DOKTORI (PHD) ÉRTEKEZÉS

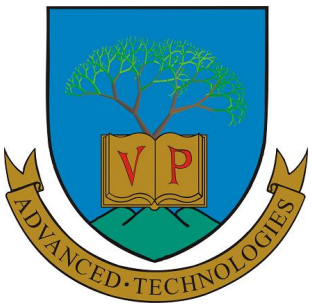
BALÁZS ISTVÁN GUSZTÁV

PANNON EGYETEM

INFORMATIKAI TUDOMÁNYOK DOKTORI ISKOLA

VESZPRÉM

2024



USE OF FORECASTING AND OPTIMIZATION TOOLS TO INCREASE THE FLEXIBILITY OF THE TRANSFORMING POWER SYSTEM

PhD Thesis

Author: IstvánGusztávBalázs

DOI:10.18136/PE.2025.917

Supervisors: Dr. AttilaFodor, Dr. AttilaMagyar

Universityof Pannonia

Doctoral School of Information Science and Technology

Veszprém

2024

Use of forecasting and optimization tools to increase the flexibility of the transforming power system

Az értekezés doktori (PhD) fokozat elnyerése érdekében készült a Pannon Egyetem Informatikai Tudományok Doktori Iskolája keretében

informatikai tudományok tudományágban

Írta:
Balázs István Gusztáv

Témavezetők:
Dr. Fodor Attila
Elfogadásra javaslom: igen / nem

.....
aláírás

Dr. Magyar Attila
Elfogadásra javaslom: igen / nem

.....
aláírás

Az értekezést bírálóként elfogadásra javaslom:

Bíráló neve: ..... igen / nem

.....
aláírás

Bíráló neve: ..... igen / nem

.....
aláírás

A jelölt az értekezés nyilvános vitáján ..... %-ot ért el.
Veszprém, ...

.....
a Bíráló Bizottság elnöke

A doktori (PhD) oklevél minősítése: .....
Veszprém, ...

.....
az EDHT elnöke

ACKNOWLEDGEMENT

First and foremost, I wish to express my deepest gratitude to my supervisors, Dr. Attila Fodor and Dr. Attila Magyar, whose guidance and encouragement were pivotal throughout this journey. Their commitment to professionalism and integrity, even when the road became challenging, was invaluable. Without their persistent support, the goal of this project would not have been realized.

I would also like to express my sincere thanks to Prof. Katalin Hangos, who was present at the crucial moments of my studies. Her insightful advice and guidance were instrumental in helping me navigate the right path.

Finally, I am deeply grateful to my wife, Zsuzsi, who provided unwavering emotional support and tirelessly managed our family life, ensuring that our home remained a place of balance and warmth. Her dedication allowed me the time and focus needed to complete this dissertation, and for that, I am profoundly appreciative. I am also thankful to my parents and sister for their support along the way. My parents' dedication to education has been a significant motivation for my PhD studies, and I deeply appreciate their encouragement throughout this journey.

ABSTRACT

The widespread deployment of distributed energy resources in the electricity system presents novel challenges for all actors in the energy value chain. The integration of climate-friendly generation technologies into the energy system is not a straightforward task, requiring more complex planning, management, and market processes than in the past. Technological advancements in information technology and telecommunications offer new tools to ensure the reliability and affordability of the electricity supply, while supporting the integration of environmentally friendly renewable energy sources at larger scales. This, in turn, contributes to a reduction in CO_2 emissions and reliance on hydrocarbon-based energy sources and their producers.

My research focuses on the integration of weather-dependent and variable distributed generation into electricity industry processes, with a particular emphasis on the use of advanced IT tools. I investigate a range of methods at three distinct levels of the electricity system, with the objective of enabling a broader range of participants to become active actors, thereby increasing the flexibility of the electricity system and the efficiency of the installed assets.

The initial issue to be addressed is the imbalance in the electricity system as a whole, which is characterized by a mismatch between production and consumption. Based on the assumption that system imbalance is correlated with measured values of system variables as well as predictions of exogenous variables, I propose a multi-step version of the Autoregressive Distributed Lag Model for the short-term forecast of system imbalance. The proposed forecasting model has been compared with a Long Short-Term Memory network-based procedure as well as with an Extra Trees Regression model using real data. The results show that the proposed multistep autoregressive forecasting model outperforms the others in all three evaluation metrics.

Management of portfolios containing a mix of generation and consumption assets involves a wide range of strategies, encompassing both internal optimizations and market participation. This analysis examines how a market-based aggregated portfolio using various technologies can realize economic benefits from imbalance forecasting and parallel activities in disparate but interconnected energy markets through automated strategies. My model generates time-series data representing energy transactions, revenues, and costs for an aggregator managing diverse assets. The data generated are based on realistic scenarios, reflecting true market pricing conditions and the dynamics of specific submarkets. The demonstration results indicate that even with a hybrid portfolio including weather-dependent, variable technologies, it is possible to achieve significant economic results through parallel participation in multiple

markets, using complex operational logic, and assuming advanced IT, measurement, and control backgrounds.

The potential of passive residential consumers to contribute to the operation of the electricity system has yet to be fully realized. I propose a framework for the estimation, as well as the prediction of the power flexibility of residential prosumers. In order to quantify the residential buildings' demand flexibility, a thermodynamic simulation model of a typical residential house was developed based on first engineering principles and integrated with the flexibility calculation of the building's electricity consumption and generation. Based on the calculated flexibility values, forecasting techniques, including ridge regression and extra trees regression, were incorporated to predict short-term flexibility. The results were validated by simulation experiments using real data for four different scenarios.

TARTALMI KIVONAT

Az elosztott energiaforrások széles körű elterjedése a villamosenergia-rendszerben új kihívások elé állítja az energetikai értéklánc valamennyi szereplőjét. A környezetbarát termelési technológiák integrálása az energiarendszerbe nem egyszerű feladat, a korábbinál összetettebb tervezési, irányítási és piaci folyamatokat igényel. Az informatika és a távközlés technológiai fejlődése új eszközöket kínál a villamosenergia-ellátás megbízhatóságának és megfizethetőségének biztosítására, miközben támogatja a környezetbarát megújuló energiaforrások nagyobb léptékű integrációját. Ez pedig hozzájárul a CO_2-kibocsátás és a szénhidrogén alapú energiaforrásoktól és azok termelőitől való függőség csökkentéséhez.

Kutatásaim középpontjában az időjárásfüggő és elosztott energiatermelés villamosenergia-ipari folyamatokba történő integrálása áll, különös hangsúlyt fektetve a fejlett informatikai eszközök használatára. A villamosenergia-rendszer három különböző szintjén vizsgálom a módszerek széles körét azzal a céllal, hogy a résztvevők minél szélesebb köre válhasson aktív szereplővé, növelve ezáltal a villamosenergia-rendszer rugalmasságát és az eszközök hatékonyságát.

Az első témám a villamosenergia-rendszer kiegyenlítetlensége, amely a pillanatnyi termelés és fogyasztás közti eltérésből adódik. Abból a feltételezésből kiindulva, hogy a rendszer kiegyenlítetlensége korrelál egyéb külső rendszerváltozók mért értékeivel, illetve előrejelzéseivel, a rendszer kiegyenlítetlenségének rövid távú előrejelzésére az autoregresszív elosztott késleltetésű modell (Autoregressive Distributed Lag Model) többlépésesre átalakított változatát használom. A javasolt előrejelzési modellt összehasonlítottam egy neurális hálózaton (Long Short-Term Memory), valamint egy döntési fákon (Extra Trees Regression) alapuló eljárással valós adatok felhasználásával. Az eredmények azt mutatják, hogy a javasolt többlépcsős autoregresszív előrejelző modell mindhárom értékelési metrika tekintetében felülmúlja a benchmark modelleket.

A termelési és fogyasztási eszközöket vegyesen tartalmazó portfóliók kezelése összetett működési logikák használatát követeli meg, beleértve az eszközökportfólió belső optimalizálását és különböző energiapiacokon történő párhuzamos részvételt is. A második terület azt vizsgálja, hogy a különböző technológiákat alkalmazó, piaci alapon működő aggregált portfólió milyen automatizált stratégiák révén működhet gazdaságosan úgy, hogy egyrészt kezeli a portfóliót alkotó eszközök eltérő műszaki és gazdasági adottságait, valamint kihasználja az eltérő, de összekapcsolt energiapiacokon realizálható lehetőségeket. A kidolgozott modell olyan idősoros kimeneteket generál, amelyek az energiatranzakciókat, bevételeket és költségeket reprezentálják egy különböző technológiájú eszközöket kezelő aggregátor számára. Az eredmények valós forgatókönyveken alapulnak, amelyek a piaci árképzési feltételeket és

az egyes részpiacok dinamikáját tükrözik. A demonstrációs eredmények azt mutatják, hogy még az időjárásfüggő, változó technológiákat tartalmazó hibrid portfólió esetén is lehetséges jelentős gazdasági eredményeket elérni több piacon történő párhuzamos részvétel, összetett működési logika alkalmazása, valamint fejlett informatikai, mérési és szabályozási háttér lehetőségeinek kiaknázása révén.

A többnyire passzív lakossági fogyasztók bevonása a villamosenergia-rendszer működtetésébe jelenleg még kezdeti szakaszban tart, a lakossági felhasználásban rejlő rugalmassági potenciál egyelőre nincs kihasználva, valójában ezen potenciál számszerűsítése sem megoldott. A lakóépületek keresleti rugalmasságának számszerűsítése érdekében egy tipikus családi ház termodinamikai szimulációs modelljét dolgoztam ki, majd a termodinamikai modellt integráltam az épület villamosenergia-fogyasztásának és -termelésének rugalmassági kalkulációjával. A számított rugalmassági potenciál alapján különböző előrejelzési technikákat, többek között ridge és extra trees regressziót használtam a rövid távú rugalmasság előrejelzésére. Az eredményeket valós adatokkal végzett szimulációval validáltam négy különböző forgatókönyvre vonatkozóan.

ZUSAMMENFASSUNG

Der weitverbreitete Einsatz dezentraler Energieressourcen im Elektrizitätssystem stellt alle Akteure in der Energiewertschöpfungskette neue Herausforderungen vor. Die Integration klimafreundlicher Erzeugungstechnologien in das Energiesystem wird nicht einfach, und erfordert komplexe Planungs-, Management- und Marktprozesse als es vorher wurde. Technologische Fortschritte in der Informationstechnologie und Telekommunikation bieten neue Instrumente an, um die Zuverlässigkeit und Erschwinglichkeit der Stromversorgung zu gewährleisten, und gleichzeitig die Integration umweltfreundlicher erneuerbarer Energiequellen in größerem Maßstab zu unterstützen. Dies wiederum trägt zu einer Verringerung der CO_2-Emissionen und der Abhängigkeit von kohlenwasserstoffbasierten Energiequellen und deren Erzeugern bei.

Meine Forschung konzentriert sich mit besonderem Schwerpunkt auf dem Einsatz moderner IT-Tools auf die Integration von wetterabhängiger, und variabler dezentraler Erzeugung in die Prozesse der Elektrizitätswirtschaft. Ich untersuche eine Reihe von Methoden auf drei verschiedenen Ebenen des Elektrizitätssystems mit dem Ziel, ein breiteres Spektrum von Teilnehmern in die Lage aktive Akteure versetzen zu werden und dadurch die Flexibilität des Elektrizitätssystems und die Effizienz der installierten Anlagen zu erhöhen.

Das erste Problem, das es zu lösen gilt, ist das Ungleichgewicht im Elektrizitätssystem als Ganzes, das durch ein Missverhältnis zwischen Erzeugung und Verbrauch gekennzeichnet ist. Ausgehend von der Annahme, dass das System Ungleichgewicht sowohl mit den gemessenen Werten der Systemvariablen als auch mit den Vorhersagen der exogenen Variablen korreliert ist, schlage ich eine mehrstufige Version des autoregressiven Distributed-Lag-Modells für die kurzfristige Prognose des Systemungleichgewichts vor. Das vorgeschlagene Prognosemodell wurde mit einem auf Netzwerk mit Langzeitgedächtnis basierenden Verfahren sowie mit einem Extrabaum-Regressionsmodell unter Verwendung wahren Daten verglichen. Die ausgewerteten Daten stellen dar, dass das vorgeschlagene mehrstufige autoregressive Prognosemodell in allen drei Bewertungsmaßstäben besser als die anderen abschneidet.

Die Verwaltung von Portfolios, die eine Mischung aus Erzeugungs- und Verbrauchsanlagen enthalten, umfasst eine breite Palette von Strategien, die sowohl interne Optimierungen, als auch die Teilnahme am Markt umfassen. In dieser Analyse wird es untersucht, wie ein marktbasiertes, aggregiertes Portfolio, welches verschiedene Technologien einsetzt, wirtschaftliche Vorteile aus Ungleichgewichtsprognosen und parallelen Aktivitäten in unterschiedlichen, aber miteinander verbundenen Energiemärkten durch automatisierte Strategien erzielen kann. Mein Modell generiert Zeitreihendaten, Energie Transaktionen, Einnahmen und Kosten für einen Aggregator darstellen, der verschiedene Anlagen verwaltet. Die

erzeugten Daten beruhen auf realistischen Szenarien, die die tatsächlichen Marktpreisbedingungen und die Dynamik bestimmter Teilmärkte widerspiegeln. Die Ergebnisse der Demonstration zeigen, dass es selbst mit einem hybriden Portfolio, das wetterabhängige, variable Technologien umfasst, möglich ist, durch die parallele Teilnahme an mehreren Märkten, die Verwendung einer komplexen Betriebslogik und die Annahme eines fortgeschrittenen IT-, Mess- und Steuerungshintergrunds bedeutende wirtschaftliche Ergebnisse zu erreichen.

Das Potenzial passiver Privatverbraucher zum Betrieb des Elektrizitätssystems beizutragen ist vollständig noch nicht ausgeschöpft. Ich schlage einen Rahmen für die Abschätzung sowie die Vorhersage der Stromflexibilität von Prosumern in Wohngebäuden vor. Um die Nachfrageflexibilität von Wohngebäuden zu quantifizieren, wurde ein thermodynamisches Simulationsmodell von einem typischen Wohnhaus auf der Grundlage erster technischer Prinzipien entwickelt, und mit der Flexibilitätsberechnung des Stromverbrauchs und mit der Stromerzeugung des Gebäudes integriert. Auf der Grundlage der berechneten Flexibilitätswerte wurden Prognosetechniken, einschließlich Ridge-Regression und Extrabaum-Regression eingesetzt, um die kurzfristige Flexibilität vorherzusagen. Die Ergebnisse wurden durch Simulationsexperimente mit wahren Daten für vier verschiedene Szenarien validiert.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Introduction

1.1. Motivation

Implications of distributed energy resources

In the contemporary landscape of power systems, the proliferation of Distributed Energy Resources (DERs) marks a paradigm shift, transitioning the electricity grid from a traditionally centralized architecture to a more decentralized and dynamic ecosystem (Figure 1.1). In the past, conventional power systems were characterized by large generation sources feeding power into the transmission grid, which was transported to distribution grids and then delivered to end users. Power flowed in one direction from the high-voltage transmission system to the end user on the low-voltage systems. Centralized, dispatchable, and predictable generation provided transmission-level flexibility to the electric system to balance generation and demand, implementing a generation-follows-load paradigm [78]. The increasing amount of distributed and renewable generation (from about 21% share of net electricity generation in 2010 to 44% in 2030 [62]) transforms the generation side into a more variable and intermittent energy source, so that the forecasting and control of solar [14] and wind [46] energy production pose a handful of problems. With the emergence and significant growth of distributed and intermittent generation, the load-following generation paradigm has become unsustainable.

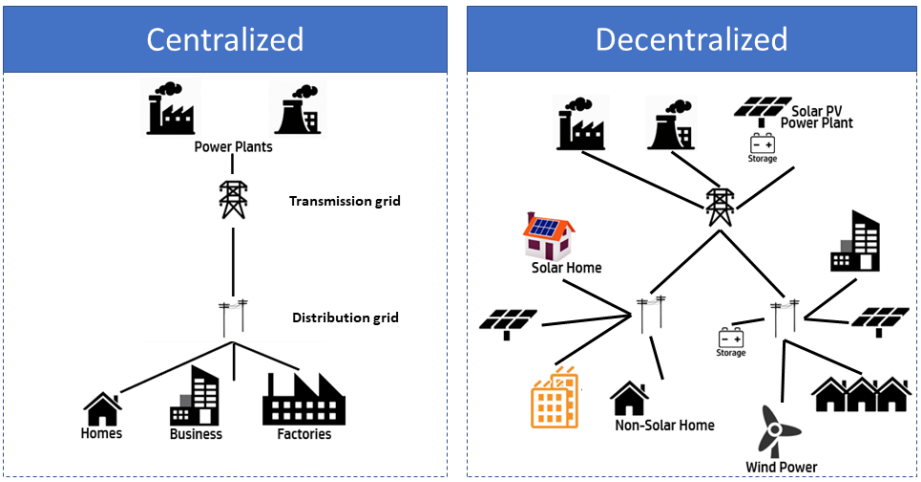


Figure 1.1: Power systems are changing from a centralized towards a distributed generation scheme

DERs, which encompass technologies such as solar panels, wind turbines, and electric vehi-

cles, represent a shift away from large, centralized power generation facilities. This decentralization democratizes energy production, enabling consumers to become both producers and consumers of energy. While this transition facilitates the integration of renewable energy sources and can result in a more sustainable grid, it introduces complexity in balancing supply and demand, as energy generation becomes more intermittent and less predictable [68].

The growth of DER generation presents new challenges to both system operators and system users. DER generation is typically weather dependent, i.e. uncertain and intermittent, and can only be planned with limited accuracy and controlled in one direction. Progress has been made to increase the predictability [73] and controllability [31, 29] of DERs, but the involvement of the demand-side and load-follows-generation models are inevitable. Variability increases the magnitude and frequency of changes in the supply-demand balance and also affects the predictability of these changes, resulting in an increased need for flexible resources to manage imbalances between supply and demand in power systems [41]. Any imbalance between supply and demand gives rise to fluctuations in frequency which must be maintained by the Transmission System Operator (TSO). Voltage and frequency regulation is critical in power systems to ensure the stability and reliability of the electrical grid. Maintaining the frequency within certain limits is essential for the safe and efficient operation of electrical equipment, frequency fluctuations can disrupt the timing and synchronization of electrical equipment, potentially causing damage or outages [50]. Distribution System Operators (DSOs) face the task of managing peak loads and preventing network congestion. The sporadic nature of DERs, such as the variability in solar and wind power generation, complicates the prediction and management of load patterns. This unpredictability can lead to periods of significant strain or underutilization of the grid infrastructure, necessitating advanced management and planning strategies [27].

The stability of the power grid, historically reliant on predictable and controllable power plants, is challenged by the stochastic output of many DERs. Ensuring a constant balance between supply and demand, a critical aspect of grid stability, becomes increasingly complex with the higher penetration of DERs. This complexity can lead to frequency fluctuations and voltage variations, potentially compromising the reliability of the power supply.

Necessity of digitalization

The effect of DERs necessitates enhanced digitalization for real-time monitoring, improved communication between TSOs and DSOs, and better integration of flexible consumers into the grid. Digitalization is crucial for grid operators and the industry to adapt to the evolving energy ecosystem, ensuring stability and efficiency in the face of increasing complexity [36].

The integration of DERs necessitates seamless communication and coordination between TSOs and DSOs. Digital platforms can bridge the informational gap between these entities, allowing for the exchange of critical data regarding energy production, consumption patterns, and grid performance. This enhanced communication is vital for maintaining grid balance and ensuring efficient energy distribution [99].

Digitalization paves the way for the integration of flexible consumers into the grid. Demand response programs, empowered by digital technologies, can adjust consumer energy usage in response to grid needs, thereby providing an additional tool for balancing supply and demand. Smart meters and home energy management systems play a crucial role in this regard, enabling consumers to participate actively in grid stabilization [67].

Digital technologies are crucial in integrating and optimizing the use of renewable energy sources. Advanced control systems, IoT devices, and AI algorithms can enhance the predictability and management of renewable energy outputs, thus mitigating the challenges posed by their variability and intermittency [84].

The rise of DERs signifies a transformative phase in the energy sector, bringing about sustainable energy production but also introducing complexity in grid management. In this context, digitalization is not just an enabler but a necessity for modern power systems. It provides the tools and capabilities required to manage the challenges of decentralization, ensuring that the grid remains stable, efficient, and capable of meeting the evolving demands of a more dynamic energy landscape. As we progress towards a more sustainable and decentralized grid, the symbiotic relationship between DERs and digitalization will be pivotal in shaping the future of energy systems [12].

1.2. Research questions

A fundamental question of the digitization of the electricity sector is whether it is possible to formulate methods that can be automated in order to increase the efficiency and utilization of the system, and make it cheaper and safer to operate. Although there are centralized actors in the electricity industry, it is largely driven by market processes and market actors that follow the operating rules of the electricity system and aim to achieve economic benefits.

Another way to make efficient use of the system is to involve previously passive actors in its operation, of which the involvement of energy consumers is a typical example. It is true for consumption in general, but for household consumption in particular, that its inherent flexibility capacity is not systematically exploited. As a first step, we need to know the amount that is available to the consumer at any given moment as a potential consumption reduction or increase.

In this thesis the following research questions will be addressed:

1. How accurately and by what methods the system imbalance, the difference between aggregated generation and consumption, can be predicted in the short term, so that the necessary measures can be taken in advance to stabilize the grid?

2. How a market-based aggregated portfolio using different technologies can derive economic benefits from imbalance forecasting and parallel activities in different but interconnected energy markets through automated strategies?

3. What methodology can be employed to quantify and predict the up-and-down flexibility of a residential house, defined as the ability of a system to adjust its power consumption or generation in response to fluctuations?

1.3. Structure of the thesis

The structure of the thesis is organized as follows. Chapter 2 provides an overview of the context, electricity industry background, key concepts and actors. Additionally, a general presentation of the mathematical tools used is discussed in this chapter. The short-term forecast of the system imbalance is presented in Chapter 3. Simulation of trading activities in Chapter 4 addresses research question No. 2. The quantification of the flexibility capability of a residential household is discussed in Chapter 5. In Chapter 6 the main results of this research are summarized in three thesis points. The relevant publications can also be found there.

CHAPTER 2

Basic notions

2.1. Electricity market

The objective of this research is not to provide a comprehensive overview of the electricity industry and markets. However, in order to contextualize the motivation behind this research, it is essential to have a clear understanding of the actors involved and the relationships between them. This chapter introduces the system operator, market players, and balancing groups. System imbalance is defined in more detail, its significance is described and the ways in which market settlement rules can be used to incentivise individual actors to reduce imbalance are briefly discussed. Grid-connected users are ultimately physical entities with generating or consuming (or storing) capability, but commercial processes on the one hand and economies of scale on the other have created more complex structures. One example of this is virtual power plants, which aggregate smaller-scale actors, such as generators, consumers, and storage, to participate in the electricity industry.

2.1.1. Structure and Participants

Transmission System Operator

TSOs play a pivotal role in the seamless functioning of the power sector. TSOs are responsible for ensuring the stable operation of the high-voltage transmission networks, which involves the balancing of electricity supply and demand, maintaining grid reliability, and overseeing the electricity market. At the core of TSO operations lies the critical responsibility of ensuring the reliable and secure transmission of electricity across the grid. TSOs are entrusted with the task of managing the transmission network's physical infrastructure, comprising high-voltage lines, substations, and interconnectors. Maintenance of this infrastructure is paramount to guaranteeing the grid's robustness and minimizing the risk of disruptions.

As the electricity industry evolves, the role of TSOs becomes increasingly complex and critical. The heterogeneity, intermittency, and decentralization of power generation sources, along with the complexity of modern energy markets, necessitate advanced IT systems for TSOs. These systems include SCADA (Supervisory Control and Data Acquisition), Energy Management Systems (EMS), balancing market operations systems, metering infrastructure, settlement systems, and telecommunications networks.

Grid planning emerges as another crucial facet of TSO responsibilities. As the energy landscape undergoes transformations, with an increasing emphasis on renewable energy integra-

tion, TSOs are charged with the strategic task of planning and expanding the transmission infrastructure to accommodate the evolving energy mix. This involves conducting thorough assessments of load forecasts, generation capacities, and technological advancements to optimize the grid's configuration for enhanced efficiency and resilience. Moreover, TSOs engage in long-term planning exercises to anticipate future energy demands and adapt the transmission network accordingly.

The coordination of electricity flows across borders represents a fundamental role of TSOs, particularly in regions with interconnected grids. Collaborative efforts with neighboring TSOs become imperative to facilitate the smooth exchange of electricity, ensure grid stability, and harness the benefits of diverse energy sources. This entails the development and implementation of cross-border system operation agreements, harmonizing operational procedures, and establishing communication protocols to address contingencies promptly.

A pivotal function of TSOs is to maintain system balance, a task demanding precision and foresight. Grid imbalances, arising from mismatches between electricity generation and consumption, can lead to frequency deviations and compromise the stability of the entire system. TSOs are responsible for maintaining the instantaneous balance between generation and consumption of electricity. The TSO activates balancing reserves provided by market participants to resolve real-time system imbalances. TSOs employ sophisticated tools, including advanced forecasting models and real-time monitoring systems, to predict load variations and generation patterns accurately. By adjusting the generation or consumption levels in response to fluctuations, TSOs ensure the equilibrium of the power system and mitigate the risk of instability. Automatic Generation Control (AGC) is a regulatory mechanism designed to maintain system frequency and tie-line power flows within prescribed limits by adjusting the output of generators in real-time to balance supply and demand.

Furthermore, TSOs actively engage in market operations, contributing to the efficient functioning of electricity markets. By participating in market mechanisms, such as energy balancing markets and ancillary services markets, TSOs ensure the economic optimization of grid operations. The development of market rules, pricing structures, and market coupling initiatives falls within the purview of TSO responsibilities, reflecting their commitment to fostering transparent and competitive energy markets.

Balance Responsible Parties

In the electrical grid, balancing groups serve as the linchpin in maintaining the balance between electricity supply and demand. These groups, composed of diverse electricity market players such as energy producers, suppliers, and large consumers, function as virtual clusters, aggregating their generation and consumption schedules. This collective approach not only

simplifies the process of grid management but also enhances the efficacy of balancing efforts. A Balance Responsible Party (BRP) is an entity that takes on the responsibility for the balance between supply and demand within a balancing group.

The core function of BRPs revolves around forecasting and scheduling. Each group is tasked with accurately predicting their electricity needs and production capabilities. For energy producers, this involves determining the amount of electricity they will generate, while suppliers and consumers must estimate their consumption levels. These forecasts are then meticulously compiled into schedules and submitted to the TSO. The accuracy of these schedules is paramount, as they form the foundation upon which the TSO assesses the overall balance of the grid and plans necessary balancing actions.

Beyond just forecasting and scheduling, BRPs engage in what is known as internal balancing. This involves managing discrepancies within their own group, balancing out over- or under-supply through measures like adjusting generation levels, shifting loads, or even intra-group trading. This proactive approach to internal imbalance management is crucial in mitigating the impact of these discrepancies on the wider grid.

The responsibilities of BRPs extend into several key areas. Paramount among these is the submission of accurate and timely data regarding their scheduled generation and consumption. This data is vital for the TSO to gauge the grid's balance and orchestrate effective balancing strategies. When deviations occur between the scheduled and actual electricity generation or consumption, balancing groups are held accountable. They are subject to imbalance charges, which are calculated based on the degree and nature of their imbalance. This not only incentivizes accurate forecasting but also ensures that groups contribute fairly to the cost of maintaining grid stability.

In order to maintain the balance of the grid and reduce the need for balancing reserve activation, the imbalance settlement framework incentivizes BRPs to maintain the aggregated balance of their generation and consumption balancing group portfolio. The difference between the nomination and the actual electricity consumption or production aggregated to a BRP is the imbalance volume, and BRPs are charged for their imbalance volume (balancing energy cost). BRPs with a significant share of variable generation in their portfolio will increasingly face imbalance charges due to schedule deviations. The market environment and trends point to shorter lead times, allowing market participants to trade on the basis of their continuously improving forecasts in both the commodity and balancing markets. With the internationalization of the balancing market in Europe, these processes are taking place in an expanding supply-demand environment.

Participants

The balancing groups comprise a diverse ensemble of energy producers, suppliers, traders, and large consumers. Energy producers, ranging from traditional power plants to renewable energy sources, are responsible for generating electricity in alignment with their forecasts. Energy suppliers and traders act as the intermediaries, navigating the complexities of the electricity market with a strategic finesse. Their role requires a deep understanding of market dynamics and consumer behavior, ensuring that their buying and selling of electricity aligns with the overall grid balance. This alignment is crucial, as it helps prevent sudden spikes or drops in electricity supply.

Large consumers, often industrial giants, play a surprisingly pivotal role. Their significant energy demands can sway the grid's balance. By participating in demand-side management (DSM), these consumers adjust their energy use in response to grid conditions. Their flexibility is especially valuable during peak times or when renewable sources are less productive, helping to ease the strain on the grid.

DERs, including solar panels, wind turbines, and energy storage systems, represent the evolving nature of the grid's composition. These resources, scattered across the grid, contribute to its diversification and resilience. Each resource adds its unique characteristics to the energy mix, enriching the grid's capacity to adapt and respond to changing conditions. Their decentralized nature poses a challenge, yet their contribution is essential for the transition towards a more sustainable and resilient energy landscape.

2.1.2. Imbalance in the energy system

Definition of system imbalance

System imbalance can be defined as the aggregate deviation from the situation in which all actors produce and consume exactly as planned. In other words, it represents the deviation of the control area from its planned state in the absence of intervention by the system operator. At its core, system imbalance arises from the unpredictable nature of both electricity production and consumption. The grid is a dynamic entity, constantly in flux due to varying demand from consumers and the irregular supply from producers, especially those reliant on renewable sources like wind and solar power. These renewables, while eco-friendly and increasingly essential, add a layer of complexity due to their intermittent nature.

