the aggregator.

Although the first aggregators on the market are beginning to offer electricity price guarantees for their dynamic tariffs (e.g., 1KOMMA5° (2025)), these guarantees are typically uniform and do not account for household endowment or preferences. Therefore, the contributions within Part IV can support more refined pricing and tariff design decisions for aggregators and utilities. The suggested price guarantees aim to address the behavioral uncertainty associated with the dynamic, price-based operation of household flexibility potentials by transferring the associated risks to aggregators. Aggregators, being experienced in managing portfolio risks, have the capacity to balance these risks across their portfolios. Additionally, a decision support system is proposed to assist aggregators in quantifying the economic value of a household's flexibility potential. This system helps mitigate the uncertainty that arises for aggregators when implementing such a product.

Finally, Part V explores the uncertainty surrounding the technical implications of dynamic tariff adoption in distribution grids, particularly the stress on transformers and lines, as well as the associated reinforcement costs. In addressing Research Question 10 (*"How does increasing adoption of dynamic tariffs by residential households impact distribution grid reinforcement costs across different grid topologies?"*), a case study involving various German distribution grid topologies and households equipped with heat pumps, electric vehicles, PV systems, and BESS shows that grid reinforcement costs increase significantly when a large share of households simultaneously respond to price signals, also denoted as avalanche effects.

However, in addressing Research Question 11 (*"How effective are alternative regulatory options for grid charges and PV feed-in remuneration in alleviating grid reinforcement costs given increasing dynamic tariff adoption?"*), it is demonstrated that

alternative grid charge policies—such as segmented tariffs—can effectively mitigate the increased grid reinforcement costs induced by dynamic tariff adoption. These regulatory options present a viable solution to alleviate the stress on distribution grids as dynamic tariffs gain traction.

While previous studies, such as Stute and Klobasa (2024), have already highlighted the potential threat of avalanche effects from increasing household adoption of dynamic tariffs, this study makes a significant contribution within the scope of Research Questions 10 and 11. It introduces an open-source framework that allows policymakers to assess the impact of policy options as dynamic tariff adoption grows. Additionally, a detailed analysis of different grid topologies, for instance, with a high penetration of PV or wind generation capacities, offers more profound insights into avalanche effects and their impact on grid reinforcement costs. Another key contribution is the examination of segmented grid charges and the introduction of rotating grid charges. These grid charge options provide a potential tool to mitigate rising grid stress caused by dynamic tariff adoption.

Many studies within this dissertation are supported by the open-source publication of code and data, thereby enabling greater transparency, reproducibility, and collaboration. The forecasting methods and underlying data presented in Chapter 4 are published in an open-source repository[13], fostering advancements in forecasting day-ahead heat pump loads for aggregators and improving operational decision-making in the context of dynamic tariffs. The synthetic heat pump profile generation approach introduced in Chapter 7 is also published as open-source code[14] and includes an easy-to-use web application[15] for generating heat pump load data. Additionally, the simulation framework designed to assess the impact of policy decisions on grid reinforcement costs under increasing adoption of dynamic tariffs from Chapter 9 is published alongside the utilized data[16].

In conclusion, this thesis addresses key barriers to the adoption of dynamic tariffs, focusing on three major areas: **policy, technology, and behavior**. Regarding policy, a significant contribution is the empirical evidence demonstrating that self-consumption-promoting regulation for household battery energy storage systems is

[13] `https://github.com/leloq/load-forecasting-with-heatpumps`
[14] `https://github.com/leloq/synthetic-heat-pump-load-profile-generator`
[15] `https://heatpump.ninja`
[16] `https://github.com/leloq/dynamic-tariffs-in-distribution-grids`

less beneficial for the energy system than dynamic tariffs. Building on this, the operational aspects of dynamic tariff adoption are supported through advancements in load forecasting and synthetic data generation, with a particular emphasis on the emerging flexibility potential of heat pumps and the thermal inertia of buildings. Additionally, the impact of dynamic tariff-induced changes in load patterns on grid infrastructure and associated costs is thoroughly analyzed. Finally, the behavioral barriers to dynamic tariff adoption, such as households' loss aversion, are addressed through an innovative proposal: household-individual electricity price guarantees offered by aggregators in exchange for the right-to-operate households' flexibility potential.

To make electricity markets more efficient and increase the adoption rates of dynamic tariffs, a coordinated interplay of novel policies, innovative technologies, and tailored solutions that align with household preferences and behaviors is essential. While this thesis tackles critical issues within this scope, it serves as a foundation for future research in these important areas, which is illustrated in the following and final chapter.

CHAPTER 11

OUTLOOK

Based on this thesis, several valuable avenues for further research emerge, which are outlined below.

First, a significant contribution of this dissertation is the introduction and analysis of household-individual electricity price guarantees, focusing on flexibility potentials, preferences, and behavior. However, a limitation lies in the fact that the results are purely simulation-based. To advance this concept, it is essential to transition from theoretical exploration to practical implementation. This would involve developing real-world applications that manage household flexibility potentials under uncertainty from the perspective of aggregators. A potential pathway is the use of Model Predictive Control algorithms, which are advanced optimization techniques that predict future system behavior and adjust control actions accordingly in real time, with a focus on assessing how effectively guaranteed prices and discounts can be delivered. The transition from theoretical advances to practical implementation also presents promising opportunities for further research on the load forecasting methods proposed in Part III. While these models reliably predict general household and heat pump day-ahead load curves, forecasting household responses to price signals remains an important area for future exploration.

Second, the findings of this thesis are based on analyses conducted at a low aggregation level, involving at most a few hundred households. While this granularity is important for dynamic tariff research, future studies should also adopt a broader energy system model perspective to account for endogenous effects. For instance, examining whether the counterproductive avalanche effects observed at the distribution grid level persist when widespread dynamic tariff adoption leads to altered price

curves would be valuable. Additionally, understanding how large-scale aggregator fleets influence market dynamics would provide integral insights for policymakers and practitioners. These research directions could be explored using agent-based energy market models to capture feedback loops and systemic effects.

Third, household behavior is a key determinant of dynamic tariff adoption. Although this thesis addresses important behavioral issues and proposes potential solutions, these aspects have not been explored from a dedicated behavioral research perspective. For example, understanding households' perceptions of and preferences for the proposed electricity price guarantees is essential for fostering widespread adoption. Behavioral research could provide further insights into how these guarantees are perceived and how they can be effectively communicated to increase household participation.

Fourth, this thesis proposes alternative grid charge designs, such as segmented or rotating tariffs, to mitigate the negative distribution grid impacts of increasing dynamic tariff options. While these approaches can potentially limit rising distribution grid reinforcement costs, further research is required to assess their cost reflectivity and determine how they should be implemented to ensure fair cost allocation. Additionally, the fairness of these measures across different socio-demographic groups must be examined. Overall, the impact of dynamic tariffs and their adoption on energy equity remains an essential area for future research, aiming to maximize the system-wide benefits of dynamic pricing without disadvantaging households that cannot afford PV panels or flexibility options such as BESS or thermal storage.

Finally, although this dissertation focuses exclusively on the adoption of dynamic tariffs by households, similar challenges are present for industrial and commercial customers. Addressing these requires separate studies, as the regulatory framework, customer preferences, and load profiles differ. Given that electricity prices for the industry are vital to maintaining the competitiveness and stability of the economy, future research should also explore and promote dynamic tariff adoption in the industrial sector.

These avenues highlight the need for interdisciplinary research to build upon the contributions of this dissertation and address the complex challenges associated with dynamic tariff adoption.

Appendices

APPENDIX A

FIELD EVALUATION OF BESS REGULATION

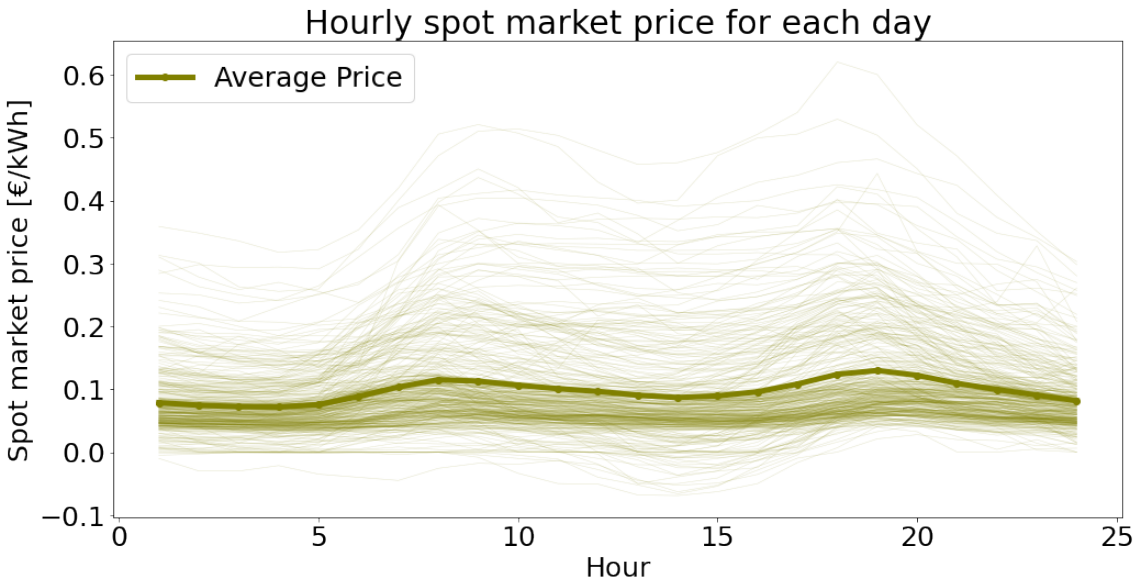


Figure A.1.: Diurnal price curves over the year 2021

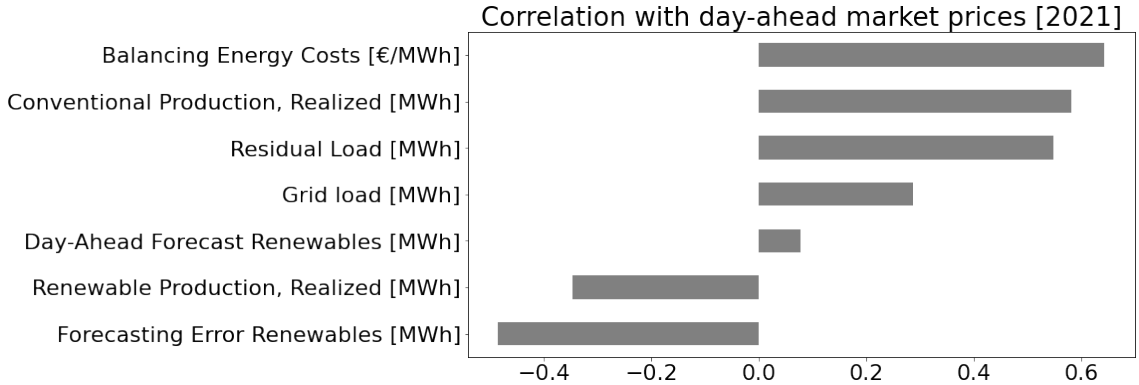
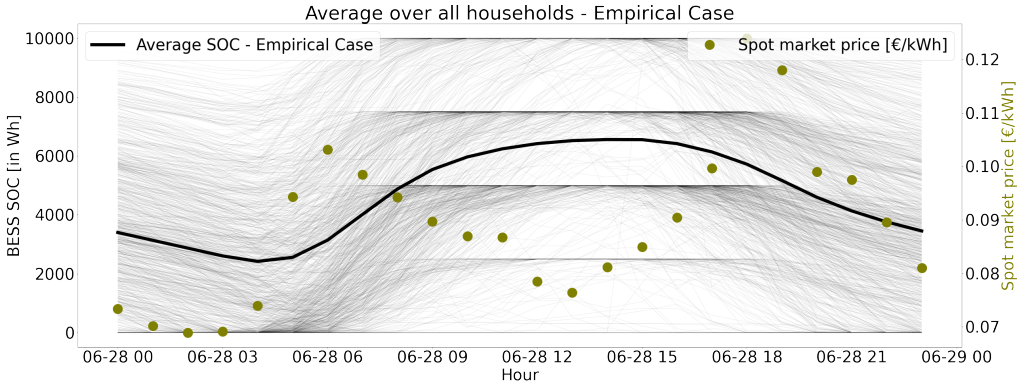
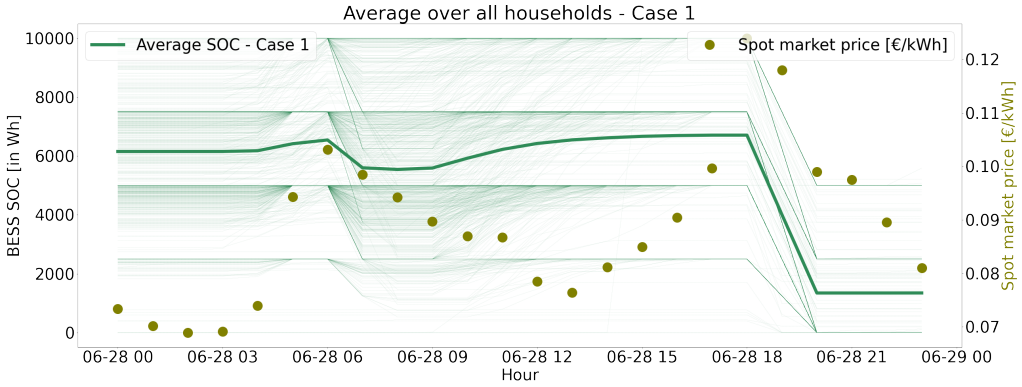


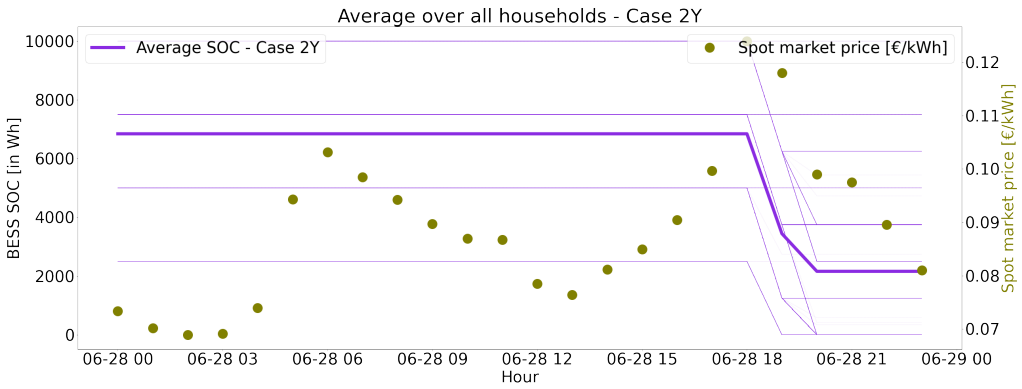
Figure A.2.: Correlation of German power market metrics with day-ahead market prices for the year 2021



(a) Battery SOC profiles for the empirical case

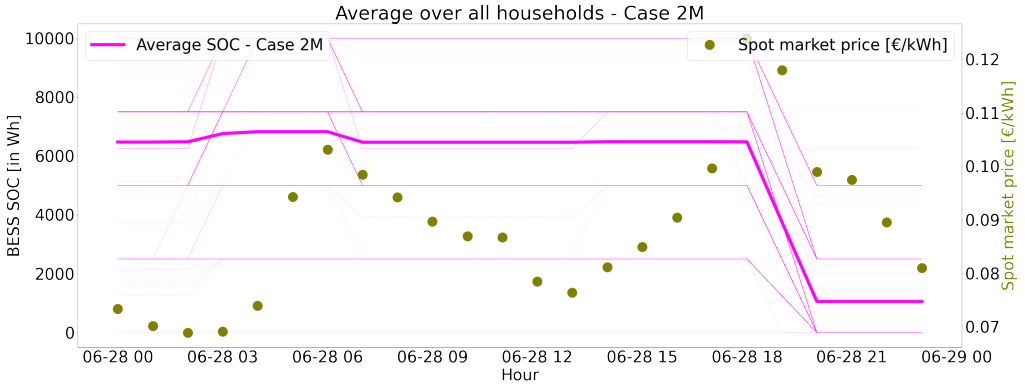


(b) Battery SOC profiles for Case 1

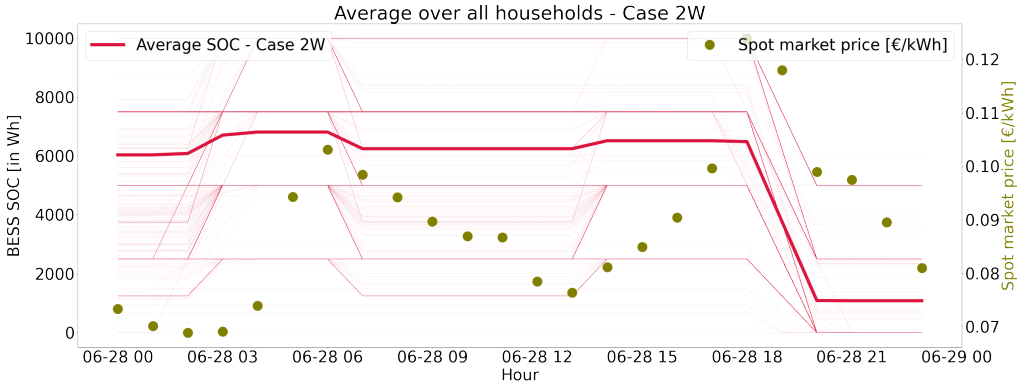


(c) Battery SOC profiles for Case 2Y

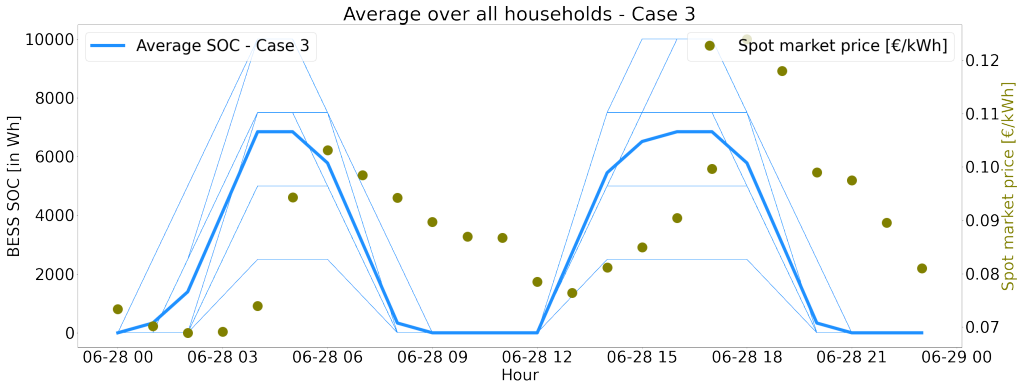
Figure A.3.: Battery SOC profiles over all cases and households



(d) Battery SOC profiles for Case 2M

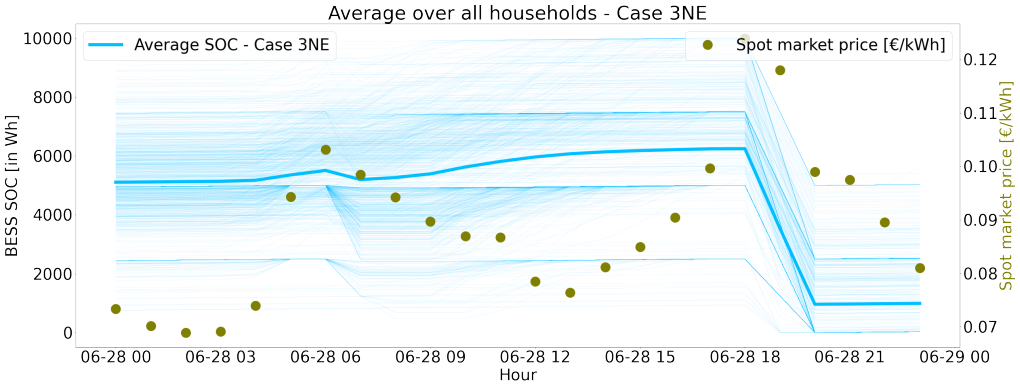


(e) Battery SOC profiles for Case 2W



(f) Battery SOC profiles for Case 3

Figure A.3.: Battery SOC profiles over all cases and households (continued)



(g) Battery SOC profiles for Case 3NE

Figure A.3.: Battery SOC profiles over all cases and households (continued)

APPENDIX B

HEAT PUMP LOAD FORECASTING

	mean	std	min	max
LSTM HH	1452.56	6.33	1447.21	1461.48
LSTM HP	2989.20	90.29	2908.37	3083.38
LSTM Comb	3096.27	58.89	3014.16	3152.67
XGB HH	1362.03	0.00	1362.03	1362.03
XGB HP	2817.45	0.00	2817.45	2817.45
XGB Comb	3256.33	0.00	3256.33	3256.33
Random Forest HH	1368.29	6.93	1362.01	1374.83
Random Forest HP	2761.41	4.09	2756.48	2765.77
Random Forest Comb	3256.02	10.81	3242.94	3269.30
Transformer HH	1324.57	13.72	1310.44	1340.74
Transformer HP	2319.38	27.72	2277.90	2335.79
Transformer Comb	2727.31	41.40	2688.11	2764.41

Table A.1.: Descriptive statistics of five runs per method

APPENDIX C

PRICE GUARANTEES

C.1 COP parameters

In Table A.1, the temperature-dependent parameters for the polynomial fitting of the heat pump COP curve are depicted, based on Verhelst et al. (2012).

