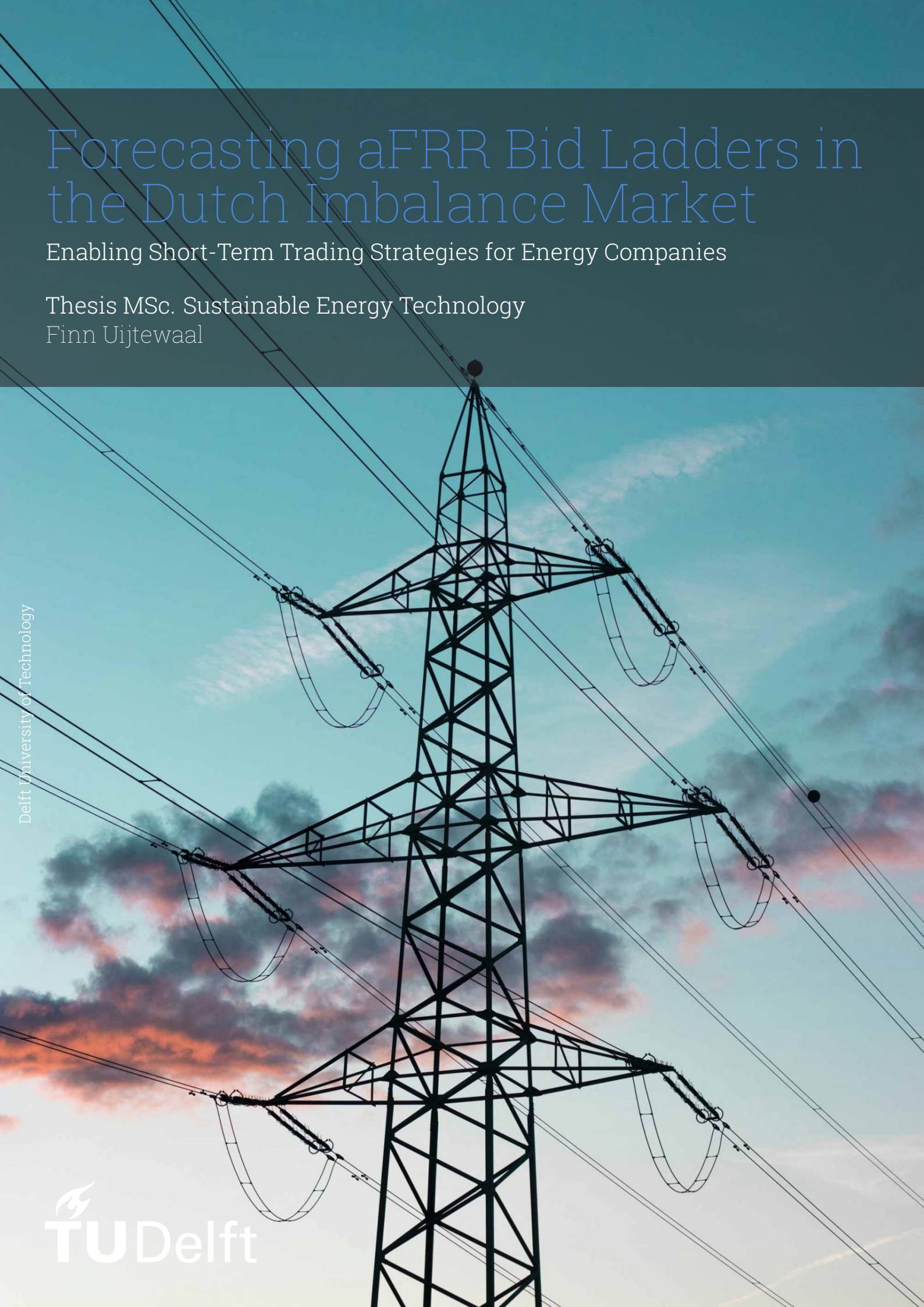
Forecasting aFRR Bid Ladders in the Dutch Imbalance Market

Enabling Short-Term Trading Strategies for Energy Companies

Thesis MSc. Sustainable Energy Technology

Finn Uijtewaal



Delft University of Technology

TU Delft

Forecasting aFRR Bid Ladders in the Dutch Imbalance Market

Enabling Short-Term Trading Strategies for Energy Companies

MSc. Thesis by

Finn Uijtewaal

Student Name	Student Number	Student Program
Finn Uijtewaal	4725182	Sustainable Energy Technology (Track: Electric Mobility, Power, Economics)

To obtain the degree of Master of Science at the Delft University of Technology,
to be defended publicly on June 27, 2025.

Thesis committee:	Dr.ir. P. (Pedro) Vergara Barrios	TU Delft, (daily) supervisor
	Dr.ir. T. (Ties) van der Heijden	TU Delft, supervisor
	Dr.ir. S.H. (Simon) Tindemans	TU Delft, (chair) supervisor
	Dr.ir. R.A. (Remco) Verzijlbergh	TU Delft, (external) supervisor
	Dr.ir. S. (Sumeyra) Demir	Essent, (company) supervisor
Faculty:	Electrical Engineering, Mathematics and Computer Science (EEMCS)	TU Delft
Project duration:	September 2024 — June 2025	

Cover: Image on Unsplash by Severin Demchuk [1]

TU Delft

Preface

This thesis concludes the master's program in Sustainable Energy Technology at the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology.

Foremost, I would like to thank my TU Delft supervisors, Pedro Vergara and Ties van der Heijden. Pedro's positive attitude and guidance were reassuring, especially when I had concerns about the planning or direction of the project. Ties shared valuable knowledge with a direct and flexible approach. His honest feedback and thoughtful challenges consistently pushed me to think critically and make more independent, well-founded choices. Lastly, I thank Simon Tindemans and Remco Verzijlbergh, who also served on my committee, for their contributions during the final phase of the project.

Also, I would like to express my gratitude to my supervisors at Essent. Sumeyra Demir provided weekly guidance and support throughout the project. Edgar Wilton recognised the practical value of the forecast's application and supported its relevance. I also want to thank Wouter Bles, who also conducted his thesis on imbalance markets. Sharing knowledge and exchanging ideas on similar topics proved very helpful and more fun throughout the project.

Finally, I am thankful to my family, friends, and girlfriend for their support and understanding throughout this project. Taking on a topic largely outside my study program's scope meant stepping into a completely new field. Coming from a family involved in a contractor's business, this project initially felt like building a house without having the right tools. However, as the knowledge and skills developed along the way, they provided me with the tools to make better informed choices. This gradually gave me confidence in the direction and outcome of the work.

Now, my academic journey comes to an end. While school never truly captured my interest in my earlier years, this gradually changed during the later stages of my bachelor's and in my master's. I grew to appreciate the value of learning and discovery, and I am grateful for the experiences, friendships, and personal growth that came with it.

Finn Uijtewaal
Rotterdam, June 2025

Abstract

Background: The decentralisation of energy supply, driven by renewables, increases volatility and creates imbalances, requiring flexible balancing to maintain grid stability. Ancillary services provide this flexibility by injecting power (upward regulation) or removing it (downward regulation). Among these, Automatic Frequency Restoration Reserve (aFRR) is key due to its large regulating volume and direct effect on imbalance prices through activation costs. As volatility rises, short-term trading in intra-day and imbalance markets becomes increasingly important, raising the demand for accurate forecasts. While forecasting for day-ahead and intra-day markets is well established in Electricity Price Forecasting (EPF), limited research on imbalance markets, particularly Dutch aFRR, highlights this study's novelty. The Dutch market applies dual pricing with separate supply curves (bid ladders) for upward and downward regulation.

Objective: This research forecasts aFRR bid ladders in the Dutch electricity market and applies them to short-term trading strategies. Forecasts are generated three hours ahead of delivery, aligning with decision windows for intra-day trading, aFRR participation, or no interaction. The study includes analyses of the aFRR market, reviews of forecasting methods, development of a tailored machine-learning framework, evaluation of model performance, and application of forecasts in short-term trading.

Methodology: A structured machine-learning pipeline is developed, consisting of pre-processing, transformation, model selection, prediction, and evaluation. In the transformation step, data is first scaled and then decomposed using Principal Component Analysis (PCA) to reduce dimensionality and capture key variance. These components serve as inputs for predictive models (LASSO, XGBOOST, and LSTM), which are benchmarked against preliminary bid ladders available three hours before delivery (BENCHMARK). Forecast performance is assessed using point metrics (sMAPE), interval metrics (PICP, PINAW), and a novel self-developed metric, the Largest Knick Volume (LKV), which measures accuracy at critical bid ladder inflection points relevant for short-term trading.

Results: The outcome consists of two parts. First, model evaluations at both PCA and reconstructed bid ladder levels indicate that all models capture general trends, but none outperform the benchmark. This is confirmed by Diebold-Mariano tests, which show superior performance at the PCA level. After bid ladder reconstruction, the benchmark yields sMAPE values of 7% (upward) and 8% (downward), outperforming LASSO (7/10%), XGBOOST (8/11%), and LSTM (7/11%). Second, these forecasts are integrated into a Battery Energy Storage System (BESS) intra-day trading profile, including one based solely on intra-day trading and three using different aFRR bid ladder positions: gas turbine marginal costs, intra-day with premium, and LKV-based. As a result, aFRR integration yields revenue gains, with high volume-price strategies increasing revenues by up to 18%.

Conclusion: The various developed and evaluated forecasting methods underperform compared to the benchmark. Incorporating available information three hours ahead already captures most predictive signals, leaving little room for additional forecasting gains. This underperformance is due to weak correlations between market fundamentals and bid outcomes, and the influence of dominant individual actors in the relatively small Dutch aFRR market. Nonetheless, combining existing market data with participation in other markets can enhance profitability. This study demonstrates that forecast-informed bidding enhances battery storage profitability and supports data-driven market participation strategic value.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Table 1: List of all abbreviations used throughout the thesis.

Abbreviation	Definition
FF	Forwards & Futures Market
DAM	Day-ahead Market
IDM	Intra-day Market
IM	Imbalance Market
FCR	Frequency Containment Reserve
aFRR	automatic Frequency Restoration Reserve
mFRR	manual Frequency Restoration Reserve
PICASSO	Platform International Coordination Automated Frequency - Restoration And Stable System Operation
DSO	Distribution System Operator
TSO	Transmission System Operator
ES	Electricity Supplier
EA	Energy Aggregator
BRP	Balance Responsible Party
BSP	Balancing Service Provider
CSP	Congestion Service Provider
OTC	Over The Counter
PPA	Power Purchase Agreement
ISP	Imbalance Settlement Period
E-program	Energy Program or Schedule
APFAS	Auction Platform For Ancillary Services
ASC	Aggregated Supply Curve
EPF	Electricity Price Forecasting
LASSO	Least Absolute Shrinkage and Selection Operator
XGBOOST	eXtreme Gradient Boosting
LSTM	Long Short-Term Memory
KDE	Kernel Density Estimation
PCA	Principal Component Analysis
CP	Conformal Prediction
ICP	Inductive Conformal Prediction = Split Conformal Prediction
TPE	Tree-structured Parzen Estimator
sMAPE	symmetric Mean Absolute Percentage Error
LKV	Largest Knick Volume
PICP	Prediction Interval Coverage Percentage
PINAW	Prediction Interval Normalised Average Width
DM	Diebold-Mariano
BESS	Battery Energy Storage System
P&L	Profit & Loss
MW	Megawatt
MWh	Megawatt-hour
PJ	Petajoule

Symbols

Table 2: List of all symbols, descriptions, and units used throughout the thesis.

Category	Symbol	Description	Unit
General	p	bid price	[€/MWh]
	q	bid volume	[MW]
	t	delivery time of energy	[ISP]
Data dimensions	x	input data	[-]
	y	output data	[-]
	$\hat{y}$	predicted output	[-]
	n	index of timestep	[-]
	N	amount of samples	[-]
	M	amount of features	[-]
Lasso Model	β	regression coefficient	[-]
	λ_{LASSO}	regularisation parameter	[-]
XGBoost Model	T	number of leaves in decision tree	[-]
	K	total number of decision trees	[-]
	γ	regularisation parameter for adding a new leaf	[-]
	λ_{XGB}	regularisation parameter for model leaf weights	[-]
	w	leaf weight	[-]
	g	first-order gradient	[-]
	h	second-order gradient (Hessian)	[-]
LSTM Model	f	forget gate	[-]
	i	input gate	[-]
	C	cell state	[-]
	o	output gate	[-]
Transformation	$y_{n,q}$	output data	[€/MWh]
	$y_{n,p}$	output data transposed	[MW]
	q	index of timestep	[-]
	p	index of price	[-]
	$S0$	raw data	[-]
	$S1$	sigmoid transformation	[-]
	$S2$	PCA dimension reduction	[-]
Conformal	C	prediction interval	[-]
	α	significance level	[-]
	$\mathbb{P}$	probability measure	[-]
	s	nonconformity score	[-]
	$\hat{q}$	quantile of nonconformity scores	[-]
Error metrics	σ	standard deviation	[-]
	μ	mean	[-]
	τ	spike location	[-]
	z	standard score	[-]
	R	range of observed values	[-]
	y^U	upper boundary	[-]
	y^L	lower boundary	[-]
Strategy	E	battery energy	[MWh]
	ΔE	aFRR compensation volume	[MWh]
	P	battery power	[MW]

1 Introduction

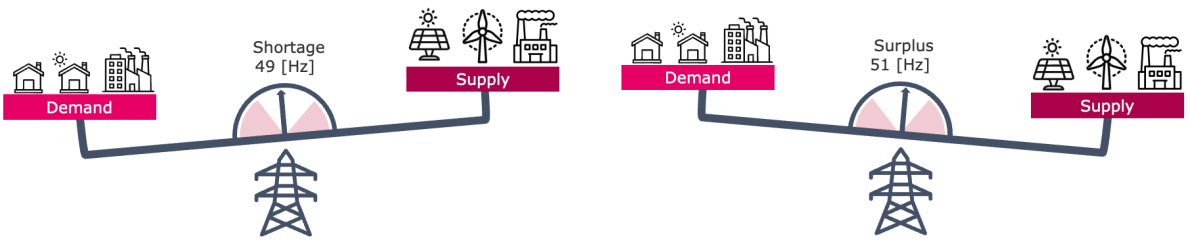
Over the past three decades, the Dutch electricity market has undergone a significant transformation due to liberalisation and integration of renewable energy sources, such as solar and wind energy [2].

Initially, the market was characterised by a few state-controlled utilities. These vertically integrated monopolies managed the entire supply chain from production, distribution, to retail [3]. This shift toward a liberalised and competitive market began with the unbundling of generation and transmission. State-owned entities were privatised and electricity markets gradually opened up to competition [2].

Over the last decade, renewable energy sources have driven a shift toward a more dynamic and decentralised market landscape. A centralised energy system relies on a few large power plants with top-down control, while a decentralised system distributes generation across many smaller power assets closer to consumption.

In recent years, the Dutch electricity grid transitioned to a more decentralised system. This shift is strongly influenced by the EU Renewable Energy Directive [4], which set a binding target of at least 42.5% renewable energy and greater cross-border cooperation by 2030. It caused a major change in energy production, with centralised generation falling from 63 to 46 Petajoules, while decentralised generation rose from 38 to 75 Petajoules [3]. In decentralised systems, a large share of electricity comes from renewables. Although these are cheaper to produce than traditional fuels, such as coal and combined cycle gas turbines (CCGT), they lack dispatchable flexibility due to weather dependence. The impact of non-flexible energy assets on grid stability is shown in Figure 1.1. The grid operates stable at 50 Hz when supply equals demand. A high share of renewables raises the risk of both surpluses and shortages due to their variability. On sunny and windy days, generation may exceed demand, raising the grid frequency. Sudden drops in wind or solar output can cause shortages and lower the frequency. These imbalances increase price volatility and complicate system balancing.

Given these developments, the Dutch electricity market presents a unique case for research. The combination of a high share of renewable energy [5] and a distinct operational design places increased pressure on grid stability. As a result, effective balancing mechanisms are crucial.



(a) Grid shortage **(b)** Grid surplus

Figure 1.1: Imbalances in the electricity grid occur when increased renewable energy output does not align with demand.

1.1. Background

To understand balancing mechanisms, it is important to first examine how grid imbalances are compensated and how balancing energy is traded through electricity markets.

1.1.1. Role of Ancillary Services

The grid is physically stabilised through ancillary services. Ancillary services ensure the proper operation of the power grid by maintaining system frequency around 50 Hz, and by stabilising voltage and power load [6]. These services operate on different timescales to correct deviations, as illustrated in Figure 1.2, which shows the sequence and response times of frequency reserves:

- **Frequency Containment Reserve (FCR)**: Activated within seconds to immediately counteract frequency deviations and stabilise the system.
- **Automatic Frequency Restoration Reserve (aFRR)**: Automatically activated within tens of seconds to minutes to restore frequency to its nominal value by adjusting generation or consumption based on real-time measurements.
- **Manual Frequency Restoration Reserve (mFRR)**: Activated manually within several minutes to address prolonged or larger imbalances not corrected by FCR or aFRR.

The importance of aFRR lies in its substantial activation volume and direct influence on imbalance pricing mechanisms. First of all, aFRR accounts for the largest volume of energy among balancing services [7]. This is because FCR is mainly used to correct small frequency deviations with minimal energy, while mFRR is typically reserved for emergency situations. Additionally, the prominence of aFRR is reinforced by its role in determining the imbalance price. This imbalance price is calculated for each 15-minute period and is derived from the activation costs of aFRR: the most expensive bid for upward regulation and the cheapest bid for downward regulation. Since upward regulation injects electricity into the grid and downward regulation removes excess supply, these prices reflect the real-time value of balancing services [6]. Energy companies are required to pay this imbalance price when their actual electricity consumption or production deviates from their scheduled plans [6].

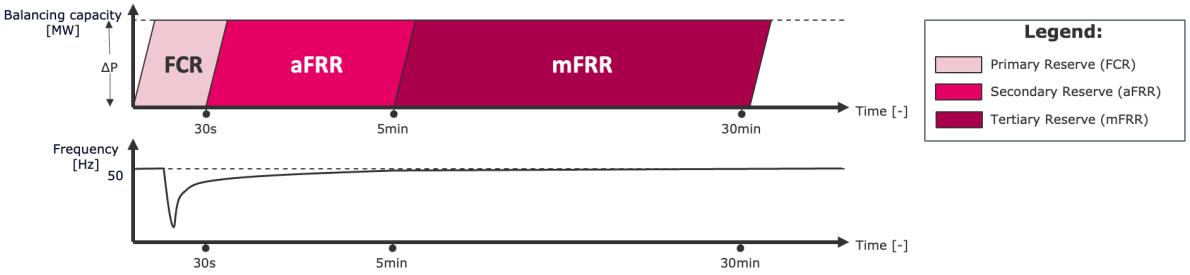


Figure 1.2: Power system frequency control reserves (FCR, aFRR, mFRR) with activation times, relative to the nominal frequency of 50 Hz.

1.1.2. Market-Based Mechanisms

The energy delivered through aFRR is not only physically essential for grid balancing, but also financially settled through dedicated electricity markets. After aFRR is activated to correct imbalances, the corresponding energy volumes are traded and priced according to market mechanisms. Two principal pricing models govern these markets: pay-as-cleared (marginal pricing) and pay-as-bid, as illustrated in Figure 1.3. In pay-as-cleared markets, all accepted bids receive the price of the highest accepted offer, promoting uniform pricing. In pay-as-bid markets, each participant is paid according to their individual submitted matched bid and sell offer.

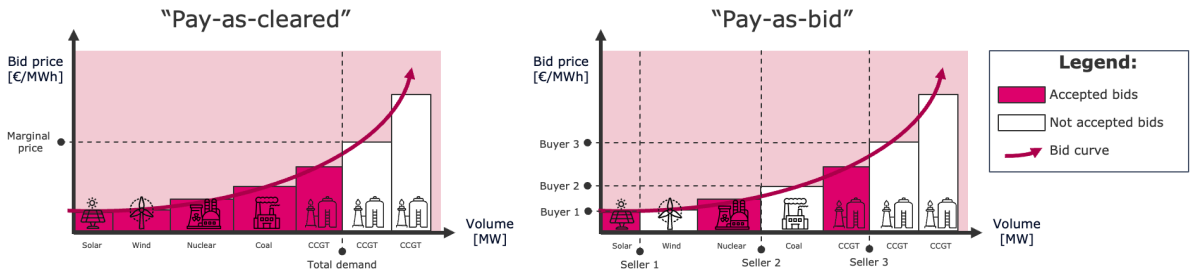


Figure 1.3: The two market-clearing mechanisms that determine electricity market prices.

1.2. Research Motivation

Three key developments shape the motivation for this research into the Dutch aFRR market: the growing role of electrification in the energy transition, the increasing importance of grid stability, and the rising significance of electricity trading closer to delivery.

Electrification is a key component for Sustainable Energy Technology, as it enables the integration of renewable energy sources and reduces dependence on fossil fuels [8]. To support this shift, accurate electricity market forecasts are essential for maintaining grid balance. Accurate predictions help anticipate future events before they occur. The aFRR market supports this by automatically correcting frequency deviations in near real-time, enabling a more reliable and flexible integration of renewable energy into the grid.

Secondly, increasing pressure on grid stability is observed, both in technical and financial terms. Over the past four years, the demand for active balancing capacity, which refers to reserve power used to correct real-time imbalances, has increased by 20 to 30% [7]. This rise is partly due to more frequent deviations from energy programs (E-programs). E-programs are schedules submitted by energy suppliers, indicating how much electricity they plan to generate or consume in each time block. Because renewable energy sources are difficult to predict, actual production or consumption often deviates from these schedules. As a result, energy companies must frequently purchase electricity at short notice and at higher prices to restore balance. This leads to higher imbalance costs and places financial pressure on utilities, grid operators, and consumers [9]. Forecasting aFRR bid ladders provides insight into expected balancing needs supporting grid stability.

Lastly, the role of intra-day and aFRR trading within the Dutch electricity market has grown significantly in recent years. This is primarily driven by increased price volatility, which has led to more frequent and extreme price fluctuations. As a result, short-term markets have become more attractive for participants aiming to optimise positions or profit from rapid changes. Consequently, intra-day and aFRR markets, once considered secondary, have gained strategic importance due to their flexibility and potential for responsive trading. This trend is further reinforced by the introduction of PICASSO (Platform for the International Coordination of Automated Frequency Restoration and Stable System Operation), which went live end 2024 [9]. By enabling cross-border aFRR activation based on a common European merit order, PICASSO increases market integration and competitiveness, further raising the complexity of strategic bidding [9]. In this context, advanced forecasting methods become even more relevant. The findings of Browell and Gilbert (2022) [10] highlight how improved data availability supports the use of machine learning and probabilistic forecasting to predict imbalance prices and reserve activations. In the aFRR market, such predictive capabilities can help optimise bid strategies and contribute to maintaining grid balance under increasingly dynamic market conditions.

These developments underline the relevance of the aFRR market and highlight the need for more advanced forecasting methods. The next section outlines the specific challenges this research aims to address.

1.3. Problem Description

Two central challenges emerge in the Dutch imbalance market: the difficulty of forecasting imbalance prices due to market complexity, and the financial risk energy companies face from unexpected deviations in generation or consumption.

One key issue is the complexity around forecasting imbalance prices. This is mainly due to the uncertainty in the market and the complex market structure.

Uncertainty in the imbalance market arises from real-time fluctuations in supply and demand. These fluctuations are often caused by unexpected outages, variable renewable generation, and the unpredictable behaviour of market participants [11]. Because imbalance prices are determined shortly before delivery, they reflect rapidly changing system conditions. As a result, forecasting must contend with limited lead times and highly uncertain inputs [12]. Crucially, the imbalance market is designed to address precisely these unforeseen changes. Its role is to ensure system stability by compensating for deviations from scheduled supply and demand. Therefore, a certain degree of unpredictability is not only unavoidable but also essential to its function. Forecast errors are a natural reflection of the market's purpose, which is to respond flexibly to real-time conditions rather than to follow predictable patterns.

In addition, the structure of the imbalance market introduces further challenges to forecasting efforts. In the Netherlands, imbalance prices are determined through a dual pricing system, with separate prices for upward regulation (grid injections) and downward regulation (grid withdrawals) [6]. These prices depend both on the outcomes of preceding market segments, such as the day-ahead and intra-day markets, and on the real-time state of the power system. This combined dependency complicates the determination of imbalance prices and increases forecasting uncertainty.

Another key concern for energy utility companies is managing imbalance price risk for electricity portfolios. Risk management involves minimising the cost of deviating from submitted E-programs. An electricity portfolio includes the generation assets and financial instruments managed by an entity. In this context, risk mitigation refers to identifying potential threats to assets and operations and implementing strategies to reduce their financial and operational impacts. To manage these risks, companies adjust portfolio deviations through intra-day trading and participation in the aFRR market. For example, negative prices in the day-ahead market may make participation there unprofitable. As a result, the use of other markets, such as aFRR, becomes more attractive [8].

Given the difficulty of forecasting imbalance prices and the associated financial risks, forecasting aFRR bid ladders offers practical means to gain market insights and enhance trading decisions.

1.4. Research Objective

This research aims to forecasting upward and downward aFRR bid ladders in the Dutch imbalance market by applying machine learning models. The forecasts support energy companies' short-term trading decisions by providing improved insights into expected balancing needs and price dynamics. Forecasts are generated three hours prior to delivery, aligning with operational decision windows during which market participants engage in intra-day trading, aFRR activation, or portfolio adjustments. A 15-minute resolution is applied, reflecting the operational granularity of the intra-day and aFRR markets.

The objective of this research is to forecast upward and downward aFRR bid ladders in the Dutch electricity market to support short-term trading strategies. To achieve this, a set of specific research objectives is formulated. First, the operational mechanisms of the Dutch electricity and aFRR market are analysed to provide a contextual foundation. Second, existing forecasting methodologies are reviewed and assessed to identify suitable approaches for bid ladder prediction. Third, a tailored forecasting framework is developed and implemented using machine learning models. Fourth, the effectiveness of these models is compared based on their predictive performance. Finally, a method is proposed to translate forecasted bid curves into actionable trading strategies for energy utility companies. These objectives ensure a structured approach towards fulfilling the overall goal of the study.

To address these research objectives, the following research question is formulated:

How to effectively forecast automatic Frequency Restoration Reserve (aFRR) bid ladders, and how can these forecasts support energy trading companies in the Dutch electricity market?

To support the main research question, this study is structured around five sub-questions that provide a comprehensive approach to achieve the research objective. The research questions are categorised into three thematic areas. RQ1 addresses the underlying market dynamics affecting aFRR bid ladders. RQ2 and RQ3 concentrate on developing and selecting forecasting models and techniques. RQ4 and RQ5 focus on the practical application of the forecasts, including identifying key thresholds and informing trading strategies.

- *RQ1: What external factors, such as weather, generation, market, time factors, and historical bid ladders, influence the price and volume of aFRR bid ladders?*
- *RQ2: Which machine learning models commonly applied in EPF can be used to forecast full aFRR bid curves?*
- *RQ3: What decomposition techniques can be used to forecast full aFRR bid curves to account for varying lengths and computational cost?*
- *RQ4: What forecasting insights derived from aFRR bid ladders can be utilised to inform short-term trading decisions between intra-day and imbalance markets?*
- *RQ5: What is the economic value of incorporating aFRR price forecasts into short-term trading strategies for profit optimisation?*

1.5. Thesis Structure

This thesis adopts a structured approach to address the research objectives: it begins with foundational electricity market knowledge, continues with model development, and ends with a model application. The structure is outlined in the following chapters. Chapter 2 outlines the Dutch electricity market. It provides a basis for understanding energy markets, with a focus on aFRR bid ladders and their price formation. Relevant academic work is reviewed in Chapter 3, where the research gaps and commonly used predictive models are explained. Chapter 4 examines the structure of aFRR bid ladders, including bid prices and volumes. It also explores underlying factors that shape these curves, such as market conditions and external variables. Chapter 5 explains the development of the forecasting models. It covers data pre-processing, data transformation, and the selection of evaluation metrics. Chapter 6 presents model outcomes and performance metrics. A battery trading strategy for different scenarios are introduced in Chapter 7. Implications, limitations, and underlying assumptions are addressed in Chapter 8. The main findings and contributions are summarised in Chapter 9, followed by recommendations for future research and practical applications.

2

Overview Dutch Electricity Market

This chapter provides an overview of the Dutch electricity market and terminology used throughout the thesis. It first describes the roles of market participants in the physical, administrative, and market domains. Next, it explains the sequence of trading in the forward & futures, day-ahead, intra-day, and imbalance markets. Special focus is placed on the imbalance market and the role of ancillary services, especially aFRR. The chapter ends with a description of how imbalance prices are set based on the aFRR bid ladder and the regulation state of the grid.