The significance of managing system imbalance cannot be understated. The electrical grid operates within a delicate balance, maintaining a frequency close to a set point, typically 50 Hz in Europe. Any significant deviation from this frequency can lead to a host of issues, ranging from minor inefficiencies to major grid failures. Thus, balancing supply and demand

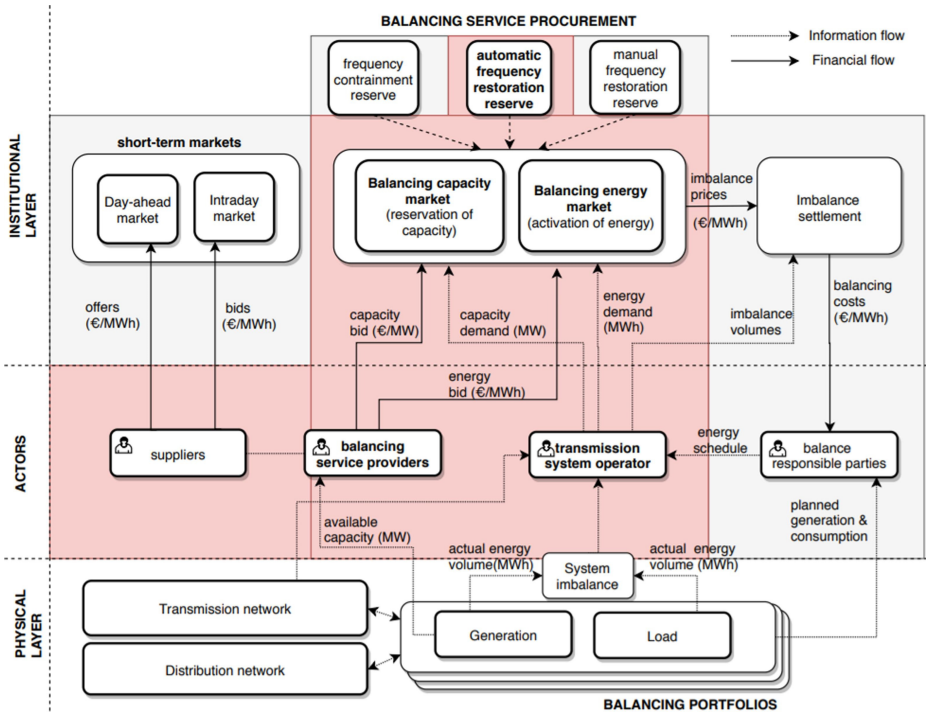


Figure 2.1: Market model [71]

in real-time becomes an imperative task.

In order to comprehend the concept of system imbalance, it is essential to gain an understanding of the roles of the various stakeholders involved in the operation of the grid. This encompasses producers, consumers, balancing groups, and, most importantly, the TSO. Each of these entities plays a distinct role in either contributing to or alleviating the imbalance, as illustrated in Figure 2.1. Producers and consumers are directly involved in creating the imbalance through their electricity generation and usage patterns, respectively. On the other hand, balancing groups and TSOs work towards mitigating these imbalances by adjusting supply, managing demand, or both.

The process of managing system imbalance involves several stages. Initially, there is a need for accurate forecasting of electricity demand and supply. This forecasting involves complex algorithms and models that take into account various factors such as weather conditions, consumer behavior patterns, and historical data. Following this, there is the scheduling stage, where electricity generation and consumption are planned out to align as closely as possible with the forecasts.

Despite these efforts, imbalances occur, and this is where the real-time management of the grid comes into play. TSOs, with the assistance of balancing groups, constantly monitor the

grid, ready to implement measures to address any imbalances. These measures can include adjusting the output of power stations, utilizing energy stored in batteries, or momentarily reducing demand through DSM strategies.

A key aspect is the definition of what precisely the system imbalance value represents. From a control point of view system imbalance can be associated to the frequency restoration control error (FRCE). FRCE means the sum of the power control error and the frequency control error [15]. This is a residual error that remains in the system after the TSO has acquired the services needed to minimize the deviation [7]. All variables of FRCE are measurable.

$$FRCE = (P - P_0) + K \cdot \Delta f, \tag{2.1}$$

where:

- P is the measured actual real-time power interchange value,
- P_0 is the the total planned position of the cross-border exchange,
- K is the a coefficient that relates the frequency deviation from its nominal value to the generation power output that needs to be adjusted to restore the frequency to its nominal value,
- Δf is the measured frequency deviation from 50 Hz nominal value.

The FRCE acts as a feedback signal in the closed-loop control system of the TSO, it provides a measure of how much additional generation or load reduction is needed to restore frequency to its set point. Based on the FRCE, the AGC system sends control signals to generation units to adjust their output. If the FRCE indicates that the frequency is below the set point (implying demand exceeds supply), the AGC system will command an increase in generation. Conversely, if the frequency is above the set point (supply exceeds demand), it will command a decrease.

However, the concept of system imbalance as used in Europe is not of interest for control systems, but rather as an input to accounting and incentive mechanisms. We need to distinguish between BRP imbalance and system imbalance. According to the European imbalance settlement harmonization methodology [1] BRP imbalance means an energy volume calculated for a balance responsible party and representing the difference between the allocated volume and the final position of that balance responsible party, including any imbalance adjustment applied. Final position equals to the sum of the BRP's external and internal commercial trade schedules and the allocated volume is the netted volume of all injections and withdrawals for which the BRP is financially responsible. Adjustment is the netted vol-

ume of activated balancing energy or all volumes activated by the TSO for purposes other than balancing (e.g. redispatching, cross-zonal curtailment).

System imbalance cannot be measured directly. The European TSO's imbalance settlement harmonization methodology [1] takes a satisfied balancing energy demand approach, summing up the activated reserves and not taking into account the residual control error in the FRCE value. This is partly because the FRCE frequency deviation component is approximately the same for the whole synchronous zone, its value does not depend only on the power deviation of the area. On the other hand, the control energy required by the TSO may not be completely accurate especially for manually activated and slow mFRR. We can assume that the TSO has issued an activation up to the extent of the actual imbalance and the residual error is actually due to the inaccurate execution of the activation, so the addition of the residual error would double the activation/execution variance in the system imbalance. Based on these, the system imbalance used is the result of the following calculation.

$$Imb = aFRR_{pos} - aFRRneg + mFRRpos - mFRRneg + IGCCpos - IGCCneg, \quad (2.2)$$

where:

- Imb is the system imbalance. Negative sign indicates a shortage, positive sign indicates a surplus,
- $aFRR_{pos}$ is the activated automatic frequency restoration reserves in the positive direction,
- $aFRR_{neg}$ is absolute value of the activated automatic frequency restoration reserves in the negative direction,
- $mFRR_{pos}$ is the activated manual frequency restoration reserves in the positive direction,
- $mFRR_{neg}$ is absolute value of the activated manual frequency restoration reserves in the negative direction,
- $IGCC_{pos}$ is the sum of settled imbalance netting volumes between neighboring TSOs in the positive direction[1],

[1]Imbalance netting is the process agreed between TSOs of two or more areas that allows avoiding the simultaneous activation of frequency restoration reserves in opposite directions by correcting the input of the involved frequency restoration processes accordingly. Instead of immediately activating local balancing resources to correct these imbalances, system operators first look for opportunities to offset these imbalances

- $IGCC_{neg}$ is the sum of settled imbalance netting volumes between neighboring TSOs in the negative direction.

Minimizing system imbalance

Imbalance settlement mechanisms provide a framework for the settlement of imbalances. According to the EB Regulation [16], an imbalance price is defined as a positive, negative or zero price defined for each settlement period for an imbalance in each direction. The literature distinguishes between single and dual pricing imbalance settlement systems. Under single pricing, a BRP that is short its scheduled energy is subject to the same imbalance price as a BRP that is over its scheduled energy. Under dual pricing, two imbalance prices are set: one for positive imbalances, which occur when a BRP has a surplus of energy, and another for negative imbalances, which occur when a BRP has a shortage of energy [44]. The choice between single and dual pricing affects the behavior and incentives of the BRP, as well as the financial options available. When using a single pricing approach, the BRP is motivated to contribute to the overall system balance, while dual pricing incentivizes the BRP to maintain a balanced position within its portfolio [4]. The European TSO's imbalance settlement harmonization methodology [1] confirms that the European TSO shall implement the use of single imbalance pricing in accordance with Article 55 of the EB Regulation.

Minimizing system imbalance through frequency regulation is an important system operation function of the TSO. Through the balancing group system, it delegates this task to the BRPs. In the single price model, the BRPs also have an interest in ensuring that their deviation from the schedule is such that the direction of the imbalance contributes to the restoration of system balance. In fact, a supportive schedule deviation may generate revenue for a BRP, depending on the imbalance pricing methodology and price forecast. The BRP must know its position, control its assets, and be able to predict system imbalance in order to take advantage of this.

BRPs can influence this imbalance volume by increasing the accuracy of their nominations, by over-nominating or under-nominating, or by applying internal balancing. Over the time horizon in which the system imbalance can be estimated and exploited, minimizing the balancing energy charge/maximizing revenues is not the only decision criterion for a BRP. The shortening of intraday lead times and the internationalization of wholesale markets during this period also provide a liquid market to cover the current position with commercial transactions in case it deviates from the day-ahead plans. A complex decision logic governs its activities. It takes into account the possibility of participating in organized intraday markets (liquidity, prices) and bilateral trading within the scheduling intervals, the schedules and position of its own aggregated portfolio and individual assets, the operating costs and

against opposite imbalances in neighboring areas [15].

their adjustments, the forecast direction of the system imbalance and the expected price of balancing energy.

The management of imbalances that are not managed by the trading markets remains the responsibility of the TSO. The TSO will ensure that the quality characteristics of the system are maintained at the required level by activating different types of reserves through the frequency control processes. While the activation of FCR (frequency containment reserve) is distributed and autonomous, with the primary controllers of the units involved in FCR control responding to local frequency measurements, the activation of FRR (frequency restoration reserve) is controlled by the TSO [72]. Automatic frequency restoration reserve (aFRR) is activated automatically by the TSO's closed-loop automatic generation controller (AGC). This system is designed to adjust the output of power plants or other resources in order to maintain the frequency of the grid within specified limits. aFRR process typically operates continuously, responding to frequency deviations within seconds to minutes. In contrast, the manual frequency restoration reserve (mFRR) necessitates manual involvement from the operator, who dispatches reserve resources with the objective of restoring frequency balance within a timeframe of minutes to up to 15 minutes. The aFRR is designed for rapid, precise adjustments, whereas the mFRR is employed for larger, more considered corrections. Both reserves are indispensable for maintaining grid stability. The aFRR addresses short-term imbalances, whereas the mFRR provides support for longer-lasting disturbances.

One of the main inputs to the AGC is the area control error (ACE). The objective of the controller is to reduce this value to 0 using the controlled units. If the imbalance can be estimated by the TSO, it will be able to reduce the area error to be handled by the aFRR controller by activating the cheaper and more available mFRR reserves in advance. Due to the growing variability of the increasing share of variable generation TSOs are expected to increase their demand for reserves [84], [33] and the importance of fast reserves continues to grow, emphasizing the importance of imbalance forecasting for TSOs as well.

2.1.3. Virtual power plants

Virtual Power Plants (VPPs) are a crucial development in the energy sector. They link multiple distributed energy resources, including renewable energy sources, storage systems, and flexible power consumers, to operate as a single, coordinated power plant. VPPs aggregate the capacities of these dispersed units through advanced software and communication technologies, enabling them to be centrally managed and optimized in real-time. VPPs enable centralized control, providing a range of grid services, including balancing, ancillary services, and grid stability enhancement. In this study, the terms "aggregator" and "VPP" are used interchangeably. It should be noted that the term aggregator encompasses a more general conceptual scope than that of the term VPP. The term aggregator refers to the pooling of

electrical capacity or flexibility of multiple energy resources and consumers, which enables the sale of the service they can provide to the grid operator or electricity market. In contrast, the term VPP is specifically focused on generators, while still performing an aggregation function.

VPPs operate in multiple electricity markets, contributing to both wholesale and balancing markets. In the wholesale market, VPPs can trade electricity, offering both generation and demand reduction. In the balancing market, VPPs provide valuable flexibility services to help transmission system operators manage real-time imbalances between supply and demand. VPPs accurately forecast system imbalances and utilize the aggregated capacity of connected resources to respond swiftly to grid needs, minimizing the requirement for traditional, carbon-intensive balancing reserves. The significance of VPPs extends beyond their ability to integrate renewable energy into the grid efficiently. They also enable consumers, including small-scale producers and flexible consumers, to actively participate in the energy market. This democratization of energy production and consumption can lead to more competitive markets, innovation, and improved energy security.

2.2. Applied tools and methods

In the following, I list and present at a high level the methodological tools that I have used in the development of the theses. I used ARIMAX (AutoRegressive Integrated Moving Average with eXogenous inputs), LSTM (Long Short-Term Memory) and ETR (Extra Trees Regression) for the imbalance forecasting in order to make the chosen ARDL (Autoregressive Distributed Lag) based time series forecasting comparable to other commonly used methods. In the following, I present the mathematical and application basics of these benchmark methods. In addition to the general descriptions, I will also present the specific settings that I used when using them as a benchmark.

2.2.1. ARIMAX

The linear ARIMAX model extends the traditional ARIMA (Autoregressive integrated moving average) model by including exogenous variables. The core of ARIMAX is the ARIMA model, which includes autoregression (AR), differencing (I) and moving average (MA) components [87]. The AR part expresses the current value of a variable as a linear combination of its own past values. The order of the AR part indicates the number of lagged terms used. The 'I' part includes differencing to ensure stationarity. The MA part models the current value of the series as a linear combination of current and past error terms. The order of the MA determines the number of lagged forecast errors in the forecast equation. The main extension of ARIMAX is the inclusion of external variables that may influence the forecast. These variables are not part of the time series itself, but are assumed to influence it. They are treated as additional regressors in the forecasting equation and their coefficients are esti-

mated alongside the ARIMA parameters. The ARIMAX(p, d, q) model can be represented as follows:

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - L)^d y_t = (1 + \sum_{j=1}^{q} \theta_j L^j)\varepsilon_t + \sum_{k=1}^{m} \beta_k x_{k,t}, \tag{2.3}$$

where:

- p is the order of the AR component,
- d is the degree of differencing required to achieve stationarity,
- q is the order of the MA component,
- ϕ_i are the parameters of the AR component for $i = 1, \ldots, p$,
- θ_j are the parameters of the MA component for $j = 1, \ldots, q$,
- L is the lag operator ($L^i y_t = y_{t-i}$),
- ε_t is the error term at time t,
- $x_{k,t}$ are the exogenous variables (external factors) at time t for $k = 1, \ldots, m$,
- β_k are the coefficients of the exogenous variables,
- m is the number of exogenous variables included in the model.

The input data for an ARIMAX model must meet several key characteristics. In Chapter 3, I checked the stationarity of the dependent variable when using the ARIMAX model. Additionally, a Granger test [34] was performed to confirm a plausible link between the independent and dependent variables. Augmented Dickey-Fuller tests [22] were performed to determine the order of differencing. Models were then fitted in the range of 0-8 for the autoregressive and moving average terms to determine the optimal parameters of the ARIMAX model.

2.2.2. Extra Trees Regression

The Extra Trees Regression (ETR) model uses an ensemble of unpruned decision trees in a process known as "extremely randomized trees" [30]. This method is particularly useful for time series forecasting tasks because it effectively captures complex interactions between variables, thereby improving the model's predictive performance. The ETR model's inherent ability to accommodate nonlinear relationships makes it well suited for this task, as it can

identify intricate patterns and dependencies among these variables [86].

ETR is an ensemble learning technique that belongs to the family of tree-based methods [30]. It builds upon the concept of random forests by introducing even more randomness into the construction of the individual trees. Unlike random forests, where the splitting decisions at each node are based on a subset of features and the best split is chosen, ETR goes a step further by using the entire dataset and selecting splits for each node from a random distribution. This approach reduces the variance of the model without significantly increasing bias, leading to a robust prediction model capable of handling overfitting more effectively.

The ETR's ability to handle complex, nonlinear relationships between variables makes it particularly suitable for time series forecasting. Time series data often exhibit intricate patterns such as seasonality, trend components, and noise, which traditional linear models might fail to capture effectively. The ETR model, with its ensemble approach, can accommodate these nonlinear interactions and dependencies among variables, enhancing predictive performance [86].

Consider a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i represents the input features and y_i the output (target) values, with i indexing the observations in the dataset. The objective of ETR is to construct an ensemble of M unpruned decision trees, $\{T_m\}_{m=1}^{M}$, and use their collective decision to predict the output for a new input x.

For each tree T_m, a bootstrap sample $\mathcal{D}_m$ is drawn from the original dataset $\mathcal{D}$. At each node of the tree, instead of evaluating all possible splits across all features, a random subset of features is selected, and for each feature in this subset, a split is chosen randomly. The split that offers the best reduction in variance (for regression tasks) is selected to partition the data at that node. This process is repeated recursively until the tree is fully grown, without pruning.

The prediction for a new input x by the ETR model is given by averaging the predictions of all M trees:

$$\hat{y}(x) = \frac{1}{M} \sum_{m=1}^{M} T_m(x), \tag{2.4}$$

where $T_m(x)$ is the prediction of the m-th tree.

2.2.3. LSTM

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) architecture first proposed by Hochreiter and Schmidhuber [38], and has been widely adopted for sequence prediction problems due to its potential to remember and learn long-term dependencies in sequential data. This property of LSTM makes it particularly well suited for multi-step time series forecasting, where the objective is to predict several future steps of a target variable given its historical observations and possibly also exogenous variables. It also plays a vital role in efficient power system planning [65].

LSTM units are a type of recurrent neural network (RNN) architecture that was developed to address the limitations of conventional RNNs, particularly the vanishing gradient problem [35], which makes it difficult for RNNs to learn long-term dependencies in sequence data. An LSTM unit is composed of a cell, an input gate, an output gate, and a forget gate (Figure 2.2). The cell is responsible for remembering values over arbitrary time intervals, and each of the three gates can be thought of as a conventional artificial neuron, as they perform operations on the data passing through them. The gates regulate the flow of information into and out of the cell, which is why they are called gates. The cell state, or the "memory" of the LSTM unit, carries information throughout the sequence processing. Hidden (h) and cell (c) states serve as short-term working memory and long-term memory, respectively. The LSTM does have a form of memory because the output of each LSTM unit is fed into the next, and changes to the cell state are made dependent on the operation of the gates. This is how the LSTM can maintain information in "memory" over time. The forget gate (f) is a sigmoid function (σ) that decides what information should be discarded or kept. The input (i) gate updates the cell state with new information. The cell state candidate gate (g) first regulates the information flow in the network by using the tanh function on the previous hidden state and the current input (x). The product of tanh is multiplied by the input gate output to calculate the candidate for the current cell state. The output gate (o) calculates the new hidden state. W and R denotes the weights of the input signal and the hidden state respectively, b denotes the bias [100].

To make multi-step imbalance forecasts from a lagged multivariate input, a sequential LSTM model was created in this work. In a sequential model, the layers are stacked sequentially, meaning the output of one layer serves as the input to the next layer. An encoder-decoder architecture [81] was implemented, that is commonly used for sequence-to-sequence prediction tasks, where the input sequence is encoded into a compressed, fixed-length representation (context vector), and then the decoder generates the output sequence based on that representation.

The first encoder LSTM layer takes the input sequence, processes it, and learns the context

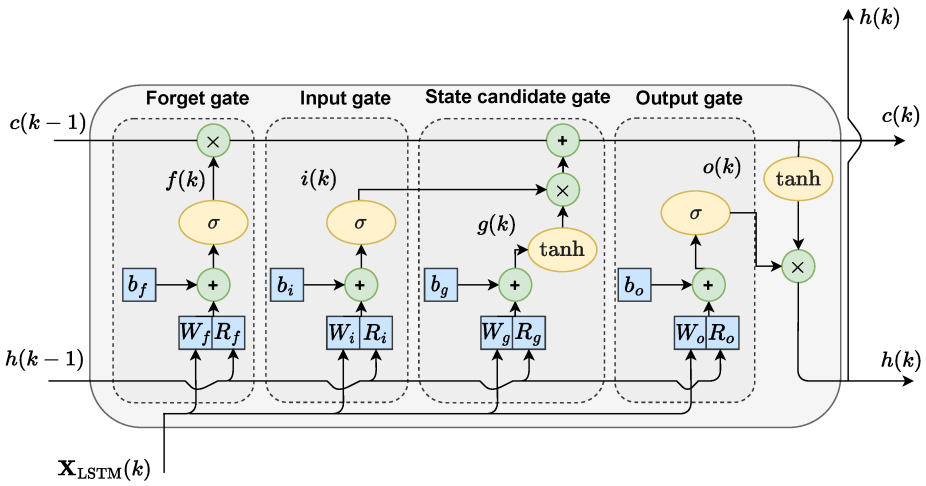


Figure 2.2: LSTM structure [101]

vector. Its final hidden state is passed to a 'repeat vector' layer that repeats the context vector as many times as the length of the output steps (imbalance forecast window). This step allows the context vector to match the desired length of the output sequence and prepares the encoded representation to be fed into the decoder part of the network. The decoder LSTM layer is then used to generate the multi-step output sequence. Then a dense layer is applied to each time step of the output sequence independently, ensuring that the output sequence matches the desired shape.

CHAPTER 3

Short-term system imbalance forecast of the power grid using autoregressive distributed lag method

The problem of imbalance in the electricity system and the need to minimize it was discussed in detail in Chapter 2. This chapter presents a forecasting method that predicts the system imbalance with a 2-hour lead time and a quarter-hour resolution from publicly available data. The forecasting uses an econometric model, ARDL (Autoregressive Distributed Lag), to forecast time-series data. The results are compared with other state-of-the-art methods, demonstrating that the mathematically simpler ARDL, used in the same context, produces results of comparable accuracy for the research problem at hand.

The chapter is organized as follows. Section 3.1 provides a review of the relevant literature. Section3.2 describes the methodology used in the study. Section 3.3 presents the results and introduces the benchmark models. Finally, Section 3.4 concludes the topic with a summary of findings and implications.

3.1. Related work

In their study, Garcia et al. [28] explore and highlight the constraints and inadequacies of commonly employed yet simplistic forecasting methods such as ARIMA and exponential smoothing, due to the non-periodic, non-stationary, and noisy nature of imbalance time-series data. To address these challenges, the researchers propose the utilization of neural networks to capture the non-linear and irregular patterns within the data, enabling accurate prediction of daily imbalance medians.

Kratochvil [49] deals with multivariate short-term imbalance forecasts from the perspective of a BRP that tries to gain profit from achieving the opposite direction of the system imbalance. The impact of the most important predictors is analyzed by applying autocorrelation analysis. Focusing on BRP's objective, the point forecast of the concrete value of the system imbalance is not calculated. 5 intervals of the system imbalance are defined and the predicted imbalance is mapped to one of them, simplifying the prediction problem to a classification. Kratochvil's ARIMA model resulted in an accuracy of 61.0 % in the Czech market. Contreras [17] is dedicated to hourly imbalance predictions using random forest regression in the Spanish market with an accuracy of 68.3 %.

In their study, Salem et al. [79] introduce an additional forecasting approach that offers the Transmission System Operator (TSO) valuable insights into the anticipated trends of imbalances within the upcoming two hours. This solution is accompanied by prediction

intervals, providing information on the reliability of the forecasts. The researchers employ a quantile regression forests ensemble method to predict imbalances in the Norwegian power system. They found that training the model with datasets spanning at least twelve months led to significant enhancements in forecast accuracy compared to using three or six-month datasets.

In [89] a comprehensive model is proposed that combines a Bidirectional Long Short-Term Memory (BLSTM) architecture, an attention mechanism, and an encoder-decoder structure. The proposed model provides valuable insights into the relative significance of features and aid in understanding the complex temporal dependencies present in the data.

To gain valuable market insight and competitive advantage, both imbalance volume and price are necessary inputs for decision-making close to delivery time. While the modeling of day-ahead and intraday electricity markets has been the subject of numerous papers, the modeling of imbalance market prices has received less attention. Klæboe et al. [47] benchmarked earlier models published before 2015 and concluded that none of the benchmarked models produced informative day-ahead point forecasts. This suggests that information available before the day-ahead market closes is efficiently reflected in the day-ahead market price rather than the balancing market price.

Dumas et al. [23] combine imbalance volume forecasts with reserve costs. The authors utilize a two-step approach, namely they first calculate the probabilities for the system imbalance and then based on that make predictions regarding the imbalance prices.

In the study conducted by Browell and Gilbert [13], it was shown that the forecast of imbalance in the UK could be achieved by utilizing a logistic regression model. The model incorporated demand, wind, solar, and total supply margin as explanatory variables. The findings indicated that the logistic regression consistently outperformed the benchmark by an impressive margin of 4 %.

A novel approach to probabilistic forecasting of German power imbalance prices is presented by Michal Narajewski [65]. This study is significant because it addresses the highly volatile nature of the imbalance market, which is characterised by frequent extreme price spikes. Narajewski uses advanced methods such as lasso with bootstrap, GAMLSS (Generalized Additive Models for Location, Scale and Shape) and probabilistic neural networks to forecast 30 minutes before delivery. These methods are compared to a naive benchmark and it is shown that while they do not significantly outperform the benchmark in terms of prediction accuracy, they do provide a significantly better empirical coverage.

Bottieau et al. [11] demonstrated the superiority of machine learning techniques compared

to conventional benchmarks when utilizing a feature set consisting of forward prices, as well as recent and forecasted data on generation and load. The authors successfully employed a one-step-ahead forecasting model for system imbalance.

Koch [48] analyzes a strategy of taking positions in the German intraday market based on expected imbalance prices and examines its impact on system stability. It uses a logistic regression model to predict the direction of the overall system balance and to apply a profitable trading strategy. Intraday trading is used to estimate the imbalance prices and decide whether to take a buy or sell position on a quarter-hourly basis. The applied strategy simulates a decision with available information during active trading considering the current and not just average market prices. The model was able to correctly classify the system balance in 68 % of all quarter hours.

3.2. Forecast model

The models presented assume that imbalance correlates with other external variables that can be measured or scheduled, or with historical values of imbalance itself. The power system can be affected by several primary sources of imbalance. The amount of power consumed by consumers can be highly variable over the course of a day. This can result in deviations from the schedule. Unplanned power outages due to equipment failure, extreme weather conditions, or other unforeseen circumstances can cause imbalances. Power market conditions, such as changes in demand, fuel, or electricity prices can also cause imbalances in the power system. Table 3.1 contains all the explanatory variables that are used to train the forecast models and forecast future values.

3.2.1. Problem statement

For the power system, we have metering data, plans and forecasts available. Measurements are historical data with quarterly resolution. They are available for the quarter preceding the forecast. The plans, which are typically schedules provided by market participants, refer to both past and future time periods. Forecasts published by system operators according to their internal methodology are also used as predictors. Given these time-series data, or goal is to provide a multi-step forecast of system imbalances at forecast execution time t for the current and subsequent quarters $t+0, t+1, ..., t+7$. This is achieved by creating separate lag structures of the feature set and training ARDL models for each forecast timestep. Both training and verification are performed on historical data, adhering to the rules of temporal availability of plan and measured data. The quality of the prediction is evaluated using metrics commonly used in the literature and compared to the performance of leading non-linear machine learning methods.

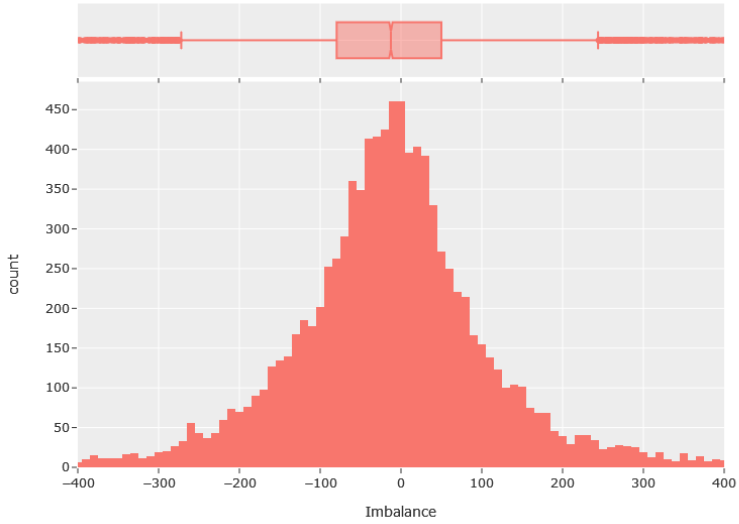


Figure 3.1: Distribution of imbalance in MW

3.2.2. Variables

The distribution of the observed imbalance of the Hungarian power grid in 2021 in Figure 3.1 is centered around 0 (median: -12.5), with slightly more weight in the positive imbalance (skew: 0.24). The imbalance distribution is leptokurtic. It has a sharper peak and fatter tails than a normal distribution. The Q1 and Q3 quartiles are -79 and 50, respectively. The imbalance time series in Figure 3.2 shows fluctuations over time with no clear seasonal pattern, suggesting that it may be stationary. A stationary time series has a constant mean and variance over time. This implies no trend or seasonality. As the lags increase, the ACF plot shows a gradual decrease in correlation. The fact that the autocorrelations remain within significance limits after the first few lags suggests that there is no strong autocorrelation in the data at higher lags. The plot of the PACF shows a spike at lag 1, and then a decline in the partial autocorrelations immediately thereafter. The sharp cutoff after the first lag in PACF, together with the gradual decline in ACF, suggests that an AR(1) model would be appropriate for the imbalance time series.

A seasonal decomposition was applied to the imbalance time series. This suggests that there is no significant seasonality in the data. There is no clear long-term upward or downward trend, and the pattern appears to be somewhat cyclical. The consistent spread of residuals suggests that the model captures trend and seasonality fairly well, leaving random noise that the model cannot explain.

An important consideration in the selection of independent variables was that they should

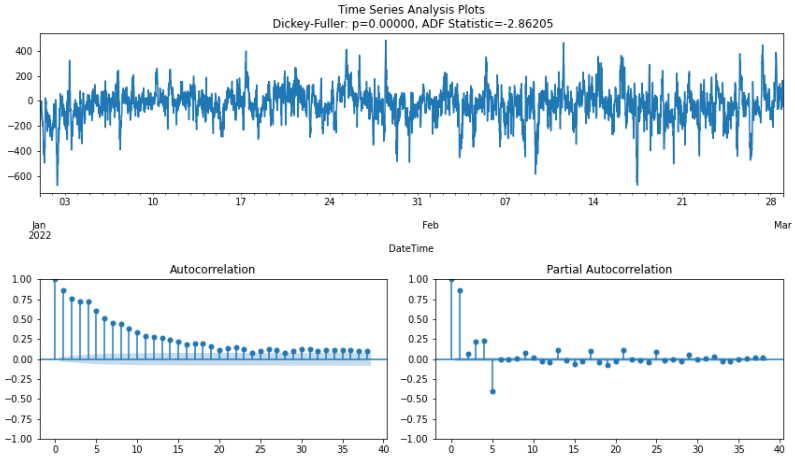


Figure 3.2: Imbalance time series analysis

be publicly available in sufficient quantity to be used to train the model. Given that my aim is a short-term forecast using a sliding window approach, it is expected that they will be available at the time of forecasting for the period in question. The third aspect is that the variables are indeed correlated with the system's imbalance.