Coefficient	Value
c0	8.24
c1	0.158
c2	-0.195
c3	0.00101
c4	0.00148
c5	-0.00233
T_{ws}	50

Table A.1.: Coefficients used in the COP calculation

C.2 PV and conditional BESS size distributions

In Figure A.1, the empirically observed size of BESS is displayed, depending on connected PV sizes.

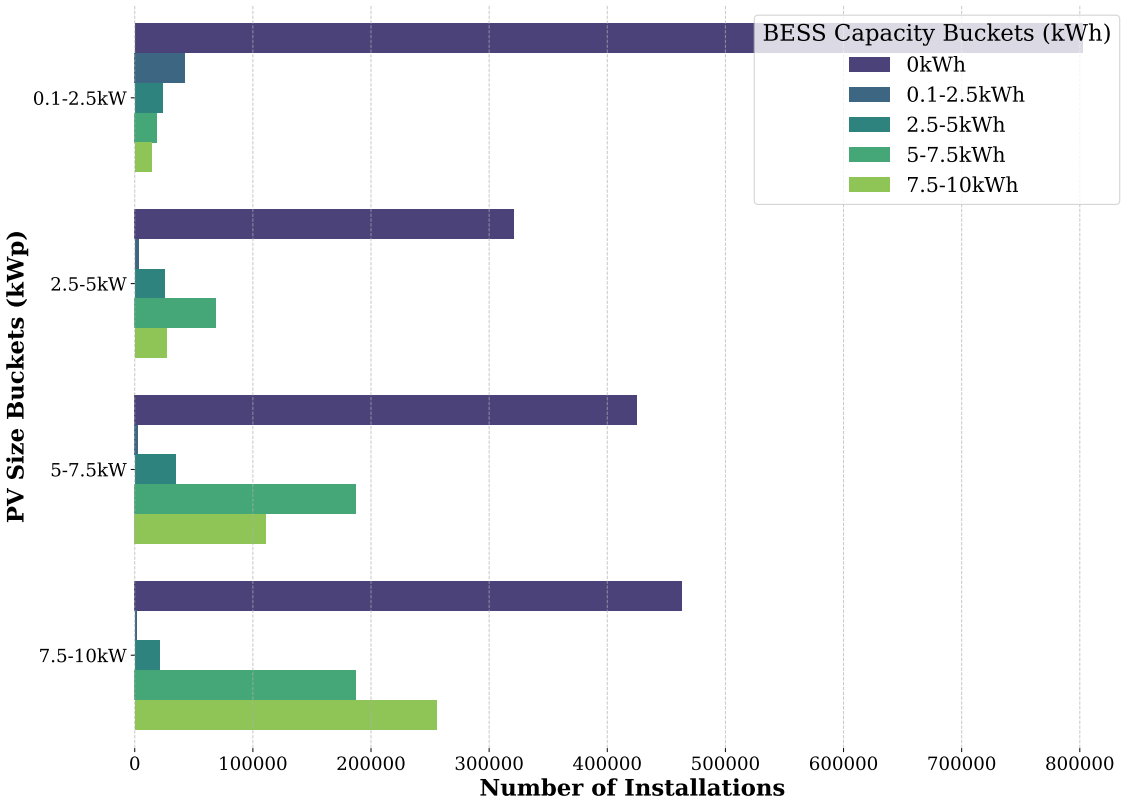


Figure A.1.: Distributions of PV size buckets and BESS capacity buckets

C.3 Price components

In Table A.2, the fees and taxes that are added on the day-ahead spot market prices are depicted, based on Bundesnetzagentur (2024c) and Bundesverband der Energie- und Wasserwirtschaft (BDEW) (2024). The volume weighted prices are calculated by weighting the day-ahead price curves with standard load profiles from (Bundesverband der Energie- und Wasserwirtschaft e.V., 2024).

Year	2018	2019	2020	2021	2022	2023
Mean yearly market price [€/kWh]	0.044469	0.037667	0.030471	0.096850	0.235446	0.095175
Grid Charges	0.072900	0.073900	0.077500	0.078000	0.080800	0.095200
EEG Umlage	0.067900	0.064100	0.067600	0.065000	0.018600	0.000000
Taxes	0.044100	0.044100	0.044100	0.044100	0.044100	0.044100
Volume Weighted Average Price [€/kWh]	0.231075	0.221882	0.221459	0.289756	0.379456	0.238699
Volume Weighted Average Price [€/kWh] with VAT	0.274979	0.264040	0.263536	0.344810	0.451553	0.284052

Table A.2.: Fees and taxes in €/kWh

C.4 Transformation of Global Horizontal Irradiance

In Figure A.2, an exemplary transformation of Global Horizontal Irradiance to Southern Vertical Irradiance is depicted.

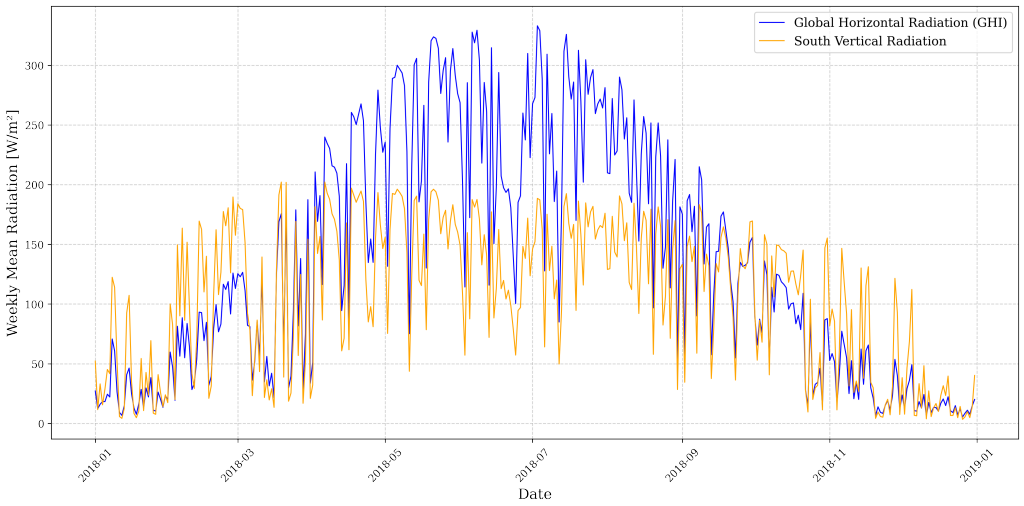


Figure A.2.: Transformation from Global Horizontal Irradiance to Southern Vertical Irradiance

C.5 Monte Carlo simulation convergence analysis based on Central Limit Theorem

In Figure A.3, the convergence of the Monte Carlo simulation is depicted according to the Central Limit Theorem, following the method of Yang (2011).

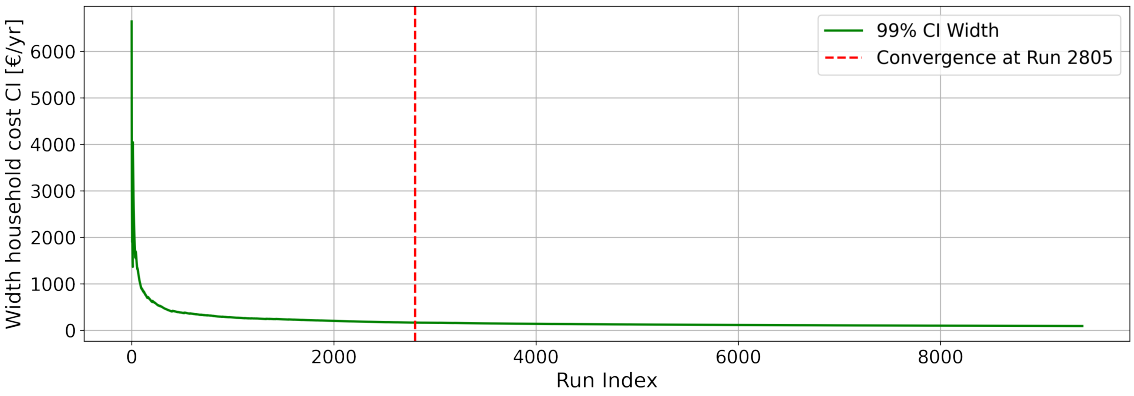
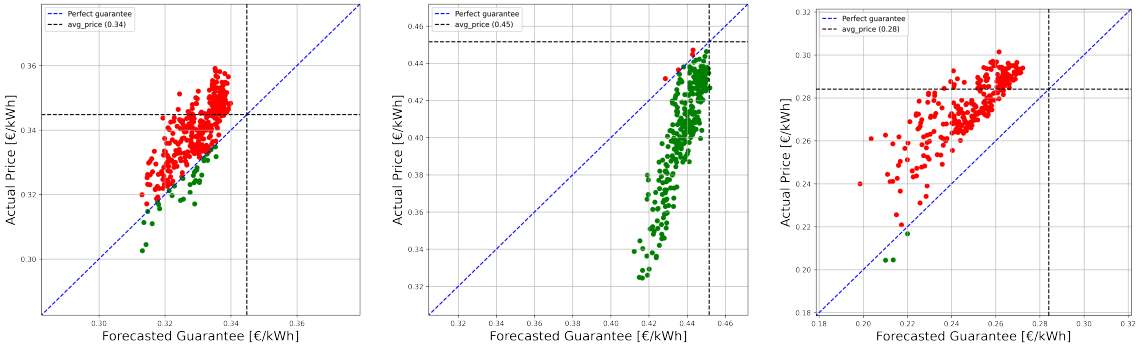


Figure A.3.: Confidence interval analysis based on the Central Limit Theorem

C.6 Quantile regression sensitivity analysis

In Figures A.4 and A.5, the quantile regression results for the 0.1 and 0.9 quantiles are visualized, comparing guaranteed and realized prices for households across 2021, 2022, and 2023. Tables A.3 and A.4 present the mean and confidence interval widths for different quantiles, respectively, highlighting the impact of different quantile levels as a tool to balance competitiveness and risk of guarantees. For instance, a 0.1 quantile leads to lower guarantees, which are often lower than the actual prices, leading to losses for the aggregator on the household level. Such a quantile level might be a reasonable choice for an aggregator that aims at winning new customers. Whereas, a 0.9 level leads to higher guarantees, which are less competitive, but also reduce the associated risk.

Table A.5 summarizes the comparison of electricity costs under aggregator-controlled and household-controlled flexibility scenarios.



(a) 2021 (b) 2022 (c) 2023

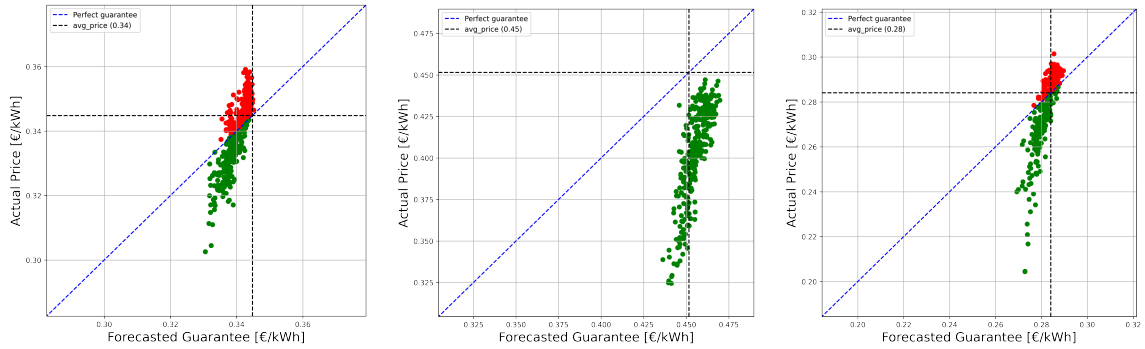
Figure A.4.: Quantile regression for 0.1 quantile: Comparison of guaranteed and realized prices per household

Target Year	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2021	-4.288654	-65.755298	-27.226418	-15.846370	-4.288654	6.833473	21.598347	33.719610	61.848928
2022	459.028291	342.463387	378.127472	414.917185	459.028291	489.397181	509.442118	527.441974	567.534031
2023	3.061538	-150.497560	-65.444641	-26.195397	3.061538	31.770486	62.237874	93.061368	129.655953

Table A.3.: (a) Mean over different quantiles

Target Year	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2021	226.374622	268.246951	223.793984	235.008560	226.374622	200.421076	181.842471	182.793065	180.343278
2022	799.164416	691.732064	722.483228	775.815889	799.164416	815.233101	843.722611	873.416549	857.152743
2023	348.326632	496.142462	397.065325	349.168304	348.326632	318.929070	313.156979	257.537927	247.326130

Table A.4.: (b) Interval Width over different quantiles



(a) 2021 (b) 2022 (c) 2023

Figure A.5.: Quantile regression for 0.9 quantile: Comparison of guaranteed and realized prices per household

Year	Aggregator Controls Flexibility	Household Controls Flexibility	Absolute Change [€/kWh]	Percent Change [%]
2021	0.339	0.358	0.019	5.652
2022	0.403	0.437	0.034	8.484
2023	0.274	0.295	0.021	7.705
All years	0.340	0.365	0.025	7.359

Table A.5.: Results per scenario and year

APPENDIX D

DYNAMIC TARIFF IMPACT ON DISTRIBUTION GRIDS

Regulatory scenarios

In Table A.2, a detailed overview of the investigated regulatory scenarios is depicted. The ramp-up from 0% to 100% households with dynamic tariffs constitutes replacing *Constant* with *Dynamic* profiles at the nodes of the distribution grid, in the scope of the respective regulatory scenario.

Peak loads of aggregated household profiles

In Table A.3, the aggregated peak loads of all 500 households within the given regulatory scenarios are depicted.

Table A.1.: Investigated policy options and resulting scenarios.

Scenario	**Feed-in Tariff** $p_t^{\text{feed-in}}$	**Grid Charges**	**Retail Tariff** $p_t^{\text{wholesale}}$	**Operation**
Constant Volumetric_FIT without flexibility	constant	volumetric	constant	constant
Constant Volumetric_FIT	constant	volumetric	constant	dynamic
Dynamic Volumetric_FIT	constant	volumetric	dynamic	dynamic
Constant Volumetric_DynFeed	dynamic	volumetric	constant	dynamic
Dynamic Volumetric_DynFeed	dynamic	volumetric	dynamic	dynamic
Constant Peak_FIT	constant	peak	constant	dynamic
Dynamic Peak_FIT	constant	peak	dynamic	dynamic
Constant Peak_DynFeed	dynamic	peak	constant	dynamic
Dynamic Peak_DynFeed	dynamic	peak	dynamic	dynamic
Constant Rotating_FIT	constant	rotating	constant	dynamic
Dynamic Rotating_FIT	constant	rotating	dynamic	dynamic
Constant Rotating_DynFeed	dynamic	rotating	constant	dynamic
Dynamic Rotating_DynFeed	dynamic	rotating	dynamic	dynamic
Constant Segmented_FIT	constant	segmented	constant	dynamic
Dynamic Segmented_FIT	constant	segmented	dynamic	dynamic
Constant Segmented_DynFeed	dynamic	segmented	constant	dynamic
Dynamic Segmented_DynFeed	dynamic	segmented	dynamic	dynamic

Table A.2.: Scenarios with varying tariffs, grid charges, and operational strategies.

Table A.3.: Aggregated peak loads [kW] per scenario on the day with the peak empirical HP load (2019-12-01), PV feed-in (2019-05-13), EV demand (2019-11-18) and the highest yearly peak load per scenario.

Name	Heat Pump Peak [kW]	Feed-in Peak [kW]	EV Peak [kW]	Yearly Peak [kW]
Constant Volumetric_FIT without flexibility	1189.9	1969.3	1183.9	2272.3
Constant Volumetric_FIT	1179.7	1865.7	1095.6	2270.2
Dynamic Volumetric_FIT	2258.9	1891.8	2352.3	2788.4
Constant Volumetric_DynFeed	1202.1	2189.0	1122.4	2496.5
Dynamic Volumetric_DynFeed	2257.1	2159.0	2351.2	2787.8
Constant Peak_FIT	996.4	1892.5	989.0	2267.1
Dynamic Peak_FIT	1693.6	1944.1	1670.4	2560.1
Constant Peak_DynFeed	1204.5	2597.8	1277.7	2652.3
Dynamic Peak_DynFeed	1699.6	2512.5	1668.6	2650.3
Constant Rotating_FIT	1227.6	1876.1	1096.4	2246.8
Dynamic Rotating_FIT	1437.6	1890.0	1449.2	2290.5
Constant Rotating_DynFeed	1229.4	2192.9	1093.1	2496.6
Dynamic Rotating_DynFeed	1442.2	2134.6	1448.5	2492.4
Constant Segmented_FIT	959.8	1827.3	937.5	2296.3
Dynamic Segmented_FIT	1264.0	1863.2	1184.8	2263.1
Constant Segmented_DynFeed	990.5	2178.3	924.0	2496.8
Dynamic Segmented_DynFeed	1248.4	2192.1	1185.0	2495.6

Peak loads with varying PV/BESS/HP/EV adoption rates

In Figure A.1, the weekly aggregated energy consumption of all 500 modelled households is depicted. Figure A.2 shows a sensitivity analysis for varying PV/BESS/HP/EV adoption rates under the status-quo $Volumetric\_FIT$ regulatory scenario.

Figure A.3 depicts the impact of varying adoption rates under the alternative $Segmented\_FIT$ regulatory scenario, showing remarkably lower peak loads and lower peak load increases through the introduction of dynamic tariffs.

Grid topologies

Table A.4 describes the properties of the investigated grid topologies in terms of nominal transformer capacity (in MW), installed renewables capacity (in MW), line length (in km) and transformer numbers. Figure A.4 depicts an exemplary grid topology consisting of transformers, lines, and connected generators and loads. In Figure A.5, an overview of the simulated grids and the types of installed loads and generation capacities is depicted.