2.1. Market Participants and Roles

The Dutch electricity market operates as a free market where both private entities and semi-governmental organisations provide oversight. Private market participants include producers, suppliers, and traders, who operate competitively based on mutual agreements and market regulations. Semi-governmental entities, such as TenneT and regional grid operators (e.g. Enexis, Stedin, Liander), ensure system reliability, grid access, and regulatory compliance. Market actors are categorised into three domains: (1) physical (generation, transmission, consumption), (2) administrative (metering, data exchange), and (3) market (trading, settlements). These actors are visually summarised in Figure 2.1, where each market participant is listed by category, and the arrows indicate how they interact 2.1 [13].

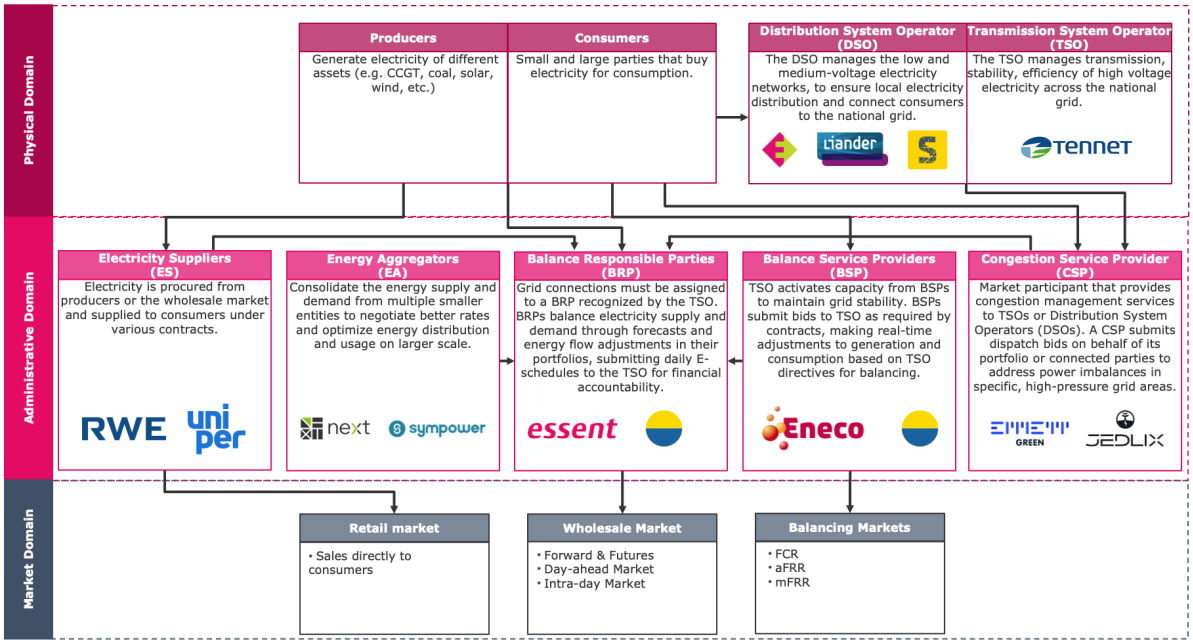


Figure 2.1: Overview of the Dutch electricity market structure, illustrating the roles of physical, administrative, and market domains.

(1) Physical Domain:
Electricity is produced and consumed within the electricity system [13]. Producers generate electricity using various assets. This electricity is then transported through the electricity grid to meet demand. Consumption occurs at different scales, ranging from households to large industrial facilities. To enable this flow, the grid connects production to consumption. The Transmission System Operator (TSO) oversees electricity transport across the high-voltage grid, ensuring system-wide balance and reliability. At the regional level, the Distribution System Operator (DSO) takes over responsibility, delivering electricity to end users through the medium- and low-voltage networks.

(2) Administration Domain:
In the administration domain, electricity is traded and settled [13]. Electricity Suppliers (ES) and Energy Aggregators (EA) sell electricity on various markets, with EA bundling smaller assets into larger tradable units. These market transactions feed into the responsibility of Balance Responsible Parties (BRP), who are financially accountable for supply-demand imbalances. To manage this, the BRP submits daily E-programs per imbalance settlement period (ISP), which remain adjustable until 10:00 the following day. When imbalances occur, Balance Service Providers (BSP) provide ancillary services to stabilise the grid by injecting or withdrawing power. In parallel, Congestion Service Providers (CSP) address local grid constraints caused by limited cable or transformer capacity by activating flexible consumption or generation in the affected area [13].

(3) Market Domain:
Electricity is exchanged through various markets [13]. In the Netherlands, these markets are broadly categorised into three parts: retail, wholesale, and balancing markets. In retail markets, electricity is directly sold from electricity suppliers to end consumers. These suppliers offer different contracts per customer, such as fixed-price or dynamic tariffs [14]. The wholesale market in the Netherlands facilitates electricity trading between generators, traders, and suppliers [14]. Generators offer their production, which traders buy and sell across various time frames. This enables suppliers to procure electricity for their end-users. Trading occurs through long-term contracts (forwards & futures), the day-ahead market, and the intra-day market, as detailed in Subsections 2.2.1, 2.2.2, and 2.2.3. Participation in electricity markets is open to parties that hold the appropriate license and meet the technical and regulatory requirements set by TSO and the market regulator. The imbalance market is operated by the Dutch TSO [14]. Subsection 2.2.4 provides further details on these balancing markets.

These participants interact by buying and selling electricity through a sequence of organised markets, each serving a specific timeframe and system function.

2.2. Long- and Short-Term Electricity Markets

Electricity markets are unique due to the requirement that supply and demand must always be balanced. To facilitate this, the Dutch market is structured into long-term and short-term segments, as shown in Figure 2.2 [14]. Long-term markets (forwards & futures) allow participants to hedge price risks and secure delivery over extended periods. Short-term markets (day-ahead, intra-day, imbalance) provide mechanisms to correct forecast errors, fine-tune trading positions, and determine the actual output of energy assets based on real-time system needs. This progressive structure reduces uncertainty, improves cost allocation, and enhances system reliability [14].

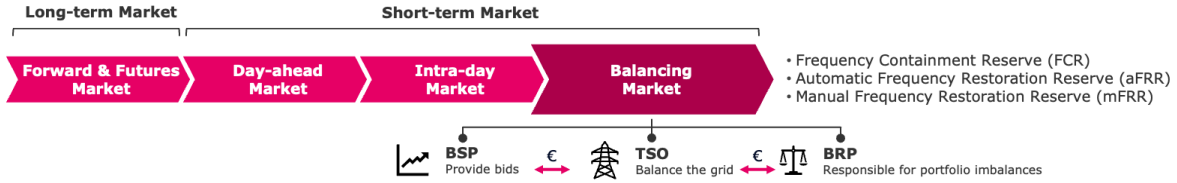


Figure 2.2: Overview of all long- and short-term markets in the Netherlands.

Figure 2.3 summarises a detailed and time-ordered overview of how the four markets interact. It visualises the continuous flow from long-term to real-time markets, showing when bids must be submitted, how market segments overlap, and through which platforms the auctions are held. This timeline highlights the operational dependencies and deadlines that shape decision-making in each trading phase.

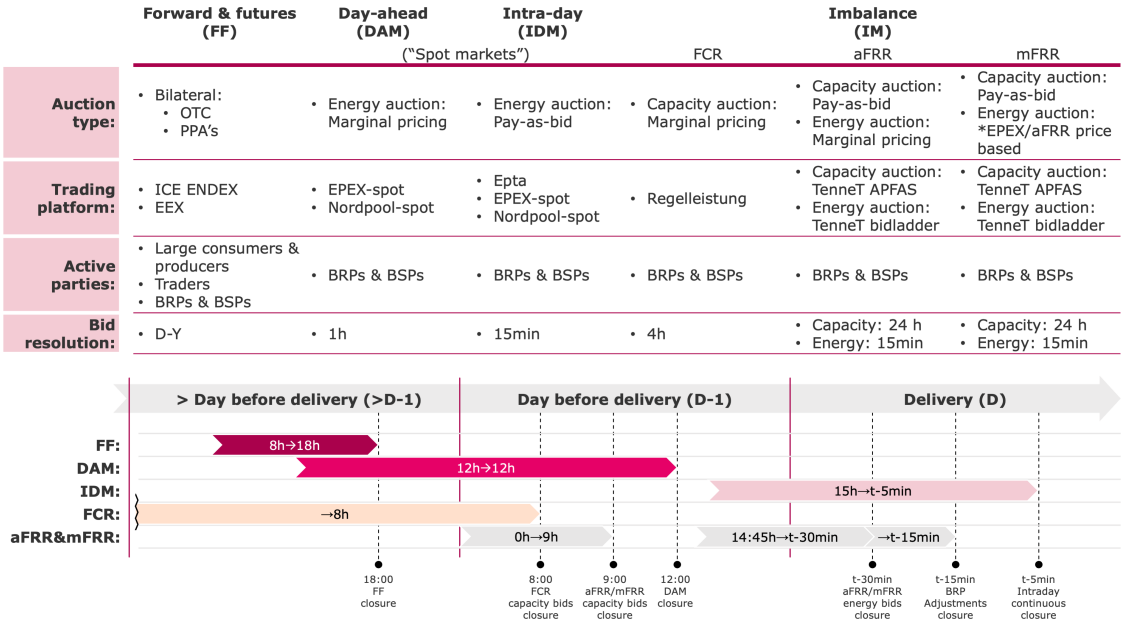


Figure 2.3: Overview of Dutch electricity markets, timelines, and bidding deadlines.

2.2.1. Forward and Futures Market

Enable participants to hedge against price risks by trading forward and futures (FF) contracts. These contracts fix the price and volume of electricity for extended delivery periods [15]. Securing prices in advance reduces exposure to long-term volatility and improves risk control. Forwards are customised, over-the-counter contracts (OTC) that allow parties to agree on specific volumes, prices, and delivery periods. Futures are standardised contracts traded on exchanges, typically for fixed products like baseload delivery at set future dates. A specific type of forward contract is a Power Purchase Agreement (PPA), in which a buyer agrees to purchase electricity directly from a producer, often from renewable sources, under long-term conditions. This market is open daily from 8:00 to 18:00. Energy trading in the forward and futures markets is carried out on ICE ENDEX and the European Energy Exchange (EEX) [16, 17]. The bid resolution in these markets can range from daily contracts to multi-year agreements, which offer flexibility in terms of contract duration. The trade volumes in the forward and futures markets are generally considered larger, with no fixed limits, due to the scale of participants and the long-term planning involved [14].

2.2.2. Day-ahead Market

Electricity is traded one day in advance on the day-ahead market (DAM), allowing market participants to plan generation and consumption by submitting bids and offers. This market is settled based on marginal pricing. Due to its high liquidity and frequent participation, the DAM price is often referred to as the ”electricity price”. The market clears daily at 12:00, after which trading the next day opens. Trading is conducted via European-wide platforms such as EPEX Spot and Nord Pool Spot, where bids and offers from market participants are automatically matched through market coupling mechanisms that optimise cross-border flows [18, 19]. To align with operational planning, bids are submitted in 1-hour intervals. A minimum bid volume of 0.1 MW is applied to allow participation from both small and large-scale market actors [14].

2.2.3. Intra-day Market

After the DAM has been cleared, the intra-day market (IDM) opens at 15:00 to allow participants to adjust their positions and correct forecasting errors closer to real-time. This market consists of four components: three intra-day auctions, IDA1 (D-1 15:00), IDA2 (D-1 22:00), and IDA3 (D 10:00), and the continuous IDM [18]. While IDA auction volumes remain relatively low in the Netherlands, the continuous market is widely used due to its flexibility, enabling transactions up to shortly before delivery. In this segment, bids are accepted immediately upon submission, following a pay-as-bid pricing mechanism [14]. Increasing uncertainties in DAM forecasts lead to higher trading volumes closer to real-time. Intra-day trading takes place on platforms such as Epta, EPEX Spot, and Nord Pool Spot, where participants can submit and match bids in real time [18] [19]. Electricity is traded in 15-minute intervals to accommodate short-term balancing needs, with bids required at least 5 minutes before delivery. The minimum bid size is 0.1 MW, ensuring accessibility for both small and large market participants [18].

2.2.4. Imbalance Market

Deviations that remain after the closure of the DAM and IDM are resolved in the imbalance market (IM). Participation in this market is limited to BRPs, BSPs, and the TSO, as its focus on real-time system balancing requires specialised operational capabilities. Table A.1 overviews all registered Dutch BSPs, highlighting the market's limited size and concentrated structure.

Frequency Containment Reserve

When the grid frequency deviates from the nominal value of 50 Hz, FCR is automatically activated. This response occurs within 2 to 30 seconds to counteract the deviation. By restoring frequency balance, FCR helps maintain system stability and prevents further instability, load shedding, or blackouts [20]. The FCR market uses a capacity auction model in which BSPs are compensated in €/MW for the capacity they make available. Bids are submitted day-ahead, before 08:00, using a pay-as-cleared mechanism, meaning all selected participants receive the price of the most expensive accepted bid for that block [20]. Only availability is remunerated. BSPs are not paid for the actual energy delivered during activation. This reflects FCR's role as a preventive service rather than an energy product. The market is facilitated by the TSO and operated via Regelleistung, a centralised European trading platform for FCR procurement. Each bid must cover a 4-hour block and be symmetric, providing equal upward and downward capacity. The delivery day is divided into six time blocks of four hours each: 00:00–04:00, 04:00–08:00, 08:00–12:00, 12:00–16:00, 16:00–20:00, and 20:00–24:00. BSPs may place separate bids for each block, enabling flexible participation throughout the day. The minimum bid size is 1 MW [20].

Automatic Frequency Restoration Reserve

After FCR has been activated to stabilise the initial frequency deviation, aFRR takes over to further restore the system balance. This secondary reserve typically responds within approximately five minutes [21]. To enable this response, the aFRR market is structured around two sequential types of bids: capacity bids (€/MW) and energy bids (€/MWh) [6], as visualised in Figure 2.4. The process begins with capacity bids, which ensure that a sufficient volume of reserve power is committed in advance. These bids must be submitted one day ahead (D-1) before 09:00. Accepted units are required to remain available for the entire delivery day. Capacity bids are cleared using a pay-as-bid mechanism. Once the required capacity has been contracted, energy bids are submitted to determine the actual activation order of reserves. These bids form the aFRR bid ladders and can be placed by both contracted participants (who secured capacity) and voluntary participants. Contracted energy bids must be submitted by 14:45 on the day before delivery, while voluntary bids can be placed independently, up to 30 minutes before delivery. Clearing of energy bids occurs through marginal pricing [21]. The full sequence of bid submission and activation is shown in Figure 2.4. BSPs submit capacity bids through the Auction Platform for Ancillary Services (APFAS), while energy bids are handled via TenneT's aFRR bid ladder. The bid resolution for aFRR capacity is set at 24 hours, meaning assets must be available throughout the entire day. In contrast, energy bids have a finer resolution of 15 minutes, allowing for more precise activation. This 15-minute interval is referred to as the Imbalance Settlement Period (ISP) [21]. Each bid must cover at least 1 MW. Furthermore, no more than three bids under 4 MW are permitted per ISP, in line with regulatory constraints [21].

Manual Frequency Restoration Reserve

When larger imbalances persist after the activation of FCR and aFRR, mFRR, is deployed by the TSO. The TSO activates mFRR in response to severe incidents or prolonged imbalances that cannot be resolved through automated mechanisms [22]. Therefore, it is also known as emergency power. The mFRR market operates in a structure similar to that of aFRR, where both capacity and energy bids are placed [22]. To ensure pricing consistency and avoid extreme bid values, upward prices are subject to three minimum conditions: €200/MWh, the aFRR price plus 10%, or the DAM price plus 200 €/MWh. Downward prices are constrained by the lower of either the aFRR shortage price minus 100 €/MWh or the DAM price minus 250 €/MWh. Bids for mFRR capacity are submitted through the Auction Platform for Ancillary Services (APFAS). The resolution of capacity bids is 24 hours, meaning assets must be available throughout the day. Energy bids, on the other hand, are submitted with a resolution of 15 minutes to reflect real-time system needs. Each bid must cover a minimum volume of 1 MW.

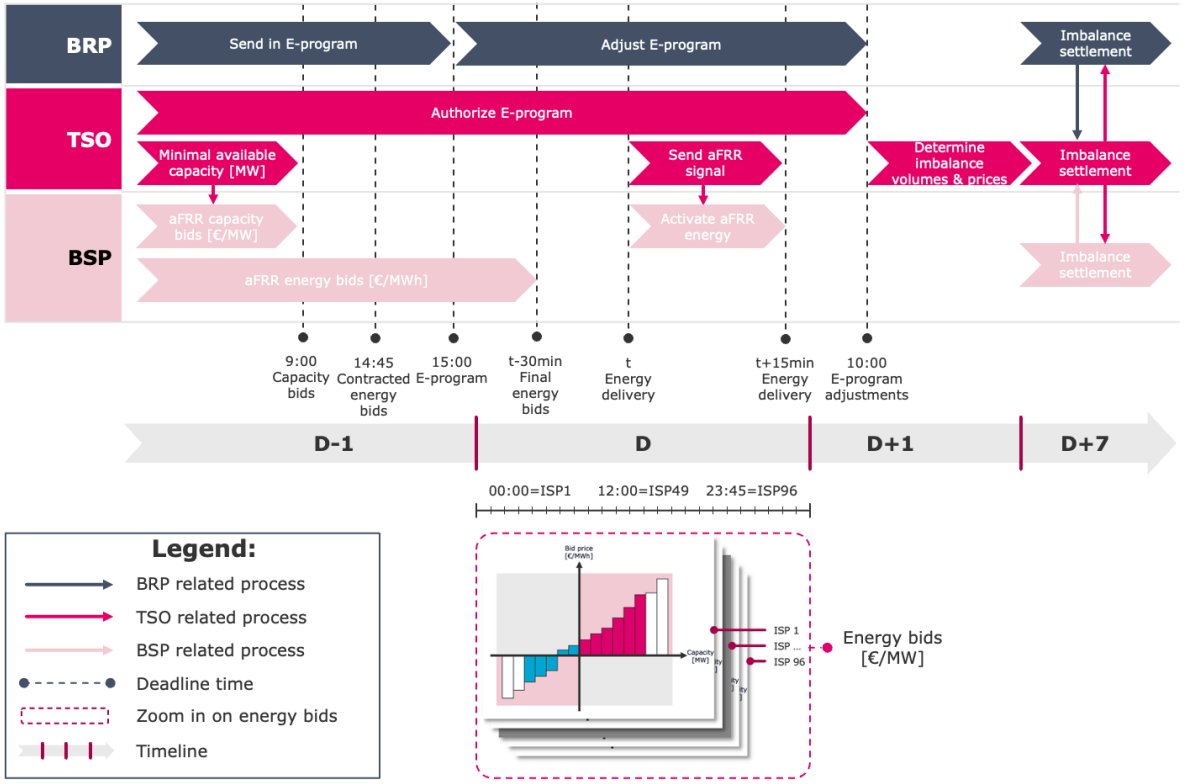


Figure 2.4: Timeline of the aFRR bidding process, including coordinated roles.

Now that the trading of electricity across successive markets from forward contracting to real-time balancing has been outlined, the final element to be addressed is the financial settlement of deviations between planned and actual delivery.

2.3. Imbalance Price Determination

When a BRP deviates from its submitted E-program, the resulting imbalance is settled financially through the imbalance price. This price is based on the activation price of aFRR market and reflects the system state (regulation state) [6]. The system state influences the applicable imbalance price, which varies depending on whether the BRP contributes to a shortage or surplus. To account for this, a distinction is made between active and passive imbalance. This depends on whether the TSO must intervene. If the deviation requires activation of balancing reserves to maintain system stability, it is referred to as active imbalance. If the system absorbs the deviation without intervention, the imbalance is considered passive. In both cases, the deviation is settled financially, but only active imbalance involves operational actions by the TSO. Table 2.1 summarises this distinction and clarifies how positive and negative signs are used for BSP and BRP actions [6].

Table 2.1: Sign convention for BSP active and BRP passive imbalance.

	BSP (Active imbalance)	**BRP (Passive imbalance)**
Positive (+)	Upward bids: BSP ensure an additional injection of electricity to the grid (e.g. CCGT increases its output to meet a sudden demand surge).	Surplus: BRP injects more electricity into the grid than in the submitted E-program (e.g. due to higher wind speeds than predicted).
Negative (-)	Downward bids: BSP ensures that extra electricity can be consumed from the grid (e.g. a wind or solar plant curtails its electricity generation output).	Shortage: BRP injects less energy into the grid than in the submitted E-program (e.g. an industrial consumer uses more energy than predicted).

aFRR Bid Ladder

The imbalance price is determined at the intersection point on the aFRR bid ladder, as illustrated in Figure 2.5. This forms the basis for the Dutch dual pricing mechanism, in which separate prices apply for upward and downward regulation [6]. As a result, financial settlements between the TSO and BRP depend on the type and direction of the imbalance. The upward price p^{up} is derived from the point where the required volume of upward regulation meets the upward bid ladder. Similarly, the downward price p^{down} is determined at the intersection of the needed downward regulation volume and the downward bid ladder. If no balancing reserves are activated, a mid-price p^{mid} is applied, calculated as the average of the highest downward and lowest upward bid. The financial direction of payments and the sign convention for BSP and BRP resulting from these prices is summarised in Table 2.1. A positive upward price leads to a payment from the TSO to the BSP, while a negative price implies a payment from the BSP to the TSO. For downward regulation, this relationship is reversed [6].

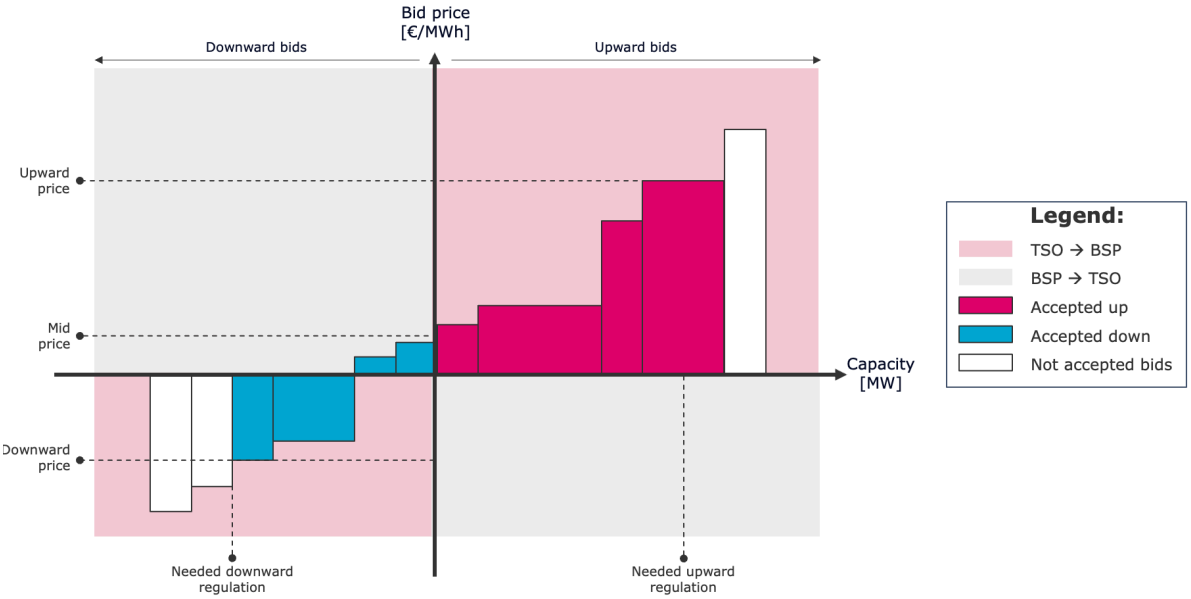


Figure 2.5: Upward and downward merit-order curves for aFRR energy bids with payment direction shown by background (light pink: TSO → BSP, grey: BSP → TSO). Accepted bids are marked as upward (dark pink), downward (blue), and non-accepted (white).

The bid ladder thus establishes the pricing framework, but the actual price applied in each settlement period depends on the system's real-time balancing condition, which is captured by the regulation state.

Regulation State

While the bid ladder defines the available price levels, the regulation state determines which of these prices applies in each market interval. The regulation state is a parameter that reflects the real-time balancing condition of the power system during each ISP. It indicates whether balancing energy is activated and in which direction. Based on this system condition, the applicable imbalance price is selected as either the upward price p^{up}, the downward price p^{down}, or the mid-price p^{mid}. Table 2.2 summarises how the regulation state guides this selection.

If no balancing energy is needed, the system is in regulation state 0. In that case, all BRP deviations are settled using the mid-price p^{mid}.

If only upward regulation is required during the ISP, or if the imbalance steadily increases, regulation state 1 applies. The imbalance price is then set to the upward price p^{up}.

If only downward regulation is needed, or the imbalance consistently decreases, regulation state minus one applies. The price used is the downward price p^{down}.

If regulation state 2 applies, both upward and downward regulation take place within the same ISP. This typically occurs under volatile system conditions where the direction of the imbalance changes during the 15-minute period. In such cases, relying directly on the actual activation prices p^{up} and p^{down} could result in extreme or inconsistent imbalance prices. This would reduce price stability and send distorted incentives to market participants to create even more imbalances in the system. To prevent this, a mechanism known as reverse pricing is applied. It is triggered when the upward regulation price is lower than the mid-price p^{mid}, or the downward regulation price is higher than the mid-price. In these cases, the mid-price ensures fair and stable settlement.

Table 2.2: Summary of regulation states and their corresponding pricing.

Regulation state	Description	Shortage price	Surplus price
0	No regulation either upward or downward (e.g. no electricity injected into or withdrawn from the grid).	p^{mid}	p^{mid}
1	Upward regulation is needed due to a shortage in the whole ISP.	p^{up}	p^{up}
-1	Downward regulation is needed due to a surplus in the whole ISP.	p^{down}	p^{down}
2	Both upward and downward regulations are needed within one ISP.	$\begin{cases} p^{\text{up}}, & \text{if } p^{\text{up}} \geq p^{\text{mid}} \\ p^{\text{mid}}, & \text{otherwise} \end{cases}$	$\begin{cases} p^{\text{down}}, & \text{if } p^{\text{down}} \leq p^{\text{mid}} \\ p^{\text{mid}}, & \text{otherwise} \end{cases}$

3

Literature Review

This chapter examines the theoretical foundation of the research methodology. To structure the existing knowledge, the literature is first categorised into four areas: DAM, IDM, IM, and bid curve studies. This classification provides a clearer understanding of the research field and serves as a basis for identifying the research gap. In response to this gap, three widely used forecasting models (LASSO, XGBOOST, and LSTM) are then introduced and explained.

3.1. Previous Work

Electricity price forecasting (EPF) represents the research domain focused on the quantitative prediction of electricity market prices to support operational and strategic decision-making. To bring structure to this diverse body of work, the existing literature is categorised according to market segment: DAM (Appendix B), IDM (Appendix B), IM (Appendix B), and bid curve forecasting (Appendix B). This categorisation is visualised in Figure 3.1, which presents a timeline of key contributions based on the historical overview in Appendix B. The figure groups studies according to their focus on point forecasts, probabilistic forecasts, or bid curve models.

In recent years, the ID and IM have received growing academic attention [23]. This shift is largely a response to increasing market volatility and the decentralisation of electricity generation. These developments have amplified the need for short-term flexibility, making accurate forecasts on shorter time scales more necessary for system stability and economic efficiency. At the same time, improved availability of high-frequency market and system data has enabled the development of more sophisticated models suited to the operational complexity of these markets [10].

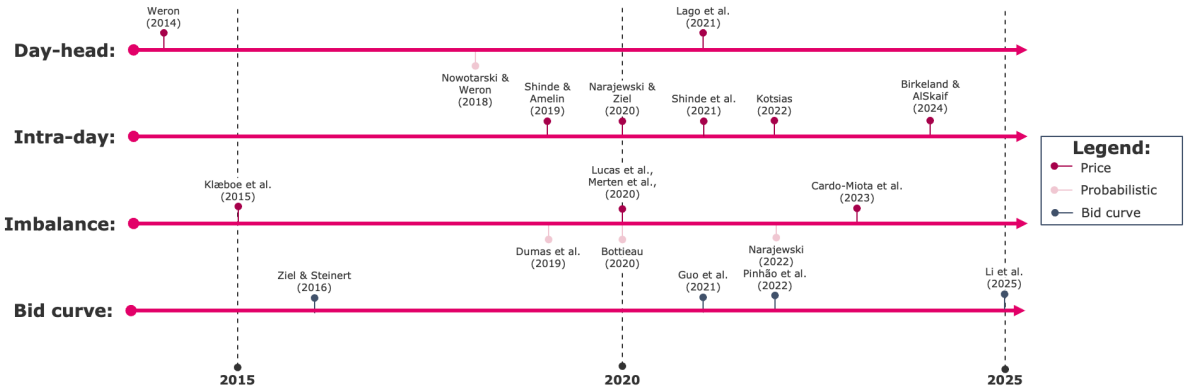


Figure 3.1: Electricity price forecasting (EPF) studies from 2014 to 2025, categorised by market and forecasting focus: price (dark pink), probabilistic (pink), and bid curve (blue).