All variables must be prepared for use in the predictive model. This included standardising the different time resolutions of the data sources and different time zones, as well as filling in missing data and dealing with outliers.

The variables used in the forecasting model are presented below, highlighting the category of use and their availability over time.

Table 3.1: Usage of predictor variables. Present refers to the period the forecast is executed in.

Variable	Type	Past	Present	Future
System imbalance	Metered dependent	X		
Scheduled solar generation	Planned exog	X	X	X
Scheduled wind generation	Planned exog	X	X	X
Scheduled load	Planned exog	X	X	X
Solar deviation	Metered exog	X		
Wind deviation	Metered exog	X		
Load deviation	Metered exog	X		
Planned change in the scheduled load	Fixed		X	X
TSO's predicted load	Fixed		X	X
Quarter hour of the day	Fixed		X	X

The 'Metered' type refers to variables that are not available in the forecast period, only

values from the past can be used. 'Planned' variables are available both for the current time interval and for the future time interval. Typical examples are schedules that are submitted by market participants as part of the day-ahead or intra-day processes, or forecasts that are published by the TSO. 'Fixed' variables are available for both past and future, however, only the forecast interval is used here.

The correlation tests support the engineering considerations that higher load and production schedules are associated with higher imbalances, and that a high proportion of weather-dependent production increases the variability. Recent deviations from schedules have a direct impact on imbalance. These can be calculated from the difference between schedules and actuals.

The change in load has a paradoxical effect on the imbalance. According to the physics-based expectations, an increase in load decreases the frequency of the system, increasing the deficit, if not coupled with an increase in production. In contrast, a comparison of load curves and measured imbalances with a resolution greater than a quarter-hour shows that, within a quarter-hour, when the consumption trend is increasing, imbalances move towards deficit, but at the quarter-hour boundary, deficit falls sharply, moving towards surplus. The underlying phenomenon is that although the load curve is continuous, the time resolution of market transactions and market scheduling is 15 minutes or more. The planned load is therefore given as a quarter-hourly average power. Generation serving consumption already tries to follow this stepped curve to mitigate balancing energy. The consequence of the mismatch between the physical operation and the market abstraction is illustrated in Figure 3.3, where after the scheduling prepared in a balanced manner, the imbalance is determined by the difference between the production approaching the stepped load and the continuous load.

Within fifteen minutes, the schedule is unchanged, so the deficit increases continuously as the load ramps up. At the fifteen-minute limit, the imbalance changes towards surplus when the schedule jumps, and it will continue to increase continuously towards deficit from that point on as long as the upward direction in load is unchanged. This means that when calculating the imbalance for a given quarter hour, both the schedule change and the planned load change should be considered.

Numerous features could be extracted from the quarter-hourly time stamp, but correlation checks have shown that the time within a day has a clear impact on the imbalance, while annual, monthly, and weekly samples are not as relevant. To capture the cyclic patterns in time-based data, timestamps are encoded using cyclic encoding. By converting the day's quarters into cyclical components, the relative distances can be preserved between quarters while introducing cyclical patterns into the model. Rather than representing the quarter as

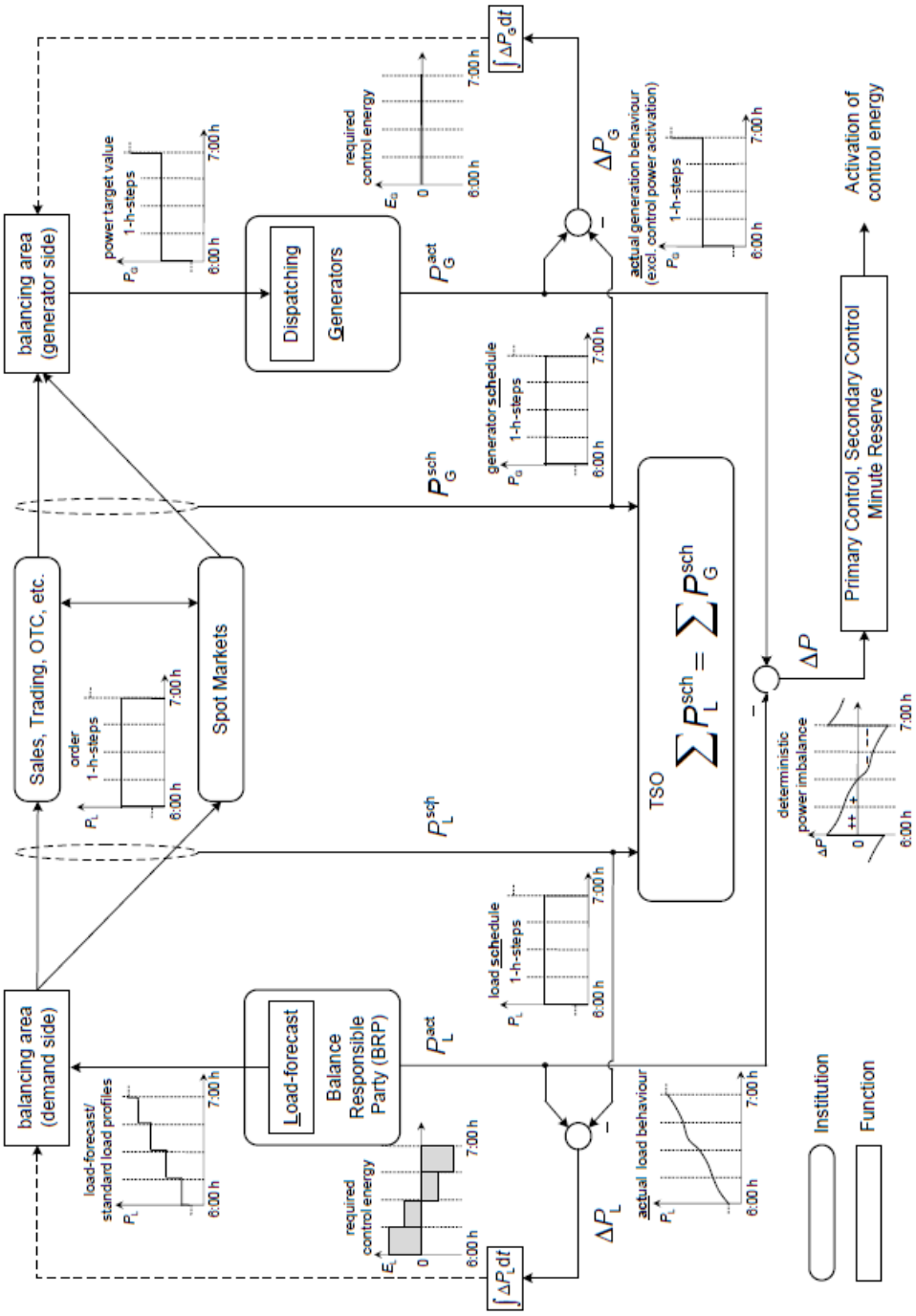


Figure 3.3: Load, generation and schedules in European power system markets [93]

a single value that increases linearly, cyclic encoding represents the feature as two separate variables as shown in Figure 3.4: one representing the sine of the angle around the circle and the other the cosine.

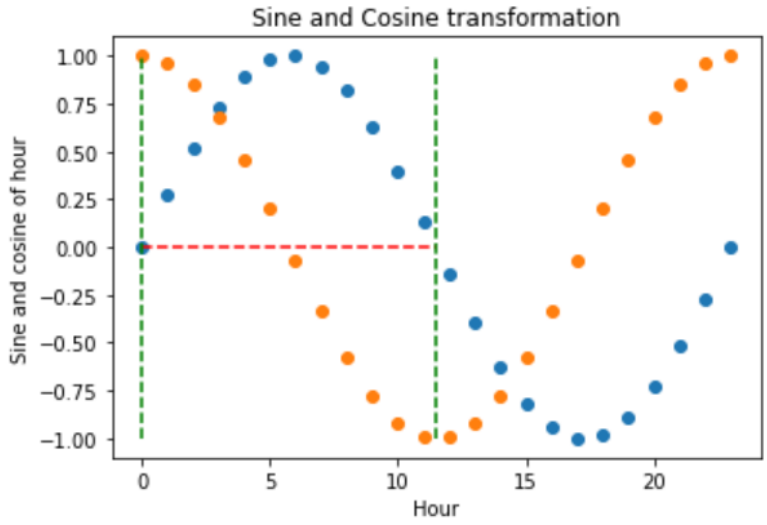


Figure 3.4: Cyclical encoding of time features. [Source: Feature-engine]

The correlation between time series can be tested with a regression causality test, called Granger test [34]. The Granger causality test is a statistical hypothesis test designed to determine whether one time series can be used to predict another time series. It measures correlation, not causation, in the sense that it can provide useful information about the relationship between two variables, but it does not provide conclusive evidence of causality. The idea is that a variable X Granger-causes Y, if past values of X can help explain Y. Granger causality is only relevant with time series variables. To test for Granger causality, the Granger causality test compares the predictive power of two models: one that includes both X and Y as predictors of Y's future values, and another that only includes Y as a predictor. If the coefficients of past values of X in the Granger model are statistically significant (p-value < 0.05) then we can reject the null hypothesis of the test (coefficients of past X values in the regression equation is zero) and conclude that X Granger-causes Y. The Granger test was carried out for 5 lags, the p-values were far below 0.05, so we can observe that all the selected variables are Granger-causes of the imbalance.

The Augmented Dickey-Fuller (ADF) test [22] is a statistical test used to determine if a time series is stationary or non-stationary. The test is based on a first-order autoregressive model of the time series. The null hypothesis of the ADF test is that there is a unit root in the AR model, implying that the data series is nonstationary because the time series mean and variance change over time. The ADF p-values for the imbalance time series are effectively

0, and the ADF statistics are below 0.05, indicating that we are confident in rejecting the null hypothesis of a unit root. This is a confirmation of the stationary nature of the data.

3.2.3. Forecast model

I propose a linear multihorizon imbalance forecasting method that uses both past values of the outcome variable and exogenous variables, and provides flexibility in the design of the lag structure. For this purpose, I have chosen the Autoregressive Distributed Lag (ARDL) model. ARDL is a time series econometric model that is used to analyze the relationship between two or more variables, where one variable is considered the dependent variable and the others are independent variables. The ARDL model allows for the examination of both short-term and long-term relationships between the variables. It is a dynamic model, as it takes into account the lagged values of the variables, which can help to capture the persistence and dynamics of the relationship. The literature review did not find any studies using the ARDL model for the prediction of imbalances, but it is considered to be a well-established tool for the analysis of time series data, and several publications have used the ARDL model in different areas. Engel and Granger [25] extended the relationship between co-integration and error correction models. Adabor et al. [2] demonstrated the effectiveness of the ARDL approach in producing reliable estimates, especially when the data sample size is relatively small. Furthermore, Tong at al. [88] highlighted the importance of the ARDL model as an essential tool in dynamic one-equation regression, which is widely used in economic time series modelling. Taken together, these references highlight the widespread application and reliability of the ARDL model in various research areas.

The basic ARDL structure includes a lagged dependent variable, a set of lagged independent variables, and possibly exogenous variables [70]. The model can be estimated using ordinary least squares (OLS) regression or other estimation techniques. My motivation was the combination of the mathematically simpler linear model with deep domain knowledge to create a prediction process that would have competitive results. This required both the target and independent variables to have regressive components. In the case of ARDL, unlike, for example, the linear ARIMAX also used as a benchmark, where the exogenous variables are not lagged, I had the flexibility to specify the lag structure of both the dependent and independent variables.

The general ARDL model applied is specified by Equation 3.1 :

$$Y_t = \underbrace{\delta_0}_{\text{Constant}} + \underbrace{\sum_{p=1}^{P} \phi_p Y_{t-p}}_{\text{Autoregressive}} + \underbrace{\sum_{k=1}^{M} \sum_{j=1}^{Q_k} \beta_{k,j} X_{k,t-j}}_{\text{Distributed Lag}} + \underbrace{Z_t \gamma}_{\text{Fixed}} + \epsilon_t, \tag{3.1}$$

where

- Y_t: the forecasted value of the dependent variable at t,
- Y_{t-p}: values of the dependent variable at $t-1, t-2, \ldots, t-p$,
- P: maximal lag of Y,
- $X_{k,t-j}$: k^{th} independent variable at $t-1, t-2, \ldots, t-j$ periods,
- Q_k: maximal lag of X_k,
- M: number of lagged variables,
- Z_t: fix non-lagged independent variable at time period t,
- ϵ_t assumed to be i.i.d.,
- $\delta, \phi, \beta, \gamma$ are estimated modell parameters.

The ARDL model according to Equation 3.1 calculates the value of the next time period after the observations. This is called single-step forecasting because we only need to forecast one step. However, in this case we are interested in several quarters of an hour, several steps need to be predicted at the same time, so this is a multi-step time series forecasting problem. There are several approaches to multi-step forecasting. The direct approach generates separate models for the forecasts $t, t+1, t+n$ with the same predictor variables, but with different dependent variables for each step. This is a simple method, but it does not allow the value of the $t+2$ prediction to be used to calculate $t+3$. The recursive method uses the same one-step model several times. The result of the previous step is used as input for the current estimation.

An ensemble model is used in this paper. For each step, the predicted value of the dependent variable from the previous step is used to train a separate model with different variables and lag structures. The model is illustrated in Figure 3.5. The forecast is made at the time t for the time intervals between t and $t+8$. I assume that for both the dependent and independent variables, observations are already available for time period $t-1$. The value of the Fixed variable (Z in Equation 3.1) is known for the given interval t. From these inputs, the ARDL model calculates the value of the dependent variable value for period $t+1$ (FC_{t+1}). In the same period t, the forecast for the interval $t+1$ is also computed (FC_{t+1}). To do this, the existing observations of the dependent variable are used as predictors, along with its forecast for the previous interval ($t+0$) and the value of the fixed variable for $t+1$. For Step $t+2$, the set of observations remains the same, but the predictors for $t+0$ and

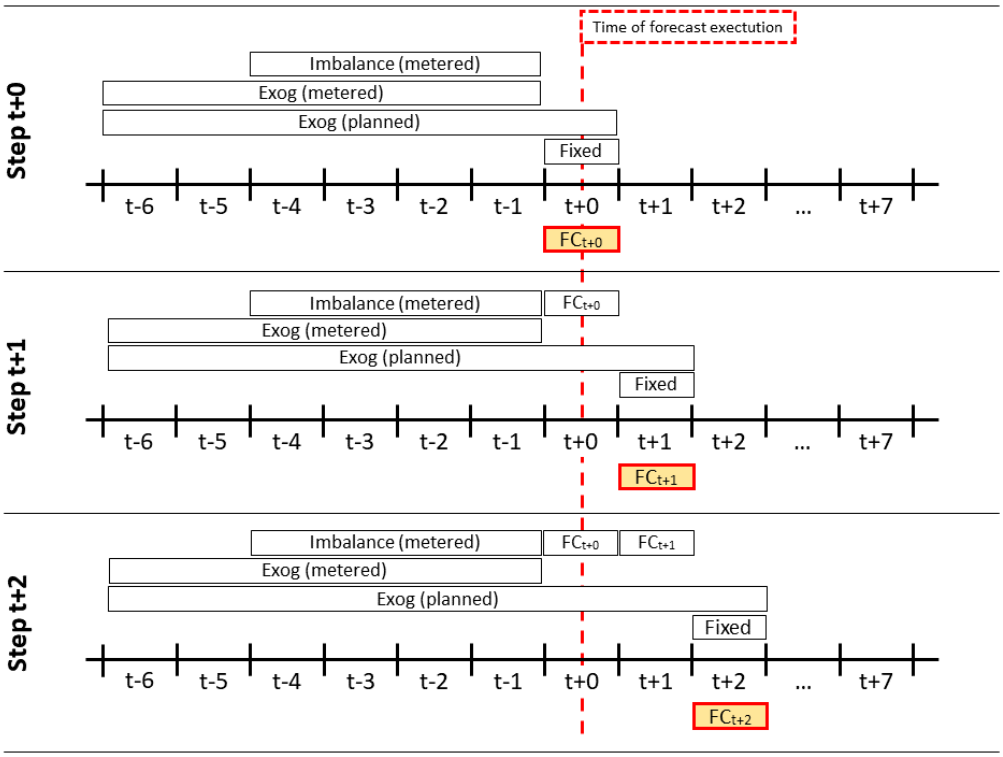


Figure 3.5: Lag structure of the predictor variables. A separate model is trained for each step. FC refers to the forecasted values of imbalance.

$t+1$ of the dependent variable are added to the set of predictors. Multi-step forecasting is implemented iteratively, with different steps using different variable and lag structures, thus training a separate ARDL model for each step.

As shown in Figure 3.5, to train the model for each step, the predictor set must be constructed by adding the prediction of the previous step to the predictor set. The algorithm for training and prediction is summarized in the Algorithm 1. The objective of the training step within the specified ARDL framework is to accurately estimate the coefficients of the model, which quantify the relationships between the dependent variable and the lagged values of both the dependent and independent variables. In order to achieve this, the ordinary least squares (OLS) method [92] is employed. The OLS method operates by identifying the set of coefficients that minimize the sum of squared residuals. This is the sum of the squared differences between the observed values and the values predicted by the model. By minimizing these residuals, it is ensured that the model provides the best linear approximation of the underlying data relationships.

After a model has been trained for a prediction step, the corresponding prediction is made,

Algorithm 1 Train multi-step ARDL and forecast

procedure TRAIN_ARDL($trainingdata, maxstep$)
  for $step \leftarrow 0, maxstep$ **do**
    if $step \neq 0$ **then**
      $trainingdata \Leftarrow trainingdata + prediction_{step-1}$
    end if
    $Model_{step} \Leftarrow Train(trainingdata)$
    $prediction_{step} \Leftarrow Predict(trainingdata, Model_{step})$
  end for
  return $Model$ ▷ Different model parameters for each step
end procedure
procedure FORECAST_ARDL($fcdata, maxstep, t$) ▷ Forecast between t and $t + maxstep$
  $fcdata_{t+0} \Leftarrow fcdata$
  for $step \leftarrow 0, maxstep$ **do**
    if $step \neq 0$ **then**
      $fcdata_{t+step} \Leftarrow fcdata_{t+step-1} + prediction_{t+step-1}$
    end if
    $prediction_{t+step} \Leftarrow Predict(fcdata_{t+step}, Model_{step})$
  end for
  return $prediction$
end procedure

and then the next step is trained by adding that prediction to the set of predictions. The forecast follows a similar logic. The forecast that is associated with a step is both the final result and the input for the forecast of the next step.

3.3. Results

3.3.1. ARDL forecast

Real data was used to verify the method described above. The data to generate model variables according to Section 3.2.2 are available on the Hungarian TSO website and the ENTSO-E Transparency Service. The train period is quarter hours between January 2022 and February 2022, the test period is between March 2022 and December 2022. In these intervals, all quarter-hour values have been taken into account. It can be said that the time stamps are independent and identically distributed (IID). The time series are stationary, the statistical properties of the time series (such as the mean and the variance) do not change over time.

A separate model is taught for each step because a given forecast event involves multiple forecast steps with different variable structures. They are also evaluated separately, so that we can see the forecasting performance for forecasting steps $t + 0, t + 1, t + 2, t + 3, ..., t + 7$

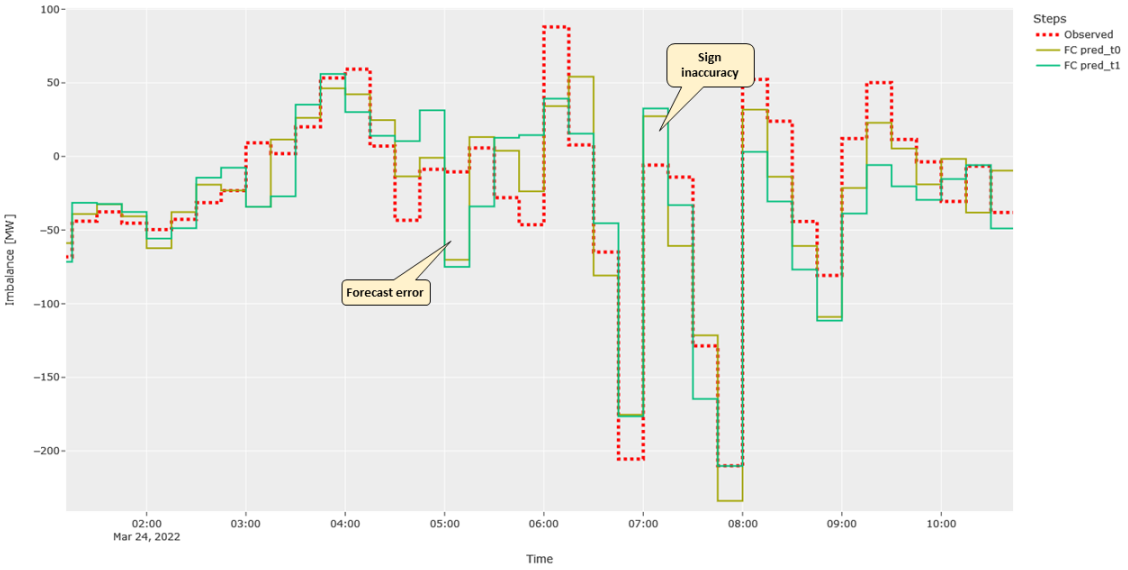


Figure 3.6: Sample of observed imbalance and prediction

and observe the deterioration of the forecast accuracy with increasing look-ahead window.

It is worth looking at the environment for which the imbalance is being studied when examining forecast errors. The peak load of the Hungarian power system was 7361 MW. The share of renewable energies was 13,66 % of the total electricity consumption. 286 MW of positive aFRR capacity has been procured in the year 2021 [61].

Figure 3.6 shows a snippet of the imbalance forecast. A solid line is the forecast value and a dashed line is the observed value for the same quarter of an hour. The time series 'FC pred_t0' represents the prediction step 'Step t+0' as shown in Figure 3.5. For a given time stamp, there are several predictions (t0-t7), but for the sake of clarity, only $t+0$ and $t+1$ are shown in the figure. The forecast error is the difference between the observed and the predicted value. In addition to the magnitude of the error, the sign accuracy is of particular importance. This is due to the fact that the methods of balancing energy management and settlement for positive and negative system imbalances are very different. Therefore, in addition to the well-known metrics, I present an independently developed metric for evaluating sign accuracy.

Figure 3.7 contains metrics (Mean Absolute Error, Root Mean Squared Error, Sign Accuracy Percentage with 40 MW threshold) used to evaluate the predictions. The Mean Absolute Error (MAE) is a measure of the average size of the errors or differences between the predicted values and the actual values, without taking into account their direction. It is calculated by

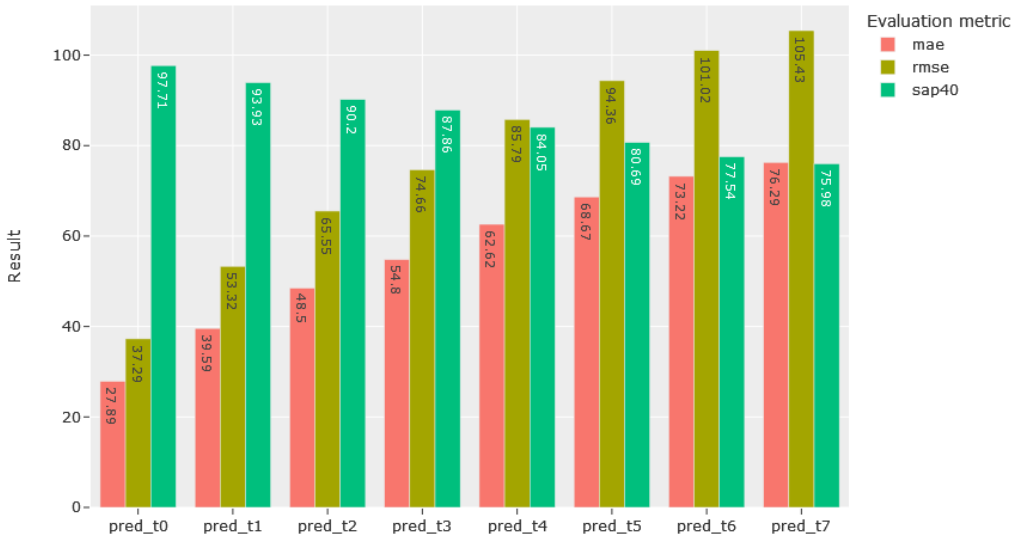


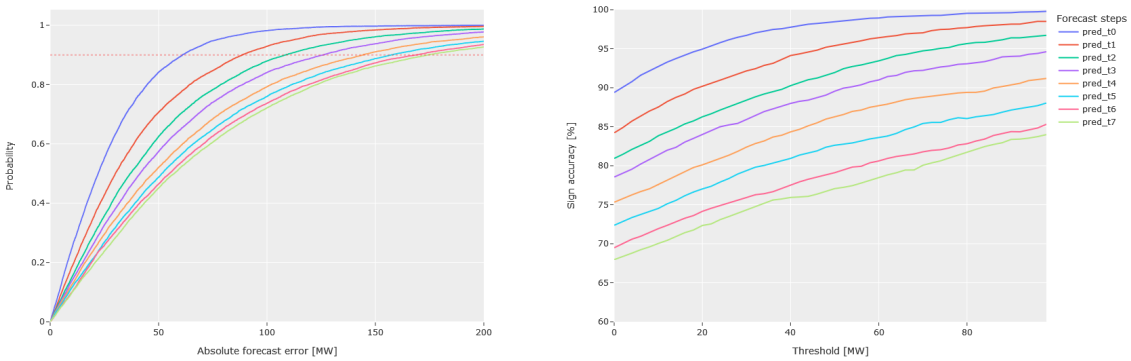
Figure 3.7: Evaluation metrics of the forecast steps

taking the absolute difference between the predicted value and the actual observed value, and then averaging these differences over all the observations [103]. The resulting value, expressed in the same units as the original variable, represents the average absolute error between the predicted and actual values.

The Root Mean Squared Error (RMSE) is a measure of the average size of the errors between the predicted value and the actual value, taking into account the direction of the errors. The RMSE is calculated by taking the square root of the average of the squared differences between the predicted and actual values [103]. Like the MAE, the RMSE is expressed in the same units as the original variable, and lower values indicate better model performance. RMSE penalizes large errors more compared to MAE, as it gives greater weight to larger errors due to the squaring operation.

As the number of prediction steps increases, all prediction metrics increase. This is as expected, since both the imbalance and the measured values of the independent variables affecting the imbalance continue to move further away from the predicted value over time.

Figure 3.8a shows the cumulative distribution of the absolute error. The error in 90 % of the projections is smaller than 60 MW at $t+0$, but at $t+7$ the error is larger than 160 MW in 10 % of the projections.



(a) Cumulative distribution of the forecast errors

(b) Sign accuracy percentage

Figure 3.8: Forecast errors of the ARDL model

Sign accuracy is important because the value, or more precisely the sign, of the system imbalance is used to build logic that encourages power system actors to reduce the system imbalance. Figure 3.1 shows that the value of the system imbalance around 0 is the most common, and we can assume that the sign is the most difficult to predict accurately for small imbalances. To illustrate, the following figure evaluates the sign accuracy as a function of a parameter (Figure 3.8b). The Sign Accuracy Percentage (SAP) is calculated by dividing the number of correct predictions of the direction (Equation 3.2) by the total number of predictions (Equation 3.3). In addition, a threshold value is applied so that SAP is only calculated for forecasts with an absolute value that is greater than the threshold value (40 MW in Figure 3.7).

$$\begin{aligned} SA^{True}_{step,\tau} &= \sum_t \Big[(FC_{t+step} \geq \tau \wedge Y_{t+step} > 0) \vee \\ &\qquad (FC_{t+step} \leq -\tau \wedge Y_{t+step} < 0) \Big] \\ SA^{False}_{step,\tau} &= \sum_t \Big[(FC_{t+step} \geq \tau \wedge Y_{t+step} < 0) \vee \\ &\qquad (FC_{t+step} \leq -\tau \wedge Y_{t+step} > 0) \Big] \end{aligned} \tag{3.2}$$

$$SAP_{step,\tau} = \frac{SA^{True}_{step,\tau}}{SA^{True}_{step,\tau} + SA^{False}_{step,\tau}} * 100, \tag{3.3}$$

where

- [*condition*] is the Iverson bracket, denoting a number that is 1 if the *condition* in square brackets is satisfied, and 0 otherwise,
- $SA^{True}_{step,\tau}, SA^{False}_{step,\tau}$: number of correct and incorrect directional predictions given a threshold value τ and a forecast window *step*,
- FC_{t+step}: imbalance forecast of interval $t + step$ executed at time interval t,
- Y_{t+step}: measured imbalance of interval $t + step$,
- τ: threshold value, $0 \leq \tau$ only forecast over τ or $-\tau$ are considered,
- $SAP_{step,\tau}$: sign accuracy percentage.

In Figure 3.7, I have set a threshold such that the accuracy of the first 1 hour estimate of the SAP is about 90%.

3.3.2. Benchmark models

In the field of short-term forecasting of power system imbalances, it is imperative to establish robust benchmarks to evaluate the effectiveness of newly developed models. In this context, my work uses three widely accepted and scientifically validated benchmark models: ARIMA(X), LSTM, and ETR. The ARIMA model, a mainstay in time series forecasting, is chosen for its proven utility in capturing linear dependencies and trends in time series data, particularly in electricity load forecasting [3]. The Long Short-Term Memory (LSTM) model, a recurrent neural network architecture, is included because of its superior ability to capture complex temporal dynamics and long-term dependencies in sequence data, making it particularly suitable for multi-step time series forecasting, as demonstrated in power load forecasting applications [65]. Finally, Extra Trees Regression (ETR), a variant of the Random Forest algorithm, is chosen for its effectiveness in handling non-linear relationships and complex interactions between variables, which is critical in time series forecasting, especially in contexts involving autocorrelated dependent variables and multiple lagged explanatory variables [11]. Together, these models provide a comprehensive framework for evaluating the performance of the ARDL model, encompassing both linear and nonlinear dynamics relevant to forecasting short-term imbalances in power systems. Although several studies use the methods selected for the benchmark, the literature on short-term imbalance forecasting is not very rich, and comparisons can be valid only if the models are run on the same assumptions and data. To this end, the benchmark models were not only presented but also applied and implemented to the problem at hand.

The results of the ARDL model have been compared with the benchmark models and the results are summarized in Table 3.2. For MAE and RMSE, a lower value indicates better

forecasting efficiency. For SAP, a higher value is favorable because the value indicates the percentage of forecasts above 40 MW and below -40 MW that correctly predict the system direction. The performance of ARDL is better than the benchmark models in the first 4 intervals, i.e. the first 1 hour. In the second hour, ARIMAX performs slightly better.