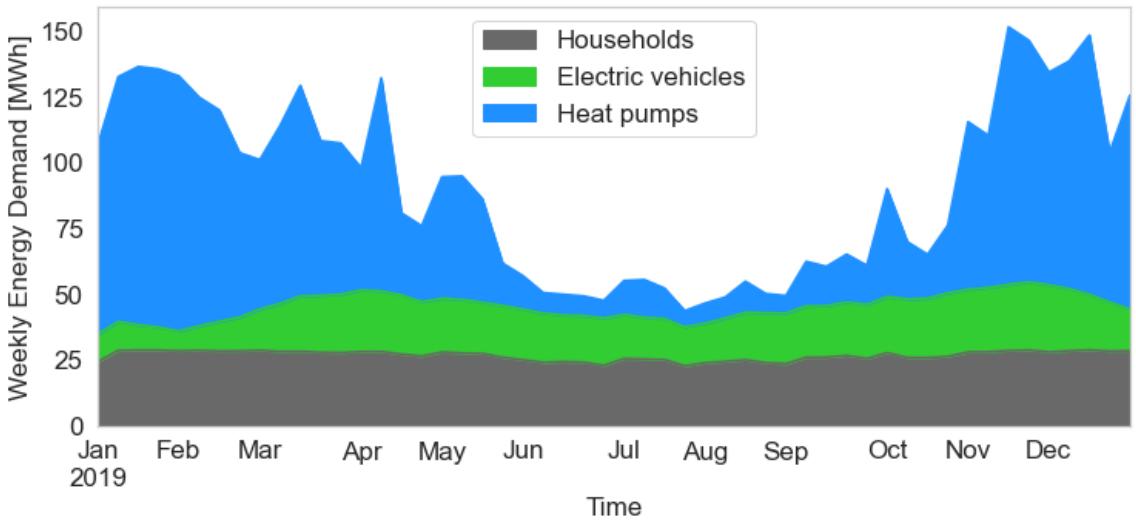


Figure A.1.: Weekly aggregated energy consumption of all households

Empirical consumption on type days

Figure A.6 shows the aggregated empirical loads on the investigated days, exhibiting particularly high peak loads on the HP and EV day.

Grid reinforcement results

In Table A.5, the grid reinforcement costs per penetration rate are split up into the analyzed grids and then aggregated for one exemplary regulatory scenario (*Volumetric FIT*). In the following experimental results per type day (Table A.6, A.7, A.8, A.9 and A.10), only the aggregated costs are considered per regulatory scenario. In Figure A.7, an overview of the grid issues that led to the necessary reinforcement measures is provided.

In Figure A.8, we show that the peak loads observed in Table A.3 are correlated with the resulting grid reinforcement costs. The Pearson correlation factor between the yearly peak loads and resulting grid reinforcement costs lies at 0.68, indicating a strong relationship between the two variables. This demonstrates that the peak loads observed in Figure A.3 can serve as a reliable initial indicator for determining whether grid reinforcement measures may be necessary. However, the coefficient of determination (R^2) of 0.47 reveals that annual peak loads alone cannot fully explain the variability in grid reinforcement costs. Consequently, this underscores the impor-

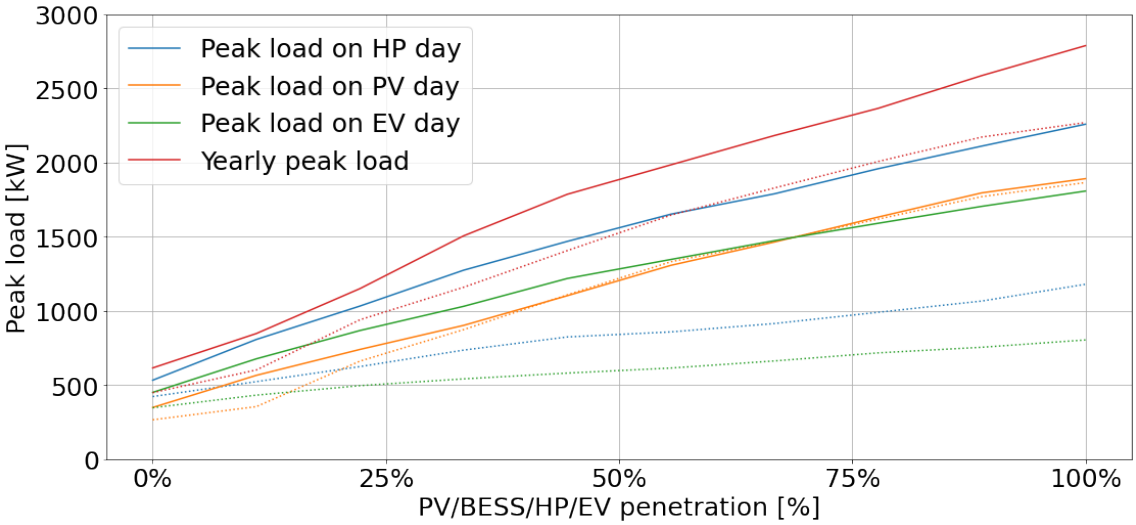


Figure A.2.: Peak loads on specific days in the scope of varying PV/BESS/HP/EV adoption rates under the $Volumetric\_FIT$ regulatory scenario

tance of conducting a more granular, scenario-specific analysis of grid reinforcement costs, as detailed in Section 9.7.2.

In Figure A.9, the correlation between the increase of grid reinforcement costs (from 0% to 100% dynamic tariff adoption under the $VolumetricFIT$ scenario) and characteristics of the 10 investigated grid topologies is investigated. Among the grid characteristics, the length of medium-voltage (MV) lines exhibits the strongest positive correlation with reinforcement costs, whereas the mean nominal power of transformers demonstrates a negative correlation.

Table A.4.: Grid characteristics

Grid name	transformers_snom_mean	installed_wind_capacity	length_lines_median	nr_lines	nr_buses	length_lines_lv	installed_pv_capacity	nr_transformers	length_lines_total	length_lines_max	length_lines_mean	length_lines_mv
urban-load-2	0.60	0.00	0.02	5776	5830	173.20	1.46	83	247.53	6.86	0.04	74.33
suburban-load-2	0.59	0.01	0.02	11224	11312	319.60	5.59	143	438.87	8.03	0.04	119.27
rural-balanced	0.52	0.00	0.02	5142	5188	149.51	11.25	79	281.10	15.65	0.05	131.58
urban-load-1	0.53	0.00	0.02	10222	10331	312.06	12.51	165	408.32	6.71	0.04	96.27
rural-wind-1	0.28	15.25	0.03	516	530	17.94	0.96	16	49.00	9.11	0.09	31.06
rural-pv-wind	0.24	45.30	0.03	13439	13730	540.74	23.39	319	1035.10	15.79	0.08	494.36
suburban-load-1	0.41	1.50	0.02	8396	8485	251.35	5.30	131	408.13	18.86	0.05	156.78
rural-pv	0.14	0.00	0.03	3668	3855	160.57	17.64	207	611.96	14.23	0.17	451.39
rural-wind-2	0.22	23.20	0.03	8728	8910	381.40	5.82	201	723.43	18.21	0.08	342.03
suburban-balanced	0.58	0.00	0.02	5197	5259	178.25	11.37	103	269.54	6.25	0.05	91.30

Table A.5.: Volumetric FIT detailed reinforcement costs [M€] per grid in the combined analysis

Dynamic Tariff Penetration [%]	urban-load-2	suburban-load-2	rural-balanced	urban-load-1	rural-wind-1	rural-pv-wind	suburban-load-1	rural-pv	rural-wind-2	suburban-balanced	Aggregated
0	0.6	2.2	1.1	1.9	0.2	8.3	0.9	1.3	4.5	1.2	22.2
25	0.6	3.6	3.3	1.9	0.2	9.5	1.0	2.3	6.4	1.2	30.0
50	0.6	5.0	5.0	2.0	0.2	9.5	1.3	2.3	6.5	1.3	33.7
75	1.2	6.1	5.0	2.3	0.2	9.9	2.0	5.6	8.0	1.2	41.3
100	1.5	8.2	4.9	2.5	0.2	10.4	3.6	10.4	11.9	1.2	54.8

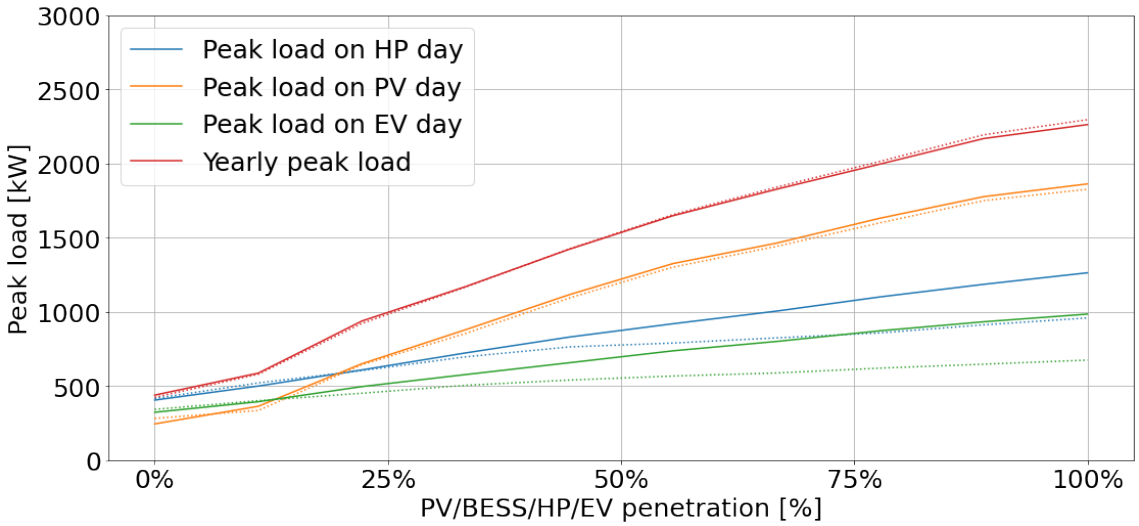


Figure A.3.: Peak loads on specific days in the scope of varying PV/BESS/HP/EV adoption rates under the *Segmented_FIT* regulatory scenario

Table A.6.: Combined analysis reinforcement costs [M€] per regulatory setting

Dynamic Tariff Penetration [%]	Volumetric_FIT	Volumetric_DynFeed	Segmented_FIT	Segmented_DynFeed	Rotating_FIT	Rotating_DynFeed	Peak_FIT	Peak_DynFeed
0	22.2	25.7	21.4	25.7	22.4	25.6	22.4	32.9
25	30.0	32.9	21.2	25.4	22.2	25.5	22.2	32.5
50	33.7	36.8	21.5	25.7	22.6	25.8	23.1	32.2
75	41.3	43.4	21.9	25.1	22.5	25.2	23.0	32.0
100	54.8	57.5	21.8	25.3	22.7	25.7	24.2	32.4

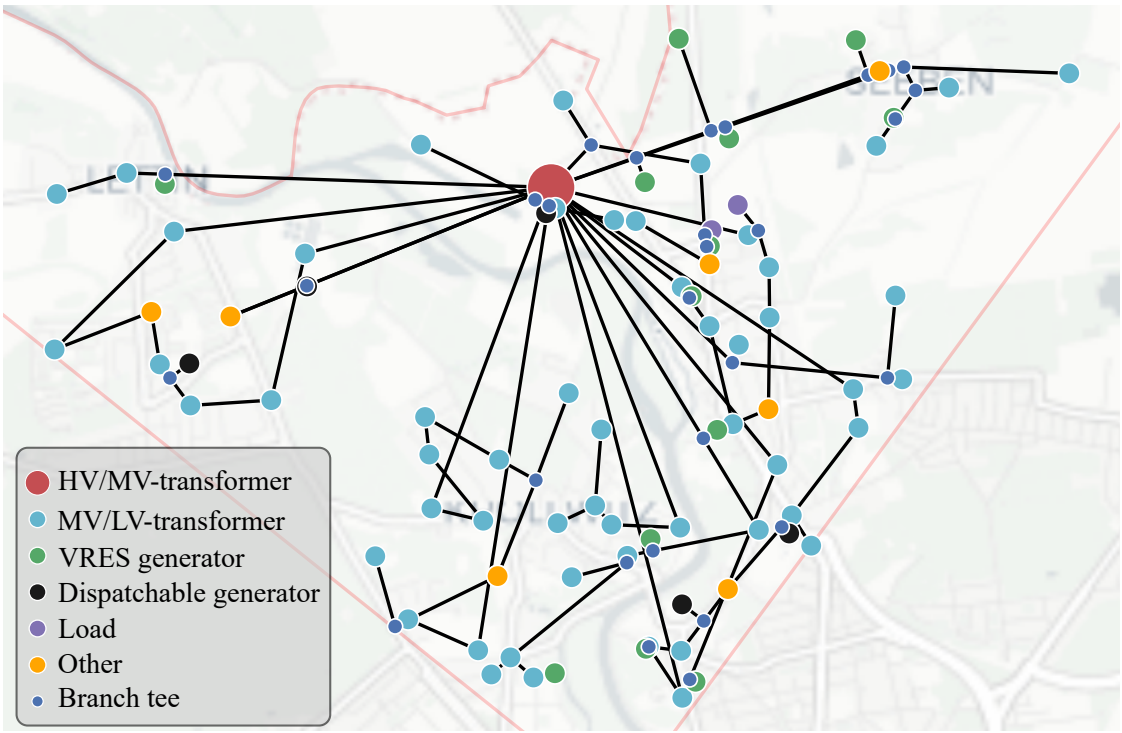


Figure A.4.: Topology of exemplary grid.

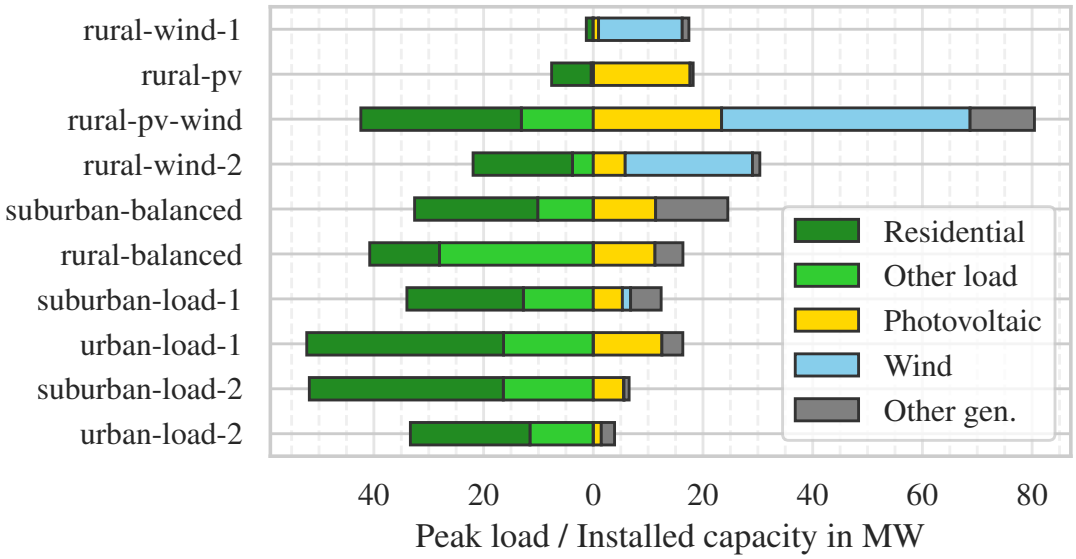


Figure A.5.: Overview of simulated grids with total peak load and installed feed-in capacities

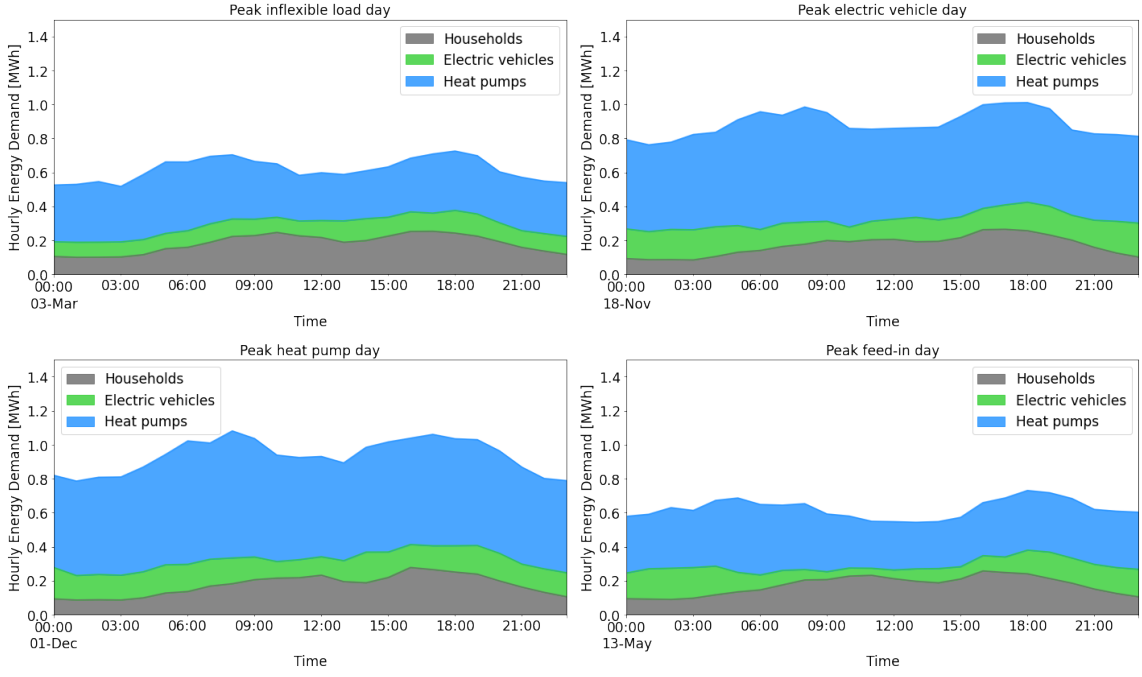


Figure A.6.: Hourly empirical aggregated energy consumption on type days

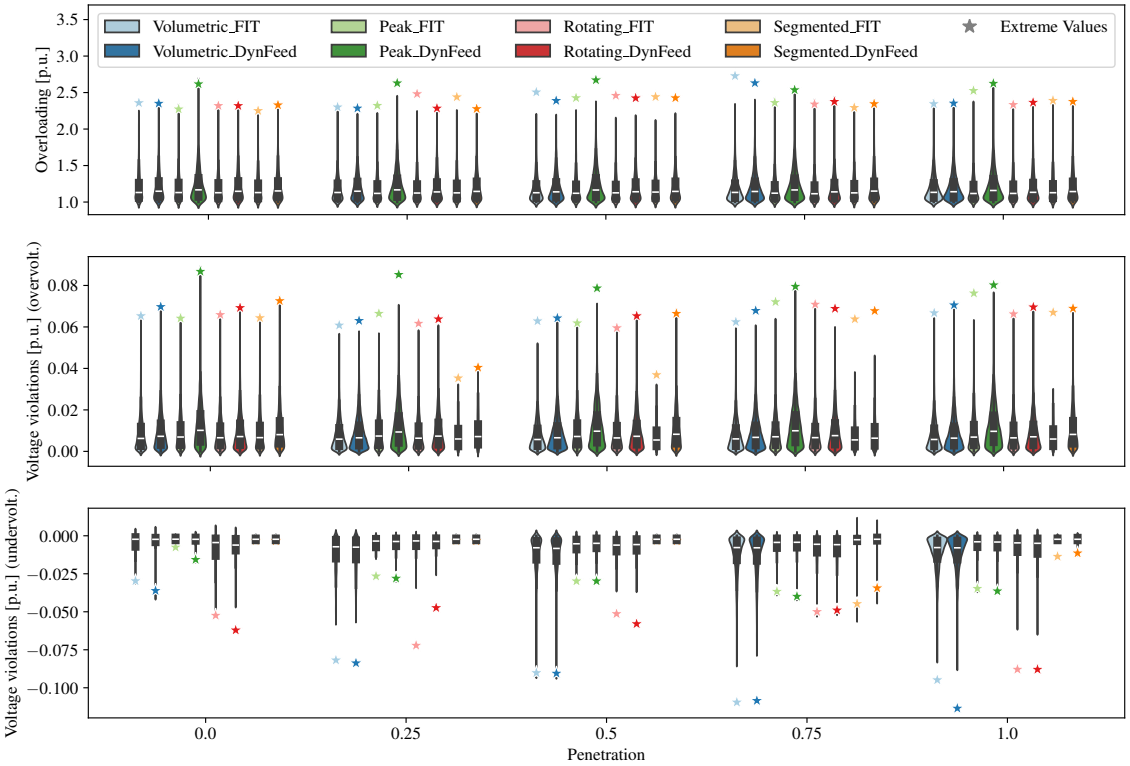


Figure A.7.: Grid issues occurring for the different policy scenarios and penetrations for one representative run