Although prior research has addressed various market segments, several areas remain underexplored, motivating the present study.

3.2. Research Gap

The research gap is identified through four key observations: (1) the predominant focus of existing literature on DAM and IDM, (2) the absence of an established status quo for imbalance price forecasting, (3) limited application and investigation, specifically within the Dutch aFRR markets, and (4) a need for continuous evaluation as market dynamics and forecasting methodologies evolve.

To start, limited research has been devoted to supply curve forecasting, that consider both price and volume. The few existing studies on this topic have predominantly targeted the DAM [24, 25, 26], reflecting a broader research trend heavily focused on DAM [27, 28, 29] and IDM [30, 31, 32, 33, 34]. Extending forecasts to complete bid curves could yield valuable insights, specifically for understanding the Dutch imbalance market.

Secondly, no established benchmark or status quo exists for imbalance price forecasting. Although several studies have investigated imbalance price forecasts directly [12, 35] or explored forecasts related to aFRR prices [36, 37], none have definitively identified a consistently reliable forecasting model. However, probabilistic forecasting approaches have demonstrated promising results [38, 39].

Thirdly, existing aFRR forecasting studies have yet to be applied specifically to the Dutch electricity market. Prior research has primarily addressed other European markets, including the German [35, 36, 39, 23], Nordic [12], Belgian [38, 39], and Iberian markets [37]. The Dutch market has unique characteristics, such as high renewable energy penetration, specific regulations, and market integration. These aspects remain underexplored, highlighting the need for focused research.

Finally, the constantly evolving nature of electricity markets requires forecasting methods to continuously adapt. Over the past decade, forecasting methodologies have significantly progressed, driven by growing market complexity, data availability, and enhanced computational capabilities. Initially, statistical approaches such as ARIMA, ARX, and SARIMAX dominated the literature [27]. Subsequently, gradient Boosting algorithms such as XGBOOST shown considerable potential [26, 35, 39]. Most recently, deep learning architectures, including Convolutional Neural Networks (CNNs) [37], Recurrent Neural Networks (RNNs) [36], and Long Short-Term Memory networks (LSTM) [25, 37, 39], have recently gained prominence. This diverse evolution underscores the complexity of electricity price forecasting and highlights the necessity for continuous methodological evaluation and adaptation to maintain forecasting accuracy amid changing market conditions.

RQ2: *LASSO, XGBOOST, and LSTM emerge as commonly used and high-performing models for bid curve forecasting. Their effectiveness is further supported by their frequent application across other EPF domains, including the DAM, IDM, and IM. These three models represent distinct modelling paradigms: statistical (LASSO), gradient boosting (XGBOOST), and deep learning (LSTM). Their selection is therefore based not only on empirical performance but also on their suitability for comparing the strengths and weaknesses of different forecasting methodologies.*

3.3. Relevant Prediction Models

As outlined in Section 3.1, regression models, such as linear models, gradient boosting, and neural networks, have been consistently identified as effective approaches in EPF. In line with these findings, LASSO [29, 32, 23, 24], XGBOOST [35, 39, 26], and LSTM [37, 39, 25] are selected as representative methods, reflecting their recurrent application in the examined literature.

3.3.1. Linear Regression

Linear regression models the linear relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data. The technique aims to find the best-fitting straight line through a set of data points by minimising the sum of squared residuals, which represents the difference between predicted and actual values.

LASSO
Least Absolute Shrinkage and Selection Operator (LASSO) is a type of linear regression that provides two main benefits: regularisation and variable selection [40]. LASSO addresses key drawbacks of traditional linear regression, particularly in situations that involve high-dimensional data where standard approaches are prone to overfitting [40]. Equation 3.1 shows the formula for the LASSO model, where the estimator is the residual sum of squares (RSS) plus a penalty term. The RSS measures the sum of the squared differences between the observed value $\boldsymbol{y}$ and the linear model $\boldsymbol{X}\boldsymbol{\beta}$. The penalty term $\lambda_{\text{LASSO}}\|\boldsymbol{\beta}\|_1$ applies L_1 regularisation, which promotes sparsity by shrinking some coefficients to zero. A higher value of λ_{LASSO} increases this effect, resulting in a simpler, faster model that automatically selects the most influential features [40]. The squared L_2-norm (sum of squares of components) is denoted by $\|\cdot\|_2^2$, and $\|\cdot\|_1$ denotes the L_1-norm (sum of absolute values of components).

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \left\{ \underbrace{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2}_{\text{RSS}} + \underbrace{\lambda_{\text{LASSO}}\|\boldsymbol{\beta}\|_1}_{\text{Penalty}} \right\} \tag{3.1}$$

$\hat{\boldsymbol{\beta}}$: Vector of estimated regression coefficients $\in \mathbb{R}^M$
$\boldsymbol{y}$: Vector of observed target values $\in \mathbb{R}^N$
$\boldsymbol{X}$: Matrix of features $\in \mathbb{R}^{N \times M}$
$\boldsymbol{\beta}$: Vector of regression coefficients $\in \mathbb{R}^M$
λ_{LASSO} : Regularisation penalty parameter controlling sparsity
N : Dimension of amount of samples
M : Dimension of amount of features

3.3.2. Gradient Boosting Regression

Gradient boosting regression is a machine learning technique that builds a predictive model by sequentially creating decision trees that combine them to create a predictive framework. Unlike linear regression, gradient boosting can capture complex, non-linear relationships in data and is particularly effective for handling high-dimensional datasets with intricate interactions between variables.

XGBOOST
eXtreme Gradient Boosting (XGBOOST) is an advanced implementation of gradient boosting that enhances computational efficiency, predictive accuracy, and model generalisation through features like regularisation, parallel processing, and optimised tree construction techniques [41]. The dataset is represented as

$$\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)\} = \{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N} \tag{3.2}$$

$\boldsymbol{x}_n$: Feature vector $\in \mathbb{R}^M$
y_n : True target value $\in \mathbb{R}$
n : Index of the training sample $n \in \{1, \ldots, N\}$

A decision tree operates by passing an input from the root node through decision nodes to a leaf node. This is described as a mapping function $q(\boldsymbol{x})$ that returns the weight (score) $w_{q(\boldsymbol{x})}$ of that leaf. This is schematically explained in Figure 3.2.

$$f(\boldsymbol{x}) = w_{q(\boldsymbol{x})} \tag{3.3}$$

$f(\boldsymbol{x})$: Prediction function for input $\boldsymbol{x}$
$\boldsymbol{x}$: Input vector $\in \mathbb{R}^M$
$w_{q(\boldsymbol{x})}$: Prediction score for input $\boldsymbol{x}$
$q(\boldsymbol{x})$: Leaf assignment function (e.g. decision rules), $q : \mathbb{R}^M \to \{1, \ldots, L\}$
L : Total number of leaves in a tree

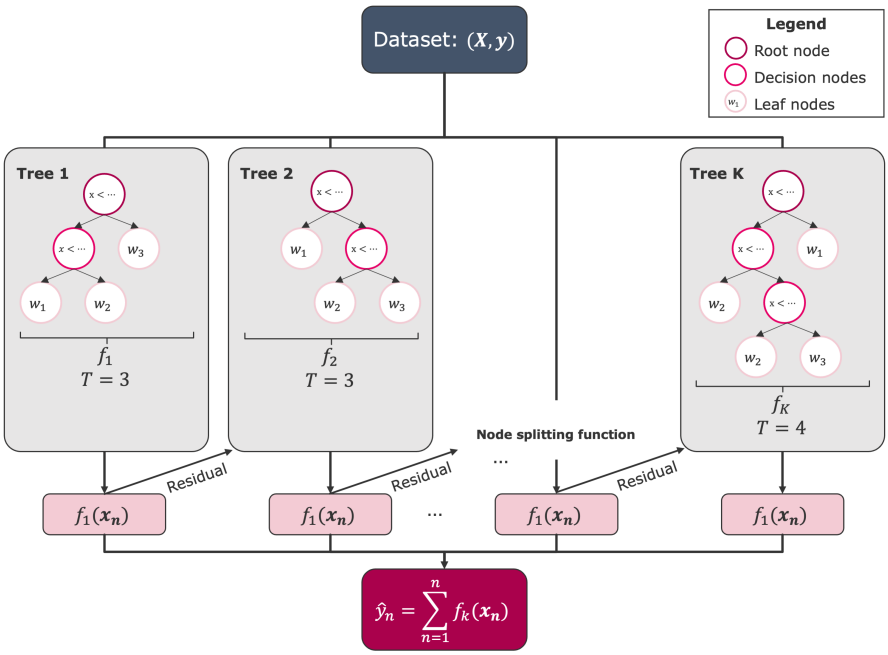


Figure 3.2: Schematic of a decision tree structure used in XGBOOST.

XGBOOST builds an ensemble of multiple (regression) decision trees. The final prediction $\hat{y}_n$ is computed as the sum of the total number of trees K

$$\hat{y}_n = \sum_{k=1}^{K} f_k(\boldsymbol{x}_n) \tag{3.4}$$

$\hat{y}_n$: Predicted value $\in \mathbb{R}$
$f_k(\boldsymbol{x}_n)$: Function where each k represents a separate decision tree structure $\in \mathcal{F}$
$\mathcal{F}$: Functional space of all regression trees
k : Index of the k-th number of trees $k \in \{1, \ldots, K\}$
K : Total number of trees (additive functions) in the ensemble

The objective of XGBOOST is to minimise the objective function $\mathcal{L}^i$ plus a regularisation term $\Omega(f_k)$ for each iteration

$$\mathcal{L}^i = \underbrace{\sum_{n=1}^{N} l(y_n, \hat{y}_n^i)}_{Loss} + \underbrace{\sum_{k=1}^{L} \Omega(f_k)}_{Tree\ complexity} \tag{3.5}$$

$\mathcal{L}^i$: Total loss at iteration i
$l(y_n, \hat{y}_n^i)$: Differentiable loss function
$\hat{y}_n^i$: Predicted value for i-th iteration $\in \mathbb{R}$
$\Omega(f_k)$: Regularisation term for each decision tree

The complexity of the tree structure is penalised through the regularisation parameters γ and λ_{XGB}, which control the tree size

$$\Omega(f) = \gamma L + \frac{1}{2}\lambda_{\text{XGB}} \|\boldsymbol{w}\|_2^2 \tag{3.6}$$

$\Omega(f)$: Regularisation term that penalizes model complexity
γ : Regularisation penalty parameter for each additional leaf in the tree
λ_{XGB} : Regularisation penalty parameter on the amount of leaf weights
$\|\boldsymbol{w}\|_2^2 = \sum_{j=1}^{L} w_j^2$: Squared $L2$-norm of the leaf weights
w_j : Leaf weight (score) on j-th index $\in \mathbb{R}^L$
j : Index of leaf score $\in \{1, \dots, L\}$

At each iteration i, XGBOOST adds a new function (tree) $f_i(\boldsymbol{x}_n)$ to improve the model and minimise the objective. Since the space of possible tree structures is discrete and non-differentiable, the loss function is approximated using a second-order Taylor expansion [41]. The gradient terms are defined as

$$g_n = \frac{\partial}{\partial \hat{y}_n^{i-1}} l(y_n, \hat{y}_n^{i-1}), \quad h_n = \frac{\partial^2}{\partial (\hat{y}_n^{i-1})^2} l(y_n, \hat{y}_n^{i-1}) \tag{3.7}$$

g_n : First-order gradient of the loss w.r.t. $\hat{y}_n^{i-1}$
h_n : Second-order gradient (Hessian) of the loss w.r.t. $\hat{y}_n^{i-1}$

The objective function is then approximated as follows

$$\begin{aligned}
\mathcal{L}^i &= \sum_{n=1}^{N} l(y_n, \hat{y}_n^{i-1} + f_i(\boldsymbol{x}_n)) + \Omega(f_i) \\
&\approx \sum_{n=1}^{N} \left(l(y_n, \hat{y}_n^{i-1}) + g_n f_i(\boldsymbol{x}_n) + \frac{1}{2} h_n f_i^2(\boldsymbol{x}_n) \right) + \Omega(f_i) \\
\tilde{\mathcal{L}}^i &= \sum_{n=1}^{N} \left(g_n f_i(\boldsymbol{x}_n) + \frac{1}{2} h_n f_i^2(\boldsymbol{x}_n) \right) + \Omega(f_i) \\
&= \sum_{n=1}^{N} \left(g_n f_i(\boldsymbol{x}_n) + \frac{1}{2} h_n f_i^2(\boldsymbol{x}_n) \right) + \gamma L + \frac{\lambda_{\text{XGB}}}{2} \sum_{j=1}^{L} w_j^2 \\
&= \sum_{j=1}^{L} \left(\left(\sum_{n \in I_j} g_n \right) w_j + \frac{1}{2} \left(\sum_{n \in I_j} h_n + \lambda_{\text{XGB}} \right) w_j^2 \right) + \gamma L
\end{aligned} \tag{3.8}$$

$\hat{y}_n^{i-1}$: Prediction of the n-th instance at iteration $i-1$
$f_i(\boldsymbol{x}_n)$: New function (tree) added at iteration i
I_j : Set of instance indices assigned to leaf j

Minimising the approximated loss with respect to w_j yields the optimal weight

$$w_j^* = -\frac{\sum_{n \in I_j} g_n}{\sum_{n \in I_j} h_n + \lambda_{\text{XGB}}} \tag{3.9}$$

Each tree structure can then be scored by substituting this into the loss

$$\tilde{\mathcal{L}}^i(q) = -\frac{1}{2} \sum_{j=1}^{L} \left(\frac{\sum_{n \in I_j} g_n}{\sum_{n \in I_j} h_n + \lambda_{\text{XGB}}} \right)^2 + \gamma L \tag{3.10}$$

In practice, enumerating all possible tree structures is computationally infeasible. Instead, trees are constructed greedily: beginning from the root, candidate splits are evaluated, and a split is added only if it improves the objective function. This process continues recursively. The quality of a split is quantified by its gain, which measures the reduction in the loss after the split. If the gain is negative, the split increases the regularised loss and is therefore not performed. This automatic rejection of unprofitable splits is referred to as ”pruning” [41], and it serves to regularise the model and prevent overfitting.

3.3.3. Deep Learning Regression

Deep learning models perform regression by learning mappings from input variables to continuous numerical outputs using neural networks [42]. A typical neural network consists of an input layer, one or more hidden layers, and an output layer. The input layer receives input data, which is passed through the hidden layers where features are extracted and transformed. Each hidden layer comprises neurons that compute weighted sums of their inputs and apply activation functions such as the Rectified Linear Unit (ReLU) or Sigmoid to introduce non-linearity. The final output layer produces a continuous prediction. Learning occurs through an iterative process involving forward and backward passes [42]. In the forward pass, inputs propagate through the network to generate a prediction, which is then compared to the target value using a loss function, such as the mean squared error. In the backward pass, gradients of the loss with respect to model parameters are computed using the chain rule and propagated backward through the network. These gradients are used to update the weights via an optimisation algorithm, commonly stochastic gradient descent [42]. This procedure is repeated over many iterations, allowing the model to progressively minimise the loss and improve its predictive accuracy

LSTM

A Long Short-Term Memory (LSTM) network is a type of recurrent neural network (RNN) designed to store and remember information over extended periods efficiently [43]. It addresses the vanishing gradient problem commonly faced in traditional RNNs, which makes it effective for tasks involving sequential data, such as time series forecasting [43]. A visual representation of an LSTM cell sequence can be found in Figure 3.3. The core idea of an LSTM is that information flows through two distinct pathways over time ($n-1, n, n+1$).

- Long-term memory (cell state C): Long-term trends, seasonality, and other persistent patterns in the data are maintained by the cell state, which ensures that information from earlier time steps is preserved throughout the sequence.
- Short-term memory (hidden state h): Short-term fluctuations and context-specific details are captured by the hidden state that provides output required to predict the next value in the series.

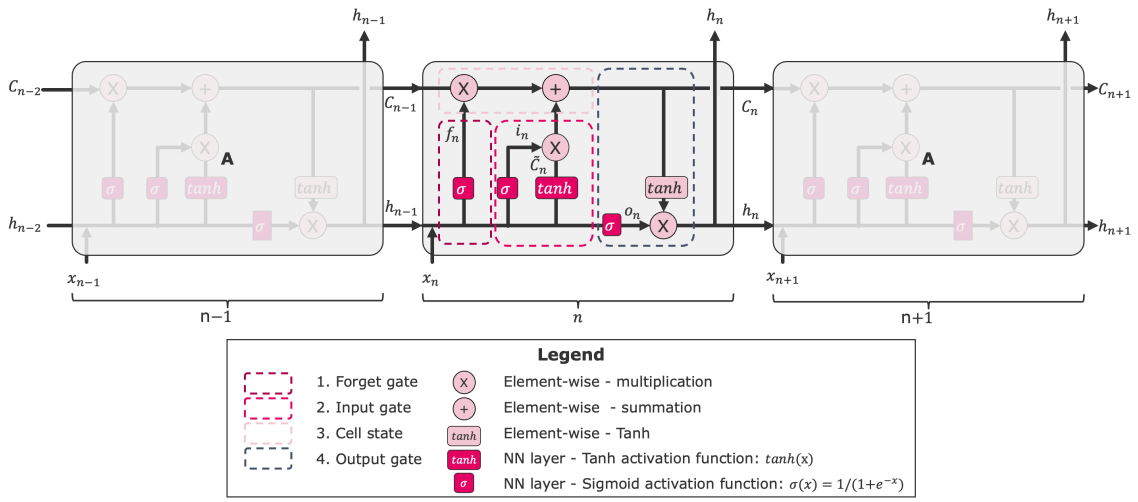


Figure 3.3: Schematic of a LSTM cell, detailing the internal flow of information across time steps. The diagram highlights key components: forget gate (red dashed), input gate (pink dashed), cell state (black dashed), and output gate (grey dashed), along with activation functions (sigmoid and tanh) and element-wise operations controlling memory retention and output generation.

An LSTM cell consists of four essential components that regulate information flow. The architecture contains three gates that control how information passes through the cell via element-wise multiplication and activation functions. These activation functions include the sigmoid function $\sigma(x) = 1/(1+e^{-x})$ and the hyperbolic tangent function $tanh(x)$. Information flows through these gates to the cell state. The inputs to these gates are formed by concatenating the previous hidden state and the current input, denoted as $[\boldsymbol{h}_{n-1}, \boldsymbol{x}_n]$.

1. Forget Gate: Equation 3.11 illustrates the forget gate, which regulates the information from the previous cell state that should be forgotten ($\boldsymbol{f}_n$)

$$\boldsymbol{f}_n = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{n-1}, \boldsymbol{x}_n] + \boldsymbol{b}_f) \tag{3.11}$$

$\boldsymbol{f}_n$: Forget gate output $\in \mathbb{R}^N$
$\boldsymbol{h}_{n-1}$: Previous hidden state $\in \mathbb{R}^N$
$\boldsymbol{x}_n$: Current input $\in \mathbb{R}^M$
$\boldsymbol{W}_f$: Forget gate weights $\in \mathbb{R}^{N \times (N+M)}$
$\boldsymbol{b}_f$: Forget gate bias, $\boldsymbol{b}_f \in \mathbb{R}^N$

2. Input Gate: The input gate determines how much new information should be added to the memory cell (Equation 3.12). The candidate cell state is calculated using Equation 3.13

$$\boldsymbol{i}_n = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{n-1}, \boldsymbol{x}_n] + \boldsymbol{b}_i) \tag{3.12}$$

$$\tilde{\boldsymbol{C}}_n = \tanh(\boldsymbol{W}_c \cdot [\boldsymbol{h}_{n-1}, \boldsymbol{x}_n] + \boldsymbol{b}_c) \tag{3.13}$$

$\boldsymbol{i}_n$: Input gate output $\in \mathbb{R}^N$
$\tilde{\boldsymbol{C}}_n$: Candidate cell state $\in \mathbb{R}^N$
$\boldsymbol{W}_i$: Input gate weights $\in \mathbb{R}^{N \times (N+M)}$
$\boldsymbol{b}_i$: Input gate bias $\in \mathbb{R}^N$
$\boldsymbol{W}_c$: Candidate state weights $\in \mathbb{R}^{N \times (N+M)}$
$\boldsymbol{b}_c$: Candidate state bias $\in \mathbb{R}^N$

3. Cell State: Equation 3.14 shows the cell state maintaining the long-term memory by combining past information ($\boldsymbol{f}_n * \boldsymbol{C}_{n-1}$) with new updates ($\boldsymbol{i}_n * \tilde{\boldsymbol{C}}_n$)

$$\boldsymbol{C}_n = \boldsymbol{f}_n * \boldsymbol{C}_{n-1} + \boldsymbol{i}_n * \tilde{\boldsymbol{C}}_n \tag{3.14}$$

$\boldsymbol{C}_{n-1}$: Previous cell state $\in \mathbb{R}^N$
$\boldsymbol{C}_n$: Updated cell state $\in \mathbb{R}^N$

4. Output Gate: The output gate determines how much of the updated cell state should be exposed as output, calculated using Equation 3.15. The hidden state is obtained using Equation 3.16

$$\boldsymbol{o}_n = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{n-1}, \boldsymbol{x}_n] + \boldsymbol{b}_o) \tag{3.15}$$

$$\boldsymbol{h}_n = \boldsymbol{o}_n * \tanh(\boldsymbol{C}_n) \tag{3.16}$$

$\boldsymbol{o}_n$: Output gate output $\in \mathbb{R}^N$
$\boldsymbol{h}_n$: Hidden state/output $\in \mathbb{R}^N$
$\boldsymbol{W}_o$: Output gate weights $\in \mathbb{R}^{N \times (N+M)}$
$\boldsymbol{b}_o$: Output gate bias $\in \mathbb{R}^N$

4

Data and Feature Engineering

This chapter provides a detailed analysis of the dataset utilised in this study. It begins by introducing the data sources, which form the basis for the subsequent examination of the bid curve dataset's structure. Building on this structural analysis, bid ladder prices and volumes are then analysed to identify the exogenous factors that influence bidding behaviour.

4.1. Data Sources

The study integrates multiple raw data sources covering the period from January—May 2025. While longer training periods are generally advised [29], this relatively short timeframe is chosen due to two main considerations: data availability and market developments. Specifically, preliminary bid ladders from TenneT are only available from January 2025 onward [44], inherently limiting the length of the training period. In addition, the introduction of PICASSO in October 2024 [9] fundamentally changed balancing market dynamics by enabling cross-border energy exchange. As a result, the selected period better reflects the current structure of the market.

Based on the available data, features are extracted and grouped into five categories: weather, capacity, market, time, and preliminary bid ladders. These features form the basis for three distinct feature vectors used in the models, as summarised in Table 4.1. The selection of features is informed by existing literature [25], empirical data analysis, and domain-specific market knowledge [45].

Table 4.1: Input features for upward and downward bidding models, including units, resolution, and data sources.

Category	Data	Unit	Resolution	Feature Vector		Source
				Upward	Downward	
Weather	Temperature forecast	[°C]	[15min]	x	x	[45]
	Wind speed forecast	[m/s]	[15min]	x	x	[45]
	Irradiance forecast	[W/m²]	[15min]		x	[45]
Capacity	Coal forecast	[MW]	[15min]	x	x	[46]
	Gas forecast	[MW]	[15min]	x	x	[46]
	Wind forecast	[MW]	[15min]	x	x	[46]
	Solar forecast	[MW]	[15min]	x	x	[46]
	Total load forecast	[MW]	[15min]	x	x	[46]
	Spinning reserve upward	[MW]	[15min]	x		[46]
	Spinning reserve downward	[MW]	[15min]		x	[46]
	Required aFRR upward	[MW]	[15min]	x		[46]
	Required aFRR downward	[MW]	[15min]		x	[46]
Market	Day-ahead price	[€/MWh]	[hourly]	x	x	[46]
	ID3 price	[€/MWh]	[15min]	x	x	[46]
	Carbon price	[€/MWh]	[15min]	x	x	[46]
	Marginal price CCGT	[€/MWh]	[daily]	x	x	[46]
Time	Day of week	[0-6]	[15min]	x	x	[45]
	Hour of day	[sin]	[15min]	x	x	[45]
Preliminary bid ladder	Provisional bids (t-3h)	[€/MWh]	[15min]	x	x	[47]

The data used in this study originates from three main sources: TenneT, Essent, and Enappsys, each providing different types of input relevant to the feature set:

- **TenneT:** Data is retrieved through their API, which provides access to the merit order list [47].
- **Essent:** Internal data from Essent includes proprietary analyses and forecasts, primarily related to weather conditions and energy demand [45].
- **Enappsys:** Is an online energy data platform offering comprehensive market and power system information as the primary data source. It includes data on ancillary services, market prices, and power forecasting [46].

4.2. Exploratory Data Analysis

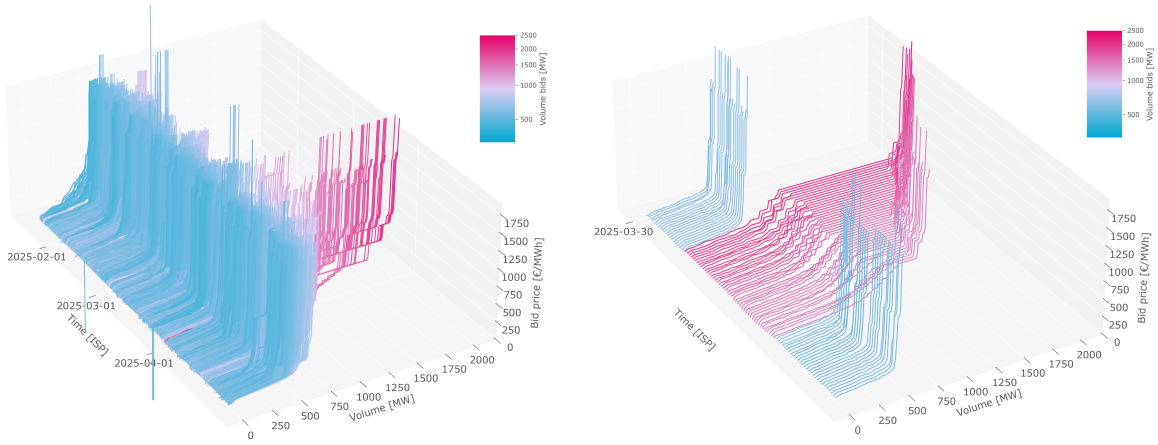
Building on the description of the data sources, the subsequent section first explores the structural patterns observed in aFRR bid curves, providing a foundation for the analysis of how external factors influence their shape. The insights gained here are further complemented by additional analyses on bid ladder statistics, distributions, correlations, and preliminary bid ladders, which are presented in Appendix C.

4.2.1. Bid Curve Structure

To analyse aFRR bid ladders, the output data are first examined to characterise the temporal structure of the combined volumes and prices. This forms the basis for exploring the underlying dynamics of the bid curves to be predicted in this study. The 2D kernel density estimation (KDE), presented in Appendix Figure C.1, illustrates the distribution of placed bids and their associated volumes, aggregated in 10 MW steps, to visualise the bidding behaviour.

Figures 4.1 and 4.2 present the full set of aFRR bid curves over time, plotted at 15-minute granularity. Each curve corresponds to a single time step, with bid volumes binned in 10 MW intervals. The x-axis represents time, the y-axis volume in MW, and the z-axis bid price in €/MWh. Bid volumes are visualised through a colourmap, where pink indicates higher volumes and blue lower volumes. To compare both markets, downward bids (Figure 4.2) are plotted with inverted price axes, aligning them visually with upward bids (Figure 4.1). This mirroring enables clearer visual comparison of curve steepness, volume spread, and price levels. Given the dataset's granularity, Figures 4.1b and 4.2b offer a more detailed view for May 30, 2025, a particularly active period with noticeable fluctuations in aFRR bidding across both regimes.

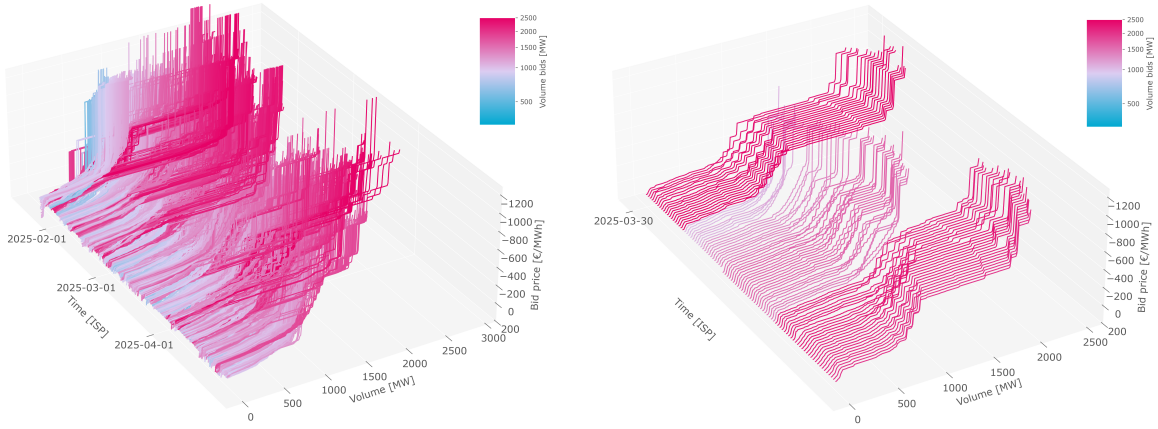
In the upward market, bid curves generally exhibit a smoother and stable shape across the selected days. Volumes increase gradually, and price levels in the lower-volume segments remain relatively linear, fluctuating around 100–150 €/MWh. Only in the higher-volume tail ends of the curves are sharp price increases observed, reaching up to 1900 €/MWh. Usually only a few 10 MW of these spikes are present per curve.