The present work is based on the assumption that the correct choice of relevant external variables and the design of a lag structure tailored to the forecasting task effectively support forecast accuracy. These preparatory steps made the mathematically simpler dynamic regression method competitive with the state-of-the-art neural network and random forest models.

Table 3.2: Comparison of prediction results using 5 months of test data

Metric	Step	ARDL	ARIMAX	ETR	LSTM
MAE	Step 0	28	54	32	46
	Step 1	40	55	45	55
	Step 2	49	58	53	64
	Step 3	55	59	59	71
	Step 4	63	66	67	73
	Step 5	69	69	73	75
	Step 6	73	71	77	79
	Step 7	76	73	80	86
RMSE	Step 0	37	67	43	62
	Step 1	53	69	61	73
	Step 2	66	75	72	84
	Step 3	75	76	81	94
	Step 4	86	84	92	98
	Step 5	94	86	101	100
	Step 6	101	92	108	107
	Step 7	105	93	113	117
SAP	Step 0	98	89	97	91
	Step 1	94	88	92	86
	Step 2	90	87	88	82
	Step 3	88	85	85	80
	Step 4	84	84	82	78
	Step 5	81	82	78	76
	Step 6	78	80	76	73
	Step 7	76	80	74	69

3.4. Conclusion

This study has demonstrated the effectiveness of the ARDL model in forecasting power system imbalances, crucial for maintaining system frequency and overall stability. Through comprehensive evaluations using real-world data, the ARDL model has shown superior performance compared to the benchmark nonlinear models, including Long Short-Term Memory networks and ETR. The introduction of a multi-step forecasting approach allows for predictions over multiple future quarters, enhancing the ability of system operators to anticipate

and manage imbalances. A further innovation presented in this work is the incorporation of sign accuracy as an evaluation metric. This metric assesses the model's ability to predict not just the magnitude but also the direction of the imbalance, which is often more critical for operational decision-making. The results indicate that the ARDL model not only predicts the quantitative aspects of the imbalance more accurately but also provides reliable directional forecasts.

The computational efficiency of the ARDL model, coupled with its high predictive accuracy, makes it a practical tool for real-time power system management. This is particularly significant in an era where timely and accurate predictions are essential for integrating intermittent renewable energy sources and maintaining grid reliability.

CHAPTER 4

Time series simulation of the operation and trading activity of aggregated power generation and consumption portfolios in a multi-market environment

Management of portfolios containing a mix of generation and consumption assets involves a wide range of strategies, encompassing both internal optimizations and market participation. However, the feasible actions during real-time operations depend on several factors, including the state of associated devices, power system demand, and external influences. This chapter presents a combined technical and market model that enables the simulation of potential operational strategies. This model generates time-series data representing energy transactions, revenues, and costs for an aggregator managing diverse assets. The data generated are based on realistic scenarios, reflecting true market pricing conditions and the dynamics of specific sub-markets.

4.1. Background

The steady penetration of distributed, weather-dependent generation over conventional large power plants has been a worldwide trend in recent years, which, together with the evolution of technology and market operation, allows the development of portfolios that make optimal use of the specificities of different generation/consumption technologies (e.g. weather-dependent and conventional generation units, controllable and non-controllable consumers, energy storage) in different electricity markets [40].

The implementation of an advanced portfolio management solution entails significant investment and operational costs and therefore a high economic risk. While overcoming the technical challenges, it is also difficult to find tangible answers to the questions of whether it is worth moving towards a heterogeneous portfolio; what level of aggregation is recommended; and what markets, what operations, what results can be achieved. Studies of the expected profitability and returns prior to investing in homogeneous assets such as stand-alone power plants or energy storage facilities usually approach the problem along the lines of some kind of industry estimate or average revenue generating capacity. However, portfolios can be very diverse in terms of composition, size and physical location, so thinking along the lines of the panels used previously can lead to much greater inaccuracy than before [20]. Gomes et al. [32] presents an optimal bid submission in a day-ahead electricity market for the problem of joint operation of wind with photovoltaic power systems having an energy storage device. Paper [95] proposes three-stage coordinated optimization scheduling strategy for a combined cooling, heating, and power microgrid. Yang et al. [97] proposes

an optimal bidding strategy model of a virtual power plant in the day-ahead market that contains energy, reserve, and regulation markets. Pei et al. [69] introduces the concept of microgrid aggregator to involve small-scale microgrids in real-time balancing market bidding via a hierarchical market framework. In [21] an optimization model is provided for participation of a distributed energy resource aggregator in the day-ahead market in the presence of demand flexibility. Wang et al. [91] proposes an optimal bidding strategy model for a load aggregator that implements a demand response program, which enables the load aggregator to reduce the risk of financial loss caused by price volatility. An optimal operation strategy for a virtual power plant participating in the day-ahead and intraday energy markets has been discussed in [53]. The aggregation function determines internal prices by evaluating its real-time responses to the day-ahead schedule and updating the proposed pricing function parameters, and adjusts its energy reserves.

In the portfolio, the combination of assets determines the revenue-generating potential. However, they can only ever work with the resources available at any given time, and in compliance with their commitments at that time, within the rules and regulations of each market. In my work, I consider a reasonably accurate model of the current functioning of the European electricity market. This model depicts market participants operating in electricity markets with different time horizons and balancing control markets. It also depicts market participants adapting to the market's process, timing requirements and settlement rules. In my research, I have created an aggregator model that includes generation, consumption and storage assets with different technologies. This model operates within a real operational framework. The assets comprising the portfolio are coordinated and operated by an aggregator role in different electricity markets with the objective of maximising economic benefits, subject to the technical constraints of the assets that make up the portfolio.

The main contribution of my work is that it addresses both the internal modelling complexity arising from the inhomogeneity of the portfolio and the multi-market nature of the modern electricity industry, while incorporating imbalance forecasting in order to profit from deliberate schedule deviations. This techno-market model provides an opportunity to study and compare portfolios of different composition and size and different scenarios. The modelling also enables the effectiveness of different market strategies and the consequences of their application to be considered, in addition to the technical composition.

4.2. Model

The starting point of the analysis is a 12-month reference period, with the technical data of the portfolio and the parameters of the market strategy to be applied inserted into the real operational framework and factual data. The simulation is run on real parameters and historical time series data, and the calculation of revenues and expenses is based on the

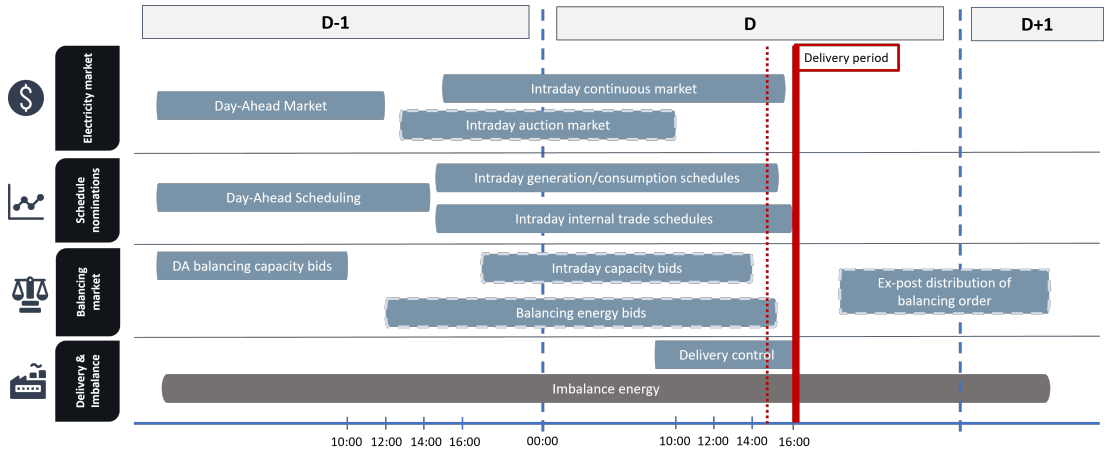


Figure 4.1: Available energy markets and trading intervals in the day-ahead and intraday timeframe in 2023.

actual functioning of each sub-market (different time horizons, timings, settlement rules) and the resulting states. We can consider the portfolio as a price taker [8], as it cannot influence the price of the commodity due to the competitive nature of the market and the small size of the aggregator's transactions relative to the size of the market. The output of the model is a time series of energy flows, which allows a transparent step-by-step representation of the operation under the set scenario, facilitating easy tracing of the logic applied in different market and technical situations.

In the following, we consider a decentralized aggregator portfolio consisting of a consumer, a weather-dependent solar (PV) generator and a battery energy storage, forming an individual balancing group. All devices provide real-time metering data to the aggregator. The latter is capable of closed-loop control of solar and battery energy storage. The priority of the consumer is to meet the instantaneous energy demand required by its core activity. In my work, I have created an operational model for the aggregator to determine how to influence the generation and consumption of each asset in the portfolio, and how to transact in the available markets, after considering the technical and financial as well as the external market framework. The model employs a time-series approach with a time resolution based on the standard quarterly intervals commonly employed in the field.

The challenge in defining the operating rules that will yield the greatest economic benefits is not solely due to the necessity of ensuring that assets with disparate technologies, technical and economic characteristics function in unison. Additionally, the fact that the operational areas and individual markets are fragmented, yet still exhibit interdependence, further complicates the matter as illustrated in Figure 4.1.

In the model, the aggregator operates in the following energy markets:

- Day-Ahead and Intraday Market (DAM, IDM),
- Imbalance Market,
- Balancing Capacity Market (aFRR),
- Balancing Energy Market (aFRR).

4.2.1. Portfolio assets

The following section describes the main characteristics and operational options of the assets that comprise the portfolio. These operational options reflect the scenarios considered in the model. Furthermore, this section provides a brief overview of the operating rules implemented by the model to control the assets during the time series simulation.

PV Generation

The flexibility of solar PV production was based on the power output values of the different time horizons of the production forecast and their estimated accuracy. The deviation from the schedule was estimated from the available historical fact data.

1. In the model, solar PV generation has the potential for pre-scheduled dispatch on a day-ahead or intraday timeframe:
   (a) The supply of the portfolio's own energy consumption without using the grid, with a predefined output, as defined in the production and consumption schedules.
   (b) Selling to an external operator on the basis of an internal trade schedule.
   (c) Charging of energy storage without using the public grid, with a predetermined quantity as fixed in the production schedule.
2. Providing balancing services with solar generation according to the transmission system operator's (TSO) needs:
   (a) Availability of negative balancing capacity up to the level of the generation forecast or the planned feed-in to the public network. The reservation of capacity does not imply its activation.
   (b) Provision of negative balancing energy in response to a TSO activation order up to the planned level of injection into the public network.
3. Real-time control options for solar PV production:

(a) Allowing over-generation compared to the schedule in order to balance the portfolio schedule or to deliberately increase balancing energy.

(b) Charging the battery without using the public grid, outside of the schedule.

(c) Curtailment of generation to balance schedules or to intentionally increase balancing energy.

Parameters

- *window*: Look ahead window (value: 5) [quarter hour]
- t: Current time interval (quarter hour)
- L_{DAM}: Trading threshold (value: 5000 kWh)
- L_{aFRR}: Balancing bid threshold (value: 15000 kWh)

Variables

- $E^{\text{DA}}_{\text{aFRR-}}$: aFRR- day-ahead energy bid from PV generation [kWh]
- $P^{\text{Gen}}_{\text{DA}}$: Day-ahead planned PV generation [kWh]
- $P^{\text{Gen}}_{\text{ID}}$: Intra-day planned PV generation [kWh]
- $S^{\text{Sale}}_{\text{DAM}}$: PV DAM market sale [kWh]
- $C^{\text{Sched}}_{\text{Batt}}$: Battery charge considered during PV scheduling [kWh]
- $C^{\text{Sched}}_{\text{EndUser}}$: End-user consumption considered during PV scheduling [kWh]
- $P^{\text{Grid}}_{\text{Scheduled}}$: PV ID scheduled grid feed-in (PV generation) [kWh]
- $T^{\text{Trans}}_{\text{IDM}}$: PV IDM market transactions [kWh]
- $E^{\text{ID}}_{\text{aFRR-}}$: aFRR- intra-day energy bid from PV generation [kWh]
- $D^{\text{Dev}}_{\text{PV}}$: Deviation of PV generation from ID schedule [kWh]
- $P^{\text{NoReg}}_{\text{PV}}$: PV production behind the connection point without down-regulation [kWh]
- $C^{\text{BattAch}}_{\text{PV}}$: Planned battery charging with PV realization [kWh]
- $C^{\text{SupplyAch}}_{\text{PV}}$: Planned supply with PV realization [kWh]
- $E^{\text{aFRR-}}_{\text{PV}}$: PV aFRR- service [kWh]

- $C_{\mathrm{PV}}^{\mathrm{BattExtra}}$: Additional PV battery charging behind the meter [kWh]
- $C_{\mathrm{PV}}^{\mathrm{SupplyExtra}}$: Additional PV supply behind meter [kWh]
- $E_{\mathrm{PV}}^{\mathrm{Reg}}$: PV down-regulation for scheduling purposes [kWh]
- $I_{\mathrm{PV}}^{\mathrm{Point}}$: Effect of PV production at the connection point [kWh]

Equations

$$
\begin{aligned}
S_{\mathrm{DAM}}^{\mathrm{Sale}} &= \max\left(P_{\mathrm{DA}}^{\mathrm{Gen}} - L_{\mathrm{DAM}}, 0\right) \\
C_{\mathrm{Batt}}^{\mathrm{Sched}} &= \max\left(S_{\mathrm{Chg}}^{\mathrm{Sched}}, -(P_{\mathrm{ID}}^{\mathrm{Gen}} + E_{\mathrm{aFRR\text{-}}}^{\mathrm{DA}})\right) \\
C_{\mathrm{EndUser}}^{\mathrm{Sched}} &= \max\left(P_{\mathrm{ID}}^{\mathrm{Cons}}, -(P_{\mathrm{ID}}^{\mathrm{Gen}} + E_{\mathrm{aFRR\text{-}}}^{\mathrm{DA}} + C_{\mathrm{Batt}}^{\mathrm{Sched}})\right) \\
P_{\mathrm{Scheduled}}^{\mathrm{Grid}} &= P_{\mathrm{ID}}^{\mathrm{Gen}} + C_{\mathrm{Batt}}^{\mathrm{Sched}} + C_{\mathrm{EndUser}}^{\mathrm{Sched}} \\
T_{\mathrm{IDM}}^{\mathrm{Trans}} &= P_{\mathrm{Scheduled}}^{\mathrm{Grid}} - S_{\mathrm{DAM}}^{\mathrm{Sale}} \\
E_{\mathrm{aFRR\text{-}}}^{\mathrm{ID}} &= \min\left(-\max\left(P_{\mathrm{Scheduled}}^{\mathrm{Grid}} - L_{\mathrm{aFRR}}, 0\right) - E_{\mathrm{aFRR\text{-}}}^{\mathrm{DA}}, 0\right) \\
D_{\mathrm{PV}}^{\mathrm{Dev}} &= P_{\mathrm{PV}}^{\mathrm{NoReg}} - P_{\mathrm{ID}}^{\mathrm{Gen}} \\
C_{\mathrm{PV}}^{\mathrm{BattAch}} &= \min\Big(\max\Big(-P_{\mathrm{PV}}^{\mathrm{NoReg}} + S_{\mathrm{DAM}}^{\mathrm{Sale}} + T_{\mathrm{IDM}}^{\mathrm{Trans}}, \\
&\qquad C_{\mathrm{Batt}}^{\mathrm{Sched}}, -(P_{\mathrm{PV}}^{\mathrm{NoReg}} + E_{\mathrm{aFRR\text{-}}}^{\mathrm{DA}} + E_{\mathrm{aFRR\text{-}}}^{\mathrm{ID}})\Big), 0\Big) \\
C_{\mathrm{PV}}^{\mathrm{SupplyAch}} &= \min\Big(\max\Big(-(P_{\mathrm{PV}}^{\mathrm{NoReg}} + C_{\mathrm{PV}}^{\mathrm{BattAch}}) + S_{\mathrm{DAM}}^{\mathrm{Sale}} + T_{\mathrm{IDM}}^{\mathrm{Trans}}, \\
&\qquad -(P_{\mathrm{PV}}^{\mathrm{NoReg}} + C_{\mathrm{PV}}^{\mathrm{BattAch}} + E_{\mathrm{aFRR\text{-}}}^{\mathrm{DA}} + E_{\mathrm{aFRR\text{-}}}^{\mathrm{ID}}), C_{\mathrm{EndUser}}^{\mathrm{Sched}}, M_{\mathrm{Cons}}\Big), 0\Big) \\
E_{\mathrm{PV}}^{\mathrm{aFRR\text{-}}} &= -\min\left(P_{\mathrm{PV}}^{\mathrm{NoReg}}, -E_{\mathrm{aFRR\text{-}}}^{\mathrm{Calc}}\right) \\
I_{\mathrm{PV}}^{\mathrm{Point}} &= P_{\mathrm{PV}}^{\mathrm{NoReg}} + C_{\mathrm{PV}}^{\mathrm{BattAch}} + C_{\mathrm{PV}}^{\mathrm{SupplyAch}} + E_{\mathrm{PV}}^{\mathrm{aFRR\text{-}}} + C_{\mathrm{PV}}^{\mathrm{BattExtra}} + C_{\mathrm{PV}}^{\mathrm{SupplyExtra}} + E_{\mathrm{PV}}^{\mathrm{Reg}}
\end{aligned}
$$

Consumer

The planned production of the solar panels will be consumed by the consumer to the extent of its availability and planned needs, with the remainder of the production plan being sold on the organized electricity market. The planned demand of the consumer that cannot be covered by the solar PV production will be purchased from the organized electricity market in a predefined and scheduled manner. The priority of the consumer is to meet its current energy needs as required by its core business. In the modeling, the consumer is not controllable and therefore plays a passive role, its deviation from the schedule is managed by the aggregator through other assets in the portfolio.

Variables

- $P_{\mathrm{ID}}^{\mathrm{Cons}}$: Intra-day planned consumption [kWh]
- $S_{\mathrm{ID}}^{\mathrm{Cons}}$: Consumption schedule (intra-day consumption) [kWh]
- M_{Cons}: Measured consumption behind the connection point [kWh]
- $I_{\mathrm{Cons}}^{\mathrm{Point}}$: Effect of consumption at the connection point [kWh]
- $P_{\mathrm{Market}}^{\mathrm{EndUser}}$: Market purchase for consumption (intra-day trading) [kWh]
- $D_{\mathrm{EndUser}}^{\mathrm{Dev}}$: Deviation of end-user consumption from schedule [kWh]

Equations

$$
\begin{aligned}
S_{\mathrm{ID}}^{\mathrm{Cons}} &= P_{\mathrm{ID}}^{\mathrm{Cons}} - C_{\mathrm{EndUser}}^{\mathrm{Sched}} \\
I_{\mathrm{Cons}}^{\mathrm{Point}} &= M_{\mathrm{Cons}} - C_{\mathrm{PV}}^{\mathrm{SupplyAch}} - C_{\mathrm{PV}}^{\mathrm{SupplyExtra}} \\
P_{\mathrm{Market}}^{\mathrm{EndUser}} &= P_{\mathrm{ID}}^{\mathrm{Cons}} - C_{\mathrm{EndUser}}^{\mathrm{Sched}} \\
D_{\mathrm{EndUser}}^{\mathrm{Dev}} &= M_{\mathrm{Cons}} - P_{\mathrm{ID}}^{\mathrm{Cons}}
\end{aligned}
$$

Energy Storage

The scheduling of energy storage only occurs during the day. For the deviation from the energy storage schedule, full discharge and full recharge can be considered as strict limits, while the desired deviation can be flexibly controlled/adjusted in the range between the two.

1. Scheduled utilization of energy storage:
   (a) Charging (purchase) or discharging (sale) to an external operator on the basis of an individual agreement or organised market prices.
   (b) Planned charging of energy storage from solar PV generation without use of the public grid, as planned in the production schedule.
2. Providing balancing services with the energy storage:
   (a) Availability of positive or negative balancing capacity based on a predefined bidding strategy or a day-ahead forecasted load level, or up to the difference between the planned and the minimum or maximum charge level.
   (b) Provision of positive or negative balancing energy up to the difference between the planned and the minimum or maximum charge level.
3. Real-time control options for energy storage:

(a) Battery charging by solar generation without using the public grid in an unscheduled way.

(b) Charging or discharging a battery from the public grid to balance deviations from schedules.

(c) Charging or discharging of a battery from the public network beyond the scope of internal balancing.

Variables

- E_{MinChg}: Minimum battery state of charge [kWh]
- P_{MaxDis}: Maximum discharge power [kW]
- P_{MaxChg}: Maximum charge power [kW]
- E_{MaxCap}: Maximum battery capacity [kWh]
- η_{Batt}: Battery efficiency per direction [%]
- P_{NetDis}: Maximum discharge power measured on the grid [kW]
- P_{NetChg}: Maximum charge power measured on the grid [kW]
- $E^{\text{DA}}_{\text{aFRR+}}$: aFRR+ day-ahead energy bid from battery [kWh]
- $t^{\text{Avail}}_{\text{Remain}}$: Remaining time until availability [quarter-hours]
- $t^{\text{End}}_{\text{Remain}}$: Remaining time until end of availability [quarter-hours]
- $E^{\text{Exp}}_{\text{MinChg}}$: Expected minimum battery charge at the end of period [kWh]
- E_{PermChg}: Permitted battery charging requirement at the end of period [kWh]
- $S^{\text{Sched}}_{\text{Chg}}$: Battery charging schedule for the target time period [kWh]
- $S^{\text{Decl}}_{\text{Chg}}$: Submitted charging schedule for the battery [kWh]
- $E^{\text{ID}}_{\text{aFRR+}}$: aFRR+ intra-day energy bid from battery [kWh]
- $S^{\text{Batt}}_{\text{Market}}$: Market purchase for battery charging (intra-day trading) [kWh]
- $D^{\text{Dev}}_{\text{Batt}}$: Deviation caused by the battery from schedule (to be balanced by MAVIR) [kWh]
- $C^{\text{Avail}}_{\text{Batt}}$: Actual down-regulation (charging) capacity of the battery [kWh]

- $C_{\text{Batt}}^{\text{UpReg}}$: Actual up-regulation (discharge) capacity of the battery [kWh]
- $C_{\text{Batt}}^{\text{ExtChg}}$: Battery charging with external scheduled ID market purchase [kWh]
- $E_{\text{Batt}}^{\text{aFRR+}}$: aFRR+ service with battery [kWh]
- $E_{\text{Batt}}^{\text{aFRR-Assist}}$: Assistance of aFRR- service with battery [kWh]
- $E_{\text{Batt}}^{\text{Charge}}$: Battery charge at the end of period [kWh]
- $C_{\text{Batt}}^{\text{OffSched}}$: Non-scheduled battery charging from the public grid (LE3) [kWh]
- $C_{\text{Batt}}^{\text{OtherDis}}$: Other battery discharges [kWh]
- $E_{\text{Batt}}^{\text{TechLossChg}}$: Technological loss during battery charging [kWh]
- $E_{\text{Batt}}^{\text{TechLossDis}}$: Technological loss during battery discharge [kWh]
- $E_{\text{Batt}}^{\text{NoLimitChg}}$: Battery charge at the end of period [kWh]
- $E_{\text{Batt}}^{\text{Balance}}$: Balancing energy due to battery limits [kWh]
- $E_{\text{Batt}}^{\text{TechLossNetChg}}$: Technological loss during battery charging [kWh]
- $E_{\text{Batt}}^{\text{TechLossNetDis}}$: Technological loss during battery discharge [kWh]

Equations

$$E^{\text{Exp}}_{\text{MinChg}} = \begin{cases} t^{\text{End}}_{\text{Remain}} \cdot \frac{P_{\text{MaxDis}}}{4}, & \text{if } t^{\text{Avail}}_{\text{Remain}} = 0 \\ \max\left(E_{\text{MaxCap}} - t^{\text{Avail}}_{\text{Remain}} \cdot \frac{P_{\text{MaxChg}}}{4}, 0\right), & \text{otherwise} \end{cases}$$

$$E_{\text{PermChg}} = \begin{cases} 0, & \text{if } E^{\text{DA}}_{\text{aFRR+}} > 0 \\ \max\left(-E_{\text{MaxCap}} - E^{\text{Charge}}_{\text{Batt}}, -\frac{P_{\text{NetChg}}}{4}\right), & \text{otherwise} \end{cases}$$

$$S^{\text{Decl}}_{\text{Chg}} = \min\left(\max\left(S^{\text{Perm}}_{\text{Chg}}, C^{\text{Avail}}_{\text{Batt}} + E^{\text{Exp}}_{\text{MinChg}} - E^{\text{Current}}_{\text{MinChg}}, -\frac{P_{\text{NetChg}}}{4}\right), 0\right)$$

$$S^{\text{Sched}}_{\text{Chg}} = S^{\text{Decl}}_{\text{Chg}}(\text{t}-2)$$

$$E^{\text{ID}}_{\text{aFRR+}} = \max\left(\min\left(\max\left(-(E^{\text{Charge}}_{\text{Batt}} - E^{\text{Exp}}_{\text{MinChg}} + \frac{P_{\text{MaxDis}}}{4} \cdot \text{window}) \cdot \eta_{\text{Batt}}, 0\right), \frac{P_{\text{NetDis}}}{4}\right) - E^{\text{DA}}_{\text{aFRR+}}, 0\right)$$

$$S^{\text{Batt}}_{\text{Market}} = \min\left(S^{\text{Sched}}_{\text{Chg}} - C^{\text{Sched}}_{\text{Batt}}, 0\right)$$

$$D^{\text{Dev}}_{\text{Batt}} = -E^{\text{Balance}}_{\text{Batt}}$$

$$C^{\text{Avail}}_{\text{Batt}} = \max\left(-\frac{P_{\text{NetChg}}}{4}, -(E^{\text{Charge}}_{\text{Batt}} + E_{\text{MaxCap}}) \cdot \frac{1}{\eta_{\text{Batt}}}\right)$$

$$C^{\text{UpReg}}_{\text{Batt}} = \min\left(\frac{P_{\text{NetDis}}}{4}, \min(E^{\text{Charge}}_{\text{Batt}} - E_{\text{MinChg}} + P_{\text{MaxDis}}, 0) \cdot \eta_{\text{Batt}}\right)$$

$$C^{\text{ExtChg}}_{\text{Batt}} = S^{\text{Batt}}_{\text{Market}}$$

$$E^{\text{aFRR+}}_{\text{Batt}} = \min\left(E^{\text{Calc}}_{\text{aFRR+}}, -E^{\text{Charge}}_{\text{Batt}}(\text{t}-1) \cdot \frac{1}{\eta_{\text{Batt}}}\right)$$

$$E^{\text{aFRR-Assist}}_{\text{Batt}} = \max\left(E^{\text{Calc}}_{\text{aFRR-}} - E^{\text{aFRR-}}_{\text{PV}}, C^{\text{Avail}}_{\text{Batt}}, \min\left(-\frac{P_{\text{NetChg}}}{4} - C^{\text{ExtChg}}_{\text{Batt}} - C^{\text{OtherDis}}_{\text{Batt}}, 0\right)\right)$$

$$E^{\text{Balance}}_{\text{Batt}} = \begin{cases} -E_{\text{MaxCap}} - E^{\text{NoLimitChg}}_{\text{Batt}}, & \text{if } E^{\text{NoLimitChg}}_{\text{Batt}} < -E_{\text{MaxCap}} \\ -P_{\text{MaxDis}} - E^{\text{NoLimitChg}}_{\text{Batt}}, & \text{if } E^{\text{NoLimitChg}}_{\text{Batt}} > P_{\text{MaxDis}} \\ 0, & \text{otherwise} \end{cases}$$

$$E^{\text{TechLossNetChg}}_{\text{Batt}} = \begin{cases} \frac{E^{\text{Balance}}_{\text{Batt}}}{\eta_{\text{Batt}}} - E^{\text{Balance}}_{\text{Batt}}, & \text{if } E^{\text{Balance}}_{\text{Batt}} > 0 \\ 0 \end{cases} + E^{\text{TechLossChg}}_{\text{Batt}}$$

$$E^{\text{TechLossNetDis}}_{\text{Batt}} = \begin{cases} -\frac{E^{\text{Balance}}_{\text{Batt}}}{\eta_{\text{Batt}}} - E^{\text{Balance}}_{\text{Batt}}, & \text{if } E^{\text{Balance}}_{\text{Batt}} < 0 \\ 0 \end{cases} + E^{\text{TechLossDis}}_{\text{Batt}}$$

$$E^{\text{Charge}}_{\text{Batt}} = E^{\text{Charge}}_{\text{Batt}}(\text{t}-1) + \left(C^{\text{ExtChg}}_{\text{Batt}} + E^{\text{aFRR+}}_{\text{Batt}} + E^{\text{aFRR-Assist}}_{\text{Batt}} + C^{\text{OffSched}}_{\text{Batt}} + C^{\text{OtherDis}}_{\text{Batt}}\right) + \left(E^{\text{TechLossNetChg}}_{\text{Batt}} + E^{\text{TechLossNetDis}}_{\text{Batt}} + E^{\text{Balance}}_{\text{Batt}}\right)$$

4.2.2. Operating rules

Market transactions are priced using the HUPX (Hungarian Power Exchange) DAM hourly and IDM quarterly average prices for the reference period. Among the balancing services, the aggregator participates in the aFRR (automatic frequency restoration reserve) market. The model takes into account the possibility of hourly bidding in the aFRR market and the possibility of intra-day (T-25 minutes) bidding in the intraday market. Reserves and activated quantities are based on a pricing strategy that can be set in the model. In our case, the offer prices are calculated based on a pre-fixed ratio of the moving average price of the market offers already known at the time, and in the case of negative aFRR energy bids, a fixed negative price was considered in order to avoid loss periods during activations. The timings used by the model correspond to real market rules for the different activities.

For each time interval, it must be ensured that the portfolio energy balance, i.e. the signed sum of generation, consumption, trade transactions, balancing energy and imbalance, is 0. All components except the imbalance can be directly influenced, and the imbalance can be understood as a residual error in the preliminary plans and in the actual operation, the reduction of which is a fundamental system objective.

In line with the European imbalance settlement rules [1], the imbalance settlement logic used in the model implies that imbalance can be not only an expense but also an income, depending on the relationship between the direction of the portfolio imbalance and the direction of the system imbalance. The model also employs a separate system imbalance prediction model [6] to generate additional revenue through deliberate deviation from the schedule in future settlement periods. Depending on the system state prediction, a deliberate schedule deviation can be chosen instead of keeping the schedule, which can help to increase the revenue from balancing energy.