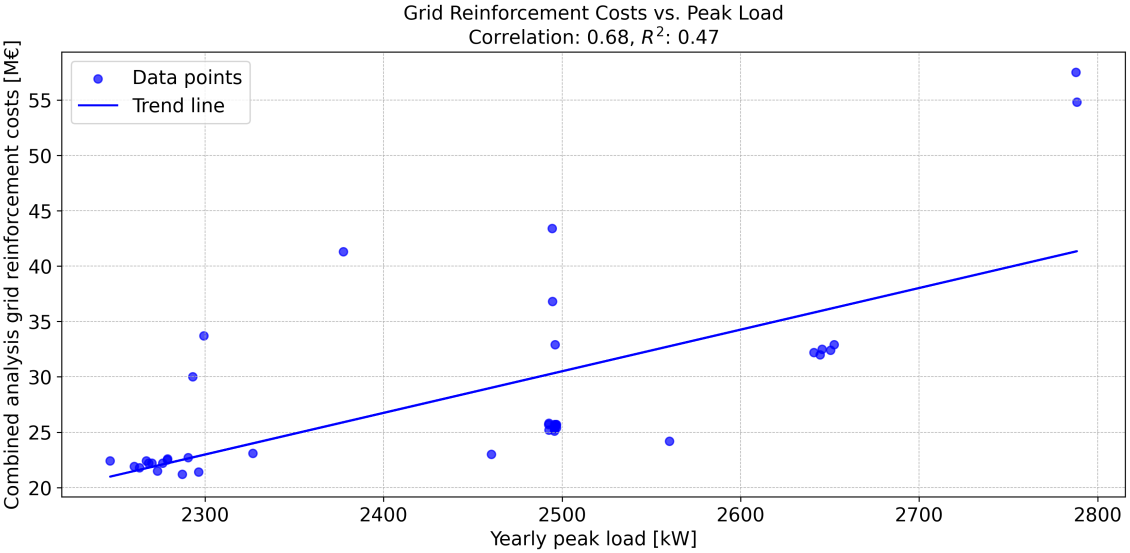


Figure A.8.: Relationship between yearly peak loads of the 500 investigated households (see Table A.3) and the "combined analysis" grid reinforcement costs (see Table A.6), for all policy scenarios, grid topologies and adoption rates.

Table A.7.: Electric vehicle type-day reinforcement costs [M€] per regulatory setting

Dynamic Tariff Penetration [%]	Volumetric_FIT	Volumetric_DynFeed	Segmented_FIT	Segmented_DynFeed	Rotating_FIT	Rotating_DynFeed	Peak_FIT	Peak_DynFeed
0	0.5	0.5	0.2	0.2	0.8	0.8	0.2	0.4
25	11.5	11.3	0.2	0.2	1.3	1.3	0.6	0.8
50	17.0	17.4	0.2	0.2	1.6	1.7	1.0	1.2
75	27.8	27.3	0.2	0.2	1.9	1.8	1.1	1.2
100	44.7	45.1	0.2	0.2	3.0	3.0	2.8	3.4

Table A.8.: Feed-in type-day reinforcement costs [M€] per regulatory setting

Dynamic Tariff Penetration [%]	Volumetric_FIT	Volumetric_DynFeed	Segmented_FIT	Segmented_DynFeed	Rotating_FIT	Rotating_DynFeed	Peak_FIT	Peak_DynFeed
0	21.9	25.5	21.2	25.6	22.1	25.3	22.2	32.8
25	21.7	25.2	21.1	25.2	21.8	25.1	22.1	32.4
50	22.4	25.5	21.4	25.5	22.3	25.5	23.0	32.1
75	21.3	25.0	21.7	25.0	22.1	24.9	22.9	31.8
100	21.9	25.1	21.6	25.2	22.2	25.2	23.4	31.4

Table A.9.: Heat pump type-day reinforcement costs [M€] per regulatory setting

Dynamic Tariff Penetration [%]	Volumetric_FIT	Volumetric_DynFeed	Segmented_FIT	Segmented_DynFeed	Rotating_FIT	Rotating_DynFeed	Peak_FIT	Peak_DynFeed
0	0.5	0.5	0.2	0.2	1.1	1.2	0.2	0.3
25	5.2	5.7	0.2	0.2	1.6	1.7	0.5	0.7
50	9.0	9.7	0.2	0.2	2.1	2.1	0.8	0.9
75	17.2	18.3	0.2	0.2	2.4	2.2	1.0	1.1
100	29.3	30.5	0.3	0.2	2.9	3.1	3.5	3.6

Table A.10.: Inflexible load type-day reinforcement costs [M€] per regulatory setting

Dynamic Tariff Penetration [%]	Volumetric_FIT	Volumetric_DynFeed	Segmented_FIT	Segmented_DynFeed	Rotating_FIT	Rotating_DynFeed	Peak_FIT	Peak_DynFeed
0	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.1
25	0.6	0.5	0.1	0.1	0.3	0.3	0.2	0.2
50	0.8	0.8	0.1	0.1	0.3	0.3	0.2	0.2
75	1.2	1.1	0.1	0.1	0.4	0.3	0.2	0.2
100	1.8	1.7	0.1	0.1	0.5	0.5	0.4	0.3

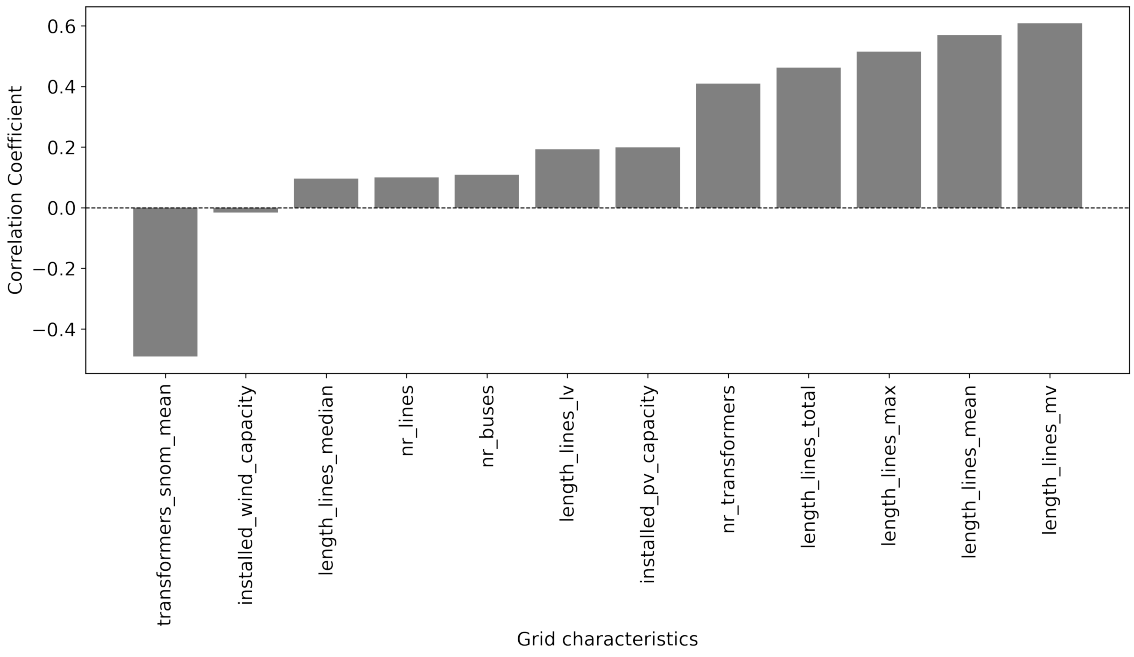


Figure A.9.: Correlation between increase in reinforcement costs and grid characteristics in the Volumetric FIT scenario

BIBLIOGRAPHY

1KOMMA5° (2025). 1KOMMA5° Strompreisgarantie: Dynamischer Stromtarif. Retrieved from `https://1komma5.com/de/magazin/pressemitteilungen/1komma5-strompreisgarantie-dynamischer-stromtarif/`. Accessed 2025-02-06.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283.

Abbasi, R. A., Javaid, N., Ghuman, M. N. J., Khan, Z. A., Ur Rehman, S., et al. (2019). Short term load forecasting using XGBoost. In *Workshops of the International Conference on Advanced Information Networking and Applications*, pages 1120–1131. Springer.

Abdin, G. C. and Noussan, M. (2018). Electricity storage compared to net metering in residential pv applications. *Journal of Cleaner Production*, 176:175–186.

Ackermann, S., Szabo, A., Bamberger, J., and Steinke, F. (2022). Design and optimization of performance guarantees for hybrid power plants. *Energy*, 239:121742.

Aghaei, J. and Alizadeh, M.-I. (2013). Demand response in smart electricity grids equipped with renewable energy sources: A review. *Renewable and Sustainable Energy Reviews*, 18:64–72.

Agnew, S. and Dargusch, P. (2017). Consumer preferences for household-level battery energy storage. *Renewable and Sustainable Energy Reviews*, 75:609–617.

Al Khafaf, N., Rezaei, A. A., Amani, A. M., Jalili, M., McGrath, B., Meegahapola, L., and Vahidnia, A. (2022). Impact of battery storage on residential energy

consumption: An Australian case study based on smart meter data. *Renewable Energy*, 182:390–400.

Alabid, J., Bennadji, A., and Seddiki, M. (2022). A review on the energy retrofit policies and improvements of the UK existing buildings, challenges and benefits. *Renewable and sustainable energy reviews*, 159:112161.

Aljohani, T. M., Ebrahim, A. F., and Mohammed, O. A. (2021). Dynamic real-time pricing mechanism for electric vehicles charging considering optimal microgrids energy management system. *IEEE Transactions on Industry Applications*, 57(5):5372–5381.

Althaher, S., Mancarella, P., and Mutale, J. (2015). Automated demand response from home energy management system under dynamic pricing and power and comfort constraints. *IEEE Transactions on Smart Grid*, 6(4):1874–1883.

Altmann, A., Toloşi, L., Sander, O., and Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347.

Ambrosio-Albala, P., Upham, P., Bale, C., and Taylor, P. (2020). Exploring acceptance of decentralised energy storage at household and neighbourhood scales: A UK survey. *Energy Policy*, 138:111194.

Amjady, N. (2001). Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on power systems*, 16(3):498–505.

Amme, J., Pleßmann, G., Bühler, J., Hülk, L., Kötter, E., and Schwaegerl, P. (2018). The eGo grid model: An open-source and open-data based synthetic medium-voltage grid model for distribution power supply systems. *Journal of Physics: Conference Series*, 977(1):012007.

Ampimah, B. C., Sun, M., Han, D., and Wang, X. (2018). Optimizing sheddable and shiftable residential electricity consumption by incentivized peak and off-peak credit function approach. *Applied Energy*, 210:1299–1309.

Anees, A., Dillon, T., Wallis, S., and Chen, Y.-P. P. (2021). Optimization of day-ahead and real-time prices for smart home community. *International Journal of Electrical Power & Energy Systems*, 124:106403.

Angeli, D., Dong, Z., and Strbac, G. (2023). Exact aggregate models for optimal management of heterogeneous fleets of storage devices. *IEEE Transactions on Control of Network Systems*.

Angenendt, G., Zurmühlen, S., Axelsen, H., and Sauer, D. U. (2018). Comparison of different operation strategies for pv battery home storage systems including forecast-based operation strategies. *Applied energy*, 229:884–899.

ANIE (2023). Osservatorio Sistemi di Accumulo. Retrieved from `https://anie.it/osservatorio-sistemi-di-accumulo/`. Accessed 2024-01-09.

Aniello, G. and Bertsch, V. (2023). Shaping the energy transition in the residential sector: Regulatory incentives for aligning household and system perspectives. *Applied Energy*, 333:120582.

Arlt, M.-L., Chassin, D., Rivetta, C., and Sweeney, J. (2024). Impact of real-time pricing and residential load automation on distribution systems. *Energy Policy*, 184:113906.

Atef, S. and Eltawil, A. B. (2020). Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. *Electric Power Systems Research*, 187:106489.

Autorità di Regolazione per Energia Reti e Ambiente (2008). Testo integrato delle modalità e delle condizioni tecnico-economiche per lo scambio sul posto (TISP). Retrieved from `https://www.arera.it/atti-e-provvedimenti/dettaglio/08/074-08arg`. Accessed 2024-03-11.

Avacon Netz GmbH (2019). Preisblätter Strom gesamt 2019. Retrieved from `https://www.avacon-netz.de/content/dam/revu-global/avacon-netz/documents/netzentgelte-strom`. Accessed 2024-03-20.

Avau, M., Govaerts, N., and Delarue, E. (2021). Impact of distribution tariffs on prosumer demand response. *Energy Policy*, 151:112116.

Avilés, C., Oliva, S., and Watts, D. (2019). Single-dwelling and community renewable microgrids: Optimal sizing and energy management for new business models. *Applied Energy*, 254:113665.

Ayón, X., Gruber, J. K., Hayes, B. P., Usaola, J., and Prodanović, M. (2017). An optimal day-ahead load scheduling approach based on the flexibility of aggregate demands. *Applied Energy*, 198:1–11.

Azad, S. A., Ali, A. S., and Wolfs, P. (2014). Identification of typical load profiles using k-means clustering algorithm. In *Asia-Pacific World Congress on Computer Science and Engineering*, pages 1–6. IEEE.

Babrowski, S., Heinrichs, H., Jochem, P., and Fichtner, W. (2014). Load shift potential of electric vehicles in Europe. *journal of power sources*, 255:283–293.

Bacher, P., Madsen, H., Nielsen, H. A., and Perers, B. (2013). Short-term heat load forecasting for single family houses. *Energy and Buildings*, 65:101–112.

Baeten, B., Rogiers, F., and Helsen, L. (2017). Reduction of heat pump induced peak electricity use and required generation capacity through thermal energy storage and demand response. *Applied Energy*, 195:184–195.

Ballarini, I., Corgnati, S. P., and Corrado, V. (2014). Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of tabula project. *Energy policy*, 68:273–284.

Bardenet, R., Brendel, M., Kégl, B., and Sebag, M. (2013). Collaborative hyperparameter tuning. In *International conference on machine learning*, pages 199–207. PMLR.

Barone, G., Buonomano, A., Forzano, C., Palombo, A., and Russo, G. (2023). The role of energy communities in electricity grid balancing: A flexible tool for smart grid power distribution optimization. *Renewable and Sustainable Energy Reviews*, 187:113742.

Bayer, B., Matschoss, P., Thomas, H., and Marian, A. (2018). The German experience with integrating photovoltaic systems into the low-voltage grids. *Renewable Energy*, 119:129–141.

Bedi, A. S., Ahmad, M. W., Swapnil, S., Rajawat, K., Anand, S., et al. (2018). Online algorithms for storage utilization under real-time pricing in smart grid. *International Journal of Electrical Power & Energy Systems*, 101:50–59.

Bergaentzlé, C., Jensen, I. G., Skytte, K., and Olsen, O. J. (2019). Electricity grid tariffs as a tool for flexible energy systems: A Danish case study. *Energy Policy*, 126:12–21.

Berger, M. and Worlitschek, J. (2018). A novel approach for estimating residential space heating demand. *Energy*, 159:294–301.

Beyertt, A., Verwiebe, P., Seim, S., Milojkovic, F., and Müller-Kirchenbauer, J. (2020). Felduntersuchung zu Behavioral Energy Efficiency Potentialen von privaten Haushalten.

Bharadiya, J. (2023). Machine learning in cybersecurity: Techniques and challenges. *European Journal of Technology*, 7(2):1–14.

Bhela, S., Kekatos, V., and Veeramachaneni, S. (2017). Enhancing observability in distribution grids using smart meter data. *IEEE Transactions on Smart Grid*, 9(6):5953–5961.

Bloch, L., Holweger, J., Ballif, C., and Wyrsch, N. (2019). Impact of advanced electricity tariff structures on the optimal design, operation and profitability of a grid-connected PV system with energy storage. *Energy Informatics*, 2(16).

Bosch Home Comfort (2024). Stromverbrauch Wärmepumpe - Wissenswertes. Retrieved from `https://www.bosch-homecomfort.com/de/de/wohngebaeude/wissen/heizungsratgeber/waermepumpe/stromverbrauch-waermepumpe`. Accessed 2024-12-25.

Bouktif, S., Fiaz, A., Ouni, A., and Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(7):1636.

Bouktif, S., Fiaz, A., Ouni, A., and Serhani, M. A. (2020). Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting. *Energies*, 13(2):391.

Bracale, A., Carpinelli, G., De Falco, P., and Hong, T. (2017). Short-term industrial load forecasting: A case study in an italian factory. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6. IEEE.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Brockway, P. E., Sorrell, S., Semieniuk, G., Heun, M. K., and Court, V. (2021). Energy efficiency and economy-wide rebound effects: A review of the evidence and its implications. *Renewable and sustainable energy reviews*, 141:110781.

Brounen, D., Kok, N., and Quigley, J. M. (2013). Energy literacy, awareness, and conservation behavior of residential households. *Energy Economics*, 38:42–50.

Brown, T., Hörsch, J., and Schlachtberger, D. (2017). PyPSA: Python for power system analysis. *arXiv preprint arXiv:1707.09913*.

Bruninx, K., Pandžić, H., Le Cadre, H., and Delarue, E. (2019). On the interaction between aggregators, electricity markets and residential demand response providers. *IEEE Transactions on Power Systems*, 35(2):840–853.

Bucher, E., Fieseler, C., and Lutz, C. (2016). What's mine is yours (for a nominal fee)–exploring the spectrum of utilitarian to altruistic motives for internet-mediated sharing. *Computers in Human Behavior*, 62:316–326.

Bundesministerium der Justiz und für Verbraucherschutz (2023). Gesetz für den Ausbau erneuerbarer Energien (Erneuerbare-Energien-Gesetz - EEG 2023) § 52 Zahlungen bei Pflichtverstößen.

Bundesministerium für Wirtschaft und Klimaschutz (2023). Final Beschluss des Smart-Meter-Gesetzes. Retrieved from `https://www.bmwk.de/Redaktion/DE/Pressemitteilungen/2023/05/20230512-smart-meter-gesetz-final-beschlossen.html`. Accessed 2024-03-13.