(a) Full view. **(b)** Zoomed-in view.

Figure 4.1: 3D visualisation of upward aFRR bid curves over time. Colour gradients indicate the total bid volume.

By contrast, the downward market is characterised by visibly larger bid volumes and significantly more volatile volume. This is particularly evident in the saturated pink regions of the curves, indicating frequent high-volume bids. Price levels in the lower-volume regions fluctuate between positive and negative values, typically ranging from 100 €/MWh to −100 €/MWh. Toward the higher end of the volume range, bid prices drop sharply, often reaching values as low as −1300 €/MWh. This behaviour suggests the presence of highly flexible participants willing to pay to reduce output or absorb excess energy, and reflects a more aggressive and opportunistic bidding strategy compared to the upward market.



(a) Full view.

(b) Zoomed-in view.

Figure 4.2: 3D visualisation of downward aFRR bid curves over time. Colour gradients indicate the total bid volume.

While the exploratory 3D visualisations provided initial insight into the structure of aFRR bid curves over time. Building on these findings, the analysis investigates how external factors explain observed patterns in bid volumes and prices. Because combining multiple feature effects in a single 3D view is impractical, the analysis is divided into two steps: first, the total bid volumes are examined, followed by an analysis of the individual bid prices.

4.2.2. Bid Volume Drivers

This subsection investigates which external features influence total aFRR bid volumes. It begins with an analysis of historical trends and distributions, followed by a correlation assessment between selected input features (Table 4.1) and both upward and downward bid volumes.

Bid Volume Over Time

Figure 4.3 illustrates the total bid volume over time. The pink line represents the upward bid volumes, while the blue line represents the downward bid volumes. The solid black lines denote the minimal required volumes for both upward and downward regulation. This minimal required capacity is determined daily and regulated by the capacity bids, as described in Subsection 2.2.4 and illustrated in Figure 2.4 [21].

The upward bid volumes generally remain close to the required volume, with only occasional spikes (e.g., at the end of March). This limited deviation indicates that the available upward flexibility is constrained, primarily because only a limited number of assets, such as CCGTs, are capable of rapidly increasing their output. This observation is consistent with the findings of [48], who also report that upward balancing capacity is often scarce due to the limited availability of suitable generation units.

In contrast, downward bid volumes often exceed required capacity, reflecting greater availability and participation. The higher volumes, particularly during daytime, suggest that more assets such as solar and other renewables are available to curtail output or absorb excess electricity. This operational variety likely explains the broader supply, consistent with [48], who highlight the wider availability and strategic diversity in the downward market.

Figure 4.3: Time series of aFRR bid volumes and required volumes. Upward (pink) and downward (blue) bid volumes are shown alongside the corresponding required volumes (black lines).

While total bid volume reveals the overall market activity, it does not capture how this volume is distributed across 15-minute periods. Analysing the density of bid volumes provides further insight into typical bidding behaviour and the variability within each market segment.

Bid Volume Distribution

Figure 4.4 shows the KDE of total aFRR bid volumes across the entire dataset. The pink line represents the distribution of upward bid volumes, while the blue line represents the downward bid volumes. The vertical dashed lines indicate the highest observed bid volumes in each direction: approximately 2500 MW for upward and −3500 MW for downward. Overall, both distributions are skewed, but the downward bids exhibit a wider spread and heavier tail, further supporting the observed structural differences between the two markets.

The upward bid distribution is narrowly concentrated, with a sharp peak around its typical bid level. This suggests that upward volumes are relatively consistent, with limited variability across the dataset. Most bids are close to the average minimal required volume of Figure 4.3, and extreme values over 1000 MW are often not present.

The downward distribution is significantly broader and flatter, indicating much greater variability. This reflects a higher degree of flexibility and availability in the downward market, where a larger number of bids are placed at varying volumes

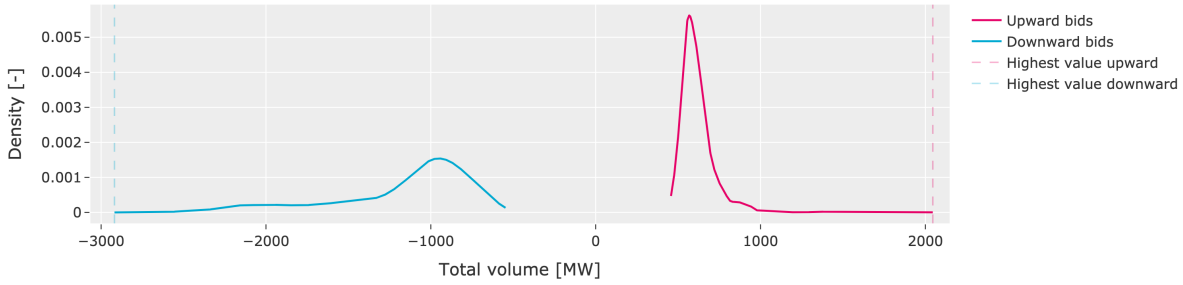


Figure 4.4: KDE distribution of total aFRR bid volumes for upward (pink) and downward (blue) directions, with dashed lines indicating maximum values.

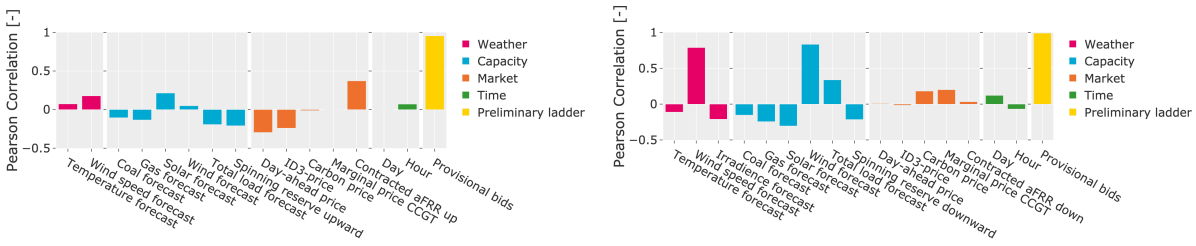
While density estimation reveals how bid volumes are typically distributed, it does not explain what drives these patterns. To better understand the underlying factors influencing bidding behaviour, the analysis next examines correlations between explanatory variables and total bid volume.

Bid Volume Correlation

Figure 4.5 shows the Pearson correlation between total bid volume and selected input features from Table 4.1, for both upward and downward regulation. Correlation values near 1 or −1 indicate strong linear relationships. Features are grouped into five categories: weather, capacity, market, calendar, and preliminary ladder, each shown in a different colour.

In the upward direction, most features show weak linear relationships with total bid volume. Preliminary ladder volume is the only variable with a notable positive correlation, suggesting earlier bids indicate final volume. The minimal available aFRR capacity also exhibits a moderate correlation. Other categories, including weather, market, and calendar features, show low correlation, implying upward volume is less driven by short-term external signals and more by internal strategies or constraints.

In contrast, downward bid volume shows stronger and more varied correlations. Preliminary ladder volumes again show a strong positive relationship, confirming their predictive value. Minimal available aFRR capacity shows only limited correlation. Most notably, wind-related features (e.g. wind speed and wind generation forecasts) exhibit high positive correlations, indicating that wind availability plays a significant role in shaping downward flexibility. Solar generation forecasts also show moderate correlation, reinforcing the idea that renewable curtailment potential drives downward bidding.



(a) Upward Volume.

(b) Downward Volume.

Figure 4.5: Pearson correlation between features and total bid volume. Features are grouped by category: weather (pink), capacity (blue), market (orange), time (green), and preliminary bid ladders (yellow).

RQ1: *Preliminary ladder volumes are the main drivers for both upward and downward aFRR bid volumes, underscoring the predictive value of early bidding activity. While other external variables show limited direct influence on total volume, downward bids exhibit greater sensitivity to exogenous factors, particularly wind and solar generation forecasts. These findings indicate that upward volumes are relatively stable and determined by the capacity market, whereas downward volumes are more flexible and closely linked to fluctuations in renewable energy supply.*

After examining bid volumes, the analysis turns to bid prices to explore the levels at which participants place their offers. The aim is to identify the main factors influencing price levels across the bid ladder.

4.2.3. Bid Price Drivers

This subsection examines how market participants set bid prices by analysing historical data across fixed volume segments, or bins. Price distributions and feature correlations are used to identify key drivers of upward and downward bidding behaviour.

Bid Price Over Time

Bid prices are examined by tracking their evolution at fixed volume levels on the bid ladder. Each bin represents the time series of prices associated with a constant bid volume. Figure 4.6 shows price trajectories for selected bins from −1000 to 1000 MW and the maximum volume bin.

For upward regulation, maximum bids often reached 1900 €/MWh. Prices in lower volume bins (<250 MW) remained relatively stable, while sharp increases appeared at higher volumes. This reflects large price differences across the curve, indicating greater uncertainty with increasing volume. Table C.1 supports this, showing higher standard deviations at larger volumes.

For downward regulation, maximum prices typically reached around −1300 €/MWh. Bids included both positive and negative values, reflecting flexibility to either pay or be paid. Like upward bids, lower bins showed stable prices, while higher volumes exhibited greater variance. This aligns with the wide price spreads and rising standard deviations in Table C.1.

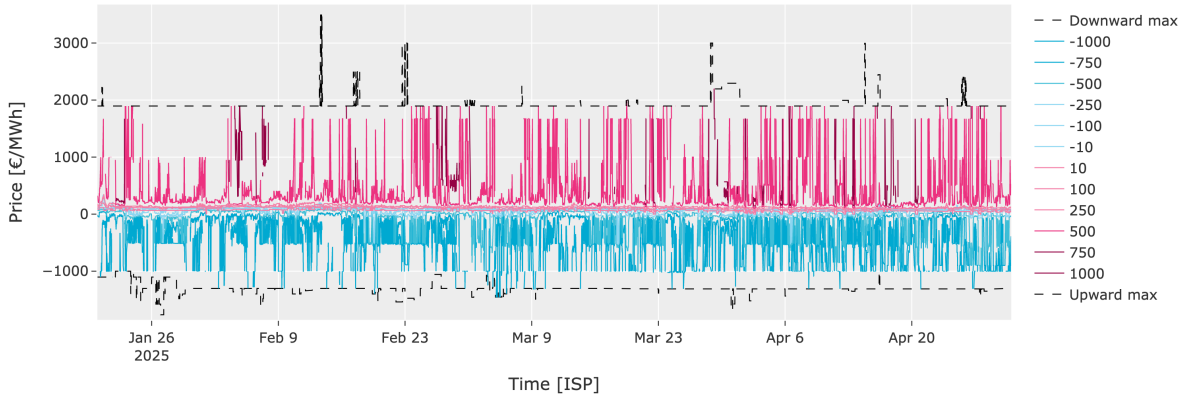


Figure 4.6: Time series of aFRR bid bins per direction, with upward (pink) and downward (blue) bids.

While the time series of price bins reveals how bid prices evolve across volumes, it does not capture how frequently individual price levels occur across the dataset.

Bid Price Distribution

To examine the distribution of all placed bids a KDE plot is constructed. Figure 4.7 shows the density of each price level in €/MWh for the upward and downward curves.

For upward regulation, prices are predominantly concentrated around 130 €/MWh, corresponding to the marginal costs of CCGT's [46]. The KDE curve also reveals additional peaks between 1600–1900 €/MWh, suggesting the presence of strategic bids at higher price levels. The maximum observed upward bid reached approximately 3500 €/MWh. Overall, upward bids are largely driven by marginal CCGT costs and strategic bidding behaviour, reflecting more structured pricing dynamics.

In contrast, the distribution of downward regulation bids is more dispersed, with several peaks around 80, −30, −500, and −1000 €/MWh. These multiple price peaks indicate that bids are placed around several preferred price levels. The maximum downward bid reached approximately −1750 €/MWh. The broader spread of downward bids reflects greater flexibility and variability compared to upward regulation, suggesting more heterogeneous bidding strategies.

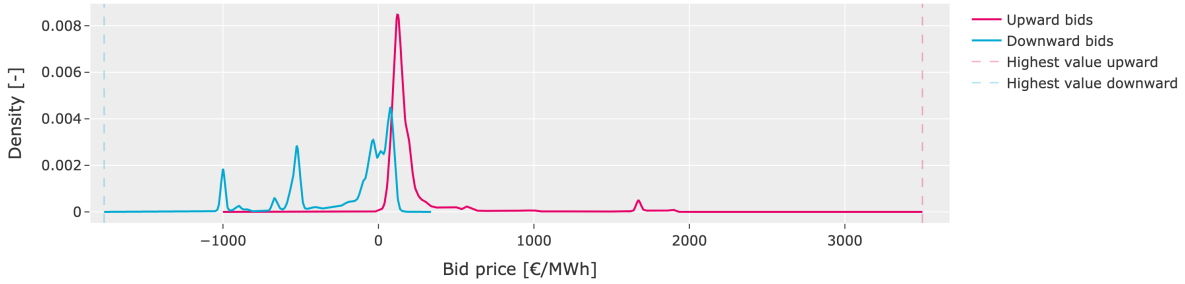


Figure 4.7: KDE of aFRR bid prices for upward (pink) and downward (blue). Maximum values are marked with dashed lines.

While the distribution of bid prices shows the range and frequency of price levels, it does not reveal which parts of the curve are affected by external factors. Examining correlations with these features helps identify their impact across different volume bins.

Bid Price Correlation

Figures 4.8a and 4.8b show the correlation between forecasted features, market variables, and bid prices across volume bins. Each bubble represents the Pearson correlation coefficient between a feature and bid price in a bin. Bubble size indicates correlation strength, and color shows sign: blue for positive, red for negative. Interpretation within a single bin must be cautious, as bid price is influenced not only by local features but also by overall shape and total volume of the bid ladder.

For upward regulation, fossil-based features such as coal, gas, and carbon price forecasts dominate the correlation patterns, especially in the lower volume bins (10–400 MW). Strong positive correlations with provisional bid ladders indicate a relationship between pre-positioned bids and clearing prices in smaller bins. However, as the volume increases, the influence of fossil-based forecasts diminishes, and other market dynamics, such as DAM prices and renewable forecasts, become more relevant. This shift suggests that in upward regulation, lower bins are primarily driven by generation costs.

For downward regulation, the correlation structure appears more consistent across volume bins. Provisional bid ladders continue to show strong positive correlations, particularly at lower volumes. Fossil-based features also play a role, but to a lesser extent compared to upward regulation. Renewable forecasts, such as wind and solar production, exhibit notable negative correlations, indicating that higher expected renewable output leads to lower downward bid prices.

(a) Upward price correlation.

(b) Downward price correlation.

Figure 4.8: Bubble chart showing correlations between explanatory features and binned aFRR bid prices, with bubble size indicating magnitude and color representing direction (blue: positive, pink: negative).

RQ1: *The analysis indicates that, in upward regulation, lower-volume prices are primarily driven by the marginal costs of fossil-based sources, including carbon prices, coal marginal prices, and CCGT costs. As the volume increases, these cost-based signals diminish, and prices become dominated by a few static high bids, leading to sharp price spikes near the curve's end. In contrast, downward regulation exhibits a broader price distribution shaped by the variability in renewable forecasts, resulting in greater price dispersion. Despite these structural differences, both markets show relatively stable price behaviour in their initial curve segments, with extreme prices emerging only toward the tail end. However, the exact location and magnitude of these higher-volume bids are difficult to forecast. This is due to two key observations. First, the placement of extreme bids varies over time, as indicated by rising standard deviations across volume bins. Second, correlation analysis shows that high-volume prices are weakly related to the input features, suggesting these bids are driven by unobserved factors or strategic behaviour.*

Based on all considered features, those with the highest correlation are selected and are presented in Table 4.1. The exploratory data analysis indicated minimal linear relationships across the dataset, with all examined features detailed in Appendix C. Nonetheless, a significant relationship is identified between preliminary bid ladders submitted three hours prior to delivery and the final bid curves, as shown in Appendix C, suggesting that early bids encapsulate important market dynamics influencing the final outcomes. These selected features and preliminary bid ladders serve as the foundation for the modelling methodology described in the next chapter.

5 Methodology

This chapter presents the methodology developed to forecast aFRR bid ladders using a structured machine learning pipeline. The process begins with data pre-processing steps, including cleaning, transposing, and splitting the raw dataset. These steps standardise the data and prepare it for transformation, where scaling and dimensionality reduction techniques are applied. The transformed data are then used to train three models: Conformal LASSO, XGBOOST, and LSTM. Their predictions are post-processed to reconstruct the original bid ladder structure. Finally, the models are evaluated using point and interval-based metrics, and a benchmark is introduced to assess their added value relative to existing preliminary bids.

The model developed in this research, as shown in Figure 5.1, builds on four key studies, each informing a specific component of the modelling approach. The structure of the machine learning pipeline, which includes pre-processing, data transformation, model fitting, prediction, and evaluation, is based on the framework proposed by [49]. This provides a clear and systematic foundation to develop a machine learning model. To apply this pipeline specifically to electricity price forecasting, the open-access framework by [29] is used as a reference. This framework offers practical guidance on aspects such as feature selection, data splitting, and model evaluation tailored to electricity market applications. Because bid curve forecasting involves variable curve lengths and shapes, the method from [50] is used to transpose the curves. This ensures a consistent input structure across all bid ladders, which is essential for enabling uniform model training and comparison. After transposition, a dimension reduction technique is applied following the approach of [25]. This step reduces the output dimension, improving computational efficiency while retaining the most relevant information for accurate forecasting.

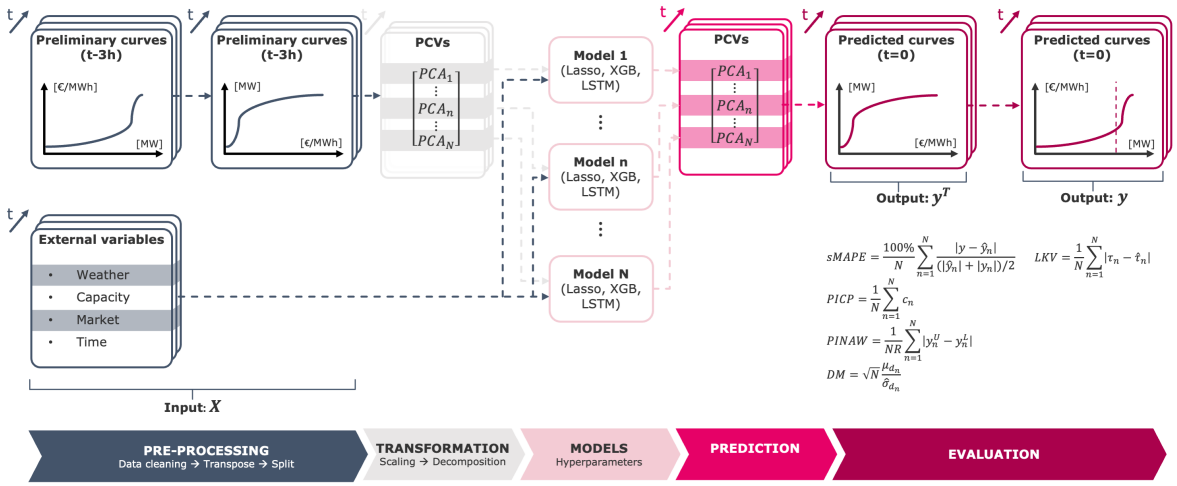


Figure 5.1: Machine learning pipeline for LASSO, XGBOOST, and LSTM models.

5.1. Pre-processing: Preparing Bid Data

The forecasting method begins with pre-processing, which involves transforming raw data into a suitable format for machine learning [49]. This step includes organising and cleaning the data, as well as partitioning it into training, validation, and test sets to support model development and evaluation.

5.1.1. Data Cleaning

The first pre-processing task involves aligning the temporal resolution of the input data. Data from multiple sources [47, 46] are initially recorded at inconsistent intervals, ranging from hourly to 15-minute resolutions. To ensure consistency, all time series are resampled to a common 15-minute interval. Hourly values are forward-filled, and remaining gaps are linearly interpolated to avoid artificial discontinuities. This alignment results in a uniform dataset structure, essential for enabling machine learning models to learn from coherent input data.

5.1.2. Data Transposing

After the dataset is cleaned, it is transposed to create a structure more suitable for model input. Originally, each bid curve consists of discrete points $y_{n,q}$, where q denotes the volume bin on the x-axis (in steps of 10 MW) and $y_{n,q}$ represents the corresponding bid price in €/MWh on the y-axis. Each index n corresponds to one bid curve submitted per ISP, resulting in a time series of bid ladders over all settlement periods

$$y_{n,q} \in \mathbb{R}^{N \times Q_{MW}} \tag{5.1}$$

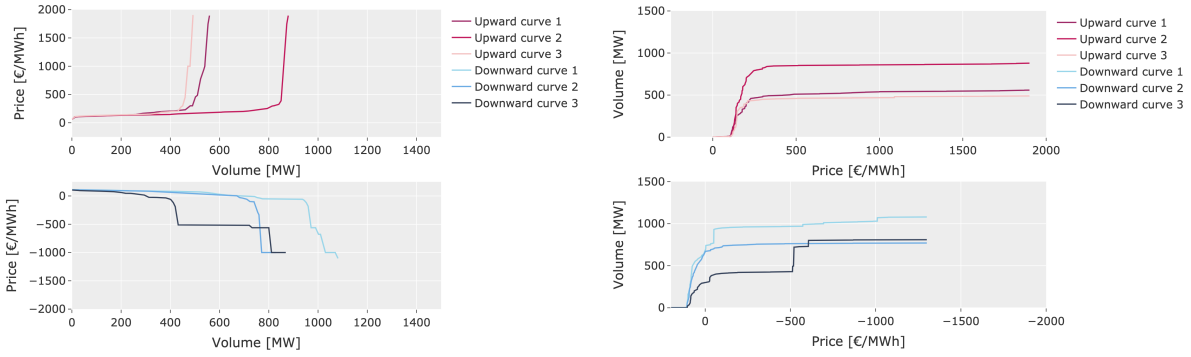
n : Index of the n-th time step, $n \in \{1, \ldots, N\}$
N : Total number of time steps
q : Index of the q-th volume, $q \in \{1, \ldots, Q_{MW}\}$
Q_{MW} : Total number of volume bins in bid ladders

The original dataset lacks a fixed dimensionality, as the total bid volume fluctuates over time. This is illustrated in Figure 5.2a. As a result, the format is not directly usable for model input. To resolve this, the method of [50] is applied, whereby the bid curves are transposed by placing volume on the y-axis and price on the x-axis, yielding a uniform structure

$$y_{n,p} \in \mathbb{R}^{N \times P_€} \tag{5.2}$$

p : Index of p-th price $\in \{1, \ldots, P_€\}$
$P_€$: Total number of prices in bid ladders

The transposition is done on a fixed price range. This range is based on the minimum and maximum volume observed across all bid curves in both directions in Figure 4.6. Upward curves typically range from 0 to 1900 €/MWh, while downward curves range from 200 to -1300 €/MWh. This fixed interval creates a uniform x-axis across all samples, as illustrated by Figure 5.2b.



(a) Normal curves

(b) Transposed curves

Figure 5.2: Original bid curves (a) are transposed (b) by switching the x-axis and y-axis for each timestep.

RQ3: *To address the varying lengths, the bid curves are first transposed by fixing the price range on the x-axis and volume on the y-axis, following the method of [50]. This transposition results in a uniform grid of price points across all samples, enabling consistent and comparable inputs for machine learning models.*

5.1.3. Data Split

The transformation to a consistent format enables a structured split of the data into training (I_{train}), validation (I_{val}), and test (I_{test}) sets to train the model. As no universally accepted standard for data splitting exists [29], an 80% training, 10% validation, and 10% test split is selected. This ratio is adopted to ensure a sufficient amount of data for both model training and evaluation. Given the limited availability of data, allocating 80% to training is considered necessary to provide the model with sufficient learning examples. This is especially important as the total training period is shorter than the one-year dataset size recommended by [29].

The data split is visualised in Figure 5.3. The first 90% of the data is used for training and validation, while the final 10% is reserved for testing. Within the training and validation portion, the data is split randomly but only on complete days, preserving the intra-day order. This approach prevents data leakage, as some market participants place bids covering full days, as shown in Figure C.6. By maintaining the intra-day sequence, the model avoids training on information also present in the validation set and is better able to capture evolving patterns, improving generalisation to unseen data.

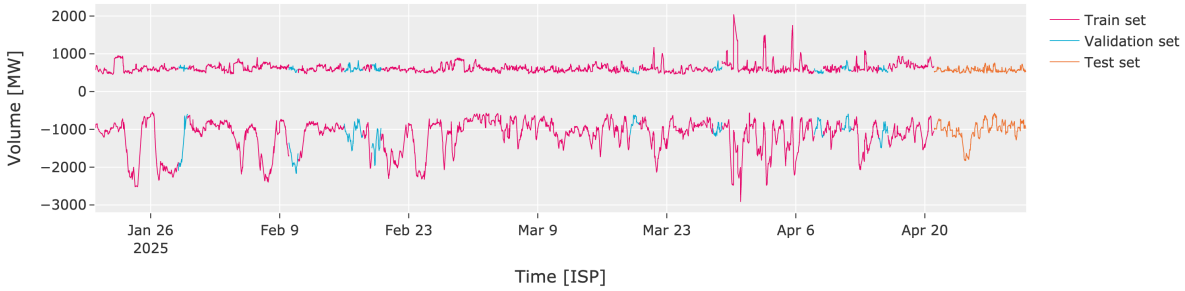


Figure 5.3: Split of bid volume MW time series for the train, validation, and test datasets from January-May 2025 .

5.2. Transformation: Scaling and Dimensionality Reduction

The data is further transformed after splitting. Each curve in the transposed dataset shows large differences between its minimum and maximum values and contains many data points. To handle these issues, the data is scaled to reduce value ranges and decomposed to reduce dimensionality.

Scaling methods, such as normalisation, standardisation, and sigmoid scaling, are used to align the magnitudes of features and outputs, mitigating the disproportionate influence of large values [29].

To address the high output size, Principal Component Analysis (PCA) is applied as a decomposition technique. Following the approach of [25], PCA transforms the original curves into a lower-dimensional representation that captures the most relevant variation. This compact representation enhances computational efficiency, and the required number of models for training is reduced.

5.2.1. Bid ladder Transformation

The output bid ladders and the preliminary bid ladders used as input are transformed through a series of steps. The transformation begins with an initialisation step ($S0$), followed by a scaling step ($S1$), and a decomposition step ($S2$).