No period should be allowed during which the reserved balancing capacity is not available or cannot be fulfilled at portfolio level, as this carries the risk of exclusion from the market. To this end, the planned use of solar PV should not jeopardize the availability of balancing reserves and the necessary recharging or discharging of energy storage should be ensured by the beginning of the settlement period.

The model calculates the energy flows for each element of the portfolio and its energy market components on the basis of the forecasts, the set parameters and the situations that occur in each period (schedule deviations, contracted reserves, activations) and their subsequent resolution. The objective function is to maximize the economic profit resulting from the balance of market transactions. The unit prices are known from historical data for the selected reference period, so the energy quantities and corresponding unit prices are used to

calculate the cash flow rate for that period. The energy flows for a given interval generate the initial data for the next interval, and the energy flows and associated cash flows for the next interval are generated based on the situation and responses to the situation.

Variables

- $P^{\text{Price}}_{\text{DAM}}$: HUPX DAM base price [Ft/kWh]
- $P^{\text{Price}}_{\text{ID60}}$: HUPX ID60 weighted average price [Ft/kWh]
- $P^{\text{Price}}_{\text{ID15}}$: HUPX ID15 weighted average price [Ft/kWh]
- $P^{\text{Bid}}_{\text{aFRR+}}$: aFRR+ bid price [Ft/kWh]
- $P^{\text{Bid}}_{\text{aFRR-}}$: aFRR- bid price [Ft/kWh]
- $P^{\text{Price}}_{\text{Balancing}}$: Balancing energy price [Ft/kWh]
- $E^{\text{MAVIR}}_{\text{aFRR+}}$: MAVIR activation according to aFRR+ MOL [kWh]
- $E^{\text{MAVIR}}_{\text{aFRR-}}$: MAVIR activation instruction according to aFRR- MOL [kWh]
- $P^{\text{Total}}_{\text{Sched}}$: Scheduled generation and consumption total [kWh]
- $S^{\text{Balance}}_{\text{Trade}}$: Scheduled commercial balance [kWh]
- $S_{\text{Imbalance}}$: Scheduling imbalance [kWh]
- $M^{\text{Measure}}_{\text{Total}}$: Total measurement [kWh]
- M_{Ref}: Reference [kWh]
- M_{Expected}: Expected measurement [kWh]
- I_{Pos}: Positive activation [kWh]
- R_{Pos}: Executed regulation + [kWh]
- $C^{\text{Settle}}_{\text{Pos}}$: Settled regulation + deviation from the market schedule [kWh]
- I_{Neg}: Negative activation [kWh]
- R_{Neg}: Executed regulation - [kWh]
- $C^{\text{Settle}}_{\text{Neg}}$: Settled regulation - deviation from the market schedule [kWh]
- E_{Balance}: Balancing energy [kWh]

- N_{NetCons}: Net consumption measured at the connection point
- $R_{\mathrm{DAM}}^{\mathrm{Rev}}$: Revenue from PV DAM sale [Ft]
- $R_{\mathrm{IDM}}^{\mathrm{Rev}}$: Revenue from PV IDM adjustment [Ft]
- $C_{\mathrm{IDM}}^{\mathrm{User}}$: Cost of user IDM supply [Ft]
- $C_{\mathrm{IDM}}^{\mathrm{Batt}}$: Cost of battery IDM charging [Ft]
- $R_{\mathrm{aFRR+}}^{\mathrm{Rev}}$: Revenue from aFRR+ regulation energy [Ft]
- $R_{\mathrm{aFRR\text{-}}}^{\mathrm{Rev}}$: Revenue from aFRR- regulation energy [Ft]

Equations

$$
\begin{aligned}
P^{\text{Total}}_{\text{Sched}} &= S^{\text{Cons}}_{\text{ID}} + P^{\text{Grid}}_{\text{Scheduled}} + C^{\text{Sched}}_{\text{Batt}} \\
S^{\text{Balance}}_{\text{Trade}} &= -(S^{\text{Sale}}_{\text{DAM}} + T^{\text{Trans}}_{\text{IDM}} + P^{\text{EndUser}}_{\text{Market}} + S^{\text{Batt}}_{\text{Market}}) \\
S_{\text{Imbalance}} &= P^{\text{Total}}_{\text{Sched}} + S^{\text{Balance}}_{\text{Trade}} \\
M^{\text{Measure}}_{\text{Total}} &= I^{\text{Point}}_{\text{Cons}} + I^{\text{Point}}_{\text{PV}} + C^{\text{ExtChg}}_{\text{Batt}} + E^{\text{aFRR+}}_{\text{Batt}} + E^{\text{aFRR-Assist}}_{\text{Batt}} + C^{\text{OffSched}}_{\text{Batt}} + C^{\text{OtherDis}}_{\text{Batt}} \\
M_{\text{Ref}} &= M^{\text{Measure}}_{\text{Total}} - R_{\text{Pos}} - R_{\text{Neg}} \\
I_{\text{Pos}} &= E^{\text{Calc}}_{\text{aFRR+}} \\
I_{\text{Neg}} &= E^{\text{Calc}}_{\text{aFRR-}} \\
M_{\text{Expected}} &= \begin{cases} M_{\text{Ref}} + I_{\text{Pos}} + I_{\text{Neg}}, & \text{if } I_{\text{Pos}} + I_{\text{Neg}} \neq 0 \\ P^{\text{Total}}_{\text{Sched}}, & \text{otherwise} \end{cases} \\
R_{\text{Pos}} &= E^{\text{aFRR+}}_{\text{Batt}} \\
C^{\text{Settle}}_{\text{Pos}} &= \begin{cases} \max(\min(M^{\text{Measure}}_{\text{Total}} - P^{\text{Total}}_{\text{Sched}}, 0), 0), & \text{if } I_{\text{Pos}} > 0 \\ 0, & \text{otherwise} \end{cases} \\
R_{\text{Neg}} &= E^{\text{aFRR-}}_{\text{PV}} + E^{\text{aFRR-Assist}}_{\text{Batt}} \\
C^{\text{Settle}}_{\text{Neg}} &= \begin{cases} \min(\max(M^{\text{Measure}}_{\text{Total}} - P^{\text{Total}}_{\text{Sched}}, 0), 0), & \text{if } I_{\text{Neg}} < 0 \\ 0, & \text{otherwise} \end{cases} \\
E_{\text{Balance}} &= M_{\text{Expected}} - M^{\text{Measure}}_{\text{Total}} \\
N_{\text{NetCons}} &= I^{\text{Point}}_{\text{Cons}} + C^{\text{ExtChg}}_{\text{Batt}} + E^{\text{aFRR-Assist}}_{\text{Batt}} + C^{\text{OffSched}}_{\text{Batt}} \\
R^{\text{Rev}}_{\text{DAM}} &= P^{\text{Price}}_{\text{DAM}} \cdot S^{\text{Sale}}_{\text{DAM}} \\
R^{\text{Rev}}_{\text{IDM}} &= P^{\text{Price}}_{\text{ID15}} \cdot T^{\text{Trans}}_{\text{IDM}} \\
C^{\text{User}}_{\text{IDM}} &= P^{\text{Price}}_{\text{ID15}} \cdot P^{\text{EndUser}}_{\text{Market}} \\
C^{\text{Batt}}_{\text{IDM}} &= P^{\text{Price}}_{\text{ID15}} \cdot S^{\text{Batt}}_{\text{Market}} \\
R^{\text{Rev}}_{\text{aFRR+}} &= P^{\text{Bid}}_{\text{aFRR+}} \cdot C^{\text{Settle}}_{\text{Pos}} \\
R^{\text{Rev}}_{\text{aFRR-}} &= P^{\text{Bid}}_{\text{aFRR-}} \cdot C^{\text{Settle}}_{\text{Neg}}
\end{aligned}
$$

4.2.3. Scenarios

The results of the operational model are presented under two scenarios. The "Basic Functions" scenario represents a traditional approach, while the "Extended Functions" scenario makes use of all the options presented in the previous sections.

Basic Functions

In the Basic Functions scenario, the planned production of solar panels is used by the consumer up to the level of its planned needs, with the remainder of the production plan being sold on the organized electricity market. The planned demand of the consumer that cannot be covered by solar PV production are purchased from the organized electricity market. The goal of the aggregator is to keep the schedule, which it tries to facilitate by controlling the battery (charging from the solar panel, buying from the utility grid, discharging to the utility grid) and the solar panel (allowing over-generation or curtailing it).

Extended Functions

In the Extended Functions scenario, the portfolio is also an active participant in the balancing market. It also tries to use its generation for internal use, but the aggregator also offers aFRR services via batteries and solar panels. Battery capacity is offered in positive direction in two-hourly increments in the next day's aFRR balancing capacity tenders according to a pre-defined pricing algorithm. Balancing energy bids are submitted for the contracted periods in accordance with the rules, also according to a pre-defined pricing algorithm. All contracted periods start at full load to ensure that TSO activations can be fully met. The battery can be recharged from solar PV generation on or off-schedule, and from the public grid on or off-schedule. If the battery is charged before the start of the contracted period, a positive aFRR offer will be made within the day.

Negative aFRR capacity bids are submitted based on the solar panel's planned generation for the next day. Energy bids are submitted for the contracted periods in accordance with market rules. If weather conditions and current portfolio circumstances allow, negative intraday balancing energy offers will also be submitted. The amount of energy fed into the public grid must be at least equal to the amount of energy offered. Negative aFRR activation may even include battery charging if necessary.

Considering that balancing energy can be an expense as well as an income for the balancing group, a system imbalance forecasting method is used by the balancing group to help balance the system during foreseeable settlement periods.

4.3. Results

The results of the two scenarios defined above are published below. The two scenarios presented cover the two extreme cases of possible operating models and market strategies, but in the space between them other intermediate operating approaches are conceivable. The large contrast is not an end in itself; rather, it demonstrates that the model developed is sen-

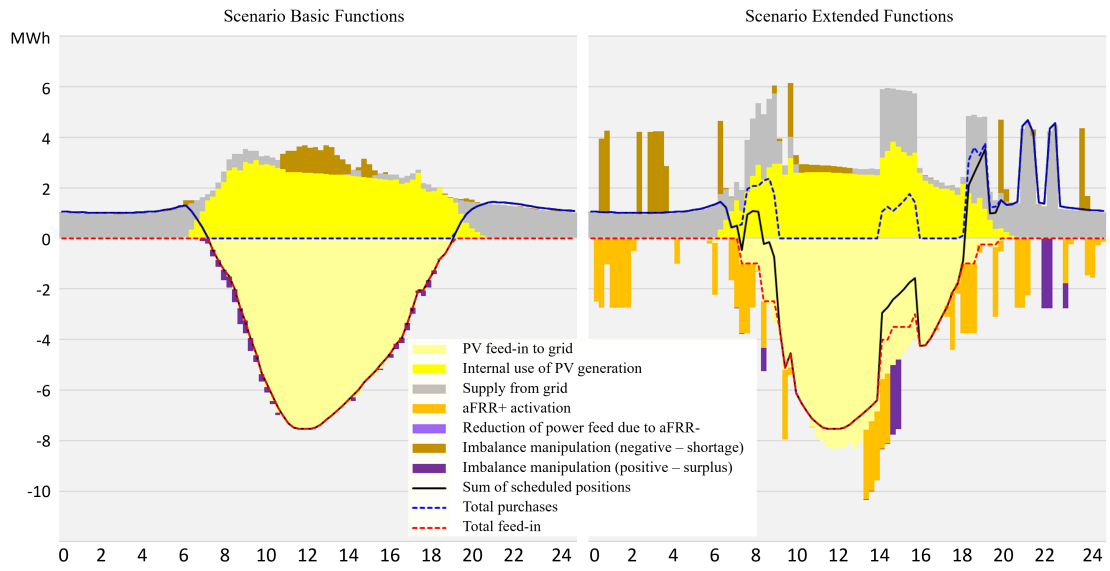


Figure 4.2: Energy flows for the scenarios

sitive to changes in the operating strategies employed and is consistent with the expectation that a more complex exploitation of assets and market opportunities will yield significant economic benefits. It is important to note that the "Extended Functionality" requires advanced control technology implementation and active commercial activity, the costs of which are not included in the financial results. Due to scope limitations, the detailed results of the simulation are illustrated with a quarter-hourly resolution plot of the events of a selected, typical but not average day, 4$^{\text{th}}$ May 2022. Size of assets in the portfolio:

- Consumer: 15 MW,
- Solar generation: 50 MW,
- Battery energy storage: 12 MW / 24 MWh.

Figure 4.2 compares the expected energy flows for the two scenarios during this day. For the Basic Functions case, it is easy to follow the process of scheduling and internal utilization of generation. During the night, consumption is entirely from the public grid, then increasingly offset by PV generation, and finally PV generation is also used for external sales. The purchase from the public grid during the daytime hours illustrates the higher consumption compared to the scheduled consumption, which is compensated by the discharge of the battery (between the 25th and 40th interval). The battery is recharged by the solar power plant, taken into account in the intraday schedules as a reduced quantity to be sold. From 10.15 am onwards, in addition to overconsumption, solar overgeneration appears to the

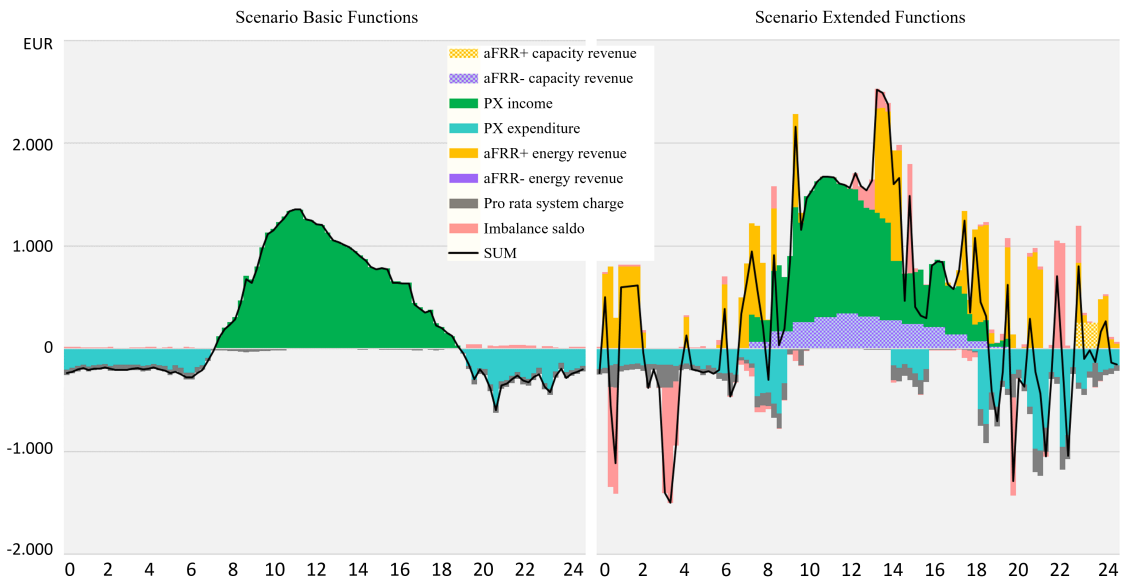


Figure 4.3: Financial results for the scenarios

extent that, in addition to covering the overconsumption, it is necessary to curtail generation. By the end of the day, the role of PV generation in supply is once again taken over by market purchases. Schedule deviations were present throughout the day, but only to a minimal extent.

In the case of the "Extended Functions", there are significant deviations from the market schedule due to the influence of both aFRR demand and balancing energy. In addition, it is worth noting the frequent surge in the level of purchase from the public grid, which is often necessary due to the urgency to recharge batteries.

Figure 4.3 summarizes the financial implications of energy flows during the day, based on simulated quarter-hourly energy flows and real quarter-hourly prices. In the case of Basic Functions, the cost of external supply of consumption and the system charges imposed on consumption dominate at the beginning of the day. There is also imbalance energy due to the limited flexibility (battery full, PV not generating), resulting in a small imbalance revenue during the period concerned. As solar PV generation ramps up, the HUPX balance becomes positive, system charges fall and the imbalance is zero for hours. At the end of the production period, HUPX expenditure and system charges increase again and imbalance reappears as a revenue item.

In the case of "Extended Functions", significant revenues are generated at the beginning of the day due to positive aFRR activations, this continues throughout the day with some

interruptions, and is supplemented at the end of the day by revenues from the contracted aFRR reserve capacity of the battery. In addition to the balancing market revenues, there are also imbalance costs due to occasional system imbalance forecasting errors. The increase in HUPX expenditure and system charges due to the frequent need to recharge the storage is striking. Every quarter hour that solar PV generation is sold, its downward balancing capacity is also contracted, resulting in additional revenue. Negative aFRR activations do not occur during the day.

While above I have shown the differences between the scenarios through selected examples, the following graphs summarize the figures for the whole year 2022 for the two scenarios. Looking at the energy flows on Figure 4.4, solar generation and behind-the-meter generation utilization show minimal differences between the two cases. The negative aFRR revenue appears as a new item, but does not have any particular spillover effects. However, the positive aFRR activation of the battery and its use for imbalance management lead to fundamental differences. While the annual energy flow of the battery is around 4-5 GWh in the base case, in is almost 25 GWh in the extended case. The increased demand for recharging energy (energy purchased for storage charging) in itself induces significant additional energy purchases, and internal balancing and supply options are also reduced. Purchases form the external market are increased by almost 19 GWh, while the use of synergies from the negative diversion of imbalance (cheap charging from external sources, internal use of solar PV generation) is doubled.

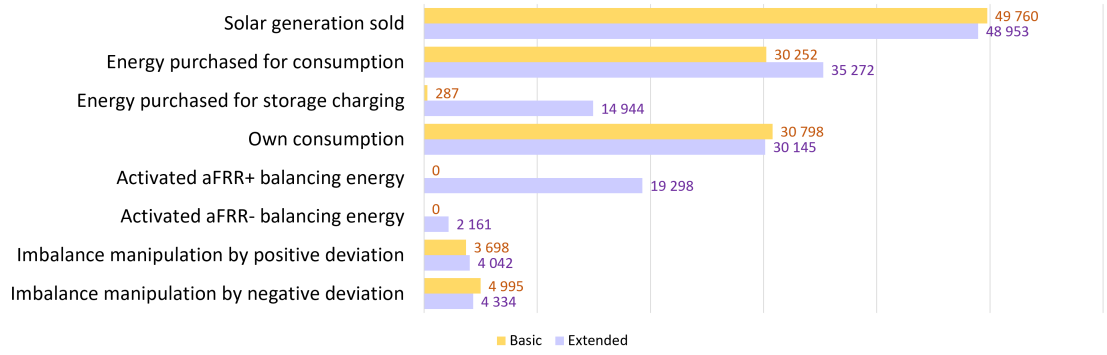


Figure 4.4: Summary of annual energy flows for the scenarios [MWh]

If we look at Figure 4.5, which summarizes the cash flows that are closely linked to the energy flows of operation under the two scenarios, the picture is even more colorful. The balance of organized electricity market transactions becomes significantly negative on an annual basis in the extended case. However, this is convincingly compensated by the revenues that can be generated by the use of balancing reserves. The frequent need to recharge the battery

almost doubles the expenditure on system charges. The positive balance of imbalance energy is also almost doubled due to system imbalance forecasts.

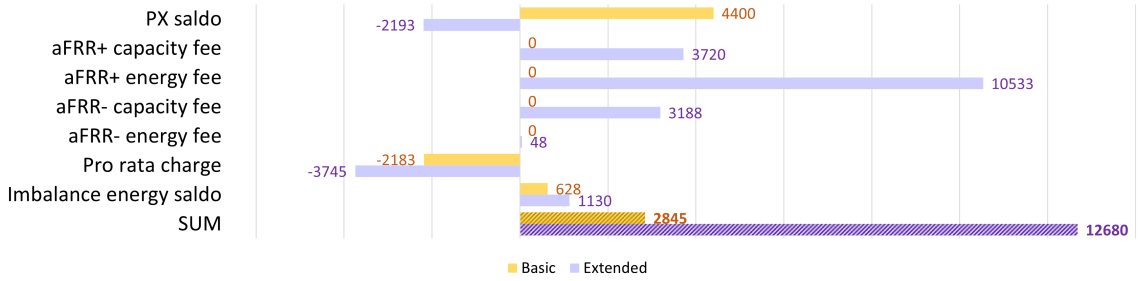


Figure 4.5: Summary of annual cash flows for the scenarios [EUR, in thousands]

The overall monetary difference is quite impressive, even though we know that this is not the total annual cash flow of the portfolio, and that there is a significant difference between the two cases for both Operational Expenditure (OPEX) and Capital Expenditure (CAPEX) - of course, the picture would be complete if these were taken into account.

It is beyond the scope of this article to discuss in detail the assumptions for additional years beyond the reference year, but some assertions can be made. If there is no change in market positioning as a strategy, similar energy flows can be expected for years to come. If the need for flexibility in the electricity system is not reduced, this assumption is in line with the principle of a conservative approach. Therefore, the cash flows for the following years can still be determined in the traditional way, based on a time series analysis of the reference period, i.e. by extrapolating the annual average changes of each item by different indices or discrete values.

4.4. Conclusion

By bringing together distributed generation and consumption units, it is possible to create a portfolio of a size and composition that can overcome barriers to entry and compete effectively in different markets. My model demonstrates how different generation and consumption technologies can work together in complex market conditions.

The purpose of analyzing different scenarios is not to identify the absolute economic optimum, that is the role of an aggregator actively engaged in real operations, using available information and accumulated operational and commercial expertise. Instead, my model is designed to provide a robust evaluation of the potential profitability from participating in a broad array of markets, considering real operational conditions and interactions at the highest feasible resolution, while acknowledging occasional inevitable inaccuracies.

The results demonstrate the potential variability in effectiveness that the same set of tools can exhibit under different operational strategies. More complex strategies typically require more intricate organizational, process, and IT support structures, leading to increased risks and higher implementation costs.

The insights from these simulations can both validate the anticipated positive outcomes of the proposed development and guide investors towards more feasible strategies (such as enhancing IT support, choosing suitable markets, adjusting portfolio composition, or involving a partner) before significant expenditures are incurred.

CHAPTER 5

Quantifying the power flexibility of a household with residential solar panels and energy storage

Flexibility is the modification of generation injection and/or consumption patterns in reaction to an external signal (price signal or activation) in order to provide a service within the energy system [26, 58]. It is the active management of an asset that can impact system balance or grid power flows on a short-term basis. The proper management of available flexibility, both on generation and the demand-side, can help to compensate for the lack of certainty of renewable sources. As variable renewable energy sources constitute a larger fraction of electricity supply, interest in flexible resources, including demand response, dispatchable generation, transmission interconnection, and storage technologies, is growing [60]. In the near future, demand-side flexibility would be required for managing power grids. Buffered heat pumps (HPs) are examples of domestic hot water (DHW) systems and stationary batteries for flexible residential loads that can store energy from the grid and local generators (e.g., solar panels). Both forms of stored energy (thermal or chemical) can be used to support self-consumption.

Building energy systems will serve as one possible source of demand-side energy flexibility. Traditionally, buildings have been consumers of energy, but they are responsible for a large share of energy demand and therefore may play a key role in improving the flexibility in the demand-side of the entire energy system.

The size of a residential household does not reach the volume to be able to bid on energy markets independently, so prosumers may share their flexibility with a pool. The flexibility pool is managed by an aggregator [64], which can be a DSO, a TSO, or an independent intermediary entity. Whatever market design is implemented, the prosumer receives a premium as compensation for the provided flexibility from the aggregator [51]. The results of [54] showed that the aggregator is able to find a match between the flexibility provisioned and the flexibility procured by the DSO. The aggregator reduces the remuneration costs paid to users for the flexibility at the same time. The aggregator bids on energy markets aggregating prosumers' flexibility, so it needs to have information about the flexibility volume of the pool. Either the prosumer provides a bid or the aggregator bears the risk; therefore, the amount of flexibility must therefore be quantified, and a function has to exist to predict the available flexibility.

In this chapter a model-based approach is applied to simulate the energy processes of a residential building, in which the electrical devices were modeled and power consump-

tion/generation was calculated for each device, so that the consumption/generation modification opportunities could be defined as the prosumer's flexibility. By configuring the external parameters of the model, the energy consumption of the house and the available flexibility could be simulated for different scenarios. A supervised prediction could be built on it, in which the simulation produced flexibility as a dependent variable. The prosumer model could be scaled to a pool of prosumers that could provide flexibility inputs for an aggregator function.

The contribution can be summarized in three main points. First, a thermodynamic model of a typical residential building energy system is developed based on first engineering principles. The simulation model was implemented in the MATLAB Simulink environment. Second, the thermodynamic model was integrated with the electrical consumption and generation assets of the house, and a calculation scheme was developed for determining the flexibility capability of the house. Finally, a linear and a non-linear prediction method was developed and implemented in Python environment to provide a 15-minute forecast of the prosumer flexibility.

The chapter is organized as follows. Section 5.1 provides a review of the most relevant literature. Modeling assumptions and the structure of the proposed residential prosumer model are detailed in Section 5.2, and based on it, a simulation-based case study is presented in Section 5.3. This is followed by Section 5.4, which summarizes and concludes the chapter.

5.1. Related work

In the example of [102], extrapolating the Belgian national level implied that domestic flexibility could equal 1.8% (upward) and 12.1% (downward) of installed generation capacity. The results presented by the authors of [24] forecast that the household sector will be able to contribute significantly to the distribution system stabilization with an average potential of 30 GW downwards and 3 GW upwards flexibility in the year 2025. This paper analyzed the potential that is made possible by technology for the provision of system services by households. Single-family and twin homes were the main focus, since those are the types of households in which all system components are available. The Electric Power Research Institute (EPRI) estimated a technical potential summer peak reduction of 175GW from demand response by 2030 in the USA [52]. The Clean Energy for All Europeans Package (CEP) [90] empowers prosumers in the EU to offer their flexibility. The volume necessary for energy markets requires a large number of prosumers to participate in energy services. Energy companies require digitalization and the utilization of advanced technologies [10], as well as market models [9] to involve prosumers in the electricity markets. Using blockchain technology can simplify the management of microgrid power transactions and realize peer-to-peer power transactions [96].

The technical potential of demand modification profiles was presented for different regions in the USA in [52]. There are six different load curves of the demands of residential users [83], for which the demand-side management (DSM) techniques are the following: peak clipping, valley filling, load shifting, load reduction, load growth, and flexible load shape. The flexible load shape technique was assumed in this paper.

Paper [5] proposed an energy storage system as a possible flexibility resource and its potential role in the future smart grid network. The potential benefits of energy storage as the flexibility resource can facilitate increased participation of storage in different electricity markets and improve the flexibility of smart grid operation with a high penetration of renewable energy sources. Paper [85] analyzed the potential of grid flexibility supply by combined heat and power systems installed at business facilities. Simulations were performed for a representative day in each season and four types of facilities. Prosumer systems can operate both an autonomous (off-grid) and grid-connected (on-grid) systems. While the energy generated in autonomous systems is consumed by the system's own consumption devices, the energy produced in grid-connected systems can be consumed by the system internally, or if there is energy surplus, it can supply the grid [7].

The flexibility of industrial prosumers is a widely researched area with several outstanding results and papers [39, 43]; however, with the advent of the smart grid, this is increasingly complemented by the flexibility of residential actors. Residential prosumers form a promising source of flexibility due to their distributed location and substantial share of the electricity market. Home energy management systems (HEMS) can reduce electricity consumption by scheduling electrical appliances [82]. There are several control strategies and methods for HEM: AI-based control (predictive control, optimization control), linear online control, and storage systems.

Quantifying flexibility is challenging due to its complex electrothermal dynamics and time delay effects in general. Methodologies to quantify the energy flexibility of buildings are affected by the definition of flexibility followed by the respective research. A summary of quantification methods for the energy flexibility of buildings was provided by [55], in which the characterization of energy flexibility was a demand increase/generation decrease as negative flexibility and a demand decrease/generation increase as positive flexibility. The methodology proposed by [66] considers the flexibility of a specific system as the ability to shift the consumption of a certain amount of electrical power in time. Reference [19] defined flexibility as the possibility to change the electricity consumption of a building from the reference scenario at a specific point in time and during a certain time span. The quantification of flexibility from a district heating system point of view was given in [94]. The district heating system was firstly decomposed into multiple parallel subsystems

with simpler topological structures. The maximum flexibility of each subsystem was then formulated as a delayed optimal control problem, and finally, the available flexibility from the original system was estimated by aggregating the flexibility of all subsystems.

In [56], household devices were categorized as shifted, but not varied, shifted and varied, and not shifted, but varied. The simulation results showed that the flexibility of houses under testing had maximum power values of 200–500 W. The authors of [77] described the customer-side time flexibility with respect to white goods (washing machine, dryer, dish washer) with two parameters: configuration time and deadline. The paper modeled customer flexibility behavior with finite mixture models.

5.2. System model

To study demand flexibility, I assumed a residential building located in Hungary. The building is equipped with an electric heating system, a separate water heater, a home energy storage system, and rooftop PV panels (see Figure 5.1).

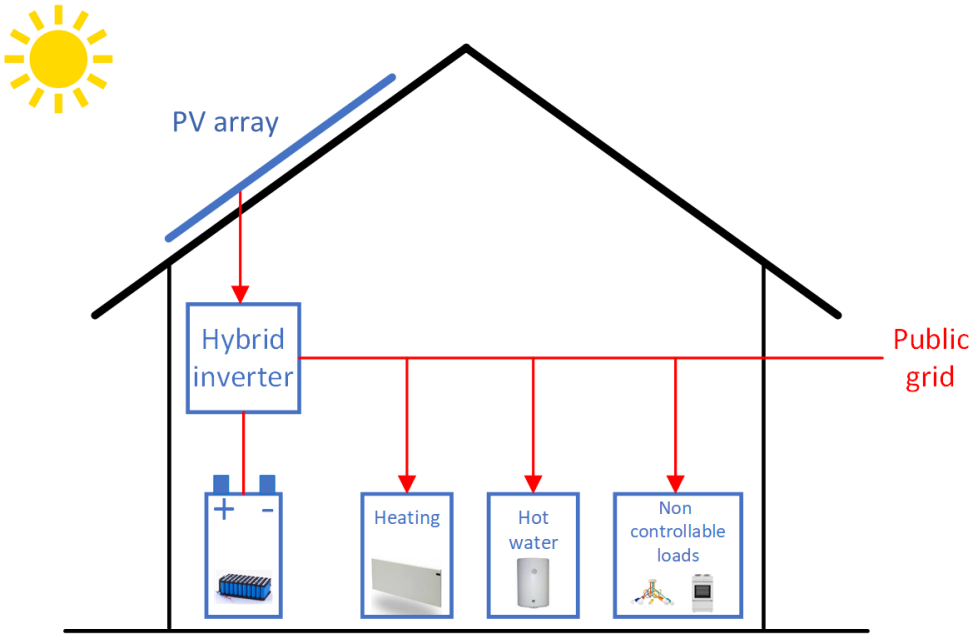


Figure 5.1: High-level model of the system. Line segments without arrow heads represent bidirectional power flow.