Bundesnetzagentur (2019). Monitoringbericht VerbraucherKennzahlen 2019. Retrieved from `https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Monitoringberichte/Monitoringbericht_VerbraucherKennzahlen2019.pdf`. Accessed 2024-03-21.

Bundesnetzagentur (2023). Bundesnetzagentur sets rules for the integration of controllable consumer devices. Accessed: 2023-12-14.

Bundesnetzagentur (2023a). Preisbestandteile und tarife. Retrieved from https://www.bundesnetzagentur.de/DE/Vportal/Energie/PreiseAbschlaege/Tarife-table.html. Accessed 2024-02-25.

Bundesnetzagentur (2023b). Speicherpapier. Retrieved from https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/ErneuerbareEnergien/Speicherpapier.pdf?__blob=publicationFile&v=5. Accessed 2024-03-06.

Bundesnetzagentur (2024a). Eeg-förderung und -fördersätze. Retrieved from https://www.bundesnetzagentur.de/DE/Fachthemen/ElektrizitaetundGas/ErneuerbareEnergien/EEG_Foerderung/start.html. Accessed 2024-02-25.

Bundesnetzagentur (2024b). Erneuerbare Energien und EEG-Foerderung - Bundesnetzagentur. Retrieved from https://www.bundesnetzagentur.de/DE/Fachthemen/ElektrizitaetundGas/ErneuerbareEnergien/EEG_Foerderung/Archiv_Vergaetze/start.html. Accessed 2024-03-20.

Bundesnetzagentur (2024c). SMARD. Retrieved from https://www.smard.de/home. Accessed 2024-12-12.

Bundesnetzagentur for Electricity, Gas, Telecommunications, Post and Railway (2024). SMARD Marktdaten Downloadcenter. Retrieved from https://www.smard.de/home/downloadcenter/download-marktdaten/. Accessed 2024-03-20.

Bundesregierung (2023). Gesetz für den Ausbau erneuerbarer Energien (EEG). Retrieved from https://www.gesetze-im-internet.de/eeg_2014/. Accessed 2023-12-01.

Bundesverband der Energie- und Wasserwirtschaft (BDEW) (2024). BDEW Strompreisanalyse. Retrieved from https://www.bdew.de/service/daten-und-grafiken/bdew-strompreisanalyse/. Accessed 2024-10-10.

Bundesverband der Energie- und Wasserwirtschaft e.V. (2024). Standardlastprofile Strom. Retrieved from https://www.bdew.de/energie/standardlastprofile-strom/. Accessed 2024-11-07.

Bundesverband Photovoltaic Austria (2023). Fact Sheet PV Branche 2023. Retrieved from https://pvaustria.at/wp-content/uploads/2023_Fact_Sheet_PV_Branche.pdf. Accessed 2024-01-09.

Bundesverband Wärmepumpe (2023). Wärmepumpenabsatz 2022: Wachstum von 53 Prozent gegenüber dem Vorjahr. Retrieved from https://www.waermepumpe.de/presse/pressemitteilungen. Accessed 2023-04-02.

Bunn and Farmer (1985). Review of Short-term Forecasting Methods in the Electric Power Industry. *IEEE Transactions on Power Systems*.

Burger, S., Chaves-Ávila, J. P., Batlle, C., and Pérez-Arriaga, I. J. (2017). A review of the value of aggregators in electricity systems. *Renewable and Sustainable Energy Reviews*, 77:395–405.

Burger, S. P., Knittel, C. R., Pérez-Arriaga, I. J., Schneider, I., and Vom Scheidt, F. (2020). The efficiency and distributional effects of alternative residential electricity rate designs. *The Energy Journal*, 41(1):199–240.

Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514.

Buryk, S., Mead, D., Mourato, S., and Torriti, J. (2015). Investigating preferences for dynamic electricity tariffs: The effect of environmental and system benefit disclosure. *Energy Policy*, 80:190–195.

Cabot, C. and Villavicencio, M. (2024). Second-best electricity pricing in france: Effectiveness of existing rates in evolving power markets. *Energy Economics*, page 107673.

Çakır, U., Çomaklı, K., Çomaklı, Ö., and Karslı, S. (2013). An experimental exergetic comparison of four different heat pump systems working at same conditions: As air to air, air to water, water to water and water to air. *Energy*, 58:210–219.

Çakmak, H. K. and Hagenmeyer, V. (2022). Using open data for modeling and simulation of the all electrical society in eASiMOV. In *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*, pages 1–6. IEEE.

California Independent System Operator (2015). Flexible Resources Help Renewable Energy Sources. Retrieved from `https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf`. Accessed 2024-08-22.

California Public Utilities Commission (2023). NEM Revisit Proceeding (R.20-08-020). Retrieved from `https://www.cpuc.ca.gov/nemrevisit`. Accessed 2024-01-07.

Castaneda, M., Zapata, S., Cherni, J., Aristizabal, A. J., and Dyner, I. (2020). The long-term effects of cautious feed-in tariff reductions on photovoltaic generation in the UK residential sector. *Renewable Energy*, 155:1432–1443.

Chapelle, O. and Chang, Y. (2011). Yahoo! learning to rank challenge overview. In *Proceedings of the learning to rank challenge*, pages 1–24. PMLR.

Chaudry, M., Abeysekera, M., Hosseini, S. H. R., Jenkins, N., and Wu, J. (2015). Uncertainties in decarbonising heat in the uk. *Energy Policy*, 87:623–640.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Chicco, D., Warrens, M. J., and Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7:e623.

Chin, J.-X., Zufferey, T., Shyti, E., and Hug, G. (2019). Load forecasting of privacy-aware consumers. In *2019 IEEE Milan PowerTech*, pages 1–6. IEEE.

Chung, W. H., Gu, Y. H., and Yoo, S. J. (2022). District heater load forecasting based on machine learning and parallel CNN-LSTM attention. *Energy*, 246:123350.

Coignard, J., Janvier, M., Debusschere, V., Moreau, G., Chollet, S., and Caire, R. (2021). Evaluating forecasting methods in the context of local energy communities. *International Journal of Electrical Power & Energy Systems*, 131:106956.

Colmenar-Santos, A., Muñoz-Gómez, A.-M., Rosales-Asensio, E., Aznar, G. F., and Galan-Hernandez, N. (2022). Adaptive model predictive control for electricity management in the household sector. *International Journal of Electrical Power & Energy Systems*, 137:107831.

Consiglio, A., Cocco, F., and Zenios, S. A. (2008). Asset and liability modelling for participating policies with guarantees. *European Journal of Operational Research*, 186(1):380–404.

Costa, C. M., Barbosa, J. C., Gonçalves, R., Castro, H., Del Campo, F., and Lanceros-Méndez, S. (2021). Recycling and environmental issues of lithium-ion batteries: Advances, challenges and opportunities. *Energy Storage Materials*, 37:433–465.

Couture, T. and Gagnon, Y. (2010). An analysis of feed-in tariff remuneration models: Implications for renewable energy investment. *Energy policy*, 38(2):955–965.

Craven, B. and Islam, S. M. (2011). Ordinary least-squares regression. *The SAGE dictionary of quantitative management research*, 1:224–228.

Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., et al. (2001). Energyplus: creating a new-generation building energy simulation program. *Energy and buildings*, 33(4):319–331.

Croatian Parliament (2021). New law stimulates solar power for households. Retrieved from `https://www.sabor.hr/en/press/news/new-law-stimulates-solar-power-households`. Accessed 2024-03-11.

da Silva Pereira, E. J., Pinho, J. T., Galhardo, M. A. B., and Macêdo, W. N. (2014). Methodology of risk analysis by monte carlo method applied to power generation with renewable energy. *Renewable Energy*, 69:347–355.

Dai, S., Meng, F., Dai, H., Wang, Q., and Chen, X. (2021). Electrical peak demand forecasting- A review.

Dai, Y., Zhou, Q., Leng, M., Yang, X., and Wang, Y. (2022). Improving the Bi-LSTM model with XGBoost and attention mechanism: A combined approach for short-term power load prediction. *Applied Soft Computing*, 130:109632.

Dallinger, D. and Wietschel, M. (2012). Grid integration of intermittent renewable energy sources using price-responsive plug-in electric vehicles. *Renewable and Sustainable Energy Reviews*, 16(5):3370–3382.

Daneshzand, F., Coker, P. J., Potter, B., and Smith, S. T. (2023). Ev smart charging: How tariff selection influences grid stress and carbon reduction. *Applied Energy*, 348:121482.

Dehler, J., Keles, D., Telsnig, T., Fleischer, B., Baumann, M., Fraboulet, D., Faure-Schuyer, A., and Fichtner, W. (2017). Self-consumption of electricity from renewable sources. In *Europe's Energy Transition*, pages 225–236. Elsevier.

Deng, G. and Newton, P. (2017). Assessing the impact of solar pv on domestic electricity consumption: Exploring the prospect of rebound effects. *Energy Policy*, 110:313–324.

Denholm, P., O'Connell, M., Brinkman, G., and Jorgenson, J. (2015). Overgeneration from solar energy in california. a field guide to the duck chart. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States).

Department for Business, Energy and Industrial Strategy (2021). Home Energy Model Consultation. Retrieved from `https://assets.publishing.service.gov.uk/media/65e1f99a2f2b3b001c7cd879/home-energy-model-consultation.pdf`. Accessed 2024-03-13.

Department for Energy Security and Net Zero (2020). Smart Export Guarantee (SEG): earn money for exporting the renewable electricity you have generated. Retrieved from gov.uk. Accessed 2024-03-06.

Destatis (2024a). Pressemitteilung Nr. 038 vom 25. Juli 2024: Neues zur Statistik der Bevölkerung. Retrieved from `https://www.destatis.de/DE/Presse/Pressemitteilungen/2024/07/PD24_N038_43.html`. Accessed 2024-12-25.

Destatis (2024b). Raumwärme in privaten Haushalten. Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/private-haushalte/Tabellen/raumwaerme-haushalte.html. Accessed 2024-12-13.

Destatis (2024c). Stromverbrauch in privaten Haushalten. Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/UGR/private-haushalte/Tabellen/stromverbrauch-haushalte.html. Accessed 2024-12-13.

DeutscheWelle (2023). How Germany Plans to Phase Out Oil and Gas Heating. Retrieved from https://www.dw.com/en/how-germany-plans-to-phase-out-oil-and-gas-heating/a-64952051. Accessed 2023-03-11.

Dijkgraaf, E., van Dorp, T. P., and Maasland, E. (2018). On the effectiveness of feed-in tariffs in the development of solar photovoltaics. *The Energy Journal*, 39(1):81–100.

Dong, S., Kremers, E., Brucoli, M., Brown, S., and Rothman, R. (2020). Impact of household heterogeneity on community energy storage in the UK. *Energy Reports*, 6:117–123.

Dong, S. and Liu, Y. (2022). A privacy-preserving electricity theft detection (petd) scheme for smart grid. In *International Conference on Science of Cyber Security*, pages 44–61. Springer.

Doostizadeh, M. and Ghasemi, H. (2012). A day-ahead electricity pricing model based on smart metering and demand-side management. *Energy*, 46(1):221–230.

Duan, Y., Zhang, J., and Wang, X. (2023). Henry Hub Monthly Natural Gas Price Forecasting Using CEEMDAN-Bagging-HHO-SVR. *Frontiers in Energy Research*, 11:1323073.

Dube, M., Awodele, K., Olayiwola, O., and Akpeji, K. (2017). Short term load forecasting using ARIMA ANN and hybrid ANN-DWT. In *Southern African Universities Power Engineering Conference*, page 6.

Dudek, G. (2015). Short-term load forecasting using random forests. In *Intelligent Systems' 2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications*, pages 821–828. Springer.

Dutta, G. and Mitra, K. (2017). A literature review on dynamic pricing of electricity. *Journal of the Operational Research Society*, 68(10):1131–1145.

Dwork, C. (2006). Differential privacy. *International colloquium on automata, languages, and programming*, pages 1–12.

Eibl, G., Bao, K., Grassal, P.-W., Bernau, D., and Schmeck, H. (2018). The influence of differential privacy on short term electric load forecasting. *Energy Informatics*, 1(1):93–113.

El Kababji, S. and Srikantha, P. (2020). A data-driven approach for generating synthetic load patterns and usage habits. *IEEE Transactions on Smart Grid*, 11(6):4984–4995.

Emami, P., Sahu, A., and Graf, P. (2023). BuildingsBench: A large-scale dataset of 900k buildings and benchmark for short-term load forecasting. *Advances in Neural Information Processing Systems (NeurIPS)*.

Emhofer, J., Marx, K., Sporr, A., Barz, T., Nitsch, B., Wiesflecker, M., and Pink, W. (2022). Experimental demonstration of an air-source heat pump application using an integrated phase change material storage as a desuperheater for domestic hot water generation. *Applied Energy*, 305:117890.

Energy Systems Integration Group (2024). Grid Planning for Building Electrification. Retrieved from `https://www.esig.energy/wp-content/uploads/2024/10/ESIG-Grid-Planning-Building-Electrification-report-2024.pdf`. Accessed 2024-12-11.

Enrich, J., Li, R., Mizrahi, A., and Reguant, M. (2024). Measuring the impact of time-of-use pricing on electricity consumption: Evidence from spain. *Journal of Environmental Economics and Management*, 123:102901.

EPEX Spot (2023). EPEX Market Data. Retrieved from https://webshop.eex-group.com/. Accessed 2023-10-11.

Erdinc, O., Paterakis, N. G., Mendes, T. D., Bakirtzis, A. G., and Catalão, J. P. (2014). Smart household operation considering bi-directional EV and ESS utilization by real-time pricing-based DR. *IEEE Transactions on Smart Grid*, 6(3):1281–1291.

Essential Services Commission (2024). Minimum Feed-in Tariff. Retrieved from https://www.esc.vic.gov.au/electricity-and-gas/electricity-and-gas-tariffs-and-benchmarks/minimum-feed-tariff. Accessed 2024-09-26.

European Parliament and Council of the European Union (2019). Directive 2019/944 on common rules for the internal market for electricity.

European Parliament and the Council of the European Union (2019). Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on the common rules for the internal market for electricity and amending Directive 2012/27/EU. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0944. Accessed 2024-03-03.

Fan, G.-F., Zhang, L.-Z., Yu, M., Hong, W.-C., and Dong, S.-Q. (2022). Applications of random forest in multivariable response surface for short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, 139:108073.

Fan, S. and Chen, L. (2006). Short-term load forecasting based on an adaptive hybrid method. *IEEE Transactions on Power Systems*, 21(1):392–401.

Faria, P. and Vale, Z. (2011). Demand response in electrical energy supply: An optimal real time pricing approach. *Energy*, 36(8):5374–5384.

Faruqui, A., Harris, D., and Hledik, R. (2010). Unlocking the€ 53 billion savings from smart meters in the EU: How increasing the adoption of dynamic tariffs could make or break the EU's smart grid investment. *Energy Policy*, 38(10):6222–6231.

Fekri, M. N., Grolinger, K., and Mir, S. (2022). Distributed load forecasting using smart meter data: Federated learning with recurrent neural networks. *International Journal of Electrical Power & Energy Systems*, 137:107669.

Feng, D., Fang, K., and Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, 56(9):e2019WR026793.

Fernández, J. D., Menci, S. P., Lee, C. M., Rieger, A., and Fridgen, G. (2022). Privacy-preserving federated learning for residential short-term load forecasting. *Applied energy*, 326:119915.

Fett, D., Fraunholz, C., and Keles, D. (2021). Diffusion and system impact of residential battery storage under different regulatory settings. *Energy Policy*, 158:112543.

Fett, D., Keles, D., Kaschub, T., and Fichtner, W. (2019). Impacts of self-generation and self-consumption on German household electricity prices. *Journal of Business Economics*, 89:867–891.

Figgener, J., Stenzel, P., Kairies, K.-P., Linßen, J., Haberschusz, D., Wessels, O., Robinius, M., Stolten, D., and Sauer, D. U. (2021). The development of stationary battery storage systems in Germany–status 2020. *Journal of Energy Storage*, 33:101982.

Finck, C., Li, R., Kramer, R., and Zeiler, W. (2018). Quantifying demand flexibility of power-to-heat and thermal energy storage in the control of building heating systems. *Applied Energy*, 209:409–425.

Finck, C., Li, R., and Zeiler, W. (2020). Optimal control of demand flexibility under real-time pricing for heating systems in buildings: A real-life demonstration. *Applied energy*, 263:114671.

Fischer, D., Bernhardt, J., Madani, H., and Wittwer, C. (2017a). Comparison of control approaches for variable speed air source heat pumps considering time variable electricity prices and PV. *Applied Energy*, 204:93–105.

Fischer, D., Härtl, A., and Wille-Haussmann, B. (2015). Model for electric load profiles with high time resolution for German households. *Energy and Buildings*, 92:170–179.

Fischer, D. and Madani, H. (2017). On heat pumps in smart grids: A review. *Renewable and Sustainable Energy Reviews*, 70:342–357.

Fischer, D., Wolf, T., Scherer, J., and Wille-Haussmann, B. (2016). A stochastic bottom-up model for space heating and domestic hot water load profiles for German households. *Energy and Buildings*, 124:120–128.

Fischer, D., Wolf, T., Wapler, J., Hollinger, R., and Madani, H. (2017b). Model-based flexibility assessment of a residential heat pump pool. *Energy*, 118:853–864.

Fitzpatrick, P., D'Ettorre, F., De Rosa, M., Yadack, M., Eicker, U., and Finn, D. P. (2020). Influence of electricity prices on energy flexibility of integrated hybrid heat pump and thermal storage systems in a residential building. *Energy and Buildings*, 223:110142.

Fraga, C., Hollmuller, P., Schneider, S., and Lachal, B. (2018). Heat pump systems for multifamily buildings: Potential and constraints of several heat sources for diverse building demands. *Applied Energy*, 225:1033–1053.

Freier, J. and von Loessl, V. (2022). Dynamic electricity tariffs: Designing reasonable pricing schemes for private households. *Energy Economics*, 112:106146.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):1–37.

Gao, J., Chen, Y., Hu, W., and Zhang, D. (2023). An adaptive deep-learning load forecasting framework by integrating transformer and domain knowledge. *Advances in Applied Energy*, 10:100142.

Gefen, D., Karahanna, E., and Straub, D. (2003). Trust and tam in online shopping: An integrated model. *MIS Quarterly*, 27(1):51–90.

Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10):2451–2471.

Ghiani, E., Giordano, A., Nieddu, A., Rosetti, L., and Pilo, F. (2019). Planning of a smart local energy community: The case of Berchidda municipality (Italy). *Energies*, 12(24):4629.

Gielen, D., Boshell, F., Saygin, D., Bazilian, M. D., Wagner, N., and Gorini, R. (2019). The role of renewable energy in the global energy transformation. *Energy strategy reviews*, 24:38–50.

Gilbert, C., Browell, J., and Stephen, B. (2023). Probabilistic load forecasting for the low voltage network: forecast fusion and daily peaks. *Sustainable Energy, Grids and Networks*, 34:100998.

Goia, A., May, C., and Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, 26(4):700–711.

Golla, A., Henni, S., Staudt, P., and Weinhardt, C. (2020). Scaling the concept of citizen energy communities through a platform-based decision support system. *European Conference on Information Systems 2020, Marakesh*.

Golmohamadi, H., Larsen, K. G., Jensen, P. G., and Hasrat, I. R. (2021). Optimization of power-to-heat flexibility for residential buildings in response to day-ahead electricity price. *Energy and Buildings*, 232:110665.

Gong, M., Zhao, Y., Sun, J., Han, C., Sun, G., and Yan, B. (2022). Load forecasting of district heating system based on Informer. *Energy*, 253:124179.

Gottwalt, S., Ketter, W., Block, C., Collins, J., and Weinhardt, C. (2011). Demand side management—a simulation of household behavior under variable prices. *Energy policy*, 39(12):8163–8174.