Step $S0$: The notation $\boldsymbol{Y}^{S0}$ is introduced to represent the complete dataset of transposed bid ladders. Equation 5.3 defines each individual bid ladder at a given time step n as $\boldsymbol{y}_n^{S0}$. These bid ladders are defined over the complete set of time steps N. Each bid ladder consists of bid volumes for a range of price points. The value $y_{n,p}$ represents the bid volume in MW that corresponds to the price index p, expressed in €/MWh

$$\boldsymbol{Y}^{S0} = \left[\boldsymbol{y}_1^{S0}, \ldots, \boldsymbol{y}_n^{S0}, \ldots, \boldsymbol{y}_N^{S0}\right] \tag{5.3}$$

$$\boldsymbol{y}_n^{S0} = \left[y_{n,1}^{S0}, \ldots, y_{n,p}^{S0}, \ldots, y_{n,P_€}^{S0}\right] \tag{5.4}$$

$\boldsymbol{Y}^{S0}$: Set of all raw bid curves $\in \mathbb{R}^{N \times P_€}$

$\boldsymbol{y}_n^{S0}$: Bid curve at n-th time step $\in \mathbb{R}^{P_€}$

$y_{n,p}^{S0}$: Bid volume at curve n and price p

$P_€$: Total number of price bins in all bid ladders

Step $S1$: As illustrated in Figure 4.4, the distribution of bid prices deviates from a perfect normal distribution. To address this, a standardisation process is applied using a sigmoid transformation. This transformation assigns greater emphasis to frequently occurring values. It also reduces the absolute differences between the highest and lowest bids. As a result, the distribution becomes more balanced, which benefits model training. The sigmoid scaling is defined using the 50th, 95th, and 5th percentiles of the data, denoted as $y_{mid}^{S0} = \mathbb{P}_{50}(y^{S0})$, $y_{max}^{S0} = \mathbb{P}_{95}(y^{S0})$, and $y_{min}^{S0} = \mathbb{P}_{5}(y^{S0})$, respectively. Asymmetry in the distribution is captured by assigning separate logistic growth rates r to the left and right sides

$$y_{n,p}^{S1} = \frac{1}{1 + e^{r(y_{n,p}^{S0} - y_{mid}^{S0})}} \tag{5.5}$$

$$r = \begin{cases} \frac{1}{\left|y_{max}^{S0} - y_{mid}^{S0}\right|}, & y_{n,p}^{S0} \geq y_{mid}^{S0} \\ \frac{1}{\left|y_{mid}^{S0} - y_{min}^{S0}\right|}, & y_{n,p}^{S0} < y_{mid}^{S0} \end{cases} \tag{5.6}$$

$y_{n,p}^{S1}$: Sigmoid scaled bid value in Step $S1$

y_{mid}^{S0} : 50th percentile value of all bids in $\boldsymbol{Y}^{S0}$

y_{max}^{S0} : 95th percentile value of all bids in $\boldsymbol{Y}^{S0}$

y_{min}^{S0} : 5th percentile value of all bids in $\boldsymbol{Y}^{S0}$

r : Scaling factor for logistic growth rate

Step $S2$: PCA reduces data dimensionality by transforming correlated variables into uncorrelated principal components [51]. To achieve this, the covariance matrix of the data is first computed, capturing the relationships between variables. Then, eigenvalue decomposition is performed on this matrix to identify directions (eigenvectors) of maximum variance. The leading eigenvectors, corresponding to the largest eigenvalues, define a lower-dimensional subspace. Finally, the data is projected onto this subspace, resulting in a dimensionality reduction $\mathbb{R}^{P_€} \rightarrow \mathbb{R}^{P_{\text{PCA}}}$

$$\boldsymbol{X} = \left[\boldsymbol{y}_1^{S2}, \ldots, \boldsymbol{y}_n^{S2}, \ldots, \boldsymbol{y}_N^{S2}\right] \tag{5.7}$$

$$\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T \tag{5.8}$$

$$\boldsymbol{C}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{5.9}$$

$$\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{P_{PCA}}] \tag{5.10}$$

$$\boldsymbol{Y}_{PCA} = \boldsymbol{X} \cdot \boldsymbol{V} \tag{5.11}$$

$$\boldsymbol{Y}^{S3} = \boldsymbol{Y}_{PCA} = \left[\boldsymbol{y}_1^{S3}, \ldots, \boldsymbol{y}_n^{S3}, \ldots, \boldsymbol{y}_N^{S3}\right] \tag{5.12}$$

$$\boldsymbol{y}_n^{S3} = \left[y_{n,1}^{S3}, \ldots, y_{n,p}^{S3}, \ldots, y_{n,PCA}^{S3}\right] \tag{5.13}$$

$\boldsymbol{y}_n^{S2}$: Standardised bid curve at n-th time step $\in \mathbb{R}^{P_€}$

$\boldsymbol{X}$: Matrix of all standardised bid curves $\in \mathbb{R}^{N \times P_€}$

$\boldsymbol{C}$: Covariance matrix of $\boldsymbol{X} \in \mathbb{R}^{P_€ \times P_€}$

$\boldsymbol{v}$: Eigenvector of covariance matrix $\boldsymbol{C} \in \mathbb{R}^{P_€}$

λ : Eigenvalue corresponding to eigenvector $\boldsymbol{v}$

$\boldsymbol{V}$: Matrix of top P_{PCA} eigenvectors $\in \mathbb{R}^{P_€ \times P_{PCA}}$

P_{PCA} : Number of selected PCA components

$\boldsymbol{Y}_{PCA} = \boldsymbol{Y}^{S3}$: PCA-transformed data matrix $\in \mathbb{R}^{N \times P_{PCA}}$

$\boldsymbol{y}_n^{S3}$: PCA-reduced bid curve at time step $n \in \mathbb{R}^{P_{PCA}}$

$y_{n,p}^{S3}$: PCA component value at time step n and component p

To evaluate the performance of the PCA-based dimensionality reduction, the eigenvalues and corresponding explained variance percentages of the principal components are presented in Table 5.1. The explained percentage reflects the proportion of total data variability captured by each component, while the cumulative percentage indicates the total variance retained up to each component.

As shown in Table 5.1b, the first upward principal component already explains 95.66% of the variance. Adding three more components increases the cumulative explained variance to over 99%, which justifies the use of four upward principal components in the modelling process.

In the case of the downward components, the first component captures 45.30% of the variance. Although the explained variance increases slightly with additional components, it begins to plateau after the fourth, indicating diminishing returns. However, at least eight components are required to reach 95% of the variance, suggesting a more complex and diffuse structure. As a result, the downward curves are harder to compress and inherently more difficult to forecast, with the maximum achievable prediction accuracy already limited at the transformation stage. Consequently, four downward principal components are still selected for further modelling as a trade-off between accuracy and dimensionality.

Table 5.1: Principal components and their corresponding explained variances and percentages.

(a) Upward PCA components

	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10
Eigenvalue	528.82	96.62	40.34	22.51	16.26	13.13	12.80	9.96	9.45	7.78
Explained %	95.66	3.19	0.56	0.17	0.09	0.06	0.06	0.03	0.03	0.02
Total %	95.66	98.86	99.41	99.59	99.68	99.74	99.79	99.83	99.86	99.88

(b) Downward PCA components

	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10
Eigenvalue	280.57	234.58	119.44	85.88	68.12	50.72	43.43	36.99	33.24	28.98
Explained %	45.30	31.67	8.21	4.25	2.67	1.48	1.09	0.79	0.64	0.48
Total %	45.30	76.97	85.18	89.43	92.10	93.58	94.67	95.45	96.09	96.57

RQ3: *To manage the high dimensionality of the transposed bid curves, PCA is applied. Each original curve includes more than one thousand price bins, leading to a large output space. PCA reduces this dimensionality by projecting the data into a lower-dimensional space of four components, while preserving the most dominant patterns in the data [51].*

5.2.2. Feature Transformation

The external variables data is subjected to separate transformation steps depending on the model type. For LASSO regression, scaling is required due to its reliance on L1 regularisation, which makes it sensitive to feature magnitudes [29]. Consequently, the input data is scaled with a StandardScaler that standardises features by removing the mean and scaling them to unit variance. XGBOOST does not strictly require scaling due to its tree-based structure [26]. LSTM networks are susceptible to input scales; therefore, a MinMaxScaler is applied to scale the data between 0 and 1 [25]. Proper scaling is crucial for LASSO and LSTM, while XGBOOST can also perform effectively without it.

5.3. Forecasting Models: Point and Interval Estimation

For each principal component, a separate model is constructed, based on the commonly used approaches described in Section 3.3. To ensure optimal performance across different components and settings, hyperparameter tuning is applied using Optuna. This framework leverages adaptive sampling and early stopping, enabling faster and more efficient optimization compared to traditional grid search methods [52]. Each model is then extended with conformal prediction to generate uncertainty-aware prediction intervals. This probabilistic extension enhances interpretability by quantifying prediction confidence and supports the detection of outliers and volatility in bid distributions, thereby facilitating more informed decision-making under uncertainty.

5.3.1. Hyperparameter Tuning

To optimise the model's hyperparameters, a Tree-structured Parzen Estimator (TPE) is applied within a Bayesian optimisation framework. This is implemented using the Python module `Optuna` [52]. Unlike grid or random search, TPE constructs probabilistic models of the objective function based on prior evaluations. It then selects new hyperparameter configurations by estimating where improvements are most likely. This allows for a more informed and efficient search of the parameter space each trial.

Conformal LASSO: The only hyperparameter for tuning in LASSO is the regularisation parameter (λ).

- Regularisation parameter (λ): Controls the L1 penalty to encourage sparsity while balancing simplicity and accuracy.

Table 5.2: Hyperparameters used for LASSO model.

	Parameters
Regularisation parameter (λ)	[0.1–10]

Conformal XGBOOST: In the case of the XGBOOST model, hyperparameter tuning focuses on optimising both the structure of the model and its regularisation behaviour.

- Learning rate: Controls the contribution of each tree to the final prediction.
- Max depth: Sets the maximum depth of individual trees, determining model complexity.
- Estimators: Specifies the number of boosting rounds (trees) in the ensemble.
- Gamma: Minimum loss reduction required to split a leaf node, acting as a regularisation parameter.
- Colsample bytree: Fraction of features randomly sampled for each tree.
- Subsample: Proportion of training data randomly sampled for each tree to manage variance and bias.
- Regression alpha: L1 regularisation term on weights to encourage sparsity.
- Regression lambda: L2 regularisation term promoting stability in model weights.

Table 5.3: Hyperparameters used for XGBOOST model.

	Parameters
Learning rate	Uniform [*0.05-0.1*]
Max depth	Uniform [*3-8*]
Estimators	Choice [*500, 1000, 1500*]
Gamma	Uniform [*0-10*]
Colsample bytree	Uniform [*0.5-1.0*]
Subsample	Uniform [*0.7-1.0*]
Regression alpha	Uniform [*0-5*]
Regression lambda	Uniform [*0-5*]

Conformal LSTM: The architecture used for forecasting aggregated supply curves (ASCs), shown in Figure 5.4, is based on the structure proposed by [25]. This design captures temporal dependencies in sequential input data of length N_I through one or more stacked LSTM layers. The final hidden state is passed to a dense layer, which transforms the learned representation to the output layer over N_O future time steps. ReLU activation is applied in the dense layers to introduce non-linearity while maintaining computational efficiency. The following hyperparameters are tuned to optimise performance:

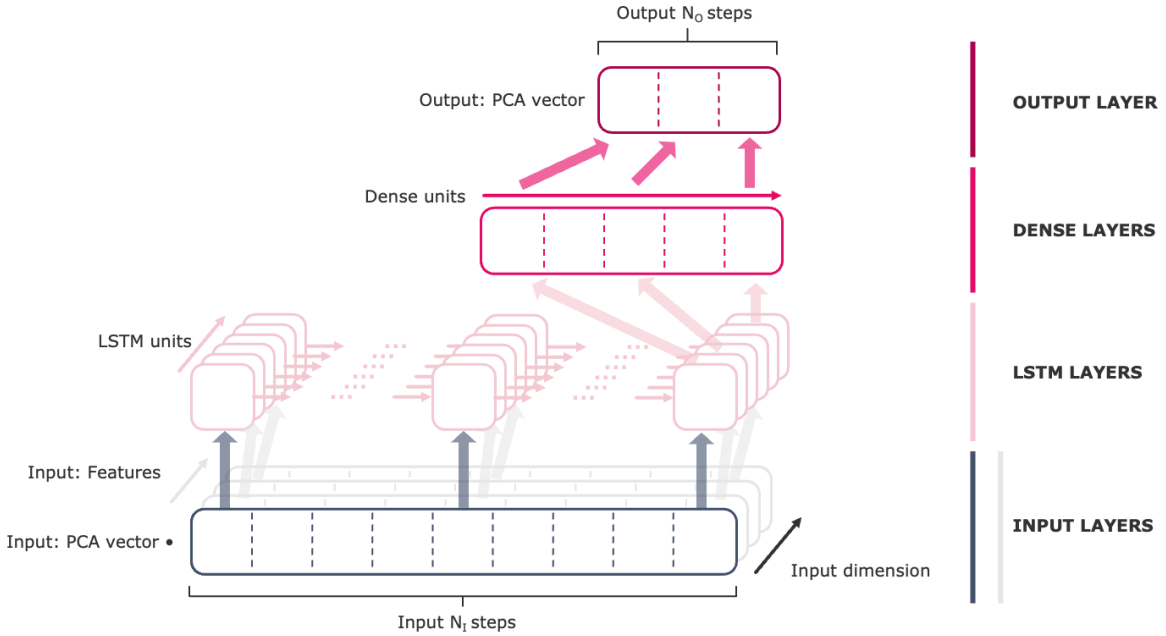


Figure 5.4: LSTM network architecture, where input features progress through multiple LSTM layers before entering dense layers and culminating in the final output layer.

The following hyperparameters are tuned to determine the best model configuration:

- Timesteps (past period): Refers to the length of historical input data used for prediction, determining how much past information the model considers.
- LSTM units: Defines the number of memory cells in the LSTM layer, influencing the model's capacity to learn temporal patterns.
- Dense units: Specifies the number of neurons in the dense layers following the LSTM, affecting model complexity.
- Learning rate: Controls the step size during gradient descent, impacting convergence speed and stability.
- Epochs: Indicates how many times the full training set is passed through the model. Early stopping is applied to prevent overfitting, halting training once performance ceases to improve.
- Early stopping patience: Sets how many epochs to wait without improvement before stopping training, helping balance learning time and overfitting prevention.

Table 5.4: Hyperparameters used for LSTM model.

	Parameters
Timesteps	Choice [0], [0, 4, 8, 12, 16], [0, 96], [0, 96, 192, 288]
LSTM units	Choice [16, 32, 64]
Dense units	Choice [16, 32, 64, 128]
Learning rate	0.01
Epochs	100
Early stopping patience	10

5.3.2. Conformal Prediction

Conformal Prediction (CP) is applied to construct prediction intervals. Conformal prediction is a user-friendly method for quantifying uncertainty in machine learning models [53]. The choice for CP is motivated by four main advantages: (1) It is distribution-free, making no assumptions about the underlying data distribution, particularly relevant as bid data are not always normally distributed [54]. (2) It is computationally efficient, requiring no additional models to be fitted to determine an interval level [54]. (3) It is model-agnostic, meaning it can be used with any model. This allows integration with any underlying model, including LASSO, XGBOOST, and LSTM [54]. (4) It provides guaranteed coverage by constructing empirical calibration based on past forecast errors [54]. This guarantee is formalised as follows

$$\mathbb{P}(\boldsymbol{y}_{new} \in C(\boldsymbol{x}_{new})) \geq 1 - \alpha \tag{5.14}$$

$\boldsymbol{y}_{new}$: New observed value (target) to predict $\in \mathbb{R}^{N_{test}}$

$\boldsymbol{x}_{new}$: New feature vector used for prediction $\in \mathbb{R}^{M}$

$C(\boldsymbol{x}_{new})$: Prediction interval or set for new test set $\in \mathbb{R}^{N_{test} \times 2}$

α : Significance level (1 minus confidence level)

$\mathbb{P}$: Probability measure

Split Conformal Prediction, also known as Inductive Conformal Prediction (ICP), is adopted in this study. ICP enables post-hoc interval construction directly on model outputs without modifying the training procedure. This makes it suited to the current setting, where predictions are decomposed for evaluation and quantile outputs cannot be reliably inverted to the desired output. The "conformalisation" workflow comprises four main steps [54].

Step 1: Divide the data into training ($I_{train} \in \mathbb{R}^{N_{train}}$), calibration ($I_{cal} = I_{val} \in \mathbb{R}^{N_{cal}}$), and test ($I_{test} \in \mathbb{R}^{N_{test}}$) sets. The validation set noted in Subsection 5.1.3 can be used as the calibration set. The model is then trained exclusively on the training set, ensuring that the calibration and test data remain unseen during fitting. After training, predictions are generated for both the calibration and test sets. These predictions are used to compute nonconformity scores (on the calibration set) and to construct prediction intervals or sets (on the test set), enabling uncertainty quantification.

Step 2: After predictions have been generated for the calibration set, nonconformity scores are calculated. Nonconformity scores quantify how "atypical" or "nonconforming" each example is relative to the model and previously observed data [54]. For regression tasks, the nonconformity score equals the absolute residual between the predicted and actual values

$$\boldsymbol{s}_{cal} = |\boldsymbol{y}_{cal} - \hat{\boldsymbol{y}}_{cal}| \tag{5.15}$$

$\boldsymbol{s}_{cal}$: Nonconformity scores of calibration set $\in \mathbb{R}^{N_{cal}}$

$\boldsymbol{y}_{cal}$: True calibration values $\in \mathbb{R}^{N_{cal}}$

$\hat{\boldsymbol{y}}_{cal}$: Predicted calibration values $\in \mathbb{R}^{N_{cal}}$

Step 3: The nonconformity scores from the calibration set are sorted to determine an empirical quantile. This quantile corresponds to the desired coverage level (e.g., 90%). The position of the quantile is computed using the ceiling function $\lceil \cdot \rceil$, which returns the smallest integer greater than or equal to the result. The selected score defines the width of the prediction interval

$$\hat{q}_{1-\alpha} = Quantile\left(\{s_1, \ldots, s_{N_{cal}}\}, \frac{\lceil (N_{cal}+1)(1-\alpha) \rceil}{N_{cal}}\right) \tag{5.16}$$

$\hat{q}_{1-\alpha}$: Quantile of sorted nonconformity scores
N_{cal} : Size of the calibration set

Step 4: Using the quantile from Step 3, prediction intervals are constructed for new, unseen test examples. For a chosen confidence level $1-\alpha$ (e.g., 90% coverage corresponds to $\alpha = 0.1$), the interval is defined so that it captures the true value with probability at least $1-\alpha$. This typically results in a symmetric interval around the model's prediction, extending by the quantile value in both directions

$$\hat{C}(\boldsymbol{x}_{new}) = \left[\hat{\boldsymbol{y}}(\boldsymbol{x}_{new}) - \hat{q}_{1-\alpha}, \hat{\boldsymbol{y}}(\boldsymbol{x}_{new}) + \hat{q}_{1-\alpha}\right] \tag{5.17}$$

$\hat{C}(\boldsymbol{x}_{new})$: Prediction interval for a new input $\in \mathbb{R}^{N_{test} \times 2}$
$\hat{\boldsymbol{y}}(\boldsymbol{x}_{new})$: Point prediction for a new input $\in \mathbb{R}^{N_{test}}$

Step 5: Bid ladders follow a monotonic structure. Therefore, prediction intervals are trimmed to preserve this property. This leads to narrower intervals while maintaining validity.

5.4. Model Evaluation: Point and Interval Metrics

To assess the performance of the proposed models, a structured evaluation framework is adopted that consists of three steps. First, a benchmark is defined to provide a practical reference. The benchmark is used to test whether the model improves on the data already available three hours in advance. Second, scoring metrics are used to evaluate both point and interval forecasts [29, 55]. Finally, a statistical test is applied to determine whether observed performance differences are statistically significant. The test helps determine whether a difference in accuracy really exists or is simply due to random differences between the predictions [56].

5.4.1. Preliminary Bid Ladder as a Practical Reference

A naive reference, hereafter referred to as BENCHMARK, is used to evaluate model performance. The benchmark consists of the preliminary bid ladders published three hours before delivery [44]. This dataset is further elaborated in Appendix C. This data serves as a practical reference to assess whether the model's predictions offer measurable improvements and to understand which inputs influence changes in price expectations at that horizon. The preliminary bid ladder is particularly suited as a benchmark due to its timeliness and availability, as it is one of the few data sources accessible before delivery that reflects actual market expectations.

5.4.2. Scoring Metrics

To compare the predicted and actual curves over time, several scoring metrics are employed. First, point forecast evaluation measures the deviation between predicted values and actual outcomes. Then, interval forecast evaluation measures the deviation between predicted and observed intervals, thus reflecting both the accuracy and the reliability of the forecasted ranges.

Point Prediction Metrics

Three metrics are employed to evaluate point forecasts. First, the Symmetric Mean Absolute Percentage Error (sMAPE) is utilised to assess the overall discrepancy between the predicted and actual curves. The sMAPE can compare the relative difference on all parts of the curve. This is important because the absolute values at the beginning of the curve are much lower than at the end of the curve. sMAPE is chosen over MAPE due to its symmetric formulation, which normalises the absolute error by the mean of predicted and actual values. This makes sMAPE more robust to zero and near-zero observations, which are usually present in the first part of the predicted curve and PCA values [29, 51]. Third, the Largest Knick Volume (LKV) is introduced to assess the model's ability to capture the volume associated with the "knick", a point of sudden price escalation. This metric is specifically designed to evaluate whether the predicted and actual values accurately represent the volume at such inflection points.

Symmetric Mean Absolute Percentage Error (sMAPE): Measures the average relative difference between predicted and actual values, normalised by their mean

$$sMAPE = \frac{100\%}{N} \sum_{n=1}^{N} \frac{|y_n - \hat{y}_n|}{(|y_n| + |\hat{y}_n|)/2} \tag{5.18}$$

y_n : Observed value at n-th timestep
$\hat{y}_n$: Predicted value at n-th timestep

Largest Knick Volume (LKV): Measures the average volume distance between the real and predicted volume at which a sudden price increase occurs. The detection of the knick volume begins by computing the standardised signal using the z-score for each separate curve. The z-score (or standard score) measures how many standard deviations a data point is from the mean of a distribution

$$z_{n,q} = \frac{y'_{n,q} - \mu'_n}{\sigma'_n} \tag{5.19}$$

$z_{n,q}$: Standardised value of the differentiated curve $y'_{n,q}$
$y'_{n,q}$: Differentiated curve at n-th timestep and q-th volume bin
μ'_n : Mean of curve at n-th timestep
σ'_n : Standard deviation of curve at n-th timestep
q : Volume of bid ladder $\in \{1, \ldots, Q_{MW}\}$

LKV is defined as the first volume point (τ_n) where the z-score of the price curve exceeds a predefined threshold. This threshold is set to one, as a z-score of 1 corresponds to one standard deviation above the mean, statistically marking the onset of abnormal behaviour. In this context, it identifies the start of the curve's volatile transition. Figure 5.5 visualises this for both upward (pink) and downward (blue) curves, with the LKV (τ is indicated by the dark pink and dark blue vertical lines, respectively

$$\tau_n = \arg\min \left\{ q \mid z_{n,1} > z^{\text{threshold}} \right\} \tag{5.20}$$

τ_n : Location where a spike occurs at n-th time
$z^{\text{threshold}}$: Predefined spike threshold for z, set to 1

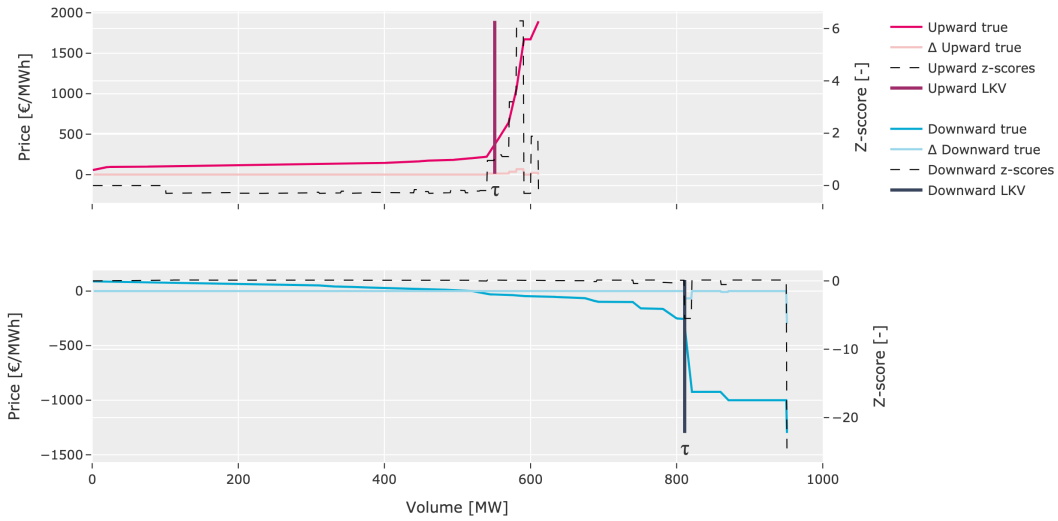


Figure 5.5: The knick point defined as the first point where the z-score exceeds 1, marking the onset of volatility in upward (pink) and downward (blue) bid curves.

Figure 5.6 illustrates the horizontal alignment between predicted and true price-volume bid curves. Solid lines represent the observed bids, while dashed lines show model predictions. The vertical markers indicate the true local kink volume (LKV) location (τ) and the corresponding point on the predicted curve with the same price, denoted $\hat{\tau}$. The horizontal distance between (τ) and ($\hat{\tau}$), shown as a black line, quantifies the distance between the most critical points of the curve

$$\hat{\tau}_n = \arg\min|\hat{y}_n - y_n(\tau)| \tag{5.21}$$

$y_n(\tau)$: True price value at spike location τ

$\hat{\tau}_n$: Spike location in prediction that best matches $y_n(\tau_n)$

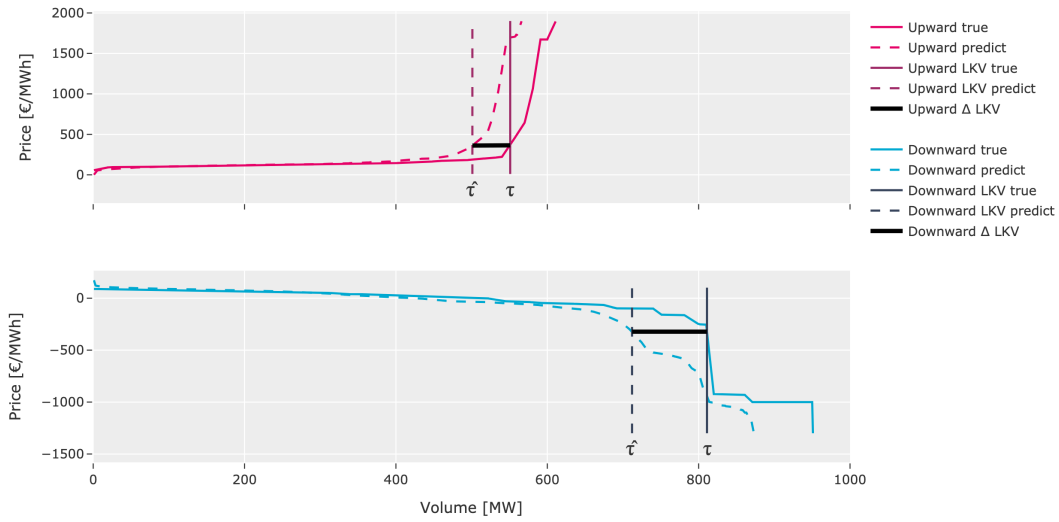


Figure 5.6: LKV is defined as the horizontal distance between the predicted and actual curve.

Finally, the LKV metric is computed as the average absolute difference between the true and predicted spike locations over all samples

$$LKV = \frac{1}{N}\sum_{n=1}^{N}|\tau_n - \hat{\tau}_n| \tag{5.22}$$

RQ4: *This work introduces the Largest Knick Volume (LKV) as a targeted metric to evaluate the model's predictive performance. It identifies the first volume point (τ_n) after which price movements become extreme. The metric assesses how well the model predicts the volume corresponding to the key price location, based on the comparison with the predicted point ($\hat{\tau}_n$).*

Interval Prediction Metrics

Interval forecasts are evaluated using the Prediction Interval Coverage Percentage (PICP) and Prediction Interval Average Width (PIAW) to assess two properties: reliability and sharpness [28]. PICP measures the proportion of observed values within the predicted intervals and reflects the reliability of the forecast. In contrast, PIAW evaluates the average interval width, providing a measure of sharpness. Narrower intervals are preferred, as they indicate more precise forecasts [28].

Prediction Interval Coverage Percentage (PICP): Quantifies the average proportion of observed values that fall within their corresponding forecasted prediction intervals. It serves as a measure of the calibration or reliability of the interval forecasts. A well-calibrated model should produce intervals that contain the true values approximately as often as the nominal coverage rate specifies (e.g., 90% true values should fall within 90% prediction intervals). Deviations from this indicate systematic under- or overestimation of uncertainty

$$PICP = \frac{1}{N}\sum_{n=1}^{N} c_n, \quad \text{with}$$
$$c_n = \begin{cases} 1, & y_n \in [\hat{y}_n^L, \hat{y}_n^U] \\ 0, & y_n \notin [\hat{y}_n^L, \hat{y}_n^U] \end{cases} \tag{5.23}$$

$\hat{y}_n^L$: Lower bound of the prediction interval

$\hat{y}_n^U$: Upper bound of the prediction interval

c_n : Heaviside indicator: 1 if true value is within prediction interval, else 0

Prediction Interval Normalised Average Width (PINAW): Measures the mean width of all prediction intervals over the forecast horizon, normalised by the range of true values. It reflects the sharpness, or informativeness, of the interval forecasts. Narrower normalised intervals indicate more confident and precise predictions, independent of the scale of the target variable.

$$PINAW = \frac{1}{NR}\sum_{n=1}^{N} |\hat{y}_n^U - \hat{y}_n^L| \tag{5.24}$$

R : Range of observed values $(\max(y_1, \ldots, y_N) - \min(y_1, \ldots, y_N))$

5.4.3. Statistical Test

To statistically assess whether the predictive performance of two competing forecasting models differs significantly, the Diebold-Mariano (DM) test is employed [56]. As recommended by [29], the test is suited for electricity price forecasting and provides statistical inference for model comparison.