The prosumer model consists of energy consumers (space heater, hot water, and non-controllable load), a producer (PV), and a battery. The house is connected to the grid (on-grid mode), and power flow is available in both directions. All the energy needs of the house are supplied by the PV panels and the grid, and there is no other source of power

(e.g., gas, central heating). Besides using energy from the grid, the PV panels and the battery can also supply the consumers. When available, self-consumption is preferred. The available flexibility is calculated by summing the instantaneous flexibility of all devices in both directions over time.

A thermal model was developed for the space heating and hot water system. It is necessary to ensure that consumers have access to certain appliances, as not all electrical appliances can be controlled without inconveniencing the consumer. The flexibility of these appliances (e.g., lights, water, kitchen devices, other household appliances) is not considered; their consumption is referred to as non-controllable load. The energy required is supplied from the PV system, energy storage, or the grid. A conventional greedy algorithm was implemented to control the operation of the energy storage system when no external flexibility regulation is applied. The objective is to minimize grid usage.

The objective of the model is to describe the dynamic behavior of the energy consumption of a typical house with a garden. The aim of the model is not to show perfect quantitative behavior with the real system; rather, it is expected that the essential dynamic transients characterizing each module will be recognizably and correctly described by the model. The key steps of the modeling procedure were performed in accordance with the methodology outlined in [37].

There is no need for a separate flexibility control block in the case of the default operation scenario, when no external flexibility activation occurs. All the dispatchable devices can operate autonomously, and the greedy implementation of battery operation requires no central control logic. Flexibility is calculated for each device, and the sum represents the available flexibility of the house for both the up and down direction. Flexibility follows the interpretation proposed by the authors of [55], it is assumed to be a signed value, positive when generation can be increased or consumption can be decreased and negative in the case when generation can be decreased or consumption can be increased.

5.2.1. Hot Water

Hot water is supplied by an electric water tank (DHW tank) containing m_w kilograms of water. A setpoint (T_w^{sp}) specifies the maximum temperature of the water, and a thermostat controls the heating cycles. The thermostat switches the heater off at the setpoint and turns it on (Equation (5.2)) when the temperature drops below the setpoint by a threshold value (T_w^{th}). When the heater is on, a heating wire warms up the water by η_w efficiently consuming a constant level of power (P_w).

A water tank model calculates the water temperature dynamics (Equation (5.1)). The water temperature change ($\frac{dT_w(t)}{dt}$) has three main inputs: heat provided by the heater ($\frac{dQ_w^g(t)}{dt}$),

hot water consumption ($\frac{dQ_w^c(t)}{dt}$), and heat losses ($\frac{dQ_w^{hl}(t)}{dt}$). When hot water is consumed, the same amount of cold water ($\frac{dm_w^c(t)}{dt}$) fills the tank. The inflow water temperature is the same as the outside temperature ($T_a^{out}(t)$); thus, consumption cools the tank. The heat outflow is proportional to the temperature difference between the cold and the tank water ($T_w(t)$). Heat loss is calculated considering a heat loss parameter (hl_w) [18] and the difference between water and room temperature (T_a^{in}). The room temperature is an output signal of the heating subsystem.

The energy balance of the hot water subsystem is expressed by the following equations:

$$\begin{aligned} Cons_w(t) &= P_w^{max} \cdot Reg_w(t) \\ \frac{dQ_w^g(t)}{dt} &= Cons_w(t) \cdot \eta_w \\ \frac{dQ_w^c(t)}{dt} &= (T_w(t) - T_a^{out}(t)) \cdot \frac{dm_w^c(t)}{dt} \cdot C_w \\ \frac{dQ_w^{hl}(t)}{dt} &= (T_w(t) - T_a^{in}) \cdot m_w \cdot hl_w \\ \frac{dT_w(t)}{dt} &= \left(\frac{dQ_w^g(t)}{dt} - \frac{dQ_w^c(t)}{dt} - \frac{dQ_w^{hl}(t)}{dt} \right) \frac{1}{m_w \cdot C_w}, \end{aligned} \tag{5.1}$$

where $Cons_w(t)$ stands for the actual water consumption of the house.

$$Reg_w(t) = \begin{cases} 1 & \text{, when the water heater is on,} \\ 0 & \text{, when the water heater is off.} \end{cases} \tag{5.2}$$

Power is a signed value. $P_w(t)$ is negative (consumption) when the heater is on ($Reg_w(t)$):

$$P_w(t) = -Cons_w(t). \tag{5.3}$$

The calculated flexibility can be defined for both the up and down direction between current power consumption and the maximum power capacity (down) or 0 (up) using (Equation (5.4)) below.

$$\begin{aligned} F_w^{up}(t) &= Cons_w(t) \\ F_w^{down}(t) &= Cons_w(t) - P_w^{max} \end{aligned} \tag{5.4}$$

The parameters of the DHW tank and the consumption of hot water are presented in Table 5.1.

Table 5.1: Exogenous parameters of hot water dynamics

Parameter	Description	Unit
m_w	mass of water in the tank	kg
$T_w(0)$	initial water temperature	°C
T_w^{sp}	water temperature setpoint	°C
T_w^{th}	water temperature threshold	°C
$dm_w^c(t)/dt$	water consumption	kg/s
C_w	heat capacity of water	J/(kg·K)
η_w	efficiency of the water heater	-
P_w^{max}	power capacity of the water heater	W
hl_w	heat loss rate	W/(kg · K)
$T_a^{in}(0)$	initial indoor air temperature	°C
$T_a^{out}(t)$	outdoor temperature	°C

5.2.2. Heating

The thermal model of a house calculates the power consumption of the heating system that keeps the indoor temperature around a defined setpoint. The heating system is equipped with a thermostat and an electric heater. Similar to the water heater, the thermostat switches the heater on and off (Equation (5.9)) when the temperature drops below and above the setpoint (T_a^{sp}) by a predefined threshold (T_a^{th}). An air-to-air heat pump supplies warm air for the house, operating at an average COP ratio (cop).

Total thermal resistance (R^{tot}) is calculated from the geometry and the material properties of the house (Equation (5.5)). The walls of the house are made of bricks, and Expanded Polystyrene (EPS) insulation is used on the walls and the slab.

$$\begin{aligned} A_{wd} &= n_{wd} \cdot h_{wd} \cdot w_{wd} \\ A_{wl} &= 2 \cdot l_{wl} \cdot h_{wl} + 2 \cdot w_{wl} \cdot h_{wl} - A_{wd} \\ A_{sl} &= w_{wl} \cdot l_{wl} \\ A &= A_{wd} + A_{wl} + A_{sl} \end{aligned} \tag{5.5}$$

The thermal resistance of thermally homogeneous components is calculated using (Equa-

tion (5.6)).

$$\begin{aligned} R_{wd} &= 1/U_{wd} \\ R_{wl} &= d_{wl}^{br}/\lambda^{br} + d_{wl}^{ins}/\lambda^{ins} \\ R_{sl} &= d_{sl}^{ins}/\lambda^{ins} \end{aligned} \tag{5.6}$$

The total thermal resistance (R^{tot}) is determined by assuming one-dimensional heat flow perpendicular to the walls. It is given (Equation (5.7)) by the method in the ISO 6946/2007 standard [42].

$$R^{tot} = \frac{A}{A_{wl}/R_{wl} + A_{wd}/R_{wd} + A_{sl}/R_{sl}} \tag{5.7}$$

A thermal model calculates the indoor air temperature dynamics of the house. Its two main inputs are the heat provided by the heating system and heat losses. Contrary to the water heater, heat gain ($\frac{dQ_h^g(t)}{dt}$) is not persistent, but proportional to the temperature difference between the room (T_a^{in}) and the constant heated air temperature (T_h). Heat loss ($\frac{dQ_h^{hl}(t)}{dt}$) is proportional to the temperature difference between the room and outdoor temperature (T_a^{out}). The indoor temperature time derivative ($\frac{dT_a^{in}(t)}{dt}$) is expressed by the following equations:

$$\begin{aligned} \frac{dQ_h^g(t)}{dt} &= \min\left((T_h - T_a^{in}(t))\frac{dm_a(t)}{dt}C_a, P_h^{max}cop\right) Reg_h(t) \\ \frac{dQ_h^{hl}(t)}{dt} &= \left(T_a^{in}(t) - T_a^{out}(t)\right)\frac{1}{R^{tot}}A \\ \frac{dT_a^{in}(t)}{dt} &= \left(\frac{dQ_h^g(t)}{dt} - \frac{dQ_h^{hl}(t)}{dt}\right)\frac{1}{(l_H \cdot w_H \cdot h_H)\cdot \rho_a \cdot C_a} \end{aligned} \tag{5.8}$$

$$Reg_h(t) = \begin{cases} 1 & \text{, when the heater is on,} \\ 0 & \text{, when the heater is off.} \end{cases} \tag{5.9}$$

$Cons_h(t)$ denotes a theoretical power consumption that is necessary to warm the indoor air

up to the constant heated air temperature.

$$Cons_h(t) = \min\left(\frac{(T_h - T_a^{in}(t))\frac{dm_a(t)}{dt}C_a}{cop}, P_h^{max}\right) \tag{5.10}$$

Power is a signed value. $P_h(t)$ is negative (consumption) when the heater is on:

$$P_h(t) = -Cons_h(t) \cdot Reg_h(t). \tag{5.11}$$

The available flexibility quantity from the heating system is calculated for both directions (Equation (5.12)). When the heater is on, its power consumption can be decreased, and the maximum upward flexibility is the difference between instantaneous power and zero. When it is off, downward regulation is available by turning the heater on.

$$\begin{aligned} F_h^{up}(t) &= Cons_h(t) \cdot Reg_h(t) \\ F_h^{down}(t) &= -Cons_h(t) \cdot (1 - Reg_h(t)) \end{aligned} \tag{5.12}$$

Table 5.2: Parameters of the heating system

Parameter	Description	Unit
l_{wl}	length of the house	m
w_{wl}	width of the house	m
h_{wl}	height of the house	m
n_{wd}	number of windows	-
w_{wd}	width of a window	m
h_{wd}	height of a window	m
λ^{br}	thermal conductivity of brick	W/(m·K)
λ^{ins}	thermal conductivity of EPS insulation	W/(m·K)
d_{wl}^{br}	brick thickness	m
d_{wl}^{ins}	insulation thickness of the walls	m
d_{sl}^{ins}	insulation thickness of the slab	m
U_{wd}	heat transfer coefficient of the windows	W/(m^2·K)
$T_a^{in}(0)$	initial indoor air temperature	°C
$T_a^{out}(t)$	outdoor temperature	°C
T_a^{sp}	air temperature setpoint	°C
T_a^{th}	air temperature threshold	°C
T_h	temperature of air exiting the heater	°C
$^{dm_a(t)}/_{dt}$	air flow rate of the heater	kg/s
C_a	heat capacity of air (273 K)	J/(kg·K)
ρ_a	density of air at sea level	kg/m^3
P_h^{max}	maximum power capacity of the heater	W
cop	heating system's coefficient of performance	-

5.2.3. Storage

Storage provides the flexibility of shifting energy over time. A conventional greedy algorithm was implemented to control the storage operation when no external flexibility regulation was applied [51] in order to prefer self-consumption and reduce feed-in power. If there is a higher consumption by the household than generation by the PV, the storage is discharged until a minimum charge level. When generation surplus occurs, the storage is charged until it is full.

The storage model (Equations (5.13)–(5.16)) processes the power requirement ($Reg_b(t)$), determines the default storage operation (charge: $Cons_b^{pos}(t)$ or discharge: $Cons_b^{neg}(t)$), and calculates current state-of-charge ($SoC_b(t)$) taking into consideration the power limits (P_b^{max}), capacity limits (SoC_b^{min}, SoC_b^{max}), and efficiency ratio (η_b).

$$Reg_b(t) = P_w(t) + P_h(t) + P_l(t) + P_{pv}(t) \tag{5.13}$$

$$Cons_b^{pos}(t) = \begin{cases} \min(\max(0, Reg_b(t)), P_b^{max}), & \text{when } SoC_b(t) < SoC_b^{max}, \\ 0, & \text{otherwise} \end{cases} \tag{5.14}$$

$$Cons_b^{neg}(t) = \begin{cases} \max(\min(0, Reg_b(t)), -P_b^{max}), & SoC_b(t) > SoC_b^{min}, \\ 0, & \text{otherwise} \end{cases} \tag{5.15}$$

$$\frac{dSoC_b(t)}{dt} = \frac{Cons_b^{pos}(t) \cdot \eta_b + Cons_b^{neg}(t) \cdot (2 - \eta_b)}{Cap_b} \tag{5.16}$$

The battery's contribution to the net power balance of the house ($P_{battery}(t)$) has the opposite sign of battery consumption.

$$P_b(t) = -Cons_b^{pos}(t) - Cons_b^{neg}(t) \tag{5.17}$$

Given the maximum charge (P_b^{max}) and discharge (P_b^{min}) power of the energy storage and the instantaneous power output, both up and down regulation capacity can be calculated ($F_b^{up}(t)$, $F_b^{down}(t)$) when the state-of-charge is between the charge limits.

$$\begin{aligned} F_b^{up}(t) &= P_b^{max} - P_b(t) \\ F_b^{down}(t) &= -P_b^{max} - P_b(t) \end{aligned} \tag{5.18}$$

Table 5.3: Storage parameters

Parameter	Description	Unit
$SoC_b(0)$	battery initial state-of-charge	%
SoC_b^{min}	battery minimum state-of-charge	%
SoC_b^{max}	battery maximum state-of-charge	%
P_b^{max}	maximum power of the battery	W
Cap_b	capacity of the battery	Wh
η_b	efficiency of the battery	-
$P_{pv}(t)$	power generation of the PV panels	W
$P_l(t)$	power consumption of the non-controllable load	W

5.2.4. non-controllable Load

The consumption of the household is made up of controllable and non-controllable loads. While the controllable load was presented in the previous sections, measured data is used for the non-controllable load time series ($P_l(t)$). No flexibility is available for the non-controllable load.

5.2.5. PV Generation

PV generation ($P_{pv}(t)$) depends only on solar radiation. When no flexibility control is applied, it is assumed that the panels always generates the maximum power, and no upward regulation is available. The PV can offer downward flexibility ($F_{pv}^{down}(t)$) between 0 and its current generation.

$$F_{pv}^{down}(t) = -P_{pv}(t) \tag{5.19}$$

5.2.6. Power Balance and Total Flexibility

The power balance of the house ($P(t)$) is calculated by summarizing the signed values of each component.

$$P(t) = P_w(t) + P_h(t) + P_b(t) + P_l(t) + P_{pv}(t) \tag{5.20}$$

The total flexibility is calculated separately for the up ($F^{up}(t)$) and down ($F^{down}(t)$) direc-

tion:

$$F^{up}(t) = F_w^{up}(t) + F_h^{up}(t) + F_b^{up}(t),$$
$$F^{down}(t) = F_w^{down}(t) + F_h^{down}(t) + F_b^{down}(t) + F_{pv}^{down}(t). \tag{5.21}$$

5.2.7. Model Verification

The simulation parameters were configured taking the geometry and materials used in the well-insulated, single-story residential house built in 2015 in Hungary. The building has a ground floor area of 172 m^2, made of insulated brick (38 + 15 cm). The ceiling has a concrete structure with 30 cm of insulation. The windows have three-pane thermal insulated glazing.

The parameters and calculated variables were compared to the values of the single-family house involved in the IEA EBC Annex 58 project [45]. The referenced building area is 100 m^2; its brick walls are insulated, and double-pane windows were built in. The benchmark building has slightly worse U-values [74], but the total conductance is lower, because of the difference in size. Interior walls were also considered in the thermodynamic calculations. The comparisons of the heat transfer coefficients are given in Table 5.4.

Table 5.4: Calculated heat transfer coefficients.

Component	**Building in the Present Study**	**Building in the IEA EBC Annex 58 Project**	**Unit of Measure**
External walls	0.16	0.2	W/(m^2·K)
Windows	0.9	1.12	W/(m^2·K)
Slab	0.12	0.17	W/(m^2·K)
Total envelope conductance	57	36.6	W/K

The model was tested with manual inputs. Figure 5.2 shows the model response for an arbitrarily chosen input set. In the beginning, there is no consumption. PV generation ramps up to 4 kW. It charges the battery and supplies the load between one and two. The power balance of the house remains zero until the battery is full. At Hour 5, the outdoor temperature drops to −2 °C from 24 °C, making the space heater turn on after 40 min. Between Hours 6 and 8, 6 l/min hot water is also used. The battery supplies the non-controllable load and the consumption of both heaters until the state-of-charge (SoC) reaches its minimum level. The building is fed from the grid after Hour 8. It can be concluded that the proposed simulation model corresponded to the engineering expectations.

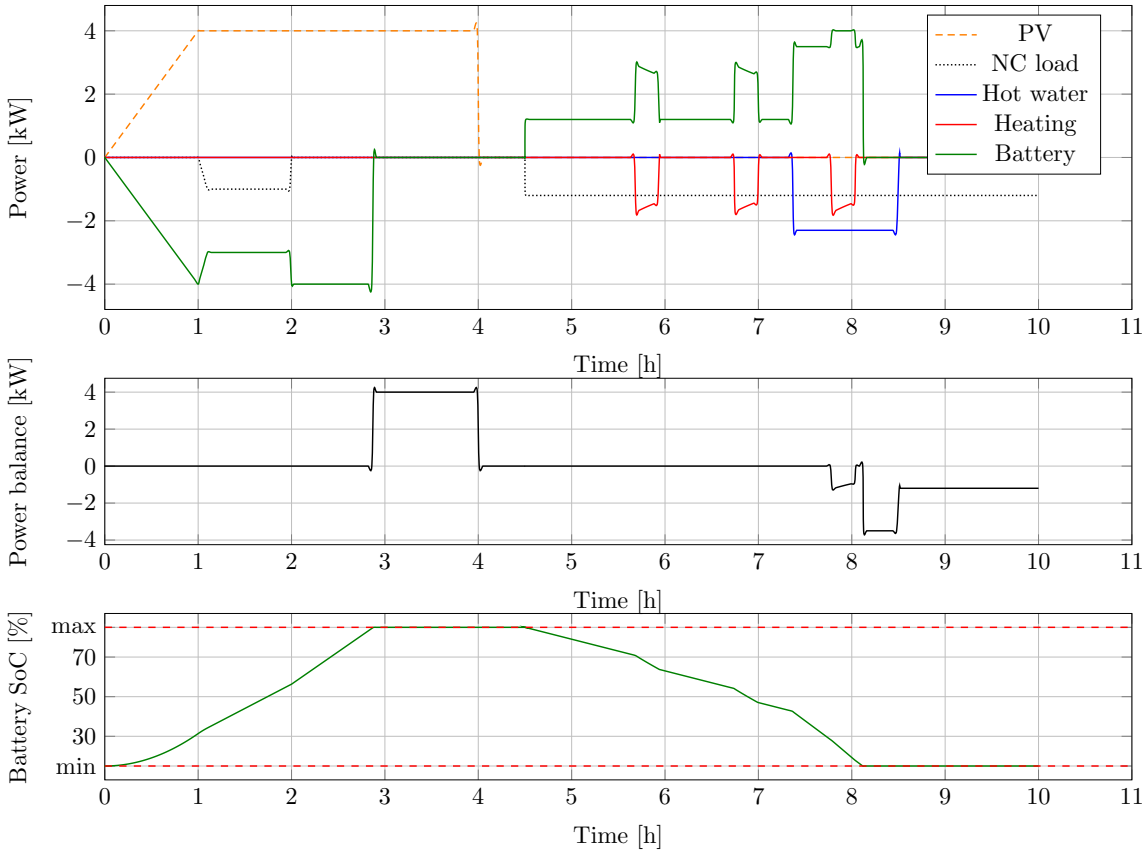


Figure 5.2: Power of the devices for the verification inputs.

5.3. Results

The energy model of the house in Section 5.2 has been implemented in MATLAB Simulink. It calculates the total energy consumption/generation and the available flexibility that it could provide. Four simulations were performed to analyze the flexibility under different weather conditions and patterns for a 24-hour period.

5.3.1. Scenarios

The scenarios differs in the PV generation and outer temperature time series data, and it is assumed that the non-controllable load and water consumption patterns are the same in each scenario. Please refer to Table A.2 to review additional parameter values.

Figure 5.3 shows the difference between the solar generation profile for the different scenarios. The 1-minute measurements from a 400 kVA PV park were collected and normalized by the peak power capacity. The relative production of each scenario was multiplied by the capacity of the modeled rooftop panels to obtain the PV generation time series ($P_{pv}(t)$).

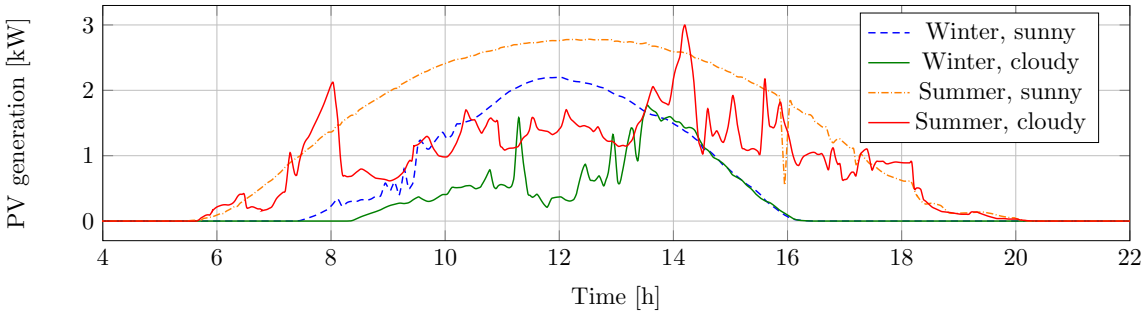


Figure 5.3: PV generation profiles for the four scenarios. The higher generation at 8 a.m. on a cloudy summer day (red line) compared to a sunny summer day (orange, dashed) comes from the efficiency increase caused by the cooling (a possible cloud passing by) at 7 a.m.

Figure 5.4 shows the environmental temperature (T_{out}) for the different scenarios. Temperature data were collected from the PV site ($T_{air}^{outdoor}(t)$).

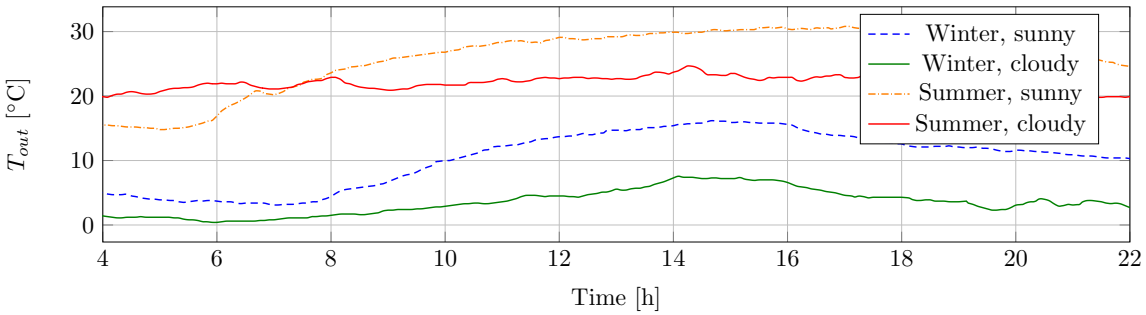


Figure 5.4: Outer temperature profiles for the four scenarios.

The Almanac of Minutely Power dataset (AMPds) [57] contains electricity, water, and natural gas measurements at one minute intervals for two years of monitoring. There is a total of twenty-one power meters and two water meters installed in a residential house similar to the one analyzed in this study. The 21 electronic submeters are assigned to groups: controllable and non-controllable devices. The controllable devices include the HVAC (heating, ventilating, and air conditioning), heat pump, and hot water heater. The non-controllable group encompasses the bedroom, basement, dining room plugs and lights, clothes washer and dryer, dishwasher, kitchen refrigerator and oven, garage, home office, entertainment, utility room, and outside plugs. All non-controllable measurements are added, and a typical 1-day time series is created by calculating the average value of the same minute for each day of the two-year monitoring period. Figure 5.5 shows the non-controllable consumption for all the scenarios ($P_{load}(t)$).

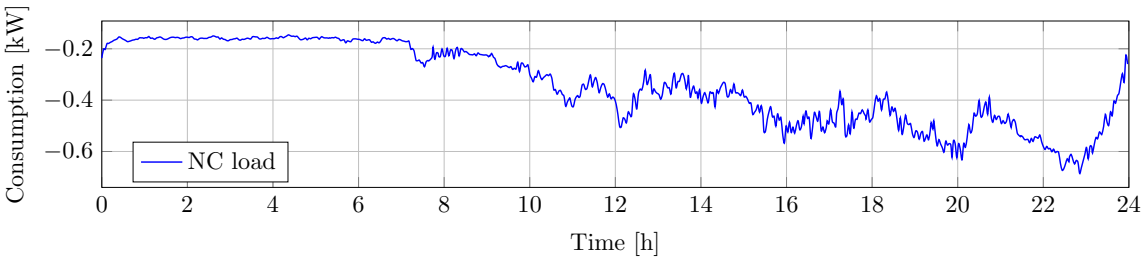


Figure 5.5: Non-controllable electrical load for the four different scenarios.

Hot water consumption (Figure 5.6) is calculated by the same method as the non-controllable load. The Almanac of Minutely Power dataset was the source of the 1-minute consumption data. The mean value was calculated for every minute of the day to generate a 1-day normal water consumption ($^{dm^{cons}_{water}(t)}/_{dt}$).

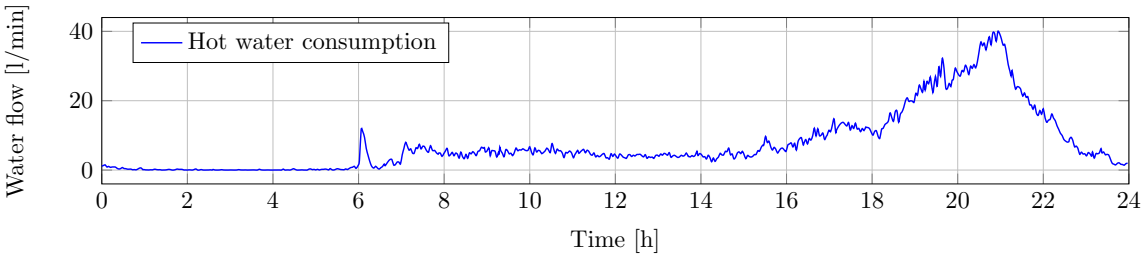


Figure 5.6: Hot water consumption for the four different scenarios.

5.3.2. Simulation Results

Winter—Sunny Day

Figure 5.7 shows the power consumption and generation of the simulated devices. Supply from the grid and PV generation is the primary sources of energy. Net power is the balance of the house; it is the volume of power consumption from or fed to the grid. When the PV generates sufficient power to feed all consumption units, the energy surplus charges the battery. The greedy battery control method discharges the storage when the PV is low. After 8 p.m., the house is supplied from the grid again, after the battery becomes empty.

The band of flexibility is not symmetric: although the house is a prosumer, generation and storage capacity is limited, and consumption is intermittent. There are more occasions when a consumption device could be turned on than off, so the downward flexibility is higher. PV generation increases the upward flexibility by charging the battery: the battery consumption can always be switched to production.

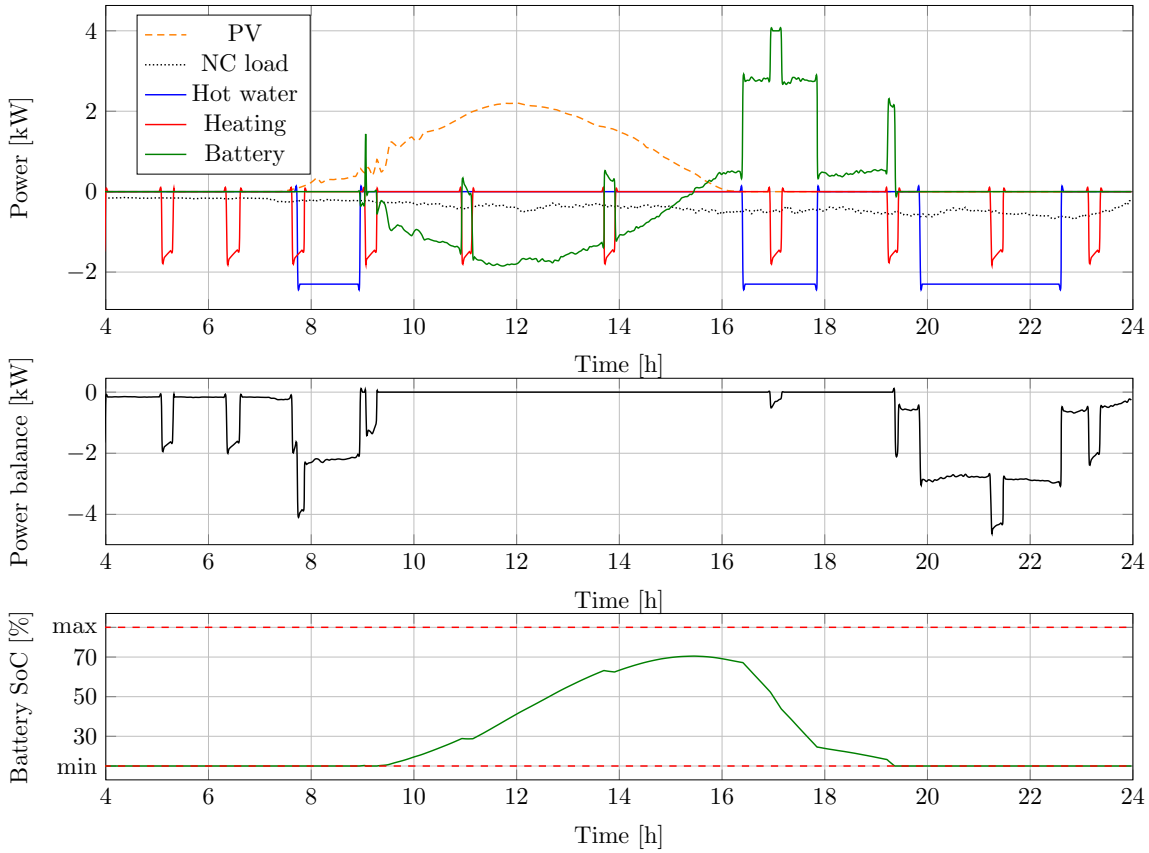


Figure 5.7: Power consumption/generation of devices for the winter sunny day scenario.