Green, R. and Staffell, I. (2017). "Prosumage" and the British electricity market. *Economics of Energy & Environmental Policy*, 6(1):33–50.

Gregoire-Zawilski, M. and Siddiki, S. (2023). Evaluating diffusion in policy designs: A study of net metering policies in the United States. *Review of Policy Research*.

Groß, A., Lenders, A., Schwenker, F., Braun, D. A., and Fischer, D. (2021). Comparison of short-term electrical load forecasting methods for different building types. *Energy Informatics*, 4(S3):1–16.

Grundmeier, N., Hahn, A., Ihle, N., Runge, S., and Meyer-Barlag, C. (2014). A simulation based approach to forecast a demand load curve for a container terminal using battery powered vehicles. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 1711–1718. IEEE.

Gunkel, P. A., Jacobsen, H. K., Bergaentzlé, C.-M., Scheller, F., and Andersen, F. M. (2023). Variability in electricity consumption by category of consumer: The impact on electricity load profiles. *International Journal of Electrical Power & Energy Systems*, 147:108852.

Günther, C., Schill, W.-P., and Zerrahn, A. (2021). Prosumage of solar electricity: Tariff design, capacity investments, and power sector effects. *Energy Policy*, 152:112168.

Guo, B. and Weeks, M. (2022). Dynamic tariffs, demand response, and regulation in retail electricity markets. *Energy Economics*, 106:105774.

Gupta, R., Pena-Bello, A., Streicher, K. N., Roduner, C., Farhat, Y., Thöni, D., Patel, M. K., and Parra, D. (2021). Spatial analysis of distribution grid capacity and costs to enable massive deployment of PV, electric mobility and electric heating. *Applied Energy*, 287:116504.

Gurobi Optimization, LLC (2023). Gurobi. Retrieved from `https://www.gurobi.com/`. Accessed 2023-02-01.

Gürses-Tran, G., Körner, T. A., and Monti, A. (2022). Introducing explainability in sequence-to-sequence learning for short-term load forecasting. *Electric Power Systems Research*, 212:108366.

Haakana, J., Haapaniemi, J., Lassila, J., Partanen, J., Niska, H., and Rautiainen, A. (2018). Effects of electric vehicles and heat pumps on long-term electricity consumption scenarios for rural areas in the nordic environment. In *2018 15th international conference on the European energy market (EEM)*, pages 1–5. IEEE.

Haas, R., Auer, H., and Biermayr, P. (1998). The impact of consumer behavior on residential energy demand for space heating. *Energy and buildings*, 27(2):195–205.

Haben, S., Arora, S., Giasemidis, G., Voss, M., and Greetham, D. V. (2021). Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy*, 304:117798.

Haben, S. and Giasemidis, G. (2016). A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting. *International Journal of Forecasting*, 32(3):1017–1022.

Hagan, M. T. and Behr, S. M. (1987). The time series approach to short term load forecasting. *IEEE transactions on Power Systems*, 2(3):785–791.

Hahn, H., Meyer-Nieberg, S., and Pickl, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 199(3):902–907.

Haida, T. and Muto, S. (1994). Regression based peak load forecasting using a transformation technique. *IEEE Transactions on Power Systems*, 9(4):1788–1794.

Hammersley, J. (2013). *Monte carlo methods*. Springer Science & Business Media.

Hanny, L., Wagner, J., Buhl, H. U., Heffron, R., Körner, M.-F., Schöpf, M., and Weibelzahl, M. (2022). On the progress in flexibility and grid charges in light of the energy transition: The case of Germany. *Energy Policy*, 165:112882.

Haq, M. R. and Ni, Z. (2019). A new hybrid model for short-term electricity load forecasting. *IEEE access*, 7:125413–125423.

Hartner, M., Mayr, D., Kollmann, A., and Haas, R. (2017). Optimal sizing of residential PV-systems from a household and social cost perspective: A case study in Austria. *Solar Energy*, 141:49–58.

Häseler, S. and Wulf, A. J. (2024). Promoting real-time electricity tariffs for more demand response from german households: a review of four policy options. *Energy, Sustainability and Society*, 14(1):59.

He, Y., Luo, F., Sun, M., and Ranzi, G. (2023). Privacy-preserving and hierarchically federated framework for short-term residential load forecasting. *IEEE Transactions on Smart Grid*.

He, Y. and Tsang, K. F. (2021). Universities power energy management: A novel hybrid model based on iCEEMDAN and Bayesian optimized LSTM. *Energy Reports*, 7:6473–6488.

Hedegaard, K. and Balyk, O. (2013). Energy system investment model incorporating heat pumps with thermal storage in buildings and buffer tanks. *Energy*, 63:356–365.

Hedegaard, R. E., Pedersen, T. H., and Petersen, S. (2017). Multi-market demand response using economic model predictive control of space heating in residential buildings. *Energy and Buildings*, 150:253–261.

Heider, A., Kundert, L., Schachler, B., and Hug, G. (2023). Grid reinforcement costs with increasing penetrations of distributed energy resources. In *2023 IEEE Belgrade PowerTech*, pages 01–06. IEEE.

Henni, S., Staudt, P., and Weinhardt, C. (2021). A sharing economy for residential communities with pv-coupled battery storage: Benefits, pricing and participant matching. *Applied Energy*, 301:117351.

Herm, L.-V., Wanner, J., Seubert, F., and Janiesch, C. (2021). I Don't Get IT, but IT seems Valid! The Connection between Explainability and Comprehensibility in (X) AI Research. In *ECIS*.

Hertel, M., Beichter, M., Heidrich, B., et al. (2023). Transformer training strategies for forecasting multiple load time series. *Energy Informatics*, 6(Suppl 1):20.

Hertel, M., Ott, S., Schäfer, B., Mikut, R., Hagenmeyer, V., and Neumann, O. (2022). Evaluation of Transformer architectures for electrical load time-series forecasting. In *Proceedings 32. Workshop Computational Intelligence*, volume 1, page 93.

Herter, K. (2007). Residential implementation of critical-peak pricing of electricity. *Energy policy*, 35(4):2121–2130.

Herter, K. and Wayland, S. (2010). Residential response to critical-peak pricing of electricity: California evidence. *Energy*, 35(4):1561–1567.

Hesse, H. C., Martins, R., Musilek, P., Naumann, M., Truong, C. N., and Jossen, A. (2017). Economic optimization of component sizing for residential battery storage systems. *Energies*, 10(7):835.

Hirth, L. and Ziegenhagen, I. (2015). Balancing power and variable renewables: Three links. *Renewable and Sustainable Energy Reviews*, 50:1035–1051.

Hledik, R. (2014). Rediscovering residential demand charges. *The Electricity Journal*, 27(7):82–96.

HM Revenue and Customs (2024). VAT Energy-Saving Materials Relief. Retrieved from `https://www.gov.uk/government/publications/vat-energy-saving-materials-relief`. Accessed 2024-03-13.

Hoarau, Q. and Perez, Y. (2019). Network tariff design with prosumers and electromobility: Who wins, who loses? *Energy Economics*, 83:26–39.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Holmgren, W. F., Hansen, C. W., and Mikofski, M. A. (2018). pvlib python: A python package for modeling solar energy systems. *Journal of Open Source Software*, 3(29):884.

Hor, C.-L., Watson, S. J., and Majithia, S. (2006). Daily load forecasting and maximum demand estimation using ARIMA and GARCH. In *2006 International Conference on Probabilistic Methods Applied to Power Systems*, pages 1–6. IEEE.

Hou, H., Liu, C., Wang, Q., Wu, X., Tang, J., Shi, Y., and Xie, C. (2022). Review of load forecasting based on artificial intelligence methodologies, models, and challenges. *Electric Power Systems Research*, 210:108067.

Hou, S., Li, H., Yang, C., and Wang, L. (2020). A new privacy-preserving framework based on edge-fog-cloud continuum for load forecasting. In *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–8. IEEE.

Hu, Z., Bao, Y., Xiong, T., and Chiong, R. (2015). Hybrid filter–wrapper feature selection for short-term load forecasting. *Engineering Applications of Artificial Intelligence*, 40:17–27.

Huang, N., Lu, G., and Xu, D. (2016). A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. *Energies*, 9(10):767.

Hubert, T. and Grijalva, S. (2012). Modeling for residential electricity optimization in dynamic pricing environments. *IEEE Transactions on Smart Grid*, 3(4):2224–2231.

Imran, K. and Kockar, I. (2014). A technical comparison of wholesale electricity markets in North America and Europe. *Electric Power Systems Research*, 108:59–67.

International Energy Agency (2021). National Survey Report of PV Power Applications in Japan, 2020. Retrieved from `https://iea-pvps.org/wp-content/uploads/2021/09/NSR_Japan_2020.pdf`. Accessed 2024-03-13.

International Energy Agency (IEA) (2024a). Heat Pumps. Retrieved from `https://www.iea.org/energy-system/buildings/heat-pumps`. Accessed 2024-12-01.

International Energy Agency (IEA) (2024b). Renewables 2024: Global Overview. Retrieved from `https://www.iea.org/reports/renewables-2024/global-overview`. Accessed 2025-01-17.

J. Liu and L. E. Brown (2019). Prediction of Hour of Coincident Daily Peak Load. In *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5.

Jahangir, H., Tayarani, H., Gougheri, S. S., Golkar, M. A., Ahmadian, A., and Elkamel, A. (2020). Deep learning-based forecasting approach in smart grids with microclustering and bidirectional LSTM network. *IEEE Transactions on Industrial Electronics*, 68(9):8298–8309.

Javadi, M. S., Gough, M., Nezhad, A. E., Santos, S. F., Shafie-khah, M., and Catalão, J. P. (2022). Pool trading model within a local energy community considering flexible loads, photovoltaic generation and energy storage systems. *Sustainable Cities and Society*, 79:103747.

Jesper, M., Pag, F., Vajen, K., and Jordan, U. (2021). Annual industrial and commercial heat load profiles: Modeling based on k-means clustering and regression analysis. *Energy Conversion and Management: X*, 10:100085.

Jessen, S. H., Ma, Z. G., Wijaya, F. D., Vasquez, J. C., Guerrero, J., and Jørgensen, B. N. (2022). Identification of natural disaster impacted electricity load profiles with k means clustering algorithm. *Energy Informatics*, 5(4):1–29.

Jia, J.-J., Guo, J., and Wei, C. (2021). Elasticities of residential electricity demand in china under increasing-block pricing constraint: New estimation using household survey data. *Energy Policy*, 156:112440.

Jiao, R., Zhang, T., Jiang, Y., and He, H. (2018). Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network. *IEEE Access*, 6:59438–59448.

Jin, X.-B., Zheng, W.-Z., Kong, J.-L., Wang, X.-Y., Bai, Y.-T., Su, T.-L., and Lin, S. (2021). Deep-learning forecasting method for electric power load via attention-based encoder-decoder with bayesian optimization. *Energies*, 14(6):1596.

Johnson, E., Beppler, R., Blackburn, C., Staver, B., Brown, M., and Matisoff, D. (2017). Peak shifting and cross-class subsidization: The impacts of solar PV on changes in electricity costs. *Energy Policy*, 106:436–444.

Juberias, G., Yunta, R., Moreno, J. G., and Mendivil, C. (1999). A new ARIMA model for hourly load forecasting. *1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333)*, 1:314–319.

Junker, R. G., Kallesøe, C. S., Real, J. P., Howard, B., Lopes, R. A., and Madsen, H. (2020). Stochastic nonlinear modelling and application of price-based energy flexibility. *Applied Energy*, 275:115096.

Kaiser, K., Kreft, M., Stai, E., Vayá, M. G., Staake, T., and Hug, G. (2023). Reducing power peaks in low-voltage grids via dynamic tariffs and automatic load control. In *27th International Conference on Electricity Distribution (CIRED 2023)*, page 1475–1479.

Kanda, I. and Veguillas, J. Q. (2019). Data preprocessing and quantile regression for probabilistic load forecasting in the GEFCom2017 final match. *International Journal of Forecasting*, 35(4):1460–1468.

Kapoor, S. and Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9).

Karijadi, I. and Chou, S.-Y. (2022). A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction. *Energy and Buildings*, 259:111908.

Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X., Sachdeva, S., et al. (2019). Evaluation of different machine learning approaches to forecasting pm2. 5 mass concentrations. *Aerosol and Air Quality Research*, 19(6):1400–1410.

Kaschub, T., Jochem, P., and Fichtner, W. (2016). Solar energy storage in german households: profitability, load changes and flexibility. *Energy policy*, 98:520–532.

Katz, J., Andersen, F. M., and Morthorst, P. E. (2016). Load-shift incentives for household demand response: Evaluation of hourly dynamic pricing and rebate schemes in a wind-based electricity system. *Energy*, 115:1602–1616.

Kell, A., McGough, A. S., and Forshaw, M. (2018). Segmenting residential smart meter data for short-term load forecasting. In *Proceedings of the Ninth International Conference on Future Energy Systems*, pages 91–96.

Khan, A. R., Mahmood, A., Safdar, A., Khan, Z. A., and Khan, N. A. (2016). Load forecasting, dynamic pricing and DSM in smart grid: A review. *Renewable and Sustainable Energy Reviews*, 54:1311–1322.

Khuntia, S. R., Rueda, J. L., and van Der Meijden, M. A. (2016). Forecasting the load of electrical power systems in mid-and long-term horizons: a review. *IET Generation, Transmission & Distribution*, 10(16):3971–3977.

Kircher, K. J. and Zhang, K. M. (2021). Heat purchase agreements could lower barriers to heat pump adoption. *Applied Energy*, 286:116489.

Klaiber, S., Bretschneider, P., Waczowicz, S., Mikut, R., Konotop, I., and Westermann, D. (2015). A contribution to the load forecast of price elastic consumption behaviour. In *2015 IEEE Eindhoven PowerTech*, pages 1–6. IEEE.

Klein, M., Ziade, A., and De Vries, L. (2019). Aligning prosumers with the electricity wholesale market–the impact of time-varying price signals and fixed network charges on solar self-consumption. *Energy Policy*, 134:110901.

Koltermann, L., Jacqué, K., Figgener, J., Zurmühlen, S., and Sauer, D. U. (2023). Operational validation of a power distribution algorithm for a modular megawatt battery storage system. *Batteries & Supercaps*, 6(3):e202200414.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., and Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE transactions on smart grid*, 10(1):841–851.

Koprinska, I., Rana, M., and Agelidis, V. G. (2015). Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*, 82:29–40.

Kotthoff, F., Muschner, C., Tepe, D., Pleßmann, G., and Hülk, L. (2024). open-mastr: A Python package to download and process the German energy registry marktstammdatenregister. *Journal of Open Source Software*, 9(100):6758.

Kou, X., Wang, R., Du, S., Xu, Z., and Zhu, X. (2024). Heat pump assists in energy transition: Challenges and approaches. *DeCarbon*, 3:100033.

Krishnamurthy, D., Uckun, C., Zhou, Z., Thimmapuram, P. R., and Botterud, A. (2017). Energy storage arbitrage under day-ahead and real-time price uncertainty. *IEEE Transactions on Power Systems*, 33(1):84–93.

Kucevic, D., Englberger, S., Sharma, A., Trivedi, A., Tepe, B., Schachler, B., Hesse, H., Srinivasan, D., and Jossen, A. (2021a). Reducing grid peak load through the coordinated control of battery energy storage systems located at electric vehicle charging parks. *Applied energy*, 295:116936.

Kucevic, D., Semmelmann, L., Collath, N., Jossen, A., and Hesse, H. (2021b). Peak shaving with battery energy storage systems in distribution grids: a novel approach to reduce local and global peak loads. *Electricity*, 2(4):573–589.

Kucevic, D., Tepe, B., Englberger, S., Parlikar, A., Mühlbauer, M., Bohlen, O., Jossen, A., and Hesse, H. (2020). Standard battery energy storage system profiles: Analysis of various applications for stationary energy storage systems using a holistic simulation framework. *Journal of Energy Storage*, 28:101077.

Kuhn, M. and Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.

Kühnbach, M., Stute, J., and Klingler, A.-L. (2021). Impacts of avalanche effects of price-optimized electric vehicle charging-does demand response make it worse? *Energy Strategy Reviews*, 34:100608.

Kuster, C., Rezgui, Y., and Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, 35:257–270.

Lahouar, A. and Slama, J. B. H. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management*, 103:1040–1051.

Lamp, S. and Samano, M. (2022). Large-scale battery storage, short-term market outcomes, and arbitrage. *Energy Economics*, 107:105786.

Le, N. K., Liu, Y., Nguyen, Q. M., Liu, Q., Liu, F., Cai, Q., and Hirche, S. (2021). Fedxgboost: Privacy-preserving XGBoost for federated learning. *arXiv preprint arXiv:2106.10662*.

Le Dréau, J. and Heiselberg, P. (2016). Energy flexibility of residential buildings using short term heat storage in the thermal mass. *Energy*, 111:991–1002.

Lechwerke AG (2021). FLAIR^2: Regional erzeugten Strom vor Ort verbrauchen. Retrieved from `https://www.lew.de/ueber-lew/zukunftsprojekte/flair2`. Accessed 2024-03-08.

Lee, C.-M. and Ko, C.-N. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, 38(5):5902–5911.

Lee, J. and Cho, Y. (2022). National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? *Energy*, 239:122366.

Lévesque, J.-C., Gagné, C., and Sabourin, R. (2016). Bayesian hyperparameter optimization for ensemble learning. *arXiv preprint arXiv:1605.06394*.

Li, C., Chen, Z., Liu, J., Li, D., Gao, X., Di, F., Li, L., and Ji, X. (2019). Power load forecasting based on the combined model of LSTM and XGBoost. In *Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence*, pages 46–51.

Li, H., Yeo, J. H., Bornsheuer, A. L., and Overbye, T. J. (2020a). The creation and validation of load time series for synthetic electric power systems. *IEEE Transactions on Power Systems*, 36(2):961–969.

Li, K., Huang, W., Hu, G., and Li, J. (2023a). Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. *Energy and Buildings*, 279:112666.

Li, N., Bruninx, K., and Tindemans, S. (2023b). Residential demand-side flexibility provision under a multi-level segmented tariff. In *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*.

Li, Q., Wu, Z., Wen, Z., and He, B. (2020b). Privacy-preserving gradient boosting decision trees. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 784–791.

Li, Y., Ou, J., and Gu, C. (2023c). Buyer guarantee and bailout in supplier finance with bankruptcy cost. *European Journal of Operational Research*, 305(1):287–299.

Li, Z., Xie, F., Zhang, H., and Zhang, H. (2024). Signaling quality through price guarantee window for technology-related products. *European Journal of Operational Research*, 313(2):669–677.

Liang, R.-H. and Cheng, C.-C. (2002). Short-term load forecasting by a neuro-fuzzy based approach. *International Journal of Electrical Power & Energy Systems*, 24(2):103–111.

Liao, H. and Chuang, A. (2004). A multilevel investigation of factors influencing employee service performance and customer outcomes. *Academy of Management journal*, 47(1):41–58.

Liao, X., Cao, N., Li, M., and Kang, X. (2019). Research on short-term load forecasting using XGBoost based on similar days. In *2019 International conference on intelligent transportation, big data & smart city (ICITBS)*, pages 675–678. IEEE.

Lin, J., Ma, J., Zhu, J., and Cui, Y. (2022). Short-term load forecasting based on LSTM networks considering attention mechanism. *International Journal of Electrical Power & Energy Systems*, 137:107818.

Linssen, J., Stenzel, P., and Fleer, J. (2017). Techno-economic analysis of photovoltaic battery systems and the influence of different consumer load profiles. *Applied Energy*, 185:2019–2025.