Diebold-Mariano Test (DM): Evaluates the null hypothesis of equal predictive accuracy between two forecasting models by analyzing the mean of the loss differential series, which is computed using a chosen loss function. For each observation, prediction errors are defined as $e_1 = y_n - \hat{y}_{n1}$ and $e_2 = y_n - \hat{y}_{n2}$ for models 1 and 2, respectively. The loss differential is given by $d_n = \|e_1\| - \|e_2\|$, with its mean and variance denoted as μ_{d_n} and σ_{d_n}. Based on this, the DM test statistic is constructed and asymptotically follows a standard normal distribution. A statistically significant outcome (e.g., $p < 0.05$) indicates that one model exhibits better predictive performance over the other

$$DM = \sqrt{N}\,\frac{\bar{\mu}_{d_n}}{\hat{\sigma}_{d_n}} \sim \mathcal{N}(0,1) \tag{5.25}$$

$\bar{\mu}_{d_n}$: Mean of d_n

$\hat{\sigma}_{d_n}$: Standard deviation of d_n

Having established the methodological foundation and detailed the model development pipeline, the next chapter presents the empirical results obtained from applying these models, evaluating their predictive accuracy and uncertainty quantification performance across multiple metrics.

6

Results I: Forecasting Models

This chapter presents the results of the proposed and benchmark models. Model performance is evaluated on two levels: first, at the level of principal component decomposed units, and second, on the fully reconstructed bid ladder curves. Lastly, the models are compared in terms of computational efficiency, based on tuning and training time.

6.1. Model Performance

A two-stage evaluation process assesses performance at intermediate and final output levels. The first stage focuses on the individual principal components (Subsection 6.1.1) to evaluate how well the models predict the decomposed signals before reconstruction. The second stage considers the fully reconstructed bid curves (Subsection 6.1.2) to determine overall forecasting accuracy in the original data space. This two-step approach follows the evaluation framework from Figure 5.1 and Section 5.4.

6.1.1. Principal Component Results

The analysis begins by evaluating the performance of the LASSO, XGBOOST, LSTM, and BENCHMARK models on the individual PCA components. The corresponding model specifications can be found in Appendix D. This includes the hyperparameter configurations (Table D.1) and the corresponding training and validation losses (Figure D.1). Furthermore, the feature importance obtained for the LASSO and XGBOOST models is presented in Figures D.2 and D.3, respectively.

Principal Component Forecasts

To analyse how each model captures the shape of the decomposed signals, Figure 6.1 shows predicted and actual trajectories of the four principal components. The y-axis represents the scaled values of the principal components, while the x-axis denotes the time index. By comparing these trajectories, it becomes clear how well each model captures the overall shape of the PCA signals, including both smooth trends and sharp peaks. These features define the structure of the PCA components. Inaccurate forecasts of these patterns can lead to poor reconstruction of the original curves after back-transformation to the bid curve domain. This, in turn, reduces the accuracy and practical usefulness of the forecasts. To complement the visual assessment, Table 6.1 reports the corresponding sMAPE values for each component and model configuration.

In the upward direction, all models reproduce the general trends. The strongest performance is observed for PCA1 and PCA2, as indicated by the lowest sMAPE scores in Table 6.1a. In contrast, forecasts for PCA3 and PCA4 show larger deviations, especially in peak magnitude and timing. These components explain less variance, making it harder for models to learn their structure.

A similar pattern is found in the downward direction. PCA1 is modelled with reasonable accuracy, but performance declines for higher-order components, as shown in Table 6.1b. Forecasts for PCA3 and PCA4 struggle to reproduce both the overall shape and local peaks. As with the upward case, the weaker signal strength and reduced structure in these components limit model accuracy and responsiveness to abrupt changes.

(a) Upward PCA's over time

(b) Downward PCA's over time

Figure 6.1: Model predictions for the first four principal components over time. LASSO (pink), XGBOOST (blue), LSTM (orange), BENCHMARK (green), and true values (dashed black).

Table 6.1: sMAPE scores for upward and downward principal components.

(a) sMAPE for upward principal components.

	PCA1	PCA2	PCA3	PCA4
LASSO	0.58	0.59	0.79	1.22
XGBOOST	0.83	0.53	0.82	1.15
LSTM	0.60	0.59	0.78	1.16
BENCHMARK	0.67	0.57	0.81	1.02

(b) sMAPE for downward principal components.

	PCA1	PCA2	PCA3	PCA4
LASSO	0.35	0.77	0.81	0.97
XGBOOST	0.38	0.69	0.71	0.96
LSTM	0.37	0.71	0.82	0.87
BENCHMARK	0.35	0.79	0.81	1.29

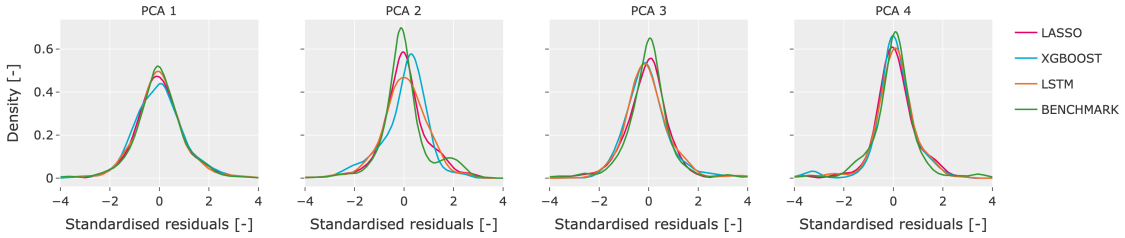
Based on the comparison between predicted and actual principal components, residual analysis is conducted as a next step to evaluate the remaining prediction errors. This analysis builds on the observed discrepancies and provides further insight into the quality and reliability of the forecasts.

Principal Component Residuals

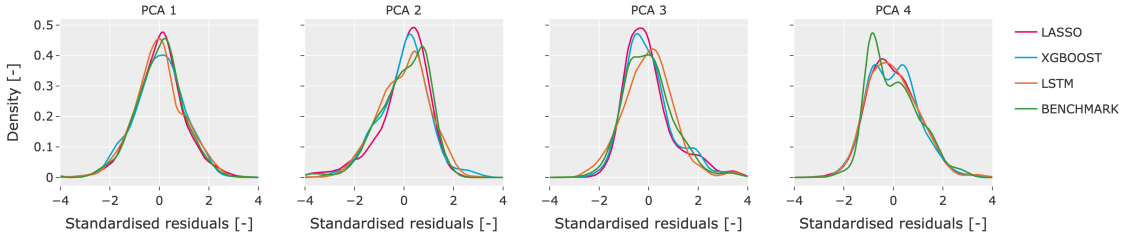
Residual analysis assesses model assumptions such as normality, constant variance, independence, and linearity [57]. Violations can reduce prediction accuracy. Figure 6.2 shows the KDE of standardised residuals for the four principal components in the downward direction. Ideally, residuals follow a standard normal distribution. Deviations such as skewness or heavy tails indicate poor model fit. Non-normally distributed residuals can also reflect a tendency for errors to be biased or behave unpredictably, undermining forecast reliability. Additional residual results are shown in Appendix Figure E.3.

The residuals in the upward direction are mostly symmetric and centred around zero. This pattern is most apparent in PCA1, PCA3, and PCA4, where all models show similarly shaped distributions with comparable peak densities. In PCA2, the residual distributions appear slightly skewed.

The downward residuals display greater variability across components and models. In PCA2, PCA3, and PCA4, the distributions differ in shape, with observable skewness and heavier tails in several cases. In particular, some distributions diverge from the expected bell-shaped form, lacking the characteristic symmetry and kurtosis of a normal distribution. PCA1 remains relatively consistent, but the remaining components exhibit broader and more irregular patterns, with increased dispersion and structural differences in the error terms.



(a) Kernel Density Estimates for each upward principal component error.



(b) Kernel Density Estimates for each downward principal component error.

Figure 6.2: Kernel density estimates of model forecast errors across principal components.

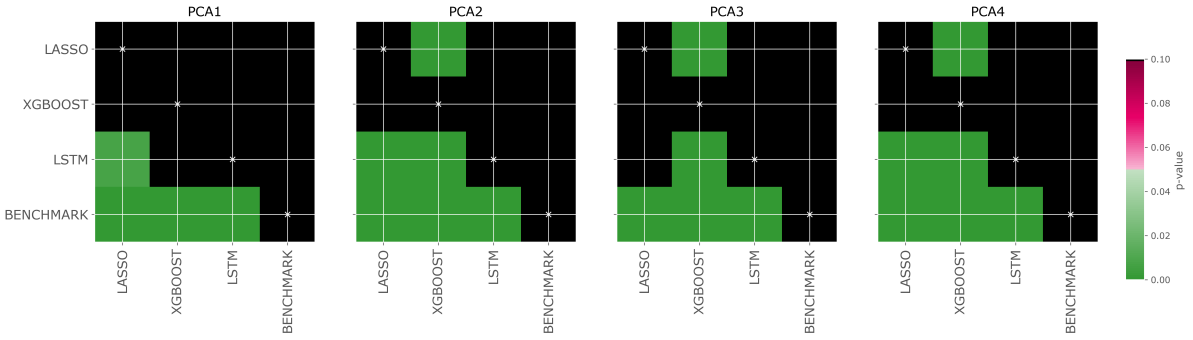
Residual analysis reveals the quality of model predictions. To compare model performance, each pair is evaluated individually through a statistical test of its predictive accuracy.

Diebold-Mariano Test

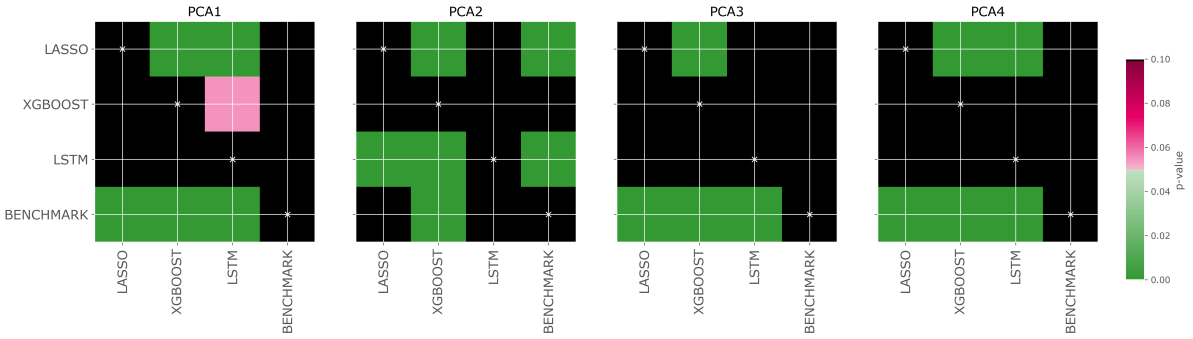
A pairwise Diebold-Mariano test evaluates the predictive accuracy of two models by comparing their forecast errors. By systematically applying this test across all model pairs, statistically grounded comparisons of predictive performance are obtained. These comparisons are summarised in Figure 6.3, which presents the corresponding p-values for each pair, evaluated separately for the four PCA components and both upward and downward directions. The colour-coding of the results aids interpretation: green indicates statistical significance at the 5% level ($p < 0.05$), implying one model significantly outperforms the other, while black ($p \geq 0.1$) denotes no significant difference. The matrix layout allows a visual comparison across all model pairs and components.

Figure 6.3a shows that the benchmark model statistically outperforms all other models in the upward direction, except for PCA3. LSTM ranks second, significantly outperforming both XGBOOST and LASSO. XGBOOST exhibits the weakest performance, as indicated by black squares denoting statistical inferiority across most comparisons.

For the downward direction in Figure 6.3b, the benchmark model again provides the best forecasts, statistically outperforming all others except in PCA2. LSTM only performs well in PCA2, while XGBOOST consistently ranks lowest across all components, consistent with the upward direction results.



(a) DM test results for upward models.



(b) DM test results for downward models.

Figure 6.3: Pairwise Diebold-Mariano test p-values for the first ten principal components. Each matrix compares forecast accuracy among LASSO, XGBOOST, and LSTM. Green cells indicate significant differences at the 5% level.

Following the evaluation at the level of individual principal components, the analysis proceeds to the fully reconstructed bid ladder curves. In this stage, the PCA-based model forecasts are assessed in the original data space and compared against the benchmark model.

6.1.2. Reconstructed Bid Ladder Results

This section presents the results of the reconstructed bid ladder predictions. First, the overall curve shape is examined to evaluate how accurately the models reflect the structure of bidding behaviour. Building on this, the analysis then considers the total predicted volume to assess whether the models also capture the magnitude of market activity. Additional results and further analysis of the forecasted bid ladders are provided in Chapter E.

Bid Ladder Forecasts

Figure 6.4 provides a visual representation of how the forecasted bid curves look compared to the true curves for each model. The upward curves are shown in the top row, and the downward direction in the bottom row. The x-axis represents bid volume in MW, and the y-axis the corresponding bid price in €/MWh. A single ISP (2025-03-10 10:00:00) is selected to depict an average-case scenario, as the high volumetric granularity limits the clarity to show all predictions. Solid lines represent the true bid curves, dashed lines the predicted curves, and shaded areas indicate the 90%, 50%, and 10% coverage intervals.

The upward bid curves (top row) display a consistent reproduction of the steep non-linear increase near the right tail of the volume axis. The prediction intervals remain narrow across most of the curve but widen around the inflection point, reflecting increased uncertainty in that region.

The downward bid curves (bottom row) also show that the general declining structure is well captured. As with the upward curves, the confidence intervals are narrow in most regions but increase around the inflection point, indicating greater uncertainty where the curve shape changes more rapidly.

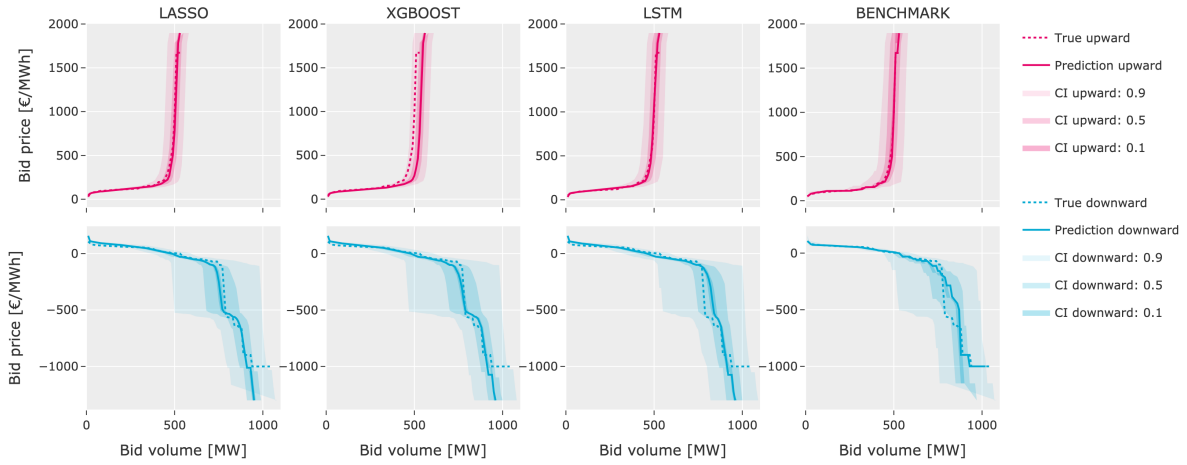


Figure 6.4: Forecast snippet for true (dash) and predicted (solid) bid curves for upward (pink) and downward (blue).

The visualisations gave a qualitative view of forecast shape and model behaviour. This is now complemented by a quantitative evaluation using error metrics to assess predictive accuracy and uncertainty.

Tables 6.2 present point-based (sMAPE, LKV) and interval-based (PICP, PINAW) error metrics for each model. The performance of the models over time and across different hours of the day is additionally reported in Figures E.1 and E.2, respectively. Together, these metrics support a structured comparison between the developed models and the benchmark in terms of shape accuracy, localisation precision, and interval quality. To be considered a better alternative, a model must significantly outperform the benchmark across these metrics. This is especially relevant for machine learning models, which add uncertainty absent in the deterministic benchmark.

For the upward curve, all models show similar sMAPE scores (0.07–0.08), indicating comparable shape accuracy. LASSO performs best on LKV, while the benchmark shows the widest intervals (highest PINAW) but also the best coverage (highest PICP). Overall, performance is comparable, but XGBOOST ranks lowest due to its higher sMAPE and lower PICP interval estimates.

For the downward curve, the benchmark achieves the lowest sMAPE and highest PICP, confirming the highest performance and coverage. LASSO performs best on LKV, while the benchmark provides the narrowest intervals (lowest PINAW). As in the upward case, the benchmark ranks highest, followed by LASSO and LSTM, with XGBOOST performing the worst.

Table 6.2: Error metrics for each model on the upward and downward bid curves.

(a) Upward bid curve metrics.

	sMAPE	**LKV**	**PICP**			**PINAW**		
			0.9	0.5	0.1	0.9	0.5	0.1
LASSO	0.07	28.15	0.89	0.53	0.12	0.34	0.13	0.03
XGBOOST	0.08	31.66	0.84	0.41	0.08	0.33	0.13	0.02
LSTM	0.07	30.54	0.89	0.52	0.12	0.36	0.14	0.03
BENCHMARK	0.07	33.55	0.91	0.62	0.20	0.41	0.17	0.03

(b) Downward bid curve metrics.

	sMAPE	**LKV**	**PICP**			**PINAW**		
			0.9	0.5	0.1	0.9	0.5	0.1
LASSO	0.10	51.59	0.95	0.63	0.13	0.46	0.16	0.02
XGBOOST	0.11	51.98	0.93	0.58	0.12	0.48	0.17	0.02
LSTM	0.11	59.22	0.94	0.57	0.13	0.46	0.14	0.02
BENCHMARK	0.08	56.74	0.97	0.59	0.22	0.30	0.08	0.02

The previous analysis evaluated how well models captured curve shape and uncertainty. The next step assesses volume prediction accuracy to reflect total market activity.

Bid Ladder Volume Forecasts

Figure 6.5 shows the total curve length, defined as the total volume of the bid ladder, evaluated over time. This enables the assessment of each model's ability to predict total bid ladder volume, where the x-axis represents time per ISP and the y-axis represents the corresponding bid volume in MW.

All upward models generally follow the main volume trends over time, but do not capture larger volume increases occurring after 3 hours before delivery. These peak events are often underestimated. No model shows consistent superiority over the entire period. The benchmark model tends to underpredict the total upward volume, in line with the distributional differences shown in Appendix Figure C.5.

Similarly, all models capture the overall trends in downward volume, with fewer outliers present in this set. During the evaluation period, only a limited amount of additional volume is bid three hours before delivery. The benchmark model underpredicts the total volume, corresponding to the distribution shown in Appendix Figure C.5.

Figure 6.5: Curve volume predictions for true values (dashed black), LASSO (pink), XGBOOST (blue), LSTM (orange), BENCHMARK (green).

6.2. Calculation Information

The proposed models are implemented using Scikit-learn [58], XGBoost [59] and TensorFlow [60] Python modules. Model training is conducted on a MacBook Pro (M1, 2020), utilising its integrated CPU for LASSO and XGBOOST and GPU through TensorFlow's Metal plugin support for Apple Silicon. These results are summarised in Table 6.3.

Table 6.3: Comparison of model training times for hyperparameter tuning and fitting across LASSO, XGBOOST, and LSTM.

	Hyperparameters	**Fit**
LASSO	5-10 [s]	<1 [s]
XGBOOST	2-3 [min]	1-3 [s]
LSTM	3-4 [h]	7-8 [min]

LASSO: Computational time for Lasso is low, with PCA hyperparameter tuning taking 5-10 seconds and model fitting requiring less than one second for both upward and downward directions.

XGBOOST: Required more computational resources, with PCA training taking 2-3 minutes on average and the fitting phase needing 1-3 seconds for both upward and downward scenarios.

LSTM: LSTM demanded more computational time, with PCA hyperparameter training requiring 3-4 hours on average and the final fitting executed 7-8 minutes for both upward and downward analyses.

In addition to assessing accuracy, bid curve forecasts can inform trading strategies by indicating market positioning and price sensitivity. This practical relevance is explored in the next section.

7

Results II: Battery Trading Strategy

This chapter investigates how aFRR bid ladder forecasts can enhance energy trading strategies for a battery operating in the IDM. It introduces the battery system and the modelling assumptions, followed by four bidding strategies, each reflecting a different pricing scenario. Central to this study is a strategy that bids at the knick point (LKV) of the merit order curve, using the benchmark model. These strategies are embedded within a trading framework that combines aFRR participation with scheduled dispatch. The chapter concludes with a comparative profit and loss analysis to evaluate the effectiveness of each strategy.

As market activity increasingly shifts toward IDM and IM, aFRR bid ladder forecasts can provide valuable support for short-term trading decisions [7]. Participating in multiple markets throughout the day enables greater profit capture by leveraging varying revenue opportunities across time and electricity markets. This requires dynamic bidding strategies that respond to real-time price signals and system conditions [61].

7.1. Battery Storage System Setup

The approach in this study is demonstrated through a case study of a Battery Energy Storage System (BESS) generating revenue via energy arbitrage. Energy arbitrage involves buying electricity at low prices and selling at higher prices to earn profit. To reflect realistic conditions, an 18 MWh battery with 8.5 MW power capacity is used, representing a system suited for both IDM and aFRR market participation [61]. The resulting power dispatch and energy level from intra-day trading are shown in Figure 7.1, with battery power output (green) varying every 15 minutes and energy level (orange) showing broader daily charge–discharge patterns. Here, the energy level is expressed as the battery's State of Charge (SoC), representing the energy currently stored in the battery in MWh. Positive power indicates discharging, which lowers the SoC, while negative power indicates charging, which increases it.

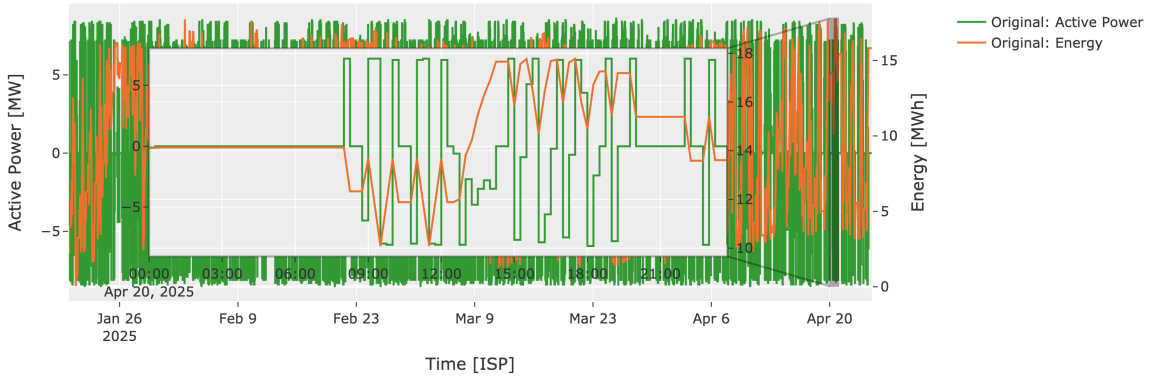


Figure 7.1: Battery active power for charging and discharging (green, left axis) and corresponding energy level or SoC (orange, right axis) during intra-day trading.

Building on this dispatch profile, a simplified model is developed to examine how market behaviour evolves under defined operational conditions.

7.1.1. Operational Assumptions

The model is based on the battery's active power and energy level, as these directly determine its trading behaviour in the IDM and aFRR markets. Charging corresponds to buying electricity, while discharging leads to selling electricity. By including only power output and energy level, the model abstracts from physical characteristics. Thereby isolating constraints such as thermal limits or contractual obligations. This simplification allows a focused analysis of market behaviour. To enable this approach, the following assumptions define the operational context:

- **Known Intra-day Schedule**: It is assumed that the battery's intra-day scheme is predetermined for the next 3 hours, allowing the available aFRR capacity to be reliably estimated for market participation.
- **ID3 Price Matching**: The intra-day price is assumed equal to the ID3 price. The ID3 price corresponds to the volume-weighted average price (VWAP) of pay-as-cleared transactions executed during the final three hours before delivery. The ID3 price serves as a proxy reflecting actual market conditions shortly before delivery. This assumption simplifies market interaction and revenue estimation, as all orders are considered matched. Mathematically defined as

$$p^{\text{ID3}} = \frac{\sum_{i=1}^{I} p_i q_i}{\sum_{i=1}^{I} q_i} \tag{7.1}$$

p^{ID3} : Volume-weighted average price over the last 3 hours before delivery
p_i : Price of intra-day trade i in the ID3 window
q_i : Volume of intra-day trade i in the ID3 window
i : Index of i-th trade $\in \{1, \ldots, I\}$
I : Total number of trades executed in the ID3 time window

- **Battery Constraints:**
  - *Preserved SoC*: Is not overwritten, which means that the energy level of the battery must stay between 0-18 MWh. This ensures that the battery does not charge beyond its maximum capacity or discharge below zero.
  - *Maximum energy capacity*: The maximum output power level of the battery is equal to the power capacity of 8.5 MW of the battery.
  - *aFRR buyback lead time*: Activated aFRR is assumed to be bought back three ISPs in advance, ensuring that the battery can still fulfil its already placed aFRR bids if they are accepted.
  - *Bid timing constraint*: Bids are only placed if the battery is expected to have sufficient capacity to deliver energy three ISPs in advance. This ensures reliable delivery upon activation and prevents market penalties for non-compliance.
  - *Battery efficiency*: Charging efficiency is fixed at 88.8%, while discharging is assumed to be lossless (100% efficiency). With this efficiency, the battery has a cumulative sum of 0, which keeps the battery within its SoC limits.
- **Full volume activation**: Entire submitted bid volume is always activated and utilised as the battery places relatively small bids compared to the total ladder size.
- **Neglected ramping**: Ramp rate limitations are neglected, therefore all activated volume can be sold as a whole [21].
- **Minimum bid granularity**: Transactions occur in minimal steps of 0.1 MWh, reflecting the minimum bid size allowed in the IDM [18].

With the modelling basis and assumptions in place, the next step is to compare different ways the battery can interact with the markets.

7.1.2. Price Scenarios

To assess the added value of aFRR market participation, four bidding strategies are proposed. Each reflects a different approach to when and how the battery engages in the aFRR market. These strategies aim to identify which price points on the aFRR bid ladder provide the most economic value and are therefore most effective for placing bids. The dispatch profile in Figure 7.1, which operates primarily in the IDM, serves as the reference for evaluating the added value of each strategy.

Strategies I and II apply predetermined price thresholds, chosen based on insights from the data analysis, such as the distribution of historical aFRR prices and the occurrence of high values in the aFRR curves. Strategy III builds on this approach by integrating the LKV forecast developed in this study, allowing for more dynamic and targeted bidding. Comparing these strategies helps determine whether incorporating aFRR forecasts improves economic performance relative to intra-day trading alone. It also highlights which bidding approach yields the highest returns in the aFRR market.

- **Strategy 0 — Intra-day profile:** The baseline strategy relies solely on intra-day trading. By comparing its results to those of strategies that include aFRR bidding, the incremental benefit of reserve market participation can be identified.
- **Strategy I — Constant price:** The first strategy uses a constant price, representing the marginal cost of a CCGT unit. This value, based on the estimated marginal price reported by [46], corresponds to the initial linear segment of the aFRR bid curves. By using a fixed price instead of dynamic bidding, the strategy aims to simulate frequent participation in the aFRR market.
- **Strategy II — Intra-day plus premium:** The second strategy builds upon the intra-day schedule by adding a price premium to aFRR bids. This ensures that any activation in the aFRR market remains profitable relative to the baseline intra-day operation. Additionally, the strategy allows the battery to benefit from downward regulation by earning revenue when charging during surplus situations. As a result, this approach reflects a more realistic bidding strategy aimed at safeguarding IDM profits while capturing extra value from reserve market participation.
- **Strategy III — LKV price:** This strategy targets price peaks by submitting bids around the knick point (LKV) of the aFRR curve, as identified in this study. These price points are based on the benchmark predictions evaluated three hours before delivery. This model is selected for the best performance compared to the other methods developed and discussed in Chapter 6. Building on this foundation, the strategy aims to test whether selectively bidding at high-value points can enhance profitability without relying on frequent market activation.

RQ4: *Forecasting insights from the aFRR bid ladder reveal distinct price levels that inform short-term trading decisions. First, the CCGT marginal cost reflects the flat part of the bid curve (Strategy I). Building on this, adding a premium to intra-day prices ensures profitability while enabling flexible market participation (Strategy II). Finally, identifying the LKV inflection point through forecasting allows targeting high-price events (Strategy III).*

The established price points define, at what prices to bid in the aFRR market. Based on this logic, a trading framework is mapped out to integrate aFRR participation into the intra-day dispatch and link bidding decisions to operational battery behaviour.

7.2. Design of Trading Framework

The trading control framework is illustrated in Figure 7.2 and consists of three steps. (1) Calculating the available energy capacity for submitting aFRR bids, or the energy required to offset previous aFRR activations. (2) Checking whether the battery can deliver or absorb energy within one ISP, based on its power and energy limits. (3) Adjusting the battery's power setpoint by modifying the intra-day schedule when feasible.