The upward and downward flexibility capabilities of each device are added, resulting in the building's maximum flexible power (Figure 5.8).

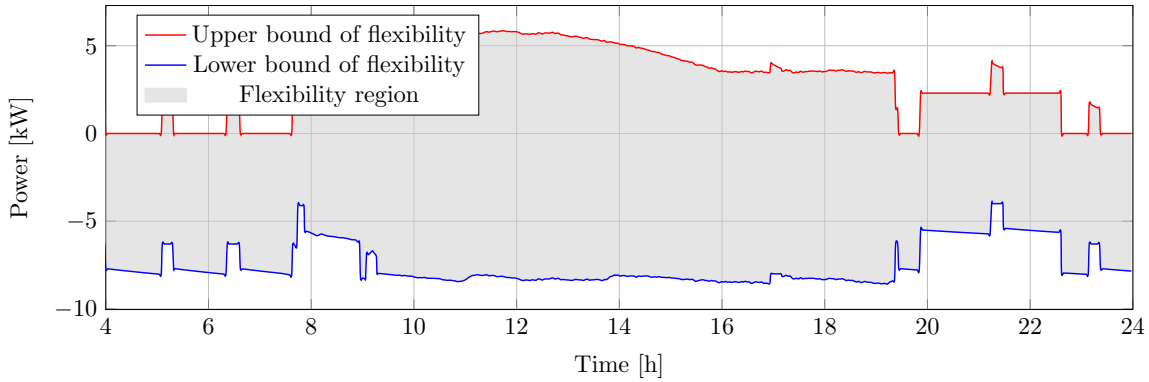


Figure 5.8: Available upward and downward flexibility for the winter sunny day scenario.

Winter—Cloudy

Figure 5.9 shows the power consumption and generation of the simulated devices. On a cloudy winter day, the effect of self-generation is limited, and upward flexibility is confined to the short periods when consumption devices operate or the PV is able to supply the house and charge the battery. The downward direction is not affected.

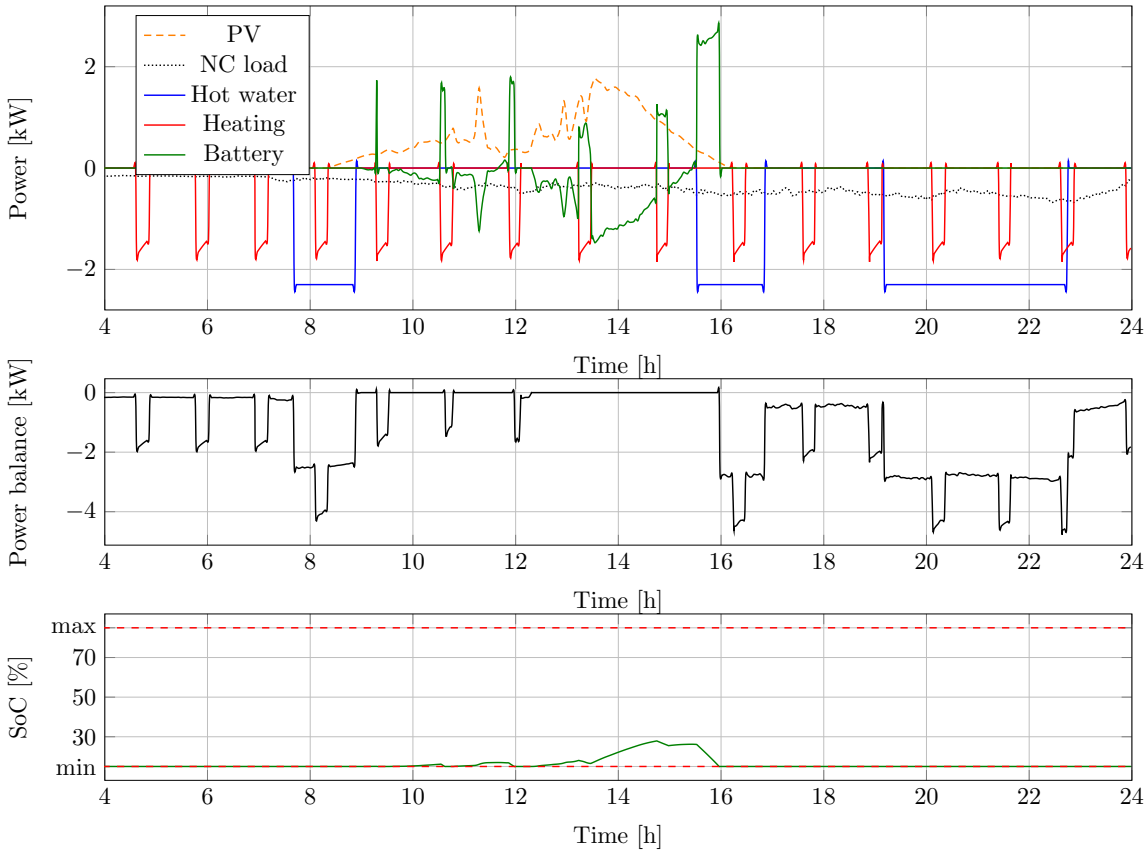


Figure 5.9: Power consumption/generation of devices for the winter cloudy day scenario.

The upward and downward flexibility capabilities of each device are added, resulting in the building's maximum flexible power (Figure 5.10).

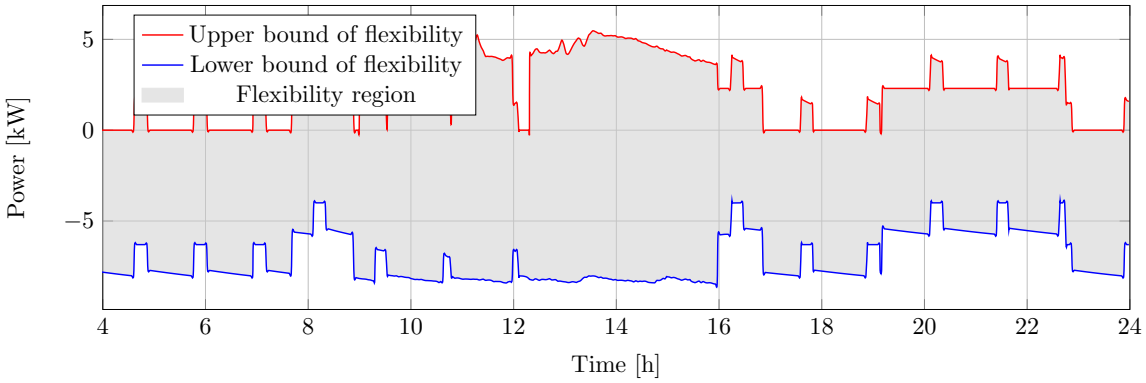


Figure 5.10: Available upward and downward flexibility for the winter cloudy day scenario.

Summer—Sunny

Figure 5.11 shows the power consumption and generation of the simulated devices. On a sunny summer day, the PV generation has a major effect: together with the storage, self-production is sufficient to supply the instantaneous power consumption of the house. After noon, the generation surplus is fed back to the grid.

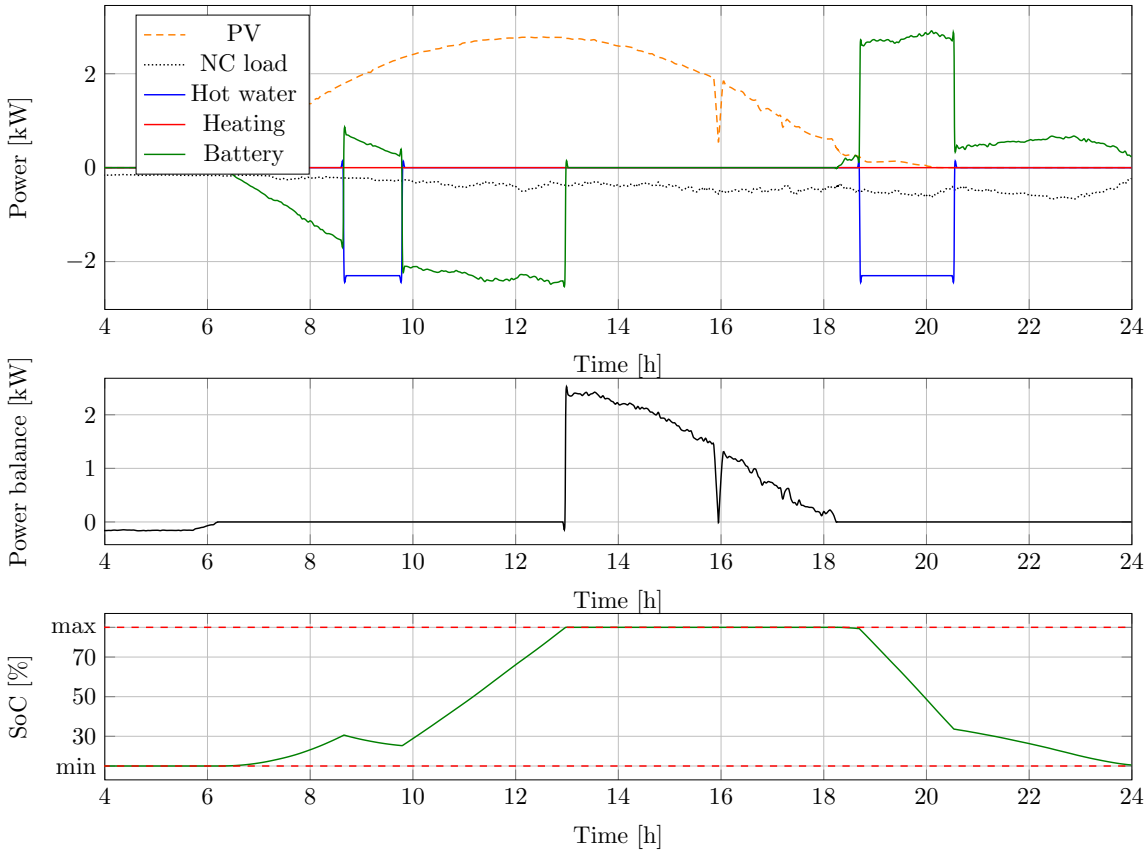


Figure 5.11: Power consumption/generation of devices for the summer sunny day scenario.

The upward and downward flexibility capabilities of each device are added, resulting in the building's maximum flexible power (Figure 5.12).

Summer—Cloudy

Figure 5.13 shows the power consumption and generation of the simulated devices. A cloudy summer day results in a more variable upward flexibility in the positive direction. The energy of PV production is not enough to supply the house all day, but 90% of the time, so there is no grid usage.

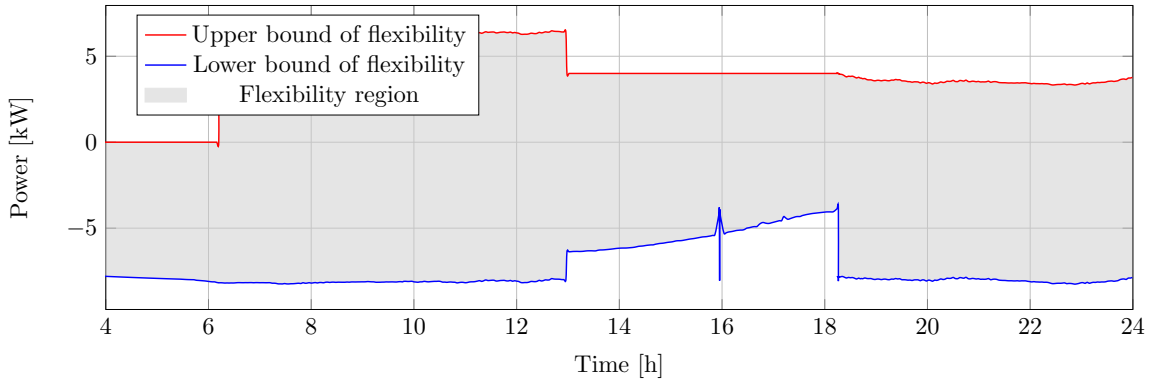


Figure 5.12: Available upward and downward flexibility for the summer sunny day scenario.

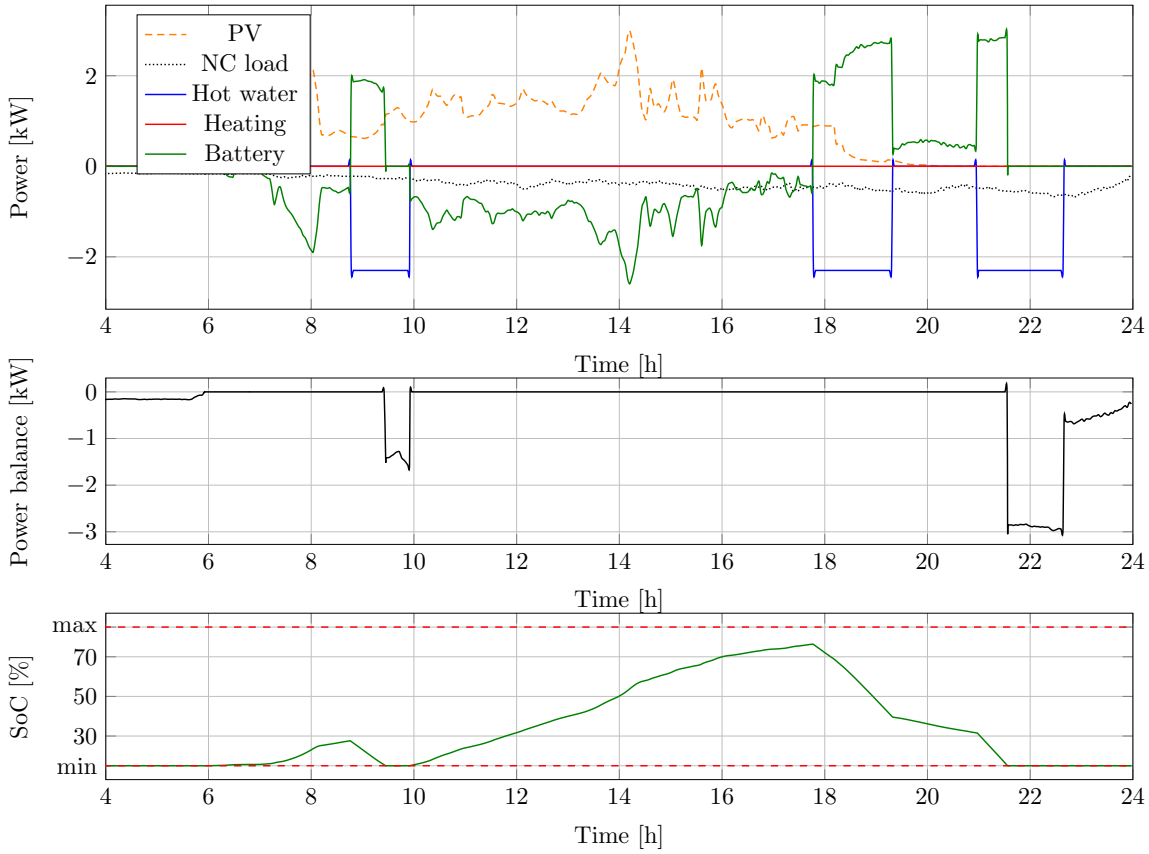


Figure 5.13: Power consumption/generation of devices for the summer cloudy day scenario.

The upward and downward flexibility capabilities of each device are added, resulting in the building's maximum flexible power (Figure 5.14).

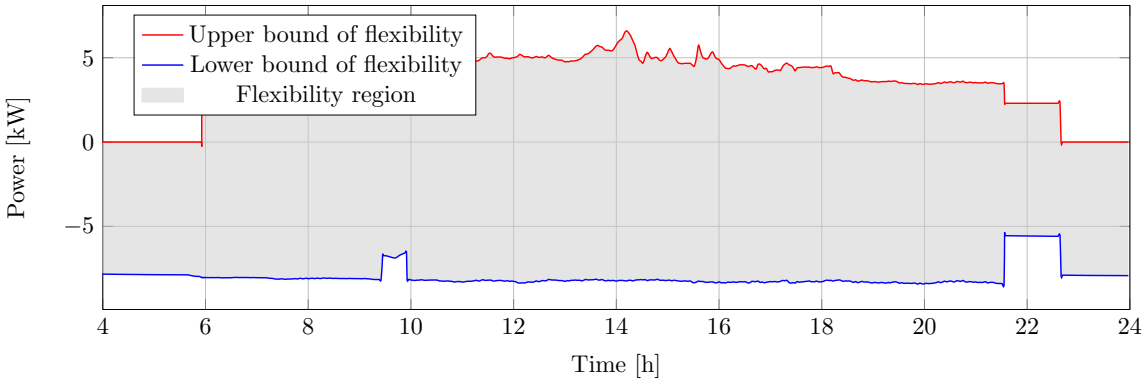


Figure 5.14: Available upward and downward flexibility for the summer cloudy day scenario.

5.3.3. Prediction

The quantification of the flexibility provides instant volumes of the up and down regulation capabilities. Instantaneous values are not valuable for flexibility buyers such as system operators or aggregators, so forecasts need to be calculated. It was not a primary objective to study the prediction methods and evaluate the results; however, a linear (Ridge Regression) and a non-linear forecast model (Extra Trees Regression) was built to show the short-term prediction opportunities.

Extra Trees Regression (ETR) method was introduced in Chapter 2. This section will commence with a brief description of ridge regression, followed by an explanation of the metrics employed to evaluate the models. Finally, the results of the predictions will be presented.

A linear model was fit to minimize the residual sum of squares between the observed features and the values predicted by the linear approximation. Ridge regression was used, which kept all predictors in the model, but performed an L2-norm regularization [80], reducing the impact of correlated predictors. Ridge regression [59] is a regularized version of linear regression where a regularization term is added to the cost function. This forces the learning algorithm to not only fit the data, but also keep the model weights smaller in magnitude [76]. The objective function of ridge regression is defined as follows [98]:

$$\min_{\mathbf{w}} \frac{1}{2} \sum_{i=1}^{n} ||\mathbf{w}^T \mathbf{x}_i - y_i||_2^2 + \lambda ||\mathbf{w}||_2^2 \tag{5.22}$$

In the formula, upper case denotes a matrix, lower case denotes a vector. $\mathbf{x}_i$ is the feature

vector of the i-th sample and y_i is the independent variable's true value. λ is a regularization parameter. Weight vector $\mathbf{w}$ is calculated by taking the derivative of Equation (5.22) and setting it to zero.

$$\mathbf{w} = (\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I})^{-1}\mathbf{X}\mathbf{y}^T \tag{5.23}$$

A 15-minute interval is the typical market time unit for settlement in the energy sector in Europe. The calculation of a 15-minute forecast of upward flexibility assumes that it has a linear relationship between upward flexibility (y_i) and the predictor variables ($\mathbf{x}_i$). PV generation, heating system/water heater consumption, power, and the state-of-charge of the battery were selected as the explanatory variables. There is no prediction of the features, so 15-minute lags are applied to construct the predictor set (Equation (5.25)). A 15-minute average of the target variable is calculated and lagged by 15 minutes (Equation (5.24)).

$$Flex^{up}_{rolling}(i) = \frac{\sum_{t=i-18001200}^{i-900300} F^{up}_{total}(t)}{900} \tag{5.24}$$

$$\mathbf{x}_i = \begin{bmatrix} P_{water}(i-900) \\ P_{heating}(i-900) \\ P_{battery}(i-900) \\ P_{pv}(i-900) \\ SoC(i-900) \\ Flex^{up}_{rolling}(i) \end{bmatrix} \tag{5.25}$$

Ridge regression puts constraints on the size of the coefficients associated with each variable. These values depend on the magnitude of each variable. If a variable is measured at a higher scale than the other variables and not centered around zero, they do not give an equal contribution to the analysis. Both the training and test set must be standardized based on the mean and standard deviation learned from the training set by removing the mean and scaling data to unit variance as follows:

$$\mathbf{z} = \frac{\mathbf{x} - \mu}{\sigma}, \tag{5.26}$$

where μ is the mean, σ is the standard deviation of $\mathbf{x}$, and $\mathbf{z}$ is the scaled predictors.

The coefficient of variation of the root-mean-squared error (CVRMSE) and the coefficient

of determination (R^2 score) were used as a set of criteria to evaluate the prediction. The CVRMSE (Equation (5.27)) measures the variability of errors between true and predicted values. It gives an indication of the model's ability to predict the overall load shape that is reflected in the data [75].

$$CVRMSE(y, \hat{y}) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} y_i} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} 100 \tag{5.27}$$

R^2 (5.28) represents the proportion of variance that has been explained by the independent variables in the model. It provides an indication of the goodness of fit and therefore a measure of how well unseen samples are likely to be predicted by the model, through the proportion of explained variance [63].

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{5.28}$$

where

- n: number of samples,
- y_i: true value,
- $\hat{y}_i$: predicted value of the i-th sample,
- $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Both predictive models were trained on the same data, i.e. one month of one-minute resolution simulated results. The simulation of flexibility represents the true value of the target variable. The 1-minute simulation of the target and the explanatory variables were resampled to 15-minute and a single period forecast was made for the flexibility in up and down direction.

The prediction results and the target value for one day are shown in Figure 5.15 and Figure 5.16, respectively. Ridge regression follows the shape of the target variable in both upward and downward directions, but in a naive manner, it reflects the previous period's value in the predictions. This occurs because despite the L2 regularization, the lagged variable of our target variable received the highest coefficient. The linear model perceives the autoregressive effect as much stronger than the explanatory power of the explanatory variables.

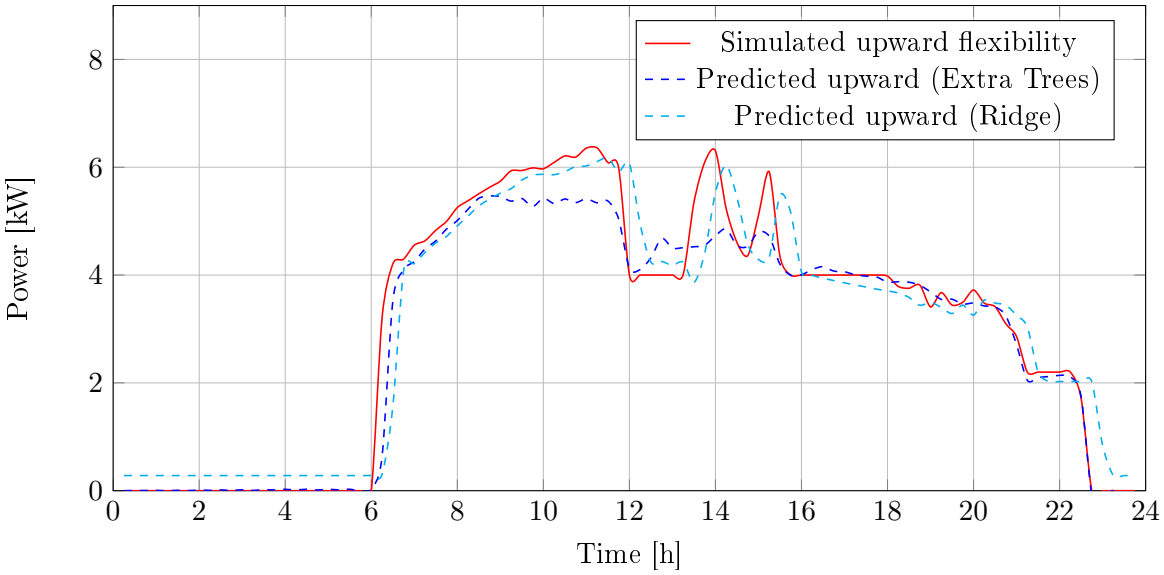


Figure 5.15: Sample from the upward flexibility prediction for a day.

In the case of ETR, it can be observed that overfitting of the target variable does not occur. However, peaks in flexibility capability are typically underestimated in absolute value in both directions.

Figure 5.16 illustrates a 24-hour sample of the prediction. It can be reasonably concluded that the results are comparable to those observed in the up direction.

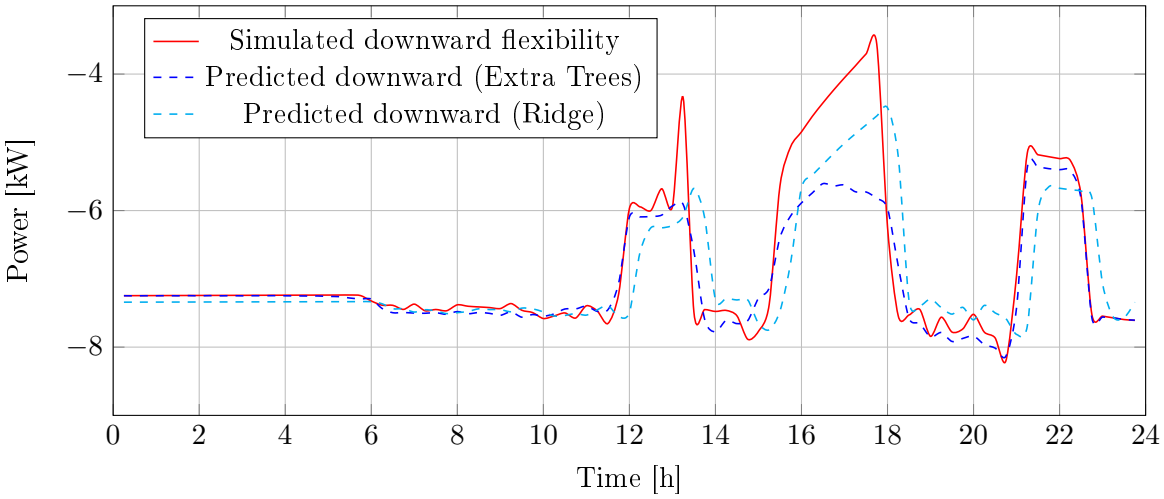


Figure 5.16: Sample from the downward flexibility prediction for a day.

Table 5.5 contains the metrics for both predictions. Despite the incorrect estimates of the

outliers, the ETR produced more accurate forecasts in both directions and by both metrics. There is potential to improve the accuracy of the predictions by increasing the size of the training set and by testing additional prediction algorithms.

Table 5.5: Prediction results

Metric	Direction	ETR	Ridge
CVRMSE	Up	16%	22%
	Down	8%	10%
R_2	Up	0.95	0.91
	Down	0.76	0.6

5.4. Conclusion

Applying measured weather and load data, the experiments were performed along four different, but typical scenarios. The results showed that the highest range of flexibility was available in the summer time, when there was no heating and the solar generation was maximal. It was also clear that the battery usage was higher in the summer, when the solar generation was not being consumed instantly. It is important to note that a properly sized air conditioner unit would balance the load between the summer and winter periods.

Demand-side flexibility can be a valuable source for system operators; however, residential prosumers do not follow a well-defined schedule, so a firm volume of available flexibility cannot be planned, but predicted. A flexibility framework was proposed for residential prosumers in a distributed generation setup. The residential prosumer of the case study was parameterized so that it described a usual residential actor of the system. To quantify building demand flexibility, the thermoelectric dynamic response of the building energy system was modeled and implemented in MATLAB Simulink. Power consumption and generation were modeled. Simulations were performed based on real world data, and the flexibility potential was calculated for both up and down flexibility. The simulation of four scenarios was executed, which covered one day in different seasons. Power consumption and generation were calculated, as well as upward and downward instantaneous flexibility.

A ridge regression and an extra trees regression-based prediction method were designed for the purpose of calculating the short-term forecast. The simulation and prediction experiments showed that the proposed methods could serve as the basis of a state estimator or prediction unit. The accuracy of the forecast was moderate, but by assessing different prediction methods on the prosumer flexibility model, I could choose the right tool to improve accuracy and confidence of the prediction.

The aggregation of prosumers' flexibility is necessary to reach the volume that a system operator can utilize. The framework presented forms a basis to analyze the flexibility prediction

opportunities on aggregated prosumer portfolios.

CHAPTER 6

New scientific results

6.1. Theses

6.1.1. Thesis 1 - Short-term system imbalance forecast of the power grid using autoregressive distributed lag method (Chapter 3)

I have conceived and constructed a data acquisition, processing, and prediction tool tailored for the short-term forecast of system imbalance within the electrical grid. I have formulated the forecasting problem, which aims to generate forecasts from publicly available time series data for the power system at a quarter-hourly resolution for 2 hours after the system imbalance forecast moment. I formulated the problem in such a way that it could be handled by an Autoregressive Distributed Lag (ARDL)-based method, then parameterized it, partially modifying the basic ARDL for the prediction purpose. To evaluate the accuracy of the estimates, I developed a custom metric (Sign Accuracy Percentage - SAP) to measure the accuracy and usability of the results. I checked the results with 10 months of test data. The procedure I developed and proposed is capable of performing the 1-hour forecasting task with a conditional predictive accuracy above 88%. I compared the results with the methods (ARIMA, LSTM, ETR) commonly used in the literature on the subject and determined that the efficacy of the devised ARDL-based approach is found to be commensurate with that of a significantly more complex neural network and decision tree-based prediction. I have developed a data collection framework that collects model input data in real time and continuously performs predictions according to time dependencies, demonstrating that the efficiency of the method does not come from a violation of causality, the short-term prediction method indeed only uses existing and available time stamps for prediction.

Relevant publications: [S1, S2]

6.1.2. Thesis 2 - Time series simulation of the operation and trading activity of aggregated power generation and consumption portfolios in a multi-market environment (Chapter 4)

I have developed a method to manage the power generation of a diverse array of assets, including producers and consumers. The proposed method is based on a discrete-time predictive model that relies on power flow equations of the underlying dynamical system. These equations are formulated to incorporate the technical parameters and economic characteristics of the assets, while adhering to the regulatory constraints and operational rules of the

market. A key innovation of this work is the ability to optimize profit maximization within a heterogeneous portfolio of assets, effectively navigating and leveraging the opportunities in a multi-market energy environment. This approach also utilizes short-term system imbalance forecasting, over a one-year simulation period, comprehensive validation demonstrates that the imbalance predictions achieve satisfactory accuracy. The results of a scenario analysis also showed that the portfolio achieved a surplus with the participation in the balancing market using the model.

Relevant publications: [S3, S5]

6.1.3. Thesis 3 - Model-based quantification of the power flexibility of a household with residential solar panels and energy storage (Chapter 5)

I have developed a discrete-time nonlinear thermodynamic simulation model of a modern residential building based on first engineering principles. The model is equipped with solar panels and energy storage, where individual dynamic models for consumers, producers, and storage devices are developed to interact in response to external disturbances. A novel aspect of this work is the integration of the thermodynamic model with the flexibility calculation of the building's electricity consumption. I have proposed a method to calculate the flexibility of each device, enabling the aggregation of the building's overall up and down flexibility. Using real-time weather data and consumption patterns, I applied the thermodynamic model in four scenarios representing real conditions to quantify flexibility potential across different time periods. Additionally, I incorporated forecasting techniques, including ridge regression and extra trees regression, to predict short-term flexibility. The results demonstrate that the developed model effectively quantifies and forecasts the flexibility potential of the system.