Liu, J. and Brown, L. E. (2019). Prediction of hour of coincident daily peak load. In *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE.

Logenthiran, T., Srinivasan, D., and Shun, T. Z. (2012). Demand side management in smart grid using heuristic optimization. *IEEE Transactions on Smart Grid*, 3(3):1244–1252.

Lokeshgupta, B. and Sivasubramani, S. (2019). Multi-objective home energy management with battery energy storage systems. *Sustainable Cities and Society*, 47:101458.

Londo, M., Matton, R., Usmani, O., van Klaveren, M., Tigchelaar, C., and Brunsting, S. (2020). Alternatives for current net metering policy for solar pv in the netherlands: A comparison of impacts on business case and purchasing behaviour of private homeowners, and on governmental costs. *Renewable Energy*, 147:903–915.

Love, J., Smith, A. Z., Watson, S., Oikonomou, E., Summerfield, A., Gleeson, C., Biddulph, P., Chiu, L. F., Wingfield, J., Martin, C., et al. (2017). The addition

of heat pump electricity load profiles to GB electricity demand: Evidence from a heat pump field trial. *Applied Energy*, 204:332–342.

Luo, F., Kong, W., Ranzi, G., and Dong, Z. Y. (2019). Optimal home energy management system with demand charge tariff and appliance operational dependencies. *IEEE Transactions on Smart Grid*, 11(1):4–14.

Luo, L., Kannan, P., and Ratchford, B. T. (2008). Incorporating subjective characteristics in product design and evaluations. *Journal of Marketing Research*, 45(2):182–194.

Luo, M. and Wu, S. (2018). A value-at-risk approach to optimisation of warranty policy. *European Journal of Operational Research*, 267(2):513–522.

Luo, Na and Hong, Tianzhen (2022). Ecobee Donate Your Data: 1,000 homes in 2017. Retrieved from `https://www.osti.gov/biblio/1854924`. Accessed 2024-08-04.

Luthander, R., Widén, J., Nilsson, D., and Palm, J. (2015). Photovoltaic self-consumption in buildings: A review. *Applied energy*, 142:80–94.

Ma, J., Cheng, J. C., Xu, Z., Chen, K., Lin, C., and Jiang, F. (2020). Identification of the most influential areas for air pollution control using XGBoost and grid importance rank. *Journal of Cleaner Production*, 274:122835.

MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297. University of California Los Angeles LA USA.

Maheshwari, A., Paterakis, N. G., Santarelli, M., and Gibescu, M. (2020). Optimizing the operation of energy storage using a non-linear lithium-ion battery degradation model. *Applied Energy*, 261:114360.

Maranghi, F., Gosselin, L., Raymond, J., and Bourbonnais, M. (2023). Modeling of solar-assisted ground-coupled heat pumps with or without batteries in remote high north communities. *Renewable Energy*.

Markard, J. (2018). The next phase of the energy transition and its implications for research and policy. *Nature Energy*, 3(8):628–633.

Märkle-Huß, J., Feuerriegel, S., and Neumann, D. (2018). Contract durations in the electricity market: Causal impact of 15 min trading on the EPEX spot market. *Energy Economics*, 69:367–378.

Marktstammdatenregister (2024). MaStR - Marktstammdatenregister. Retrieved from `https://www.marktstammdatenregister.de/MaStR`. Accessed 2024-12-24.

Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., and Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting. *Energy*, 214:118874.

Mateo, C., Cossent, R., Gómez, T., Prettico, G., Frías, P., Fulli, G., Meletiou, A., and Postigo, F. (2018). Impact of solar pv self-consumption policies on distribution networks and regulatory implications. *Solar Energy*, 176:62–72.

Matisoff, D. C., Beppler, R., Chan, G., and Carley, S. (2020). A review of barriers in implementing dynamic electricity pricing to achieve cost-causality. *Environmental Research Letters*, 15(9):093006.

Mayrink, V. and Hippert, H. S. (2016). A hybrid method using exponential smoothing and gradient boosting for electrical short-term load forecasting. In *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pages 1–6. IEEE.

McKenna, R., Hernando, D. A., ben Brahim, T., Bolwig, S., Cohen, J. J., and Reichl, J. (2021). Analyzing the energy system impacts of price-induced demand-side-flexibility with empirical data. *Journal of cleaner production*, 279:123354.

McPherson, M. and Stoll, B. (2020). Demand response for variable renewable energy integration: A proposed approach and its impacts. *Energy*, 197:117205.

Meier, H. (2000). Practical application of the VDEW standard load profiles. analytical vs. synthetic load profiles; anwendung der VDEW-lastprofile. analytisches versus synthetisches verfahren. *ET, Energiewirtschaftliche Tagesfragen*, 50.

Miletić, M., Gržanić, M., Pavić, I., Pandžić, H., and Capuder, T. (2022). The effects of household automation and dynamic electricity pricing on consumers and suppliers. *Sustainable energy, grids and networks*, 32:100931.

Mishra, A., Irwin, D., Shenoy, P., Kurose, J., and Zhu, T. (2012). Smartcharge: Cutting the electricity bill in smart homes with energy storage. In *Proceedings of the 3rd International Conference on Future Energy Systems: Where Energy, Computing and Communication Meet*, pages 1–10.

Moon, J. W. and Han, S.-H. (2011). Thermostat strategies impact on energy consumption in residential buildings. *Energy and Buildings*, 43(2-3):338–346.

Moshövel, J., Kairies, K.-P., Magnor, D., Leuthold, M., Bost, M., Gährs, S., Szczechowicz, E., Cramer, M., and Sauer, D. U. (2015). Analysis of the maximal possible grid relief from pv-peak-power impacts by using storage systems for increased self-consumption. *Applied Energy*, 137:567–575.

Müller, M., Blume, Y., and Reinhard, J. (2022). Impact of behind-the-meter optimised bidirectional electric vehicles on the distribution grid load. *Energy*, 255:124537.

Mulleriyawage, U. and Shen, W. (2020). Optimally sizing of battery energy storage capacity by operational optimization of residential PV-Battery systems: An Australian household case study. *Renewable Energy*, 160:852–864.

Mundaca, L. and Samahita, M. (2020). What drives home solar PV uptake? subsidies, peer effects and visibility in Sweden. *Energy Research & Social Science*, 60:101319.

Munem, M., Bashar, T. R., Roni, M. H., Shahriar, M., Shawkat, T. B., and Rahaman, H. (2020). Electric power load forecasting based on multivariate LSTM neural network using bayesian optimization. In *2020 IEEE Electric Power and Energy Conference (EPEC)*, pages 1–6. IEEE.

Muratori, M. (2018). Impact of uncoordinated plug-in electric vehicle charging on residential power demand. *Nature Energy*, 3(3):193–201.

Muzaffar, S. and Afshari, A. (2019). Short-term load forecasts using lstm networks. *Energy Procedia*, 158:2922–2927.

Nakai, M., Von Loessl, V., and Wetzel, H. (2024). Preferences for dynamic electricity tariffs: A comparison of households in germany and japan. *Ecological Economics*, 223:108239.

Naseri, N., Ghiassi-Farrokhfal, Y., Ketter, W., and Collins, J. (2023). Understanding and managing the participation of batteries in reserve electricity markets. *Decision Support Systems*, 165:113895.

Natras, R., Soja, B., and Schmidt, M. (2022). Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. *Remote Sensing*, 14(15):3547.

Naumann, M., Karl, R. C., Truong, C. N., Jossen, A., and Hesse, H. C. (2015). Lithium-ion battery cost analysis in pv-household application. *Energy Procedia*, 73:37–47.

Navarro-Espinosa, A. and Mancarella, P. (2014). Probabilistic modeling and assessment of the impact of electric heat pumps on low voltage distribution networks. *Applied Energy*, 127:249–266.

Netze BW (2021). Netze BW veröffentlicht vorläufige Netzentgelte für 2021. Retrieved from `https://www.netze-bw.de/News/netze-bw-veroeffentlicht-vorlaeufige-netzentgelte`. Accessed 2023-10-11.

Nicolson, M., Huebner, G., and Shipworth, D. (2017). Are consumers willing to switch to smart time of use electricity tariffs? the importance of loss-aversion and electric vehicle ownership. *Energy research & social science*, 23:82–96.

Nicolson, M. L., Fell, M. J., and Huebner, G. M. (2018). Consumer demand for time of use electricity tariffs: A systematized review of the empirical evidence. *Renewable and Sustainable Energy Reviews*, 97:276–289.

Nie, H., Liu, G., Liu, X., and Wang, Y. (2012). Hybrid of ARIMA and SVMs for short-term load forecasting. *Energy Procedia*, 16:1455–1460.

Nie, Y., H. Nguyen, N., Sinthong, P., and Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with Transformers. In *International Conference on Learning Representations*.

Nolting, L. and Praktiknjo, A. (2019). Techno-economic analysis of flexible heat pump controls. *Applied Energy*, 238:1417–1433.

Nti, I. K., Teimeh, M., Nyarko-Boateng, O., and Adekoya, A. F. (2020). Electricity load forecasting: a systematic review. *Journal of Electrical Systems and Information Technology*, 7(1):1–19.

Oldewurtel, F., Ulbig, A., Morari, M., and Andersson, G. (2011). Building control and storage management with dynamic tariffs for shaping demand response. In *2011 2nd IEEE PES international conference and exhibition on innovative smart grid technologies*, pages 1–8. IEEE.

Olk, C., Sauer, D. U., and Merten, M. (2019). Bidding strategy for a battery storage in the German secondary balancing power market. *Journal of Energy Storage*, 21:787–800.

Olsthoorn, M., Schleich, J., and Klobasa, M. (2015). Barriers to electricity load shift in companies: A survey-based exploration of the end-user perspective. *Energy Policy*, 76:32–42.

Ordóñez, Á., Sánchez, E., Rozas, L., García, R., and Parra-Domínguez, J. (2022). Net-metering and net-billing in photovoltaic self-consumption: The cases of Ecuador and Spain. *Sustainable Energy Technologies and Assessments*, 53:102434.

Ortega-Vazquez, M. A. (2014). Optimal scheduling of electric vehicle charging and vehicle-to-grid services at household level including battery degradation and price uncertainty. *IET Generation, Transmission & Distribution*, 8(6):1007–1016.

Ossenbrink, J. (2017). How feed-in remuneration design shapes residential PV prosumer paradigms. *Energy Policy*, 108:239–255.

Palensky, P. and Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics*, 7(3):381–388.

Pallante, A., Adacher, L., Botticelli, M., Pizzuti, S., Comodi, G., and Monteriu, A. (2020). Decision support methodologies and day-ahead optimization for smart building energy management in a dynamic pricing scenario. *Energy and Buildings*, 216:109963.

Pallonetto, F., Oxizidis, S., Milano, F., and Finn, D. (2016). The effect of time-of-use tariffs on the demand response flexibility of an all-electric smart-grid-ready dwelling. *Energy and Buildings*, 128:56–67.

Panapakidis, I. P. and Christoforidis, G. C. (2017). Implementation of modified versions of the k-means algorithm in power load curves profiling. *Sustainable Cities and Society*, 35:83–93.

Parra, D., Norman, S. A., Walker, G. S., and Gillott, M. (2017). Optimum community energy storage for renewable energy and demand load management. *Applied Energy*, 200:358–369.

Parra, D. and Patel, M. K. (2016). Effect of tariffs on the performance and economic benefits of PV-coupled battery systems. *Applied energy*, 164:175–187.

Patteeuw, D., Henze, G. P., and Helsen, L. (2016). Comparison of load shifting incentives for low-energy buildings with heat pumps to attain grid flexibility benefits. *Applied energy*, 167:80–92.

Pedro, A., Krutnik, M., Yadack, V. M., Pereira, L., and Morais, H. (2023). Opportunities and challenges for small-scale flexibility in european electricity markets. *Utilities Policy*, 80:101477.

Pena-Bello, A., Burer, M., Patel, M. K., and Parra, D. (2017). Optimizing pv and grid charging in combined applications to improve the profitability of residential batteries. *Journal of Energy Storage*, 13:58–72.

Peper, D., Längle, S., Muhr, M. C., and Reuther, T. (2022). Photovoltaik-und Batteriespeicherzubau in Deutschland in Zahlen-Auswertung des Markstammdatenregisters.

Pereira, D. S. and Marques, A. C. (2023). Are dynamic tariffs effective in reducing energy poverty? empirical evidence from us households. *Energy*, 282:128848.

Pergantis, E. N., Al Theeb, N., Dhillon, P., Ore, J. P., Ziviani, D., Groll, E. A., Kircher, K. J., et al. (2024). Field demonstration of predictive heating control for an all-electric house in a cold climate. *Applied Energy*, 360:122820.

Peters, D., Völker, R., Schuldt, F., and von Maydell, K. (2020). Are standard load profiles suitable for modern electricity grid models? In *2020 17th International Conference on the European Energy Market (EEM)*, pages 1–6. IEEE.

Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., and Staffell, I. (2017). The importance of open data and software: Is energy research lagging behind? *Energy Policy*, 101:211–215.

Pfenninger, S. and Staffell, I. (2016). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy*, 114:1251–1265.

Pillai, G. G., Putrus, G. A., and Pearsall, N. M. (2014). Generation of synthetic benchmark electrical load profiles using publicly available load and weather data. *International Journal of Electrical Power & Energy Systems*, 61:1–10.

Pinceti, A., Kosut, O., and Sankar, L. (2019). Data-driven generation of synthetic load datasets preserving spatio-temporal features. In *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE.

Poponi, D., Basosi, R., and Kurdgelashvili, L. (2021). Subsidisation cost analysis of renewable energy deployment: A case study on the Italian feed-in tariff programme for photovoltaics. *Energy Policy*, 154:112297.

Powell, K. M., Sriprasad, A., Cole, W. J., and Edgar, T. F. (2014). Heating, cooling, and electrical load forecasting for a large-scale district energy system. *Energy*, 74:877–885.

Protopapadaki, C. and Saelens, D. (2017). Heat pump and pv impact on residential low-voltage distribution grids as a function of building and district properties. *Applied Energy*, 192:268–281.

Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., and O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, 2:927312.

PyPi (2022). Bayesian Optimization PyPi Package: `bayesian-optimization` 1.4.3. Retrieved from `https://pypi.org/project/bayesian-optimization/`. Accessed 2022-12-04.

Rajabi, M. M. (2022). Dilemmas of energy efficiency: A systematic review of the rebound effect and attempts to curb energy consumption. *Energy Research & Social Science*, 89:102661.

Rajapaksha, D. and Bergmeir, C. (2022). Limref: Local interpretable model agnostic rule-based explanations for forecasting, with an application to electricity smart meter data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12098–12107.

Ran, P., Dong, K., Liu, X., and Wang, J. (2023). Short-term load forecasting based on CEEMDAN and Transformer. *Electric Power Systems Research*, 214:108885.

Reihani, E., Motalleb, M., Ghorbani, R., and Saoud, L. S. (2016). Load peak shaving and power smoothing of a distribution grid with high renewable energy penetration. *Renewable Energy*, 86:1372–1379.

Reijnders, V. M., van der Laan, M. D., and Dijkstra, R. (2020). Energy communities: a dutch case study. In *Behind and Beyond the Meter*, pages 137–155. Elsevier.

Reiner Lemoine Institut (2024). Ein offenes netzebenen- und sektorenübergreifendes Planungsinstrument zur Bestimmung des optimalen Einsatzes und Ausbaus von Flexibilitätsoptionen in Deutschland - Projektabschlussbericht.

Reiner Lemoine Institut gGmbH (2021). eDisGo - Electricity distribution grid optimization. Retrieved from `https://github.com/openego/eDisGo`. Accessed 2024-03-03.

Reis, I. F., Lopes, M. A., and Antunes, C. H. (2021). Energy literacy: an overlooked concept to end users' adoption of time-differentiated tariffs. *Energy Efficiency*, 14(4):39.

Rinaldi, A., Soini, M. C., Streicher, K., Patel, M. K., and Parra, D. (2021). Decarbonising heat with optimal pv and storage investments: A detailed sector coupling modelling framework with flexible heat pump operation. *Applied Energy*, 282:116110.

Roccatello, E., Prada, A., Baggio, P., and Baratieri, M. (2023). Impact of startup and defrosting on the modeling of hybrid systems in building energy simulations. *Journal of Building Engineering*, 65:105767.

Roozbehani, M., Dahleh, M. A., and Mitter, S. K. (2012). Volatility of power grids under real-time pricing. *IEEE Transactions on Power Systems*, 27(4):1926–1940.

Ruhnau, O., Bannik, S., Otten, S., Praktiknjo, A., and Robinius, M. (2019a). Direct or indirect electrification? a review of heat generation and road transport decarbonisation scenarios for germany 2050. *Energy*, 166:989–999.

Ruhnau, O., Hirth, L., and Praktiknjo, A. (2019b). Time series of heat demand and heat pump efficiency for energy system modeling. *Scientific data*, 6(1):189.

Rystad Energy (2023). New battery storage capacity to surpass 400 gwh per year by 2030 – 10 times current additions. Retrieved from https://www.rystadenergy.com. Accessed 2023-12-11.

Sachs, O., Tiefenbeck, V., Duvier, C., Qin, A., Cheney, K., Akers, C., and Roth, K. (2012). Field evaluation of programmable thermostats. Technical report, Fraunhofer Center for Sustainable Energy Systems (CSE), Cambridge, MA.

Saini, L. M. (2008). Peak load forecasting using bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks. *Electric power systems research*, 78(7):1302–1310.

Sarduy, J. R. G., Di Santo, K. G., and Saidel, M. A. (2016). Linear and non-linear methods for prediction of peak load at university of são paulo. *Measurement*, 78:187–201.

Savolainen, M. K. and Svento, R. (2012). Real-time pricing in the nordic power markets. *Energy economics*, 34(4):1131–1142.

Say, K. and John, M. (2021). Molehills into mountains: Transitional pressures from household pv-battery adoption under flat retail and feed-in tariffs. *Energy Policy*, 152:112213.

Say, K., John, M., and Dargaville, R. (2019). Power to the people: Evolutionary market pressures from residential PV battery investments in Australia. *Energy Policy*, 134:110977.

Say, K., Schill, W.-P., and John, M. (2020). Degrees of displacement: The impact of household pv battery prosumage on utility generation and storage. *Applied Energy*, 276:115466.

Schelly, C., Louie, E. P., and Pearce, J. M. (2017). Examining interconnection and net metering policy for distributed generation in the united states. *Renewable Energy Focus*, 22:10–19.

Schibuola, L., Scarpa, M., and Tambani, C. (2015). Demand response management by means of heat pumps controlled via real time pricing. *Energy and Buildings*, 90:15–28.

Schimpe, M., von Kuepach, M. E., Naumann, M., Hesse, H. C., Smith, K., and Jossen, A. (2018). Comprehensive modeling of temperature-dependent degradation mechanisms in lithium iron phosphate batteries. *Journal of The Electrochemical Society*, 165(2):A181.

Schittekatte, T., Mallapragada, D., Joskow, P. L., and Schmalensee, R. (2023). Reforming retail electricity rates to facilitate economy-wide decarbonization. *Joule*, 7(5):831–836.

Schittekatte, T., Momber, I., and Meeus, L. (2018). Future-proof tariff design: Recovering sunk grid costs in a world where consumers are pushing back. *Energy economics*, 70:484–498.

Schlemminger, M., Ohrdes, T., Schneider, E., and Knoop, M. (2022). Dataset on electrical single-family house and heat pump load profiles in Germany. *Scientific data*, 9(1):56.

Schlösser, T., Dunning, D., Johnson, K. L., and Kruger, J. (2013). How unaware are the unskilled? empirical tests of the "signal extraction" counterexplanation for the dunning–kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, 39:85–100.