The adjustment $\Delta E(n)$ represents the activated aFRR energy and determines deviation from scheduled battery power. This energy imbalance informs the framework's actions: submitting a new bid or compensating earlier activations through energy buyback or resale. These adjustments apply to the power profile per ISP (n), defined by baseline power $P(n)$ and energy level $E(n)$ (Figure 7.1). Depending on the direction and magnitude of $\Delta E(n)$, the framework distinguishes three possible imbalance states: $E = 0$, $E > 0$, and $E < 0$.

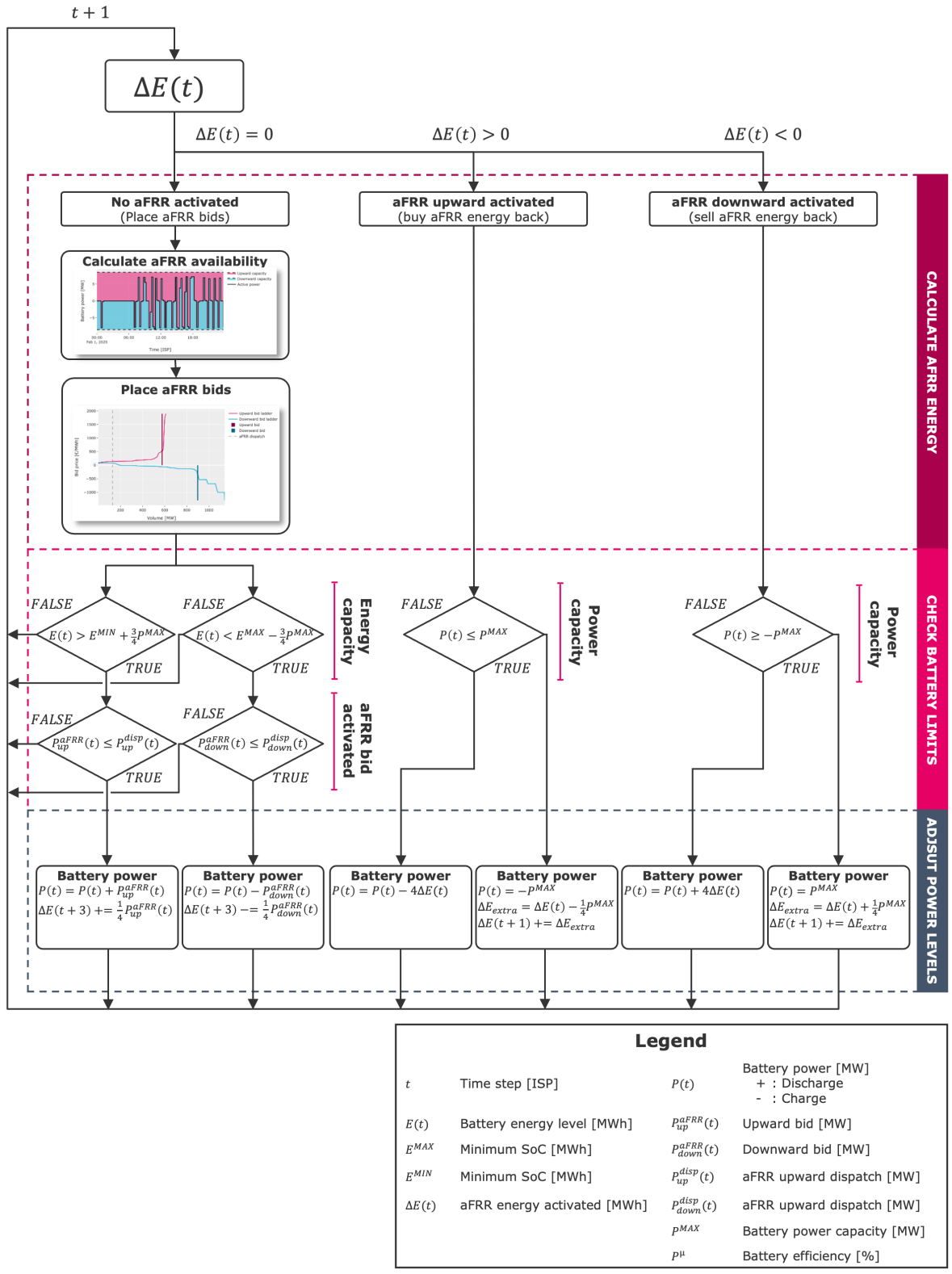


Figure 7.2: Control flow diagram illustrating battery dispatch logic where aFRR bid behaviour dynamically adjusts on top of an intra-day schedule based on activation signals and system constraints.

No aFRR Activation (ΔE = 0)
When no aFRR activation is recorded and the battery has remaining capacity, the available bid volume is calculated using Equation 7.2. This ensures that only the remaining headroom and footroom are offered in the upward and downward markets, respectively. A visual illustration of this is provided in Appendix Figure F.1.

$$P_{\text{up}}^{\text{aFRR}}(n) = P^{\text{MAX}} - P(n) \tag{7.2a}$$
$$P_{\text{down}}^{\text{aFRR}}(n) = -P^{\text{MAX}} - P(n) \tag{7.2b}$$

Bids are submitted only if the battery holds enough energy to deliver maximum power for the next three consecutive ISPs. Since bids must be placed 30 minutes (i.e., 3 ISPs) in advance, energy recovery is deferred to $n + 3$. This constraint ensures compliance and preserves availability for potential future activations

$$E(n) > E^{\text{MIN}} + \frac{3}{4}P^{\text{MAX}} \tag{7.3a}$$
$$E(n) < E^{\text{MAX}} - \frac{3}{4}P^{\text{MAX}} \tag{7.3b}$$

When aFRR bids are activated, the battery adjusts its power output accordingly: upward activations lead to an increase in output (equation 7.4), while downward activations result in a decrease (equation 7.5). These deviations from the scheduled output lead to an exchanged energy quantity $\Delta E(n)$, which is subsequently compensated through intra-day trading to restore the original baseline schedule. For an illustration of a bid location placement, refer to Appendix Figure F.2

$$P(n) = P(n) + P_{\text{up}}^{\text{aFRR}}(n) \tag{7.4a}$$
$$\Delta E(n+3) \mathrel{+}= \frac{1}{4}P_{\text{up}}^{\text{aFRR}}(n) \tag{7.4b}$$

$$P(n) = P(n) - P_{\text{down}}^{\text{aFRR}}(n) \tag{7.5a}$$
$$\Delta E(n+3) \mathrel{-}= \frac{1}{4}P_{\text{down}}^{\text{aFRR}}(n) \tag{7.5b}$$

Upward aFRR Activation (ΔE > 0)
When upward regulation is activated, energy must be absorbed, effectively requiring the battery to buy back this energy volume. If the power margin allows, the energy is fully recovered at once. The battery checks whether the energy deviation ($\Delta E(n)$) can be restored within one ISP. This condition is evaluated as

$$4\Delta E(n) \leq P^{\text{MAX}} \tag{7.6}$$

If the condition is satisfied, the battery adjusts its output accordingly

$$P(n) = P(n) - 4\Delta E(n) \tag{7.7}$$

If the condition is not satisfied, the battery restores as much energy as possible using its maximum output, and shifts the remaining energy to the next ISP. In this case, no new aFRR bids are placed until the energy balance has been restored and the battery is back on its intra-day schedule

$$P(n) = -P^{\text{MAX}} \tag{7.8a}$$
$$\Delta E_{\text{extra}} = \Delta E(n) - \frac{1}{4}P^{\text{MAX}} \tag{7.8b}$$
$$\Delta E(n+1) \mathrel{+}= \Delta E_{\text{extra}} \tag{7.8c}$$

Downward aFRR Activation ($\Delta E(n) < 0$)
In the case of downward aFRR activation, the battery absorbs energy by magnifying its discharge power. This process is symmetric to the upward activation but in the opposite direction. As in the upward case, energy recovery is delayed by 3 ISPs due to bid timing constraints. To determine if the full energy volume can be compensated within one ISP, the following constraint is checked

$$4\Delta E(n) \geq -P^{\text{MAX}} \tag{7.9}$$

If the required power adjustment lies within the battery's discharge capacity, the recovery is performed immediately

$$P(n) = P(n) + 4\Delta E(n) \tag{7.10}$$

If this is not feasible, the battery discharges at maximum power. In this case, the remaining energy is shifted to the next ISP, and no new aFRR bids are placed until the energy balance is fully restored in the intra-day schedule

$$P(n) = P^{\text{MAX}} \tag{7.11a}$$

$$\Delta E_{\text{extra}} = \Delta E(n) + \frac{1}{4} P^{\text{MAX}} \tag{7.11b}$$

$$\Delta E(n+1) \mathrel{+}= \Delta E_{\text{extra}} \tag{7.11c}$$

The full trading logic, covering bid placement, activation, and energy recovery, is defined. Based on this implementation, the performance of each strategy can be assessed using the following results.

7.3. Results of Trading Framework

The battery's corresponding energy dispatch profiles are shown in Appendix Figure F.3. As an illustrative example, particular focus is given to the performance of Strategy III, which is developed in this study using the LKV-based bidding approach. To illustrate the impact of aFRR activations in more detail, a zoomed-in view of Strategy III is provided in Figure 7.3. In this figure, the activated aFRR bids are marked by vertical pink and blue dashed lines. These activations result in deviations from the baseline schedule, shown by the dashed energy and power profiles, and lead to additional energy requirements indicated by the red line.

The analysis begins by examining aFRR bid activation, focusing on its impact on additional revenue, incurred costs, and energy deployment. These indicators capture the economic returns and operational effort associated with each strategy. Understanding this is important for assessing the overall Profit & Loss (P&L) performance of the combined trading approaches, as detailed in Subsection 7.3.2.

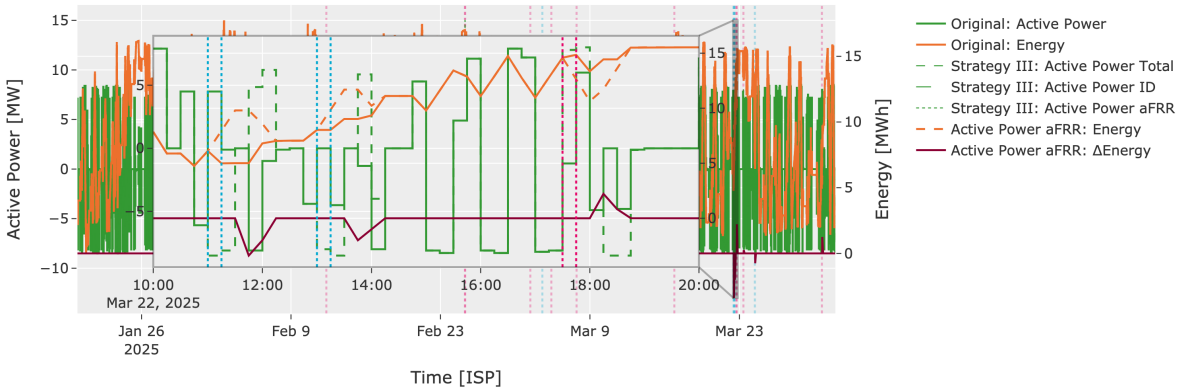


Figure 7.3: Battery energy levels under pricing strategy III: power output (green), battery energy (orange), and energy delta (red). Solid lines show strategy 0 versus the dashed strategy III.

To illustrate how these energy adjustments translate into market actions, the following Subsection examines aFRR bid activation and the resulting energy arbitrage.

7.3.1. aFRR Bid Activation

Energy arbitrage under Strategy III is illustrated in Figure 7.4, which visualises the battery's additional trades in the IDM (ID3) and the aFRR market, based on shortage and surplus prices. These prices reflect the direction of activation: shortage prices apply when the battery delivers energy (upward dispatch), and surplus prices apply when it absorbs energy (downward dispatch). The resulting trades are marked with triangles. Upward-pointing triangles represent upward market actions, while downward-pointing triangles indicate downward actions. Green triangles show energy sold, and red triangles show energy repurchased. Together, they illustrate how market signals drive battery behaviour.

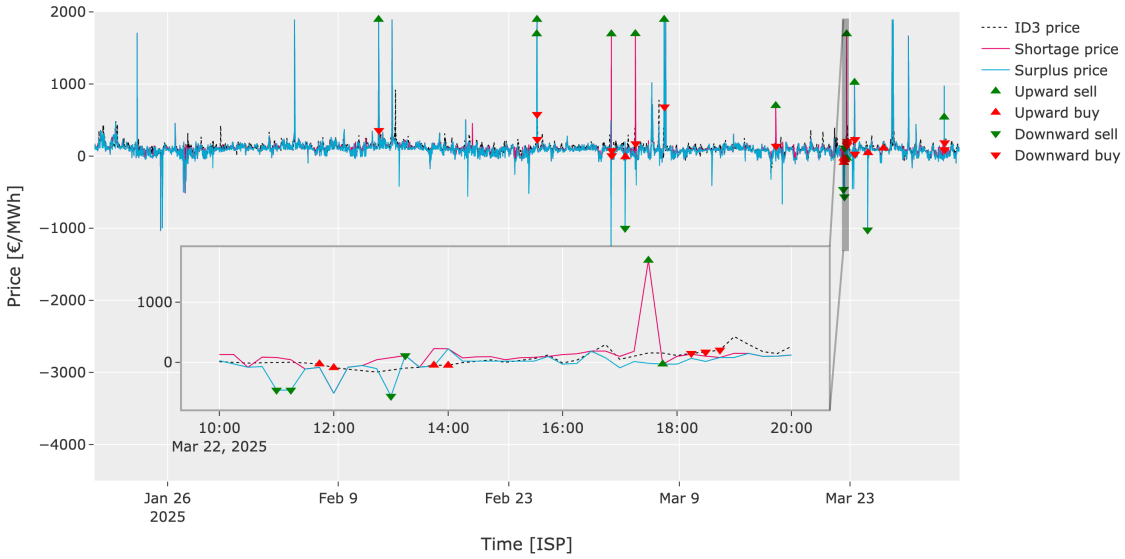


Figure 7.4: Energy arbitrage under Strategy III: ID3 (black dotted), shortage price (pink), surplus price (blue), with triangles for energy sold (green) and bought (red).

This Subsection considers only the additional impact of aFRR participation. It shows how extra trades influence total revenue, costs, and energy deployment over the full dataset. Appendix Figure F.4 displays the cumulative effects, with lollipop markers indicating activation times and their corresponding contributions. These cumulative results is are shown in Table 7.1.

Table 7.1: Cumulative aFRR participation revenue, cost, and energy deployment for each strategy, with results shown separately for upward, downward, and total.

	Revenue [k€]			**Cost [k€]**			**Energy [MWh]**		
	Upward	Downward	Total	Upward	Downward	Total	Upward	Downward	Total
Strategy I	142	7	150	143	0	143	905	17	922
Strategy II	54	9	63	33	-26	6	266	1073	1338
Strategy III	37	8	45	7	0	7	26	14	41

aFRR Revenue

Appendix Figure F.4a presents the shortage and surplus prices and total cumulative aFRR revenue for Strategies I to III. Strategy I shows the highest total revenue, driven by frequent upward activations yielding €142k, though often at moderate prices. Downward revenue remains limited at €7k. Strategy II shifts focus toward downward participation, resulting in €54k from upward and €9k from downward dispatch. Strategy III effectively captures extreme price events, generating €37k upward and €8k downward revenue. These results highlight the trade-off between frequent participation and selective, high-value trading. Limited revenue is observed in the downward market at the given prices, attributed to the more moderate downward price trends and the reduced occurrence of regulation state -1.

aFRR Cost

Appendix Figure F.4b shows the cumulative aFRR costs for Strategies I to III. Strategy I, with frequent participation and moderate buyback prices, results in high costs: €143k upward and €0k downward, leaving a limited margin relative to revenue. Strategy II also participates frequently, particularly downward, incurring €33k in upward costs but gaining €26k from negative downward costs, indicating profit from selling energy during downward activations. Strategy III engages selectively at low prices, resulting in minimal costs: €7k upward and €0k downward. These outcomes highlight how selective participation and price sensitivity can reduce exposure to high trading costs. Moreover, it is observed that selling back into the downward market can yield negative costs, thereby providing an additional source of revenue.

aFRR Energy

Appendix Figure F.4c illustrates the total aFRR energy deployed for Strategies I to III. Strategy I relies heavily on upward dispatch, selling 905 MWh while absorbing only 17 MWh, reflecting its frequent upward participation. Strategy II shows the opposite trend, with significant downward activation 1073 MWh and moderate upward energy 266 MWh, aligning with its focus on downward trading. Strategy III uses the least energy overall, with only 26 MWh upward and 14 MWh downward, demonstrating a selective and energy-efficient approach. More frequent bidding leads to higher energy dispatch. Selective participation reduces overall energy usage, resulting in a more efficient and targeted operational profile that can extend asset longevity. This is also visualised in Figure 7.5, where the higher value per unit of energy traded can be observed. The bar chart compares upward, downward, and total marginal profits across three different bidding strategies, highlighting the differences in profitability per unit of energy traded.

RQ5: *These results highlight the trade-off between frequent participation that leverages steady price levels and selective, high-value trading that capitalises on price spikes. The limited revenue in the downward market results from moderate downward price trends and the infrequent occurrence of regulation state -1. Still, the findings show that high profits are possible with low energy deployment when trading targets extreme price events. Strategy III illustrates this clearly: it activates infrequently, uses little energy, yet captures a large share of revenue by focusing on high-value moments. On average, it earns around 900 €/MWh per cycle, compared to only 8 €/MWh for Strategy I and 42 €/MWh for Strategy II. This makes it not only the most conservative operationally, but also the most effective in converting price volatility into profit.*

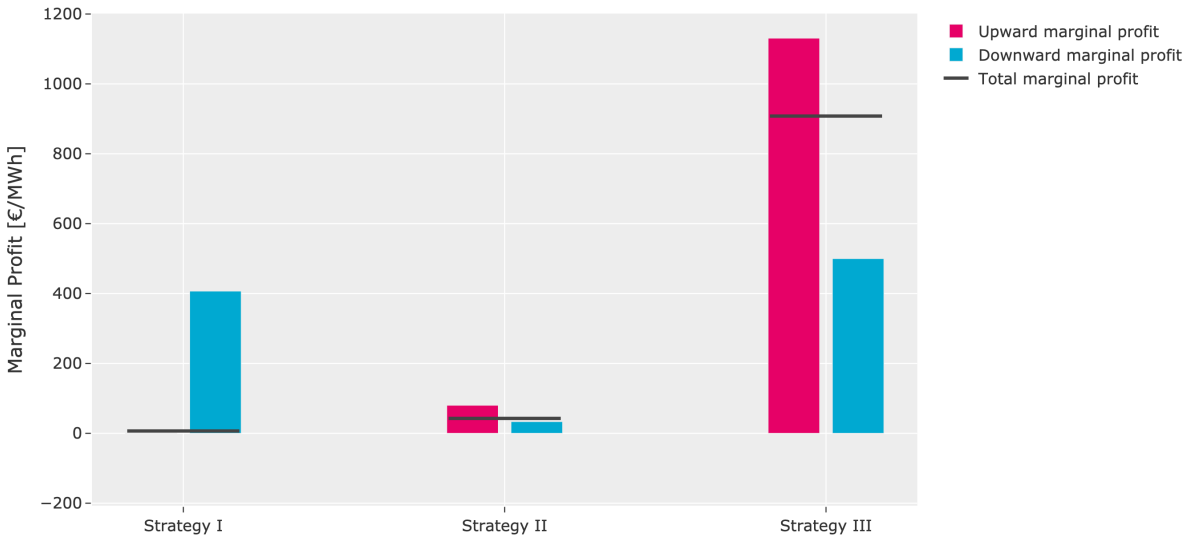


Figure 7.5: Marginal profit by strategy, calculated as the difference between selling and repurchase prices, with bars showing upward (pink), downward (blue), and total (black).

The preceding Subsection quantified the standalone impact of aFRR participation on revenue, costs, and energy deployment. The following part evaluates the total financial effect by integrating these results with the underlying intra-day trading performance.

7.3.2. Profit and Loss

The total profit and loss of the four strategies are illustrated in Figure 7.6, which shows the cumulative evolution of revenue, cost, and profit over time. This provides insight into how each strategy performs financially throughout the simulated period. Table 7.2 complements this by presenting the total cumulative values, allowing for a direct comparison of the overall financial performance.

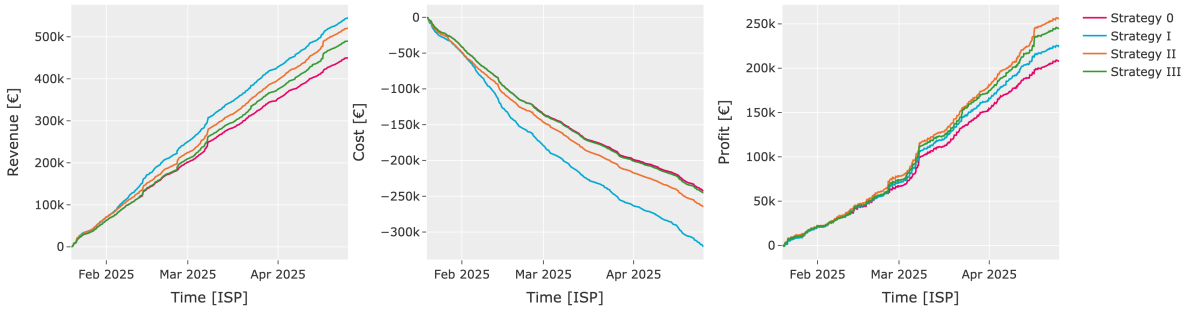


Figure 7.6: Revenue, cost, and profit for all proposed strategies, with Strategy 0 (pink), Strategy I (blue), Strategy III (orange), and Strategy IV (green).

Table 7.2: Total cumulative revenue, cost and profit of all four proposed strategies.

	Revenue [k€]	Cost [k€]	Profit [k€]
Strategy 0	450	-242	209
Strategy I	546	-321	226
Strategy II	520	-264	257
Strategy III	490	-245	246

Starting with the baseline, Strategy 0 yields the lowest profit (€209k), serving as a reference point without dynamic reserve market participation.

Strategy I achieves the highest revenue (€546k) by aggressively engaging in aFRR activations, but this comes with the highest cost (€321k), which limits its net profit to a moderate €226k. This reflects a high-volume, high-cost approach where frequent market participation drives revenue but also increases operational expenses and risk exposure.

Strategy II adopts a more balanced approach, generating slightly lower revenue (€520k) but substantially reducing costs to €264k. This results in the highest profit (€257k) among all strategies, demonstrating that pricing aFRR bids with a premium to moderate activation frequency can optimise the balance between market exposure and cost efficiency, thereby maximising profitability.

Strategy III focuses on selective bidding near peak prices using the LKV model, achieving a profit of €246k with a revenue of (€490k) and the lowest cost (€245k) among the aFRR-integrated strategies. This indicates that targeting high-value activations can deliver competitive profits while minimising costs and operational wear, offering a more conservative and cost-efficient alternative to frequent participation.

RQ5: *All three aFRR-integrated strategies outperform the baseline in terms of profit, confirming that dynamic reserve market participation enhances the economic viability of battery operation. Strategy II and Strategy III achieve similar profit levels, but Strategy III does so with notably lower cost. This indicates that targeting peak prices with selective bidding, as done in Strategy III, offers a more conservative and cost-efficient approach while maintaining strong financial performance. This resulted in an overall profit increase of 18 percent.*

8

Discussion

This chapter presents the key findings from the model development and trading strategy, which serve as the basis for examining their implications for the Dutch energy market. Building on this analysis, the findings are placed within a broader system context to assess their practical relevance. Finally, the chapter reflects on the methodological limitations that have influenced the results, including model simplifications and assumptions about market behaviour.

8.1. Interpretation of Results

This section summarises the key findings from both the forecasting model evaluation and the proposed trading strategy, providing a concise overview of the main results. For each analysis, it explains the main results, highlights what approaches worked well, and discusses what did not perform as expected.

Model Development

Chapter 6 demonstrates the application of the developed models to forecast aFRR bid ladders in the Dutch electricity market. Based on these forecasts, it then presents the results on multiple evaluation metrics, statistical tests, and computational efficiency.

Although the developed models (LASSO, XGBOOST, LSTM) did not outperform the benchmark, the results offer valuable insights into aFRR price predictability. Since the benchmark relies solely on the preliminary bid ladder available three hours before delivery, it already incorporates most relevant predictive information. As a result, real-time market transparency dominates aFRR price formation, limiting the added value that more sophisticated models can provide. This holds for both principal component prediction and after bid ladder reconstruction. As a result, relying on actual market data offers a more reliable and effective basis to support operational decision-making

The benchmark model delivers the best performance across all evaluation metrics, including sMAPE, LKV, and the Diebold-Mariano test. As shown in Tables 6.2a and 6.2b, it achieves the lowest sMAPE scores for both upward and downward bid curves, with values of 0.07 and 0.08, respectively. Furthermore, the Diebold-Mariano test results in Figure 6.3 confirm that the benchmark statistically outperforms the other models across most principal components. Taken together, these findings indicate that the benchmark provides forecasts that are consistently closer to the actual values.

Compared to all developed models, XGBOOST performed the worst. It showed higher forecast errors and less precise curve localisation. The Diebold-Mariano test revealed no statistically significant advantage. In contrast, while the LSTM model achieves forecast accuracy comparable to simpler methods such as LASSO, it does so at the expense of substantially increased model complexity and computational cost. As detailed in Table 6.3, hyperparameter tuning for LSTM can require up to four hours, highlighting a trade-off between accuracy and efficiency.

Although the proposed models did not yield any performance gains, several modelling decisions contributed positively to forecasting full aFRR bid curves:

- *Transposing the bid ladder curves:* Following the approach proposed by [50], this transformation addressed the issue of variable curve lengths across time, as shown in Figure 4.3. It enabled the standardisation of input features to a fixed volume-price domain, which in turn facilitated more stable and consistent model training.
- *Scaling of input features:* Applying feature scaling significantly improved performance by mitigating the effects of large absolute deviations in the curves, as discussed in Subsection 4.2.3 and illustrated in Figure 4.7. Similar improvements through feature scaling are also reported by [25].
- *Inclusion of preliminary bid ladders:* The inclusion of preliminary bid ladders proved a valuable data source. This value is reflected by the high feature importance for LASSO and XGBOOST models (see Appendix D.2 and D.3). The benchmark also demonstrates strong performance across nearly all evaluation metrics, achieving lower sMAPE scores (Table 6.2a) and statistical superiority (Figure 6.3).

As previously mentioned, the developed models did not outperform the benchmark. This outcome can be attributed to several limitations that most strongly affected model performance:

- *Low correlation between input features and the target:* One key issue is the weak relationship between input features and the target variable. This limitation reduced the predictive power of the models, as shown in Figures 4.5 and 4.8. The limited added value of such models is particularly evident when the input data lacks strong explanatory features. Similar observations have been made by [36] for the German aFRR market and by [35] for the British balancing market.
- *Reduced explainability due to PCA:* Although it effectively reduced dimensionality, it also obscured relationships between individual features and the target variable by transforming the original bid ladder inputs into increasingly abstract components. As shown in Table 5.1, interpretability diminished with higher component indices. Since these components captured less variance, including more led to declining predictive performance (see Figure 6.1), consistent with the findings of [25].
- *Overfitting to preliminary bid ladders:* Another critical drawback is overfitting to the preliminary bid ladders. Due to the high level of market transparency three hours in advance, the models increasingly relied on these inputs, which in turn limited their ability to generalise. This overreliance is reflected in the feature importance of both the LASSO and XGBOOST models (Figures D.2 and D.3), where the preliminary PCA components exhibit large magnitudes.

Trading Strategy

Chapter 7 demonstrates the economic potential of using aFRR bid ladder forecasts in short-term trading. Four strategies are evaluated: a baseline intra-day trading strategy (Strategy 0); a constant-price aFRR bidding strategy based on marginal CCGT costs (Strategy I); an intra-day strategy with a premium ensuring profitable activations by bidding at the IDM price plus a premium (Strategy II); and a peak-targeting strategy placing bids around the LKV to capture rare but highly profitable upward price spikes identified by BENCHMARK forecasts (Strategy III). Strategies involving aFRR participation outperform the baseline, confirming that reserve market bidding adds value beyond intra-day trading.

The evaluation of the trading framework in Section 7.3 identified Strategies II and III as the most effective approaches. These strategies achieve strong performance by targeting high-value segments of the aFRR bid ladder.

As shown in Figure F.4a, Strategy III often sells energy around €1900 to €1300 per MWh. Following such activations, the battery typically recharges on the IDM at much lower prices, commonly between €200 and €300 per MWh. The balance between targeted aFRR participation and intra-day flexibility makes Strategy III effective. Its low energy deployment preserves flexibility for intra-day trading, enabling the battery to benefit from arbitrage without incurring degradation costs from frequent aFRR activations.

In contrast, Strategy I performs the worst among the tested approaches. It places static bids based on estimated marginal CCGT costs, which leads to frequent activations. However, by ignoring IDM price uncertainty, the strategy lacks control over the eventual cost of recharging the battery. This results in an uncertain and often narrow profit margin, making many activations only marginally beneficial or unprofitable, despite the high level of participation.