Relevant publications: [S4]

6.2. Application of the results

In recent publications, the forecast of the volume of imbalances has usually been considered as a preliminary step to the price forecast. I believe that the joint prediction of imbalance volume and price provides market participants with the information they need to take the most advantageous market position by trading on intraday markets or by controlling their assets. The direction of the system imbalance is most important from a market position perspective. However, the forecasted volume helps to judge how accurate the sign prediction is. On the other hand, TSOs publish anonymous balancing bids in near real time, so that the resulting price of balancing energy can be approximated from the balancing energy demand and the merit order of the bids, based on knowing the settlement rules.

The single price model encourages the BRPs to take positions that contribute to the balance

of the system, but this is strongly influenced by their position, the market opportunities and the cost structure of their portfolio. In case of negative imbalance (shortage), the price of balancing energy ($Price_{BE}$) is derived from the average price of upward regulation, which is usually high above spot market prices (Intraday price: $Price_{ID}$). In this case, the direction of payment is determined by the imbalance position of the BRP, the supporting or causing state (non-aggravating or aggravating imbalance) determines whether at this high price the TSO pays the BRP or the BRP pays the TSO. Here the BRP should take a position above the schedule, even by purchasing energy on the Intraday market (if $Price_{ID} < Price_{BE}$), because the positive deviation from the schedule is rewarded with a high price by the single price settlement model. In real time, the potential BRP deficit should be reduced, since the cost of the upward regulation is assumed to be lower than the predicted upward balancing energy price (if $SRMC_{UP} < Price_{BE}$). Whether it is demand generated in the ID market or self-balancing of the portfolio, production is still being increased in processes outside the use of balancing reserves. As a result, the system imbalance decreases, along with the balancing energy demand and the corresponding price. In the case of a positive system imbalance (surplus), there is not so much difference between the two cases of payment direction, as the balancing energy price is derived from the negative direction regulation, which can be positive or negative, but in any case closer to zero than the upward direction.

For a portfolio of gas engines, it is worth controlling up for negative imbalance and down for positive imbalance if its short-run marginal cost is between the negative and positive balancing energy prices. A solar portfolio with a short run marginal cost around 0 should not be controlled downward in any case in case of positive system imbalance. When there is a surplus in the system, PVs may be controlled downward only if the downward balancing energy price is negative. Thus, for a given BRP position, the prediction of the imbalance direction is of great importance, because the positive and negative direction require completely different reactions.

A necessary element of integrating renewable technologies and increasing the flexibility of the electricity system is to create portfolios that together effectively exploit the benefits of different technologies and eliminate operational disadvantages. I have shown that, based on extensive data collection and knowledge of the current state of the market and the portfolio, it is possible to influence the market and control activities of a complex portfolio through automated decision logics in a way that serves both the economic interests of the BRP and the interests of the system.

The initial development of data collection and control functions for large central power plants was driven by the need to optimize the generation side with a limited number of plants. However, the decentralization trend has rendered this approach inadequate in the

context of the present distributed energy system. The involvement of end-consumers is essential to resolve local grid issues. However, the emergence of a small players represents a significant challenge for market players. My work demonstrates the intervals within which the time-series energy balance of a modern residential building can be influenced and the constraints under which it can be predicted.

6.3. Future research directions

Future work regarding system imbalance forecast should focus on further refining the ARDL model by incorporating additional predictors and exploring advanced ensemble techniques to improve both accuracy and reliability under varying system conditions. Additionally, extending the evaluation to include more diverse datasets and operational scenarios will help generalize the model's applicability and robustness across different grid environments. This research provides a solid foundation for advancing forecasting methodologies in power system operations, contributing significantly to the field's knowledge and practice.

In the area of complex portfolios, the addition of the tools presented in my work with residential households and the management of the uncertainties arising from their control can significantly increase the economic performance of the portfolio and leads to a difficult optimization problem. It is necessary to generalize the method to a higher number of households, for example a local transformer area, in order to give an estimate of the flexibility of a group of prosumers. Another step in the development of the proposed method is to use novel prediction methods from the field of data science to enhance the short-term prediction performance for flexibility. In the market dimension, the inclusion of instantaneous price information can further increase the complexity of the model.

REFERENCES

[1] ACER. Decision No 18/2020 of the European Union Agency for the Cooperation of Energy Regulators (ACER) of 15 July 2020 on the harmonisation of the main features of imbalance settlement, 2020.

[2] O. Adabor, E. K. Ayesu, and E. Nana-Amankwaah. The causal link between electricity transmission, distributional losses and economic growth in ghana. *OPEC Energy Review*, 47(2):101–117, 2022.

[3] Md Irfan Ahmed and Ramesh Kumar. Nodal Electricity Price Forecasting using Exponential Smoothing and Holt's Exponential Smoothing. *Distributed Generation & Alternative Energy Journal*, 38(05):1505–1530, July 2023.

[4] Jens Baetens, Joannes Laveyne, Greet Van Eetvelde, and Lieven Vandevelde. Imbalance Pricing Methodology in Belgium: Implications for Industrial Consumers. In *2020 17th International Conference on the European Energy Market (EEM)*, pages 1–6, 2020.

[5] Mohamed Bahloul, Ankur Majumdar, Mohamed Daoud, and Shafi Khadem. Energy storage system: A potential "Flexibility Resources" to accelerate the decarbonisation of smart grid network. *Proc. Medit. Conf. Power Gener. Transmiss. Distrib. Energy Convers*, pages 1–6, 2021. Publisher: IET.

[6] István Balázs, Attila Fodor, and Attila Magyar. Short-term system imbalance forecast using linear and non-linear methods. *Energy Systems*, pages 1–22, March 2024.

[7] Abdul Basit, Anca D. Hansen, Poul E. Soerensen, and Georgios Giannopoulos. Real-time impact of power balancing on power system operation with large scale integration of wind power. *Journal of Modern Power Systems and Clean Energy*, 5(2):202–210, 2017.

[8] John Black, Nigar Hashimzade, and Gareth Myles. *A dictionary of economics*. Oxford University Press, USA, 2012.

[9] Piotr F. Borowski. Zonal and nodal models of energy market in european union. *Energies*, 13(16), 2020. tex.article-number: 4182.

[10] Piotr F. Borowski. Digitization, digital twins, blockchain, and industry 4.0 as elements of management process in enterprises in the energy sector. *Energies*, 14(7), 2021. tex.article-number: 1885.

[11] Jérémie Bottieau, Louis Hubert, Zacharie De Grève, François Vallée, and Jean-François Toubeau. Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement. *IEEE Transactions on Power Systems*,

35(2):1218–1230, 2019. Publisher: IEEE.

[12] John Bower and Derek Bunn. Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the England and Wales electricity market. *Journal of economic dynamics and control*, 25(3-4):561–592, 2001. Publisher: Elsevier.

[13] Jethro Browell and Ciaran Gilbert. Predicting Electricity Imbalance Prices and Volumes: Capabilities and Opportunities. *Energies*, 15(10), 2022.

[14] Roland Bálint, Attila Fodor, and Attila Magyar. Model-based Power Generation Estimation of Solar Panels using Weather Forecast for Microgrid Application. *Acta Polytechnica Hungarica*, 16(7), 2019.

[15] European Commission. Commission Regulation (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation. *Off. J. Eur. Union*, 2017.

[16] European Commission. Commission Regulation (EU) 2017/2195 of 23 November 2017 establishing a guideline on electricity balancing. *Off. J. Eur. Union*, 312:6–53, 2017.

[17] Carolina Contreras. *System imbalance forecasting and short-term bidding strategy to minimize imbalance costs of transacting in the Spanish electricity market*. PhD thesis, Universidad Pontificia Comillas, Escuela Técnica Superior de Ingeniería (ICAI), 2016.

[18] Cynthia A Cruickshank and Stephen J Harrison. Heat loss characteristics for a typical solar domestic hot water storage. *Energy and buildings*, 42(10):1703–1710, 2010. Publisher: Elsevier.

[19] Roel De Coninck and Lieve Helsen. Bottom-up quantification of the flexibility potential of buildings. In *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*, pages 3250–3258, 2013.

[20] Vanderson Delapedra-Silva, Paula Ferreira, Jorge Cunha, and Herbert Kimura. Methods for Financial Assessment of Renewable Energy Projects: A Review. *Processes*, 10(2), 2022.

[21] Marialaura Di Somma, Giorgio Graditi, and Pierluigi Siano. Optimal Bidding Strategy for a DER Aggregator in the Day-Ahead Market in the Presence of Demand Flexibility. *IEEE Transactions on Industrial Electronics*, 66(2):1509–1519, 2019.

[22] David A. Dickey and Wayne A. Fuller. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74(366a):427–431, 1979. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.1979.10482531.

[23] Jonathan Dumas, Ioannis Boukas, Miguel Manuel de Villena, Sébastien Mathieu, and Bertrand Cornélusse. Probabilistic Forecasting of Imbalance Prices in the Belgian Context. In *2019 16th International Conference on the European Energy Market (EEM)*, pages 1–7, 2019.

[24] Elias Dörre, Sebastian Pfaffel, Alexander Dreher, Pedro Girón, Svenja Heising, and Kay Wiedemann. Flexibility Reserve of Self-Consumption Optimized Energy Systems in the Household Sector. *Energies*, 14(11), 2021.

[25] R. F. Engle and C. W. J. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251, 1987.

[26] European Parliament and of the Council. Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU, 2019. Published: Official Journal of the European Union.

[27] Hassan Farhangi. The path of the smart grid. *IEEE power and energy magazine*, 8(1):18–28, 2009. Publisher: IEEE.

[28] Maria P Garcia and Daniel S Kirschen. Forecasting system imbalance volumes in competitive electricity markets. *IEEE Transactions on Power Systems*, 21(1):240–248, 2006.

[29] Yimy E García Vera, Rodolfo Dufo-López, and José L Bernal-Agustín. Energy management in microgrids with renewable energy sources: A literature review. *Applied Sciences*, 9(18):3854, 2019. Publisher: Multidisciplinary Digital Publishing Institute.

[30] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006.

[31] Giorgio Castagneto Gissey, Dina Subkhankulova, Paul E Dodds, and Mark Barrett. Value of energy storage aggregation to the electricity system. *Energy Policy*, 128:685–696, 2019. Publisher: Elsevier.

[32] I. L. R. Gomes, H. M. I. Pousinho, R. Melício, and V. M. F. Mendes. Stochastic coordination of joint wind and photovoltaic systems with energy storage in day-ahead market. *Energy*, 124:310–320, 2017.

[33] Shadi Goodarzi, H. Niles Perera, and Derek Bunn. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. *Energy Policy*, 134:110827, 2019.

[34] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438, 1969. Publisher: [Wiley, Econometric

Society].

[35] Alex Graves. Long Short-Term Memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*, pages 37–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[36] Vehbi C Gungor, Bin Lu, and Gerhard P Hancke. Opportunities and challenges of wireless sensor networks in smart grid. *IEEE transactions on industrial electronics*, 57(10):3557–3564, 2010. Publisher: IEEE.

[37] Katalin Hangos and Ian Cameron. Process modelling and model analysis. 2001. Publisher: Academic press.

[38] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. Publisher: MIT Press.

[39] Nicole Hopper, Charles Goldman, and Bernie Neenan. Demand response from day-ahead hourly pricing for large customers. *The Electricity Journal*, 19(3):52–63, 2006. Publisher: Elsevier.

[40] H Husin, M Zaki, and others. A critical review of the integration of renewable energy sources with various technologies. *Protection and control of modern power systems*, 6(1):1–18, 2021. Publisher: PSPC.

[41] NREL IEA. Status of power system transformation 2018. *Int. Energy Agency, Paris, France, Tech. Rep*, 2018.

[42] International Organization for Standardization. Building components and building elements - Thermal resistance and thermal transmittance - Calculation method, 2017.

[43] Dongsik Jang, Jiyong Eom, Moon Gyu Kim, and Jae Jeung Rho. Demand responses of Korean commercial and industrial businesses to critical peak pricing of electricity. *Journal of Cleaner Production*, 90:275–290, 2015. Publisher: Elsevier.

[44] Stefan S.O. Kermer. Imbalance Settlement Design with Demand Side Management and Combined Heat and Power. In *2019 IEEE Texas Power and Energy Conference (TPEC)*, pages 1–6, 2019.

[45] M Kersken, I Heusler, and P Strachan. Full scale empirical validation for building energy simulation programs. In *Liège: 9th international conference on system simulation in buildings*, 2014.

[46] Ismail Kirbas and Alper Kerem. Short-term wind speed prediction based on artificial neural network models. *Measurement and Control*, 49(6):183–190, 2016. Publisher: SAGE Publications Sage UK: London, England.

[47] Gro Klæboe, Anders Lund Eriksrud, and Stein-Erik Fleten. Benchmarking time series based forecasting models for electricity balancing market prices. *Energy Systems*, 6:43–61, 2015. Publisher: Springer.

[48] Christopher Koch. Intraday imbalance optimization: incentives and impact of strategic intraday bidding behavior. *Energy Systems*, 13(2):409–435, 2022. Publisher: Springer.

[49] Stepán Kratochvil. *System imbalance forecast*. PhD thesis, Czech Technical University, 2016.

[50] Benjamin Kroposki, Brian Johnson, Yingchen Zhang, Vahan Gevorgian, Paul Denholm, Bri-Mathias Hodge, and Bryan Hannegan. Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy. *IEEE Power and energy magazine*, 15(2):61–73, 2017. Publisher: IEEE.

[51] Daniel Kucevic, Benedikt Tepe, Stefan Englberger, Anupam Parlikar, Markus Mühlbauer, Oliver Bohlen, Andreas Jossen, and Holger Hesse. 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, 2020.

[52] Jared Langevin, Chioke B Harris, Aven Satre-Meloy, Handi Chandra-Putra, Andrew Speake, Elaina Present, Rajendra Adhikari, Eric JH Wilson, and Andrew J Satchwell. US building energy efficiency and flexibility as an electric grid resource. *Joule*, 2021. Publisher: Elsevier.

[53] Jinho Lee and Dongjun Won. Optimal Operation Strategy of Virtual Power Plant Considering Real-Time Dispatch Uncertainty of Distributed Energy Resource Aggregation. *IEEE Access*, 9:56965–56983, 2021.

[54] Fernando Lezama, Joao Soares, Bruno Canizes, and Zita Vale. Flexibility management model of home appliances to support DSO requests in smart grids. *Sustainable Cities and Society*, 55:102048, 2020. Publisher: Elsevier.

[55] Rui Amaral Lopes, Adriana Chambel, João Neves, Daniel Aelenei, and João Martins. A literature review of methodologies used to assess the energy flexibility of buildings. *Energy Procedia*, 91:1053–1058, 2016.

[56] Alexandre Lucas, Luca Jansen, Nikoleta Andreadou, Evangelos Kotsakis, and Marcelo Masera. Load Flexibility Forecast for DR Using Non-Intrusive Load Monitoring in the Residential Sector. *Energies*, 12(14), 2019.

[57] Stephen Makonin, Bradley Ellert, Ivan V. Bajić, and Fred Popowich. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific Data*, 3(1):160037, June 2016.

[58] P Mandatova and O Mikhailova. Flexibility and Aggregation: Requirements for their interaction in the market. *Eurelectric: Brussels, Belgium*, 2014.

[59] Gary C McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100, 2009. Publisher: Wiley Online Library.

[60] Madeleine McPherson and Samiha Tahseen. Deploying storage assets to facilitate variable renewable energy integration: The impacts of grid flexibility, renewable penetration, and market structure. *Energy*, 145:856–870, 2018.

[61] MEKH and MAVIR. Data of the Hungarian Electricity System, 2021.

[62] Julia Merino, Inés Gómez, Elena Turienzo, and Carlos Madina. Ancillary service provision by RES and DSM connected at distribution level in the future power system. Technical report, SmartNet project D, 2016. Publication Title: SmartNet project D Volume: 1.

[63] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[64] Ricardo Moura and Miguel Centeno Brito. Prosumer aggregation policies, country experience and business models. *Energy Policy*, 132:820–830, 2019.

[65] Neeraj, Jimson Mathew, and Ranjan Kumar Behera. Power load forecasting based on long short term memory-singular spectrum analysis. *Energy Systems*, 13(3):789–811, August 2022.

[66] Thomas Nuytten, Bert Claessens, Kristof Paredis, Johan Van Bael, and Daan Six. Flexibility of a combined heat and power system with thermal energy storage for district heating. *Applied energy*, 104:583–591, 2013.

[67] Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics*, 7(3):381–388, 2011. Publisher: IEEE.

[68] Yael Parag and Benjamin K Sovacool. Electricity market design for the prosumer era. *Nature energy*, 1(4):1–6, 2016. Publisher: Nature Publishing Group.

[69] Wei Pei, Yan Du, Wei Deng, Kun Sheng, Hao Xiao, and Hui Qu. Optimal bidding strategy and intramarket mechanism of microgrid aggregator in real-time balancing market. *IEEE Transactions on Industrial Informatics*, 12(2):587–596, 2016. Publisher: IEEE.

[70] M Hashem Pesaran, Yongcheol Shin, and Richard J Smith. Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics*, 16(3):289–326,

2001. Publisher: Wiley Online Library.

[71] Ksenia Poplavskaya, Jesus Lago, and Laurens de Vries. Effect of market design on strategic bidding behavior: Model-based analysis of European electricity balancing markets. *Applied Energy*, 270:115130, 2020.

[72] Ksenia Poplavskaya and Laurens de Vries. Distributed energy resources and the organized balancing market: A symbiosis yet? Case of three European balancing markets. *Energy Policy*, 126:264–276, 2019.

[73] Zheng Qian, Yan Pei, Hamidreza Zareipour, and Niya Chen. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Applied Energy*, 235:939–953, February 2019. Publisher: Elsevier BV.

[74] Glenn Reynders. Quantifying the impact of building design on the potential of structural storage for active demand response in residential buildings. 2015.

[75] Germán Ramos Ruiz and Carlos Fernández Bandera. Validation of calibrated energy models: Common errors. *Energies*, 10(10):1587, 2017. Publisher: Multidisciplinary Digital Publishing Institute.

[76] Mostafa A Rushdi, Ahmad A Rushdi, Tarek N Dief, Amr M Halawa, Shigeo Yoshida, and Roland Schmehl. Power prediction of airborne wind energy systems using multivariate machine learning. *Energies*, 13(9):2367, 2020. Publisher: Multidisciplinary Digital Publishing Institute.

[77] Nasrin Sadeghianpourhamami, Thomas Demeester, DF Benoit, Matthias Strobbe, and Chris Develder. Modeling and analysis of residential flexibility: Timing of white good usage. *Applied energy*, 179:790–805, 2016. Publisher: Elsevier.

[78] S. A. Saleh, Petrus Pijnenburg, and Eduardo Castillo-Guerra. Load aggregation from generation-follows-load to load-follows-generation: Residential loads. *IEEE Transactions on Industry Applications*, 53(2):833–842, 2017.

[79] Tárik S Salem, Karan Kathuria, Heri Ramampiaro, and Helge Langseth. Forecasting intra-hour imbalances in electric power systems. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 9595–9600, 2019.

[80] Mark Schmidt. Least squares optimization with L1-norm regularization. *CS542B Project Report*, 504(2005):195–221, 2005. Publisher: University of British Columbia.

[81] Ljubisa Sehovac and Katarina Grolinger. Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention. *IEEE Access*, 8:36411–36426, 2020.

[82] Mohammad Shakeri, Jagadeesh Pasupuleti, Nowshad Amin, Md. Rokonuzzaman, Foo Wah Low, Chong Tak Yaw, Nilofar Asim, Nurul Asma Samsudin, Sieh Kiong Tiong, Chong Kok Hen, and Chin Wei Lai. An Overview of the Building Energy Management System Considering the Demand Response Programs, Smart Strategies and Smart Grid. *Energies*, 13(13), 2020.

[83] Amit Shewale, Anil Mokhade, Nitesh Funde, and Neeraj Dhanraj Bokde. An Overview of Demand Response in Smart Grid and Optimization Techniques for Efficient Residential Appliance Scheduling Problem. *Energies*, 13(16), 2020.

[84] Selahattin Murat Sirin and Berna N. Yilmaz. The impact of variable renewable energy technologies on electricity markets: An analysis of the Turkish balancing market. *Energy Policy*, 151:112093, 2021.

[85] Takuma Takeshita, Hirohisa Aki, Kotaro Kawajiri, and Masayoshi Ishida. Assessment of utilization of combined heat and power systems to provide grid flexibility alongside variable renewable energy systems. *Energy*, 214:118951, 2021.

[86] Chang Wei Tan, Christoph Bergmeir, François Petitjean, and Geoffrey I Webb. Time series extrinsic regression: Predicting numeric values from time series data. *Data Mining and Knowledge Discovery*, 35:1032–1060, 2021. Publisher: Springer.

[87] Kais Tissaoui, Taha Zaghdoudi, Abdelaziz Hakimi, and Mariem Nsaibi. Do gas price and uncertainty indices forecast crude oil prices? Fresh evidence through XGBoost modeling. *Computational Economics*, 62(2):663–687, 2023. Publisher: Springer.

[88] L. Tong, L. Luo, and F. Li. Research on autoregressive distribution lag modeling method for power load forecasting. *Proceedings of the 2019 2nd International Conference on Mathematics, Modeling and Simulation Technologies and Applications (MMS*, 2019.

[89] Jean-François Toubeau, Jérémie Bottieau, Yi Wang, and François Vallée. Interpretable Probabilistic Forecasting of Imbalances in Renewable-Dominated Electricity Systems. *IEEE Transactions on Sustainable Energy*, 13(2):1267–1277, 2022.

[90] D Validzic. Clean energy for all europeans-european union's new legislative framework. 2017.

[91] Fei Wang, Xinxin Ge, Kangping Li, and Zengqiang Mi. Day-Ahead Market Optimal Bidding Strategy and Quantitative Compensation Mechanism Design for Load Aggregator Engaging Demand Response. *IEEE Transactions on Industry Applications*, 55(6):5564–5573, 2019.

[92] Sanford Weisberg. *Applied linear regression*, volume 528. John Wiley & Sons, 2005.

[93] Tobias Weißbach, Simon Remppis, and Hendrik Lens. Impact of current market developments in Europe on deterministic grid frequency deviations and frequency restauration reserve demand. In *2018 15th International Conference on the European Energy Market (EEM)*, pages 1–6. IEEE, 2018.

[94] Xiandong Xu, Quan Lyu, Meysam Qadrdan, and Jianzhong Wu. Quantification of flexibility of a district heating system for the power grid. *IEEE Transactions on Sustainable Energy*, 11(4):2617–2630, 2020. Publisher: IEEE.

[95] Yan Xu, Zhao Luo, Zhendong Zhu, Zhiyuan Zhang, Jinghui Qin, Hao Wang, Zeyong Gao, and Zhichao Yang. A Three-Stage Coordinated Optimization Scheduling Strategy for a CCHP Microgrid Energy Management System. *Processes*, 8(2), 2020.

[96] Lei Xue, Yunlong Teng, Zhenyuan Zhang, Jian Li, Kunbing Wang, and Qi Huang. Blockchain technology for electricity market in microgrid. In *2017 2nd international conference on power and renewable energy (ICPRE)*, pages 704–708, 2017.

[97] Chen Yang, Xiao Du, Dan Xu, Junjie Tang, Xingyu Lin, Kaigui Xie, and Wenyuan Li. Optimal bidding strategy of renewable-based virtual power plant in the day-ahead market. *International Journal of Electrical Power & Energy Systems*, 144:108557, 2023.

[98] Yugen Yi, Yuqi Chen, Jiangyan Dai, Xiaolin Gui, Chunlei Chen, Gang Lei, and Wenle Wang. Semi-Supervised Ridge Regression with Adaptive Graph-Based Label Propagation. *Applied Sciences*, 8(12), 2018.

[99] Agustin Zaballos, Alex Vallejo, and Josep M Selga. Heterogeneous communication architecture for the smart grid. *IEEE network*, 25(5):30–37, 2011. Publisher: IEEE.

[100] Krzysztof Zarzycki and Maciej Ławryńczuk. LSTM and GRU Neural Networks as Models of Dynamical Processes Used in Predictive Control: A Comparison of Models Developed for Two Chemical Reactors. *Sensors*, 21(16), 2021.

[101] Krzysztof Zarzycki and Maciej Ławryńczuk. Long Short-Term Memory Neural Networks for Modeling Dynamical Processes and Predictive Control: A Hybrid Physics-Informed Approach. *Sensors*, 23(21), 2023.

[102] Lingxi Zhang, Nicholas Good, and Pierluigi Mancarella. Building-to-grid flexibility: Modelling and assessment metrics for residential demand response from heat pump aggregations. *Applied Energy*, 233-234:709–723, 2019.

[103] Lingling Zhou, Ping Zhao, Dongdong Wu, Cheng Cheng, and Hao Huang. Time series model for forecasting the number of new admission inpatients. *BMC medical informatics and decision making*, 18:1–11, 2018. Publisher: Springer.

OWN PUBLICATIONS

[S1] István Balázs, Attila Fodor, and Attila Magyar. Short-term system imbalance forecast using linear and non-linear methods. *Energy Systems*, pages 1-22, 2024. Publisher: Springer.

[S2] István Balázs, Attila Fodor, and Attila Magyar. Short-Term System Imbalance Forecast Using Autoregressive Distributed Lag Method. *2023 IEEE 6th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE)*, pages 83-88, 2023. Publisher: IEEE.

[S3] István Balázs, Gábor Oláh, Imre Pácsonyi, Attila Fodor, and Attila Magyar. Simulating Operation and Trading of Aggregated Energy Portfolios in Multi-Market Environments. *PEMC 2024*, pages X-Y, 2024. Publisher: IEEE.

[S4] István Balázs, Attila Fodor, and Attila Magyar. Quantification of the flexibility of residential prosumers. *Energies*, 14(16):4860, 2021. Publisher: MDPI.

[S5] István Balázs, Attila Fodor, and Attila Magyar. Aggregation of Heterogeneous Flexibility Resources Providing Services for System Operators and the Market Participants. *Hungarian Journal of Industry and Chemistry*, 47(1):65–70, 2019.

APPENDIX

A.1. Abbreviations

Table A.1 contains the abbreviations used in the document.

Table A.1: Abbreviations.

Abbreviation	Meaning
ACE	Area Control Error
ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller test
aFRR	Automatic Frequency Restoration Reserve
AGC	Automatic Generation Control
AI	Artificial Intelligence
AMPds	Almanac of Minutely Power Dataset
ARDL	Autoregressive Distributed Lag
ARIMA	Autoregressive Integrated Moving Average
ARIMAX	(AutoRegressive Integrated Moving Average with eXogenous inputs
BLSTM	Bidirectional Long Short-Term Memory
BRP	Balance Responsible Party
CAPEX	Capital Expenditure
CEP	Clean Energy for All Europeans Package
COP	Coefficient of performance
CVRMSE	Coefficient of variation of the root-mean-squared error
DAM	Day-Ahead Market
DER	Distributed Energy Resources
DR	Demand Response
DG	Distributed Generation
DSM	Demand-side Management
DSO	Distribution System Pperator
DHW	Domestic Hot Water
EB Regulation	Guideline on Electricity Balancing
EMS	Energy Management System
EPS	Expanded Polystyrene
ETR	Extra Trees Regression
FCR	Frequency Containment Reserve

Abbreviation	Meaning
FRCE	Frequency Restoration Control Error
FRR	Frequency Restoration Reserve
GAMLSS	Generalized Additive Models for Location, Scale and Shape
HEM	Home Energy Management
HEMS	Home Energy Management System
HP	Heat Pump
HUPX	Hungarian Power Exchange
HVAC	Heating, ventilating, and air conditioning
IDM / ID market	Intraday Market
IID	identically distributed
IT system	Information Technology System
LSTM	Long Short-Term Memory
MAE	Mean Average Error
mFRR	Manual Frequency Restoration Reserve
OLS	Ordinary Least Squares
OPEX	Operational Expenditure
PACF	Partial Autocorrelation Function
PV	Photovoltaic
PX	Power Exchange
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SAP	Sign Accuracy Percentage
SoC	State-of-charge
SCADA	Supervisory Control and Data Acquisition
TSO	Transmission system operator
VPP	Virtual Power Plant

A.2. Residential house parameters

The simulation parameters in Chapter 5 are taken from the energy audit of a residential building located in Hungary. Together with the house geometry, they are referred to as the "house parameters", which are detailed in A.2.

Table A.2: Simulation parameters.

Parameter	Value	Unit of Measure	Source
l_H	22	m	House parameter
w_H	10	m	House parameter
h_H	3	m	House parameter
n_W	7	m	House parameter
w_W	1	m	House parameter
h_W	1.6	m	House parameter
λ_{Brick}	0.179	W/(m·K)	House parameter
λ_{EPS}	0.035	W/(m·K)	House parameter
d_{Brick}	0.38	m	House parameter
d_{InsW}	0.15	m	House parameter
d_{InsS}	0.3	m	House parameter
$U_{Windows}$	0.9	W/(m^2*K)	House parameter
m_{water}	120	kg	Determined by the authors
$T_{water}(0)$	55	°C	Determined by the authors
T_{water}^{sp}	65	°C	Determined by the authors
T_{water}^{th}	15	°C	Determined by the authors
C_{water}	4181	J/(kg·K)	Physical constant
η_{water}	0.9	-	Determined by the authors
P_{water}^{max}	2300	W	Determined by the authors
hl_{water}	0.0069	W/(kg·K)	Derived from the results of [18]
$T_{air}^{indoor}(0)$	24	°C	Determined by the authors
T_{air}^{sp}	22	°C	Determined by the authors
T_{air}^{th}	3	°C	Determined by the authors
$T_{heating}$	50	°C	Determined by the authors
$^{dm_{air}(t)}/_{dt}$	Constant 0.2	kg/s	Determined by the authors
C_{air}	1005.4	J/(kg·K)	Physical constant
ρ_{air}	1.225	kg/m^3	Physical constant
$P_{heating}^{max}$	10,000	W	Determined by the authors
$cop_{heating}$	3.6	-	Determined by the authors
$SoC(0)$	15	%	Determined by the authors
$MinSoC$	15	%	Determined by the authors
$MaxSoC$	85	%	Determined by the authors
$P_{battery}^{max}$	4000	W	Determined by the authors
Cap_b	11000	Wh	Determined by the authors
$\eta_{battery}$	0.9	-	Determined by the authors