Schlund, J., Pflugradt, N., Steber, D., Muntwyler, U., and German, R. (2018). Benefits of virtual community energy storages compared to individual batteries based on behaviour based synthetic load profiles. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6. IEEE.

Schneider, S. F., Novák, P., and Kober, T. (2020). Rechargeable batteries for simultaneous demand peak shaving and price arbitrage business. *IEEE Transactions on Sustainable Energy*, 12(1):148–157.

Schopfer, S., Tiefenbeck, V., and Staake, T. (2018). Economic assessment of photovoltaic battery systems based on household load profiles. *Applied energy*, 223:229–248.

Semmelmann, L., Henni, S., and Weinhardt, C. (2022). Load forecasting for energy communities: a novel LSTM-XGBoost hybrid model based on smart meter data. *Energy Informatics*, 5(S1):1–21.

Semmelmann, L., Jaquart, P., and Weinhardt, C. (2023a). Generating synthetic load profiles of residential heat pumps: a k-means clustering approach. *Energy Informatics*, 6(Suppl 1):37.

Semmelmann, L., Konermann, M., Dietze, D., and Staudt, P. (2024). Empirical field evaluation of self-consumption promoting regulation of household battery energy storage systems. *Energy Policy*, 194:114343.

Semmelmann, L., Resch, O., Henni, S., and Weinhardt, C. (2023b). Privacy-preserving peak time forecasting with learning to rank XGBoost and extensive feature engineering. *IET Smart Grid*.

Semmelmann, L., Schmid, D., Henni, S., Heider, A., Schachler, B., and Weinhardt, C. (2023c). On the impact of heat pump installations and peak blocking strategies

on grid expansion costs. In *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, pages 1–6. IEEE.

SENEC GmbH (2024). Stromspeicher: Alles, was Sie wissen müssen. Retrieved from `https://senec.com/de/magazin/stromspeicher`. Accessed 2024-06-29.

Sevlian, R. and Rajagopal, R. (2018). A scaling law for short term load forecasting on varying levels of aggregation. *International Journal of Electrical Power & Energy Systems*, 98:350–361.

Shaqour, A., Ono, T., Hagishima, A., and Farzaneh, H. (2022). Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building. *Energy and AI*, 8:100141.

Sheeran, P. and Orbell, S. (1999). Implementation intentions and repeated behaviour: Augmenting the predictive validity of the theory of planned behaviour. *European journal of social psychology*, 29(2-3):349–369.

Shen, X., Qiu, Y. L., Bo, X., Patwardhan, A., Hultman, N., and Dong, B. (2023). The impact of co-adopting electric vehicles, solar photovoltaics, and battery storage on electricity consumption patterns: Empirical evidence from Arizona. *Resources, Conservation and Recycling*, 192:106914.

Shrestha, A., Bishwokarma, R., Chapagain, A., Banjara, S., Aryal, S., Mali, B., Thapa, R., Bista, D., Hayes, B. P., Papadakis, A., et al. (2019). Peer-to-peer energy trading in micro/mini-grids for local energy communities: A review and case study of nepal. *IEEE Access*, 7:131911–131928.

Shwartz-Ziv, R. and Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90.

Sioshansi, R., Denholm, P., Jenkin, T., and Weiss, J. (2009). Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects. *Energy economics*, 31(2):269–277.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25.

Snoke, J., Raab, G. M., Nowok, B., Dibben, C., and Slavkovic, A. (2018). General and specific utility measures for synthetic data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 181(3):663–688.

So, K. C. and Song, J.-S. (1998). Price, delivery time guarantees and capacity selection. *European Journal of operational research*, 111(1):28–49.

Soini, M. C., Parra, D., and Patel, M. K. (2020). Impact of prosumer battery operation on the cost of power supply. *Journal of Energy Storage*, 29:101323.

SolarPower Europe (2022). New analysis reveals over 1 million European homes are solar battery-powered. Retrieved from Retrieved from SolarPower Europe. Accessed 2023-12-11.

Somers, M. and Whittaker, J. (2007). Quantile regression for modelling distributions of profit and loss. *European Journal of Operational Research*, 183(3):1477–1487.

Song, Y., Peskova, M., Rolando, D., Zucker, G., and Madani, H. (2023). Estimating electric power consumption of in-situ residential heat pump systems: A data-driven approach. *Applied Energy*, 352:121971.

Sørensen, Å. L., Lindberg, K. B., Sartori, I., and Andresen, I. (2021a). Analysis of residential EV energy flexibility potential based on real-world charging reports and smart meter data. *Energy and Buildings*, 241:110923.

Sørensen, Å. L., Lindberg, K. B., Sartori, I., and Andresen, I. (2021b). Residential electric vehicle charging datasets from apartment buildings. *Data in Brief*, 36:107105.

Sørensen, Å. L., Sartori, I., Lindberg, K. B., and Andresen, I. (2023). A method for generating complete ev charging datasets and analysis of residential charging behaviour in a large Norwegian case study. *Sustainable Energy, Grids and Networks*, 36:101195.

Sperber, E., Frey, U., and Bertsch, V. (2020). Reduced-order models for assessing demand response with heat pumps–insights from the German energy system. *Energy and Buildings*, 223:110144.

Srinivasan, D. and Lee, M. (1995). Survey of hybrid fuzzy neural approaches to electric load forecasting. In *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, volume 5, pages 4004–4008 vol.5.

Stecca, M., Elizondo, L. R., Soeiro, T. B., Bauer, P., and Palensky, P. (2020). A comprehensive review of the integration of battery energy storage systems into distribution networks. *IEEE Open Journal of the Industrial Electronics Society*, 1:46–65.

Stokke, A. V., Doorman, G. L., and Ericson, T. (2010). An analysis of a demand charge electricity grid tariff in the residential sector. *Energy Efficiency*, 3:267–282.

Strbac, G. (2008). Demand side management: Benefits and challenges. *Energy policy*, 36(12):4419–4426.

Stromnetz Berlin GmbH (2019). Standardlastprofil Haushalt 2019 (Berlin). Retrieved from `https://daten.berlin.de/datensaetze/standardlastprofil-haushalt-2019-berlin`. Accessed 2022-01-15.

Stute, J. and Klobasa, M. (2024). How do dynamic electricity tariffs and different grid charge designs interact?-implications for residential consumers and grid reinforcement requirements. *Energy Policy*, 189:114062.

Stute, J. and Kühnbach, M. (2023). Dynamic pricing and the flexible consumer–investigating grid and financial implications: A case study for Germany. *Energy Strategy Reviews*, 45:100987.

Stute, J., Pelka, S., Kühnbach, M., and Klobasa, M. (2024). Assessing the conditions for economic viability of dynamic electricity retail tariffs for households. *Advances in Applied Energy*, 14:100174.

SunWiz (2023). Australia's Battery Market Grows by 55%. Retrieved from `https://www.sunwiz.com.au/australias-battery-market-grows-by-55/`. Accessed 2024-02-05.

Syed, D., Abu-Rub, H., Zainab, A., Houchati, M., Bouhali, O., Ghrayeb, A., and Refaat, S. S. (2021). Investigation on optimizing cost function to penalize underestimation of load demand through deep learning modeling. In *IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–7. IEEE.

Syed, D., Zainab, A., Ghrayeb, A., Refaat, S. S., Abu-Rub, H., and Bouhali, O. (2020). Smart grid big data analytics: Survey of technologies, techniques, and applications. *IEEE Access*, 9:59564–59585.

Tang, X., Chen, H., Xiang, W., Yang, J., and Zou, M. (2022). Short-term load forecasting using channel and temporal attention based temporal convolutional network. *Electric Power Systems Research*, 205:107761.

Tang, X., Dai, Y., Wang, T., and Chen, Y. (2019). Short-term power load forecasting based on multi-layer bidirectional recurrent neural network. *IET Generation, Transmission & Distribution*, 13(17):3847–3854.

Tarsitano, A. and Amerise, I. L. (2017). Short-term load forecasting using a two-stage SARIMAX model. *Energy*, 133:108–114.

Tibber (2023). Dynamischer Stromtarif. Retrieved from `https://tibber.com/de/stromtarif/dynamischer-stromtarif`. Accessed 2024-03-13.

Tidemann, C., Engerer, N., Franklin, E., Hussey, K., and Pezzey, J. C. (2018). Promoting behind-the-meter battery storage: Options for more effective government support and regulation. *International Journal of Technology Intelligence and Planning*, 12(1):77–98.

Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE international Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4144–4147. IEEE.

Torriti, J. (2012). Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in northern italy. *Energy*, 44(1):576–583.

Tran, H.-Y., Hu, J., and Pota, H. R. (2022). Smart meter data obfuscation with a hybrid privacy-preserving data publishing scheme without a trusted third party. *IEEE Internet of Things Journal*, 9(17):16080–16095.

Trierweiler Ribeiro, G., Guilherme Sauer, J., Fraccanabbia, N., Cocco Mariani, V., and dos Santos Coelho, L. (2020). Bayesian optimized echo state network applied to short-term load forecasting. *Energies*, 13(9):2390.

Uddin, M., Romlie, M. F., Abdullah, M. F., Abd Halim, S., Kwang, T. C., et al. (2018). A review on peak load shaving strategies. *Renewable and Sustainable Energy Reviews*, 82:3323–3332.

Unternehmensserviceportal (USP) (2024). Elektrizitätsabgabe. Retrieved from `https://www.usp.gv.at/steuern-finanzen/verbrauchsteuern_und_energieabgaben/elektrizitaetsabgabe.html`. Accessed 2024-03-11.

Urban, T. L. (2009). Establishing delivery guarantee policies. *European Journal of Operational Research*, 196(3):959–967.

Vaillant Group (2024). Pufferspeicher allSTOR exclusiv. Retrieved from `https://www.vaillant.at/privatanwender/produkte/pufferspeicher-allstor-exclusiv-9792.html`. Accessed 2024-12-12.

Van Der Stelt, S., AlSkaif, T., and Van Sark, W. (2018). Techno-economic analysis of household and community energy storage for residential prosumers with smart appliances. *Applied Energy*, 209:266–276.

Van Lint, J. and van Zuylen, H. J. (2005). Monitoring and predicting freeway travel time reliability: Using width and skew of day-to-day travel time distribution. *Transportation Research Record*, 1917(1):54–62.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Vehviläinen, I. and Keppo, J. (2003). Managing electricity market price risk. *European Journal of Operational Research*, 145(1):136–147.

Venkataramanan, V., Kaza, S., and Annaswamy, A. M. (2022). Der forecast using privacy-preserving federated learning. *IEEE Internet of Things Journal*, 10(3):2046–2055.

Verbraucherzentrale Bundesverband e.V. (2024). Dynamische Stromtarife 2024: Repräsentative Befragung im Auftrag der Marktbeobachtung Energie. Retrieved from `https://www.vzbv.de/`. Accessed 2024-10-11.

Verhelst, C., Logist, F., Van Impe, J., and Helsen, L. (2012). Study of the optimal control problem formulation for modulating air-to-water heat pumps connected to a residential floor heating system. *Energy and buildings*, 45:43–53.

Vidović, K., Tomičić, I., Slovenec, K., Mikuc, M., and Brajdić, I. (2021). Ranking network devices for alarm prioritisation: Intrusion detection case study. In *2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–5. IEEE.

Visual Crossing (2024). Visual Crossing Weather. Retrieved from `https://www.visualcrossing.com/`. Accessed 2024-12-13.

vom Scheidt, F., Medinová, H., Ludwig, N., Richter, B., Staudt, P., and Weinhardt, C. (2020). Data analytics in the electricity sector–a quantitative and qualitative literature review. *Energy and AI*, 1:100009.

Walther, J., Spanier, D., Panten, N., and Abele, E. (2019). Very short-term load forecasting on factory level–a machine learning approach. *Procedia CIRP*, 80:705–710.

Wang, C., Chen, X., Jin, W., and Fan, X. (2022a). Credit guarantee types for financing retailers through online peer-to-peer lending: Equilibrium and coordinating strategy. *European Journal of Operational Research*, 297(1):380–392.

Wang, C., Wang, Y., Ding, Z., Zheng, T., Hu, J., and Zhang, K. (2022b). A transformer-based method of multienergy load forecasting in integrated energy system. *IEEE Transactions on Smart Grid*, 13(4):2703–2714.

Wang, H., Alattas, K. A., Mohammadzadeh, A., Sabzalian, M. H., Aly, A. A., and Mosavi, A. (2022c). Comprehensive review of load forecasting with emphasis on intelligent computing approaches. *Energy Reports*, 8:13189–13198.

Wang, K., Lai, X., Wen, F., Singh, P. P., Mishra, S., and Palu, I. (2023). Dynamic network tariffs: Current practices, key issues and challenges. *Energy Conversion and Economics*, 4(1):23–35.

Wang, P., Liu, B., and Hong, T. (2016). Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, 32(3):585–597.

Wang, S., Wang, X., Wang, S., and Wang, D. (2019a). Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, 109:470–479.

Wang, W., Shi, Y., Lyu, G., and Deng, W. (2017). Electricity consumption prediction using xgboost based on discrete wavelet transform. *DEStech Trans. Comput. Sci. Eng.*

Wang, X., Fang, F., Zhang, X., Liu, Y., Wei, L., and Shi, Y. (2019b). Lstm-based short-term load forecasting for building electricity consumption. In *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, pages 1418–1423. IEEE.

Wang, Y., Das, R., Putrus, G., and Kotter, R. (2020a). Economic evaluation of photovoltaic and energy storage technologies for future domestic energy systems–A case study of the UK. *Energy*, 203:117826.

Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., Guo, Y., and Wang, T. (2021). Short-term load forecasting of industrial customers based on SVMD and XGBoost. *International Journal of Electrical Power & Energy Systems*, 129:106830.

Wang, Z., Chen, Y., and Li, Y. (2019c). Development of rc model for thermal dynamic analysis of buildings through model structure simplification. *Energy and Buildings*, 195:51–67.

Wang, Z., Hong, T., and Piette, M. A. (2020b). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy*, 263:114683.

Wankmüller, F., Thimmapuram, P. R., Gallagher, K. G., and Botterud, A. (2017). Impact of battery degradation on energy arbitrage revenue of grid-level energy storage. *Journal of Energy Storage*, 10:56–66.

Wen, L., Zhou, K., Yang, S., and Lu, X. (2019). Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. *Energy*, 171:1053–1065.

Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, 30(1):79–82.

Winzer, C., Ramírez-Molina, H., Hirth, L., and Schlecht, I. (2024). Profile contracts for electricity retail customers. *Energy Policy*, 195:114358.

Wistuba, M., Schilling, N., and Schmidt-Thieme, L. (2015). Hyperparameter search space pruning–a new component for sequential model-based hyperparameter optimization. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part II 15*, pages 104–119. Springer.

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., and Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology*, 17(1):26–40.

Wu, Y., Liu, Z., Li, B., Liu, J., and Zhang, L. (2022). Energy management strategy and optimal battery capacity for flexible pv-battery system under time-of-use tariff. *Renewable Energy*, 200:558–570.

Wu, Z., Tazvinga, H., and Xia, X. (2015). Demand side management of photovoltaic-battery hybrid system. *Applied Energy*, 148:294–304.

Xu, C., Chen, H., Xun, W., Zhou, Z., Liu, T., Zeng, Y., and Ahmad, T. (2019). Modal decomposition based ensemble learning for ground source heat pump systems load forecasting. *Energy and Buildings*, 194:62–74.

Xu, X., Jin, X., Jia, H., Yu, X., and Li, K. (2015). Hierarchical management for integrated community energy systems. *Applied Energy*, 160:231–243.

Yang, J. (2011). Convergence and uncertainty analyses in monte-carlo based sensitivity analysis. *Environmental Modelling & Software*, 26(4):444–457.

Yang, Q. and Fang, X. (2017). Demand response under real-time pricing for domestic households with renewable DGs and storage. *IET generation, transmission & distribution*, 11(8):1910–1918.

Yang, Y., Javanroodi, K., and Nik, V. M. (2021). Climate change and energy performance of European residential building stocks–a comprehensive impact assessment using climate big data from the coordinated regional climate downscaling experiment. *Applied Energy*, 298:117246.

Yang, Z., Yang, F., Min, H., Tian, H., Hu, W., Liu, J., and Eghbalian, N. (2023). Energy management programming to reduce distribution network operating costs in the presence of electric vehicles and renewable energy sources. *Energy*, 263:125695.

Young, S., Bruce, A., and MacGill, I. (2019). Potential impacts of residential pv and battery storage on Australia's electricity networks under different tariffs. *Energy policy*, 128:616–627.

Yousefi, M., Hajizadeh, A., Soltani, M. N., and Hredzak, B. (2020). Predictive home energy management system with photovoltaic array, heat pump, and plug-in electric vehicle. *IEEE Transactions on Industrial Informatics*, 17(1):430–440.

Zafirakis, D., Chalvatzis, K. J., Baiocchi, G., and Daskalakis, G. (2016). The value of arbitrage for energy storage: Evidence from European electricity markets. *Applied energy*, 184:971–986.

Zakeri, B., Cross, S., Dodds, P. E., and Gissey, G. C. (2021). Policy options for enhancing economic profitability of residential solar photovoltaic with battery energy storage. *Applied Energy*, 290:116697.

Zeng, A., Chen, M., Zhang, L., and Xu, Q. (2023). Are Transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128.

Zhang, C., Kuppannagari, S. R., Xiong, C., Kannan, R., and Prasanna, V. K. (2019a). A cooperative multi-agent deep reinforcement learning framework for real-time residential load scheduling. In *Proceedings of the International Conference on Internet of Things Design and Implementation*, pages 59–69.

Zhang, L., Good, N., and Mancarella, P. (2019b). Building-to-grid flexibility: Modelling and assessment metrics for residential demand response from heat pump aggregations. *Applied Energy*, 233:709–723.

Zhang, Y., Tian, X., Zhao, Y., Zhang, C., Zhao, Y., and Lu, J. (2024). A prior-knowledge-based time series model for heat demand prediction of district heating systems. *Applied Thermal Engineering*, page 123696.

Zheng, H. and Wu, Y. (2019). A xgboost model with weather similarity analysis and feature engineering for short-term wind power forecasting. *Applied Sciences*, 9(15):3019.

Zheng, H., Yuan, J., and Chen, L. (2017). Short-term load forecasting using emd-lstm neural networks with a Xgboost algorithm for feature importance evaluation. *Energies*, 10(8):1168.

Zhou, K., Cheng, L., Lu, X., and Wen, L. (2020). Scheduling model of electric vehicles charging considering inconvenience and dynamic electricity prices. *Applied Energy*, 276:115455.

Zhou, L., Zhang, Y., Lin, X., Li, C., Cai, Z., and Yang, P. (2018). Optimal sizing of PV and BESS for a smart household considering different price mechanisms. *IEEE access*, 6:41050–41059.

Zhou, S. and Brown, M. A. (2017). Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes. *Journal of cleaner production*, 144:22–32.

Zhu, J., Dong, H., Zheng, W., Li, S., Huang, Y., and Xi, L. (2022). Review and prospect of data-driven techniques for load forecasting in integrated energy systems. *Applied Energy*, 321:119269.

Zhu, Z., Wang, Y., Yuan, M., Zhang, R., Chen, Y., Lou, G., and Sun, Y. (2023). Energy saving and carbon reduction schemes for families with the household pv-bes-ev system. *Energy and Buildings*, page 113007.

Ziel, F. (2018). Modeling public holidays in load forecasting: a German case study. *Journal of Modern Power Systems and Clean Energy*, 6(2):191–207.

Zufferey, T., Ulbig, A., Koch, S., and Hug, G. (2016). Forecasting of smart meter time series based on neural networks. In *International workshop on data analytics for renewable energy integration*, pages 10–21. Springer.