Several aspects of the proposed trading Strategy III are found to be successful:

- *Generation of economic value beyond intra-day trading:* Most notably, the approach demonstrated the potential to generate economic value that extends beyond the conventional intra-day trading scheme. The participation in the aFRR market led to a significant revenue increase, while the associated costs remained comparatively modest. As a result, overall profits increased by 18%.
- *Monetisation of peak revenue periods:* Moreover, this extra value capture is achieved by monetising peak revenue periods, which emerged as key contributors to overall profitability. The ability to anticipate these peaks allowed for a strategic positioning of assets during the most lucrative intervals. Monetising these extreme price peaks allows the strategy to generate significant additional revenue with only a limited number of extra battery cycles.

Although the combination of aFRR bids and IDM trading increased overall economic value, Strategy III also contained instances of weak performance.

- *Limited profits in the downward trading strategy:* Specifically, the downward trading strategy contributed little to revenue generation. This is primarily due to its focus on capturing negative price spikes, which occurred less frequently and with lower financial impact. An alternative approach could involve using the downward strategy to opportunistically charge assets at favourable prices, rather than solely attempting to monetise negative price tails.

These findings not only inform model development and trading strategy design but also carry broader relevance for market participants and system operators.

8.2. Implications to the Dutch Energy Market

A novel, data-driven methodology for forecasting aFRR bid ladders in the Dutch electricity market is introduced, representing the first known attempt to model these bid curves. This approach lays the groundwork for more accurate and dynamic market analysis.

To start, the study provides detailed insights into the Dutch aFRR market by investigating how bid ladders are formed and priced. It places particular emphasis on the Netherlands' unique dual pricing system. By using live market data for forecasting, the research reveals how real-time bidding behaviour interacts with market fundamentals.

In addition, the forecasting approach supports market actors in their trading decisions. It delivers forecasts up to three hours before market closure, which matches the operational timelines of BSPs and BRPs. This advance information allows BRPs to optimise portfolio balancing and reduce imbalance costs. At the same time, it enables BSPs to adjust their bidding strategies in the aFRR market, such as repositioning flexible assets such as batteries to provide reserve capacity when price signals are favourable. The lead time also provides flexibility to reposition assets across DAM, IDM, and IM, helping participants to capture value from multiple market segments.

From a system perspective, aFRR bid ladder forecasting facilitates the integration of intermittent renewables by providing earlier insight into balancing needs. As wind and solar capacity increase, the variability in supply grows, leading to a higher risk of supply-demand mismatches. Earlier forecasts allow market participants to anticipate these imbalances and dispatch flexible assets, such as batteries, prior to real-time delivery. Consequently, this thesis demonstrates that three-hour-ahead forecasts enable the scheduling of asset behaviour across aFRR and IDMs, thereby shifting surplus renewable electricity into markets where it retains economic value. By aligning asset operations with the forecasted bid ladders, asset utilisation is optimised, resulting in improved system efficiency. For example, renewable curtailment can be reduced by charging batteries during periods of excess generation instead of wasting the surplus energy. This, in turn, contributes to national decarbonisation objectives and supports secure grid operation.

While the results offer valuable insights and practical applications, they are subject to several methodological and contextual limitations that should be considered.

8.3. Limitations

The forecasting framework and proposed trading strategy are subject to modelling and trading strategy constraints.

Model

Despite the provided result, the developed forecasting framework is subject to several limitations:

- *Data window:* The dataset covers only four months (January–May 2025), limiting model reliability. First, summer and autumn are missing, so seasonal effects such as solar peaks or holiday demand shifts are not captured. Second, the short timeframe yields a small validation and test set. As a result, rare but potentially important conditions, such as large imbalance volumes or mFRR activations, can be underrepresented, reducing model robustness.
- *Feature set and input uncertainty bias:* The current model operates with a static feature set, using the same predefined input variables for all models. This restricts the model's ability to capture complex or nonlinear combinations of features that could contain additional predictive information and improve model performance. Moreover, the input data is based on forecasted variables, primarily sourced from [46], such as weather and production predictions. As these sources are not cross-validated against independent data, any systematic errors in the input forecasts can propagate through the model, compounding uncertainties and limiting forecasting accuracy.
- *LSTM hyperparameter space:* The hyperparameter configuration for the LSTM model follows the reference study [25] to ensure comparability. Forecasting bid ladders requires high-dimensional outputs, as each curve contains many price-volume points, which increases training time and memory usage. Since forecasts are generated for every ISP, the hyperparameter search space was kept relatively small compared to the model's complexity to maintain computational feasibility. As a result, the model cannot be fully optimised, and some better-performing configurations may be missed.
- *Prediction interval widths:* While conformal prediction provides valid uncertainty estimates, the intervals are often too wide for practical use. This is mainly due to the shifting knick point, which introduces irregular patterns across price levels. The model struggles to capture these fluctuations, resulting in wider intervals, especially in the volatile middle section of the price curve (see Figure 6.4). Moreover, ICP produces constant-width intervals, limiting adaptation to local variability. In principle, quantile regression offers more adaptive intervals, but since estimates are derived in the reduced PCA space, the inverse transformation distorts them, making quantile modelling in the original price space impossible.

Market

Beyond methodological considerations, the Dutch aFRR market itself presents structural and regulatory limitations that affect the forecasting application:

- *Limited market liquidity:* The Dutch aFRR market is relatively small in scale, with a limited number of active BSPs. The total bid volume is low compared to larger balancing markets, such as those in Germany or the cross-border PICASSO platform. As a result, single large bids can significantly influence clearing prices and shape the structure of the bid ladder. These dynamics introduce discontinuities and irregularities that are difficult to model and capture through standard features. Moreover, such bidding behaviour is often driven by internal portfolio considerations of individual energy companies, which are not publicly observable and therefore cannot be included in the forecasting framework.
- *ID3 price approximation:* The model uses the ID3 price as a proxy for IDM outcomes. While the ID3 reflects the volume-weighted average price over the last three hours before delivery, it does not represent the exact transaction price achievable in practice. Since the Dutch IDM operates on a pay-as-bid basis, participants may buy or sell electricity at prices above or below the ID3, depending on bid timing and market liquidity. As a result, the use of the ID3 in revenue and cost calculations introduces a structural simplification. This could overestimate or underestimate the actual profitability of battery trading strategies based on the available bids in the market.
- *Shortage and surplus price approximation:* In this study, these prices are estimated based on the regulation state, reconstructed bid ladders, and historical dispatch volumes. Due to technical constraints such as minimum bid sizes and ramping limits, TenneT activates more volume than strictly needed, resulting in higher absolute price levels than those modelled. As TenneT's internal dispatch logic remains undisclosed, these effects cannot be fully captured, leading to a structural approximation.

9 Conclusion

This chapter presents a summary of the main findings by addressing the research questions, outlining the contributions to the field, and proposing directions for future research.

9.1. Answer to the Research Questions

In this thesis, forecasting models (Conformal Lasso, XGBoost, and LSTM) are developed to support informed and profitable energy trading strategies in the Dutch electricity market. The following central research question guides the study:

How to effectively forecast automatic Frequency Restoration Reserve (aFRR) bid ladders, and how can these forecasts support energy trading companies in the Dutch electricity market?

The first step involves identifying the key external and market-specific factors that influence aFRR bid volumes and prices:

RQ1: What external factors, such as weather, generation, market, time factors, and historical bid ladders, influence the price and volume of aFRR bid ladders?

The key factors influencing prices and volumes for aFRR bid ladder include preliminary ladder data, wind generation forecasts, and capacity market conditions.

For volume, preliminary ladder volumes serve as significant predictors for both upward and downward bid volumes (see Appendix C). In the upward market, preliminary ladder volumes closely reflect final bid volumes, indicating early bidding intentions. The minimal required capacity further signals expected volume levels. Overall, upward volumes are mainly driven by internal operational strategies rather than external factors. In contrast, downward bid volumes are additionally sensitive to weather conditions, particularly wind forecasts (see Subsection 4.2.2).

For price, in both directions, early provisional bids show strong correlations with final price levels, underlining their predictive relevance (see Subsection 4.2.3). Upward bids at lower volumes are primarily driven by fossil-based marginal costs, such as coal, gas, and carbon price forecasts. At higher volumes, strategic bidding behaviour causes higher price bids. In contrast, downward bid prices are more influenced by renewable generation forecasts, especially wind and solar, leading to a broader and more variable price distribution.

These insights into price and volume determinants provide a foundation for forecasting model development and support the informed selection of input features for machine learning algorithms:

RQ2: Which machine learning models commonly applied in EPF can be used to forecast full aFRR bid curves?

LASSO, XGBOOST, and LSTM models in combination with conformal prediction are implemented to forecast entire aFRR bid curves. These models are commonly used in EPF literature, representing standard approaches for time series and market prediction tasks (see Section 3.1). This selection also allows for a comparison between different types of machine learning methods, including linear, tree-based, and deep learning techniques.

However, forecasting full bid curves is challenged by varying curve lengths and high output dimensionality. To ensure consistent input size and computational efficiency, preprocessing and decomposition techniques are required:

RQ3: What decomposition techniques can be used to forecast full aFRR bid curves to account for varying lengths and computational cost?

To forecast full aFRR bid curves while accounting for varying lengths and computational cost, a two-step decomposition approach is applied.

In the first step, transposition is used to align the bid curves along a common reference, which standardises the input despite differences in curve length (see Subsection 5.2.1). The upward curve is transposed to the [0-1900] €/MWh domain, and the downward curve to the [−1300,0] €/MWh domain. This preprocessing ensures matched dimensions before applying dimensionality reduction.

In the second step, PCA is employed as a decomposition technique to manage the high output size associated with forecasting full aFRR bid curves. PCA reduces dimensionality by capturing significant variance with fewer components, enabling forecasting across varying curve lengths. Both upward and downward curves are reduced to four principal components, capturing over 90% of the explained variance (Subsection 5.2.1 and Subsection 6.1.1).

After forecasting the bid curves, the most relevant information must be extracted to support decision-making. For this purpose, a new technique termed the Largest Knick Volume is developed to identify the point where prices begin to show extreme behaviour:

RQ4: What forecasting insights derived from aFRR bid ladders can be utilised to inform short-term trading decisions between intra-day and imbalance markets?

The LKV metric is introduced to identify the volume threshold at which bid curves begin to exhibit extreme price behaviour. This transition is detected by analysing the standardised price profile, where the LKV corresponds to the first volume point at which the z-score exceeds a value of one. Such a statistical deviation indicates the onset of significant volatility in the curve. Beyond this point, additional volumes are increasingly associated with disproportionately high or low prices. The LKV, therefore, marks a structural breakpoint that distinguishes normal bidding behaviour from extreme market responses (see Subsection 5.4.2).

With the ability to forecast full bid curves and detect critical thresholds such as the LKV, the developed models provide direct input for trading strategies:

RQ5: What is the economic value of incorporating aFRR price forecasts into short-term trading strategies for profit optimisation?

Incorporating aFRR price forecasts into short-term trading strategies offers additional economic value. These forecasts enable more flexible and profitable operation of assets like batteries, optimising arbitrage by responding to expected prices and activation signals. Participating in both intra-day and aFRR markets unlocks additional value through price spreads and activation payments that static strategies often miss. As aFRR prices spike to 1900 or -1300 €/MWh at the curve's extremes, high-price bids yield substantial margins when activated. Advanced strategies, especially Strategy III, can increase revenues by up to 18% per unit of traded energy (see Subsection 7.3.2).

In summary, this thesis develops a structured forecasting framework by identifying key bid ladder drivers, selecting suitable machine learning models, applying dimensionality reduction, and introducing the LKV metric. While forecasting models are applicable, their added value decreases close to delivery, as most relevant information is already available within three hours. Still, bid curve data remains valuable for trading strategies. Bidding at higher prices can yield substantial aFRR revenues, often surpassing intra-day gains, with Strategy III showing strong potential and minimal additional cycling. Beyond enhancing short-term trading performance, the framework also contributes to the broader integration of renewable energy by improving the anticipation of system imbalances. Thereby the dependence on fossil-based reserves is reduced and a more stable and sustainable electricity system is supported.

9.2. Contributions to the Field

Building on the answers to the research questions, the following academic contributions highlight how this work advances the understanding and application of forecasting in the aFRR market:

- *Application of forecasting models in the Dutch aFRR market*: This study develops a forecasting model (LASSO, XGBOOST, and LSTM) for the Dutch aFRR market. It addresses a significant research gap due to limited academic attention on this specific market, as detailed in Section 3.2. The scarcity of existing literature on forecasting methodologies enables this work to make a novel contribution. This contribution advances understanding of ancillary services market dynamics and predictive modelling approaches.
- *Forecasting entire bid curves instead of point prices*: By forecasting complete bid ladders rather than individual price points, the approach enables a more granular and comprehensive understanding of market dynamics, proving that the market exhibits limited variation three hours ahead of delivery.
- *Integration of principal component analysis with machine learning models*: PCA is employed to reduce the dimensionality of high-resolution bid curve data output. This dimension reduction improves computational efficiency. However, it introduces a trade-off, as it decreases the interpretability of individual curve components, making the learning task more abstract and potentially more challenging for the model.
- *Development of a novel evaluation metric ("Largest Knick Volume")*: A metric is introduced to assess how well models capture critical inflection points in bid curves. It focuses on the segment where prices start changing sharply. This adds value beyond standard error-based metrics.
- *Strategic relevance for energy companies*: The forecasting outputs are directly applicable to intra-day trading strategies. They help energy utilities anticipate price movements and manage operational uncertainty. Accurate aFRR forecasts also enable strategic participation in ancillary service markets. By combining positions across intra-day and aFRR markets, companies can diversify their trading activities. This increases potential profits and spreads risk across multiple revenue streams, improving overall portfolio stability.

While these contributions advance the current state of research and practice, several opportunities remain to further develop and refine the proposed forecasting and trading framework.

9.3. Future Work

Two directions for future research are proposed: model enhancements and trading integration. These aim to improve forecast performance and dive more into the technological constraints for applying the proposed strategy.

9.3.1. Model Enhancements

Several avenues exist to further enhance the developed forecasting framework.

Multi-output models: Forecasting the full bid curve in its original form preserves the interpretability of price-volume relationships. Dimensionality reduction methods like PCA simplify modelling but transform the data into latent components that lose this direct interpretation. Direct forecasting avoids reconstruction and keeps the bid structure fully transparent. It also allows for quantile-based forecasting, providing full predictive distributions. Neural network architectures are particularly suited for this task, as market constraints and bidding rules can be embedded directly into the loss function, producing realistic and rule-compliant bid curves.

Apply to other markets: The applicability of the proposed framework could be extended by applying the models to more liquid and data-rich ancillary markets, such as the German aFRR market or the European PICASSO platform. These markets offer higher data availability and trading volumes, which would allow for a more rigorous evaluation of model performance. In addition, such an extension would test the transferability of the approach across different market structures and regulatory environments.

Longer training periods: Model training can be extended to a longer training period. In the current study, the model is trained on a relatively short timeframe, which may not encompass the full range of market conditions. As a result, variations such as seasonal effects or periods of volatility may not have been adequately represented in the training set. By incorporating a longer training window, the model would be exposed to a more comprehensive set of market scenarios, potentially enhancing its generalisability and robustness.

Intra-day feature engineering: A more dynamic approach would incorporate more recent or intraday input data, such as updated energy deltas, cross-border transmission constraints, or real-time system imbalances, allowing the model to better capture short-term fluctuations in supply and demand. To further automate and optimise the feature generation process, an additional calibration step, applying feature selection techniques such as recursive feature elimination or tree-based importance measures, could automatically identify and retain the most informative features, reducing manual intervention and improving the feature selection procedure.

9.3.2. Trading Integration

With regard to trading applications, further research is recommended to operationalise the proposed strategy.

Implement strategy III: This strategy emerged as the most promising result in simulation. Therefore, this strategy can already be implemented to batteries acting only on the IDM. This implementation would allow for the assessment of practical constraints and economic feasibility under live market conditions. Additionally, deploying the strategy in a physical system would enable validation of simulation assumptions, account for real-time data variability, and reveal operational limitations. Insights gained could facilitate further refinement and optimisation of the strategy.

Investigate technical limitations: It is important to ensure that the developed strategies are not only theoretically sound but also practically feasible and operationally possible. Battery systems are subject to technical constraints, such as ramp rates, state-of-charge limits, and intra-day clearing prices. By integrating them into the forecasting and trading framework, the resulting strategies become more applicable to real-world conditions.

Optimise buy-back time: Additional research should focus on optimising the timing of energy repurchase during imbalance periods. The current strategy repurchases electricity immediately after use, without considering the intra-day price at that ISP. As a result, potential cost savings are overlooked. By timing the buy-back based on market prices, the battery can repurchase energy at lower prices. An optimisation model can be formulated where the objective is to minimise repurchase costs over the imbalance settlement period, subject to battery constraints (e.g., state-of-charge limits, ramp rates). For instance, linear programming or mixed-integer linear programming (MILP) solvers can be employed. The model can incorporate forecasts of intra-day prices and dynamically schedule buy-back operations when prices are expected to be lowest. This allows for capturing additional margins and improving overall profitability.

Combine strategies: Employing multiple forecasting-based strategies can enhance portfolio diversification, as it allows risks to be distributed across various predictive methods and market segments. This diversification is beneficial because each strategy reacts differently to market fluctuations. As a result, combining strategies inherently creates a hedge against volatility. If one approach underperforms due to prediction errors or unforeseen market developments, other strategies can help offset these losses. This complementary behaviour among strategies leads to more stable and consistent returns across different market conditions. For example, Strategy III performs well in the upward regulation regime due to its responsiveness to sharp price spikes. However, its profitability diminishes under downward market conditions. This limitation justifies the integration of a complementary strategy that performs more favourably in such downward markets, thereby reinforcing portfolio robustness through strategic complementarity.

References

[1] Severin Demchuk. *Photo of black transmission tower photo – Free Blue Image on Unsplash*. 2017. URL: https://unsplash.com/photos/photo-of-black-transmission-tower-60NulquhzoI (visited on 10/18/2024).

[2] Tennet. *Tennet: Ons verhaal*. nl. URL: https://www.tennet.eu/nl/over-tennet/over-tennet/ons-verhaal (visited on 09/30/2024).

[3] Centraal Bureau voor de Statistiek. *1886: Elektriciteit*. nl. Last Modified: 20-12-2024T06:30:00. Dec. 2024. URL: https://www.cbs.nl/nl-nl/visualisaties/tijdlijn-historische-energie/1886-elektriciteit (visited on 01/31/2025).

[4] European Commision. *Renewable Energy Directive*. en. URL: https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-directive_en (visited on 04/23/2025).

[5] *IEA: Netherlands Energy Mix*. en-GB. URL: https://www.iea.org/countries/the-netherlands (visited on 09/30/2024).

[6] Esther Bos. *Tennet: Onbalansprijssystematiek - hoe komen de geldstromen tot stand?* nl. Tech. rep. TenneT TSO B.V., Mar. 2022.

[7] Sanne de Boer. *Rabobank: Elektriciteitssector*. nl. URL: https://www.rabobank.nl/kennis/d011430987-de-nederlandse-elektriciteitssector-deel-4-kansen-en-risicos-voor-bedrijven-en-huishoudens-bij-veranderende-elektriciteitsmarkten (visited on 09/04/2024).

[8] Tennet. *Tennet: Integrated Annual Report 2023*. en. Tech. rep. 2023, p. 230. (Visited on 09/30/2024).

[9] Karolin Schaps. *Picasso to Have Limited Impact on Power Prices: Dutch TSO*. 2024. URL: https://montelnews.com/news/a2193d4e-de80-439f-8d8f-9edd20ac92e9/picasso-to-have-limited-impact-on-power-prices-dutch-tso (visited on 10/03/2024).

[10] Jethro Browell and Ciaran Gilbert. "Predicting Electricity Imbalance Prices and Volumes: Capabilities and Opportunities". en. In: *Energies* 15.10 (Jan. 2022). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 3645. ISSN: 1996-1073. DOI: 10.3390/en15103645. URL: https://www.mdpi.com/1996-1073/15/10/3645 (visited on 01/29/2025).

[11] Edge Branding. *Montel - Blog - Product update: Dutch balancing markets*. 2024. URL: https://montel.energy/blog/product-update-dutch-balancing-markets (visited on 01/31/2025).

[12] Gro Klæboe, Anders Lund Eriksrud, and Stein-Erik Fleten. "Benchmarking time series based forecasting models for electricity balancing market prices". en. In: *Energy Systems* 6.1 (Mar. 2015), pp. 43–61. ISSN: 1868-3975. DOI: 10.1007/s12667-013-0103-3. URL: https://doi.org/10.1007/s12667-013-0103-3 (visited on 01/29/2025).

[13] Holding B.V. Tennet. *Marktrollen*. nl. URL: https://www.tennet.eu/nl/de-elektriciteitsmarkt/nederlandse-markt/marktrollen (visited on 09/06/2024).

[14] Holding B.V. Tennet. *What kind of markets are there and how do they work? - TenneT*. English. URL: https://netztransparenz.tennet.eu/electricity-market/about-the-electricity-market/what-kind-of-markets-are-there-and-how-do-they-work/ (visited on 09/06/2024).

[15] Sanne de Boer. *Rabobank:* nl. URL: https://www.rabobank.nl/kennis/d011424506-de-nederlandse-elektriciteitssector-deel-2-hoe-werken-de-verschillende-elektriciteitsmarkten (visited on 10/03/2024).

[16] ICE ENDEX. *ICE Endex energy exchange for gas and power*. en. URL: https://www.ice.com/endex (visited on 05/08/2025).

[17] EEX Futures. *Power Futures*. en. URL: https://www.eex.com/en/markets/power/power-futures (visited on 05/08/2025).

[18] Epex Spot. *Trading Products | EPEX SPOT*. en. URL: https://www.epexspot.com/en/tradingproducts (visited on 05/08/2025).

[19] Trading Nordpool. *Day-ahead and intraday power trading*. URL: https://www.nordpoolgroup.com/en/trading/ (visited on 05/08/2025).

[20] TenneT TSO B.V. *Tennet: FCR handboek voor BSPs*. nl. Tech. rep. SOP-SYS FCR HB V3.3. Tennet B.V., Mar. 2023. URL: https://www.tennet.eu/nl/elektriciteitsmarkt/ondersteunende-diensten/documenten-fcr/.

[21] TenneT TSO B.V. *Tennet: aFRR Handboek voor BSPs*. Dutch. Tech. rep. SOP-SYS aFRR NL V2.0. Tennet B.V., Aug. 2023. URL: https://www.tennet.eu/nl/de-elektriciteitsmarkt/ondersteunende-diensten-nederland/documenten-afrr.

[22] TenneT TSO B.V. *Tennet: mFRR noodvermogen Handboek voor BSPs*. nl. Tech. rep. Versie 1.3. Tennet B.V., May 2024. URL: https://www.tennet.eu/nl/de-elektriciteitsmarkt/ondersteunende-diensten-nederland.

[23] Michał Narajewski. "Probabilistic Forecasting of German Electricity Imbalance Prices". en. In: *Energies* 15.14 (Jan. 2022). Number: 14 Publisher: Multidisciplinary Digital Publishing Institute, p. 4976. ISSN: 1996-1073. DOI: 10.3390/en15144976. URL: https://www.mdpi.com/1996-1073/15/14/4976 (visited on 01/29/2025).

[24] Florian Ziel and Rick Steinert. "Electricity price forecasting using sale and purchase curves: The X-Model". In: *Energy Economics* 59 (Sept. 2016), pp. 435–454. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2016.08.008. URL: https://www.sciencedirect.com/science/article/pii/S0140988316302080 (visited on 01/23/2025).

[25] Hongye Guo et al. "Forecast Aggregated Supply Curves in Power Markets Based On LSTM Model". In: *IEEE Transactions on Power Systems* 36.6 (Nov. 2021). Conference Name: IEEE Transactions on Power Systems, pp. 5767–5779. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2021.3079923. URL: https://ieeexplore.ieee.org/document/9430706 (visited on 01/27/2025).

[26] Miguel Pinhão, Miguel Fonseca, and Ricardo Covas. "Electricity Spot Price Forecast by Modelling Supply and Demand Curve". en. In: *Mathematics* 10.12 (Jan. 2022). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 2012. ISSN: 2227-7390. DOI: 10.3390/math10122012. URL: https://www.mdpi.com/2227-7390/10/12/2012 (visited on 01/23/2025).

[27] Rafał Weron. "Electricity price forecasting: A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30.4 (Oct. 2014), pp. 1030–1081. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2014.08.008. URL: https://www.sciencedirect.com/science/article/pii/S0169207014001083 (visited on 01/27/2025).

[28] Jakub Nowotarski and Rafał Weron. "Recent advances in electricity price forecasting: A review of probabilistic forecasting". In: *Renewable and Sustainable Energy Reviews* 81 (Jan. 2018), pp. 1548–1568. ISSN: 1364-0321. DOI: 10.1016/j.rser.2017.05.234. URL: https://www.sciencedirect.com/science/article/pii/S1364032117308808 (visited on 01/27/2025).

[29] Jesus Lago et al. "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark". In: *Applied Energy* 293 (July 2021), p. 116983. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.116983. URL: https://www.sciencedirect.com/science/article/pii/S0306261921004529 (visited on 09/30/2024).

[30] Priyanka Shinde and Mikael Amelin. "A Literature Review of Intraday Electricity Markets and Prices". In: (June 2019), pp. 1–6. DOI: 10.1109/PTC.2019.8810752. URL: https://ieeexplore.ieee.org/abstract/document/8810752 (visited on 01/28/2025).

[31] Priyanka Shinde, Iasonas Kouveliotis-Lysikatos, and Mikael Amelin. "Analysing Trading Trends in Continuous Intraday Electricity Markets". In: *2021 56th International Universities Power Engineering Conference (UPEC)*. Aug. 2021, pp. 1–6. DOI: 10.1109/UPEC50034.2021.9548168. URL: https://ieeexplore.ieee.org/abstract/document/9548168 (visited on 01/28/2025).

[32] Michał Narajewski and Florian Ziel. "Econometric modelling and forecasting of intraday electricity prices". In: *Journal of Commodity Markets* 19 (Sept. 2020), p. 100107. ISSN: 2405-8513. DOI: 10.1016/j.jcomm.2019.100107. URL: https://www.sciencedirect.com/science/article/pii/S2405851319300728 (visited on 01/28/2025).

[33] Panagiotis-Christos Kotsias. *Forecasting Electricity Prices for Intraday Markets with Machine Learning : An exploratory comparison of the state of the art*. eng. 2022. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-325056 (visited on 01/28/2025).

[34] Dane Birkeland and Tarek AlSkaif. "Research areas and methods of interest in European intraday electricity market research—A systematic literature review". In: *Sustainable Energy, Grids and Networks* 38 (June 2024), p. 101368. ISSN: 2352-4677. DOI: 10.1016/j.segan.2024.101368. URL: https://www.sciencedirect.com/science/article/pii/S2352467724000973 (visited on 01/28/2025).

[35] Alexandre Lucas et al. "Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression". en. In: *Energies* 13.20 (Jan. 2020). Number: 20 Publisher: Multidisciplinary Digital Publishing Institute, p. 5420. ISSN: 1996-1073. DOI: 10.3390/en13205420. URL: https://www.mdpi.com/1996-1073/13/20/5420 (visited on 01/29/2025).

[36] Michael Merten et al. "Automatic frequency restoration reserve market prediction: Methodology and comparison of various approaches". In: *Applied Energy* 268 (June 2020), p. 114978. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2020.114978. URL: https://www.sciencedirect.com/science/article/pii/S0306261920304906 (visited on 01/29/2025).

[37] Javier Cardo-Miota, Emilio Pérez, and Hector Beltran. "Deep learning-based forecasting of the automatic Frequency Reserve Restoration band price in the Iberian electricity market". In: *Sustainable Energy, Grids and Networks* 35 (Sept. 2023), p. 101110. ISSN: 2352-4677. DOI: 10.1016/j.segan.2023.101110. URL: https://www.sciencedirect.com/science/article/pii/S2352467723001182 (visited on 01/29/2025).

[38] Jonathan Dumas et al. "Probabilistic Forecasting of Imbalance Prices in the Belgian Context". In: *2019 16th International Conference on the European Energy Market (EEM)*. ISSN: 2165-4093. Sept. 2019, pp. 1–7. DOI: 10.1109/EEM.2019.8916375. URL: https://ieeexplore.ieee.org/abstract/document/8916375 (visited on 01/29/2025).

[39] Jérémie Bottieau et al. "Very-Short-Term Probabilistic Forecasting for a Risk-Aware Participation in the Single Price Imbalance Settlement". In: *IEEE Transactions on Power Systems* 35.2 (Mar. 2020). Conference Name: IEEE Transactions on Power Systems, pp. 1218–1230. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2019.2940756. URL: https://ieeexplore.ieee.org/abstract/document/8832203 (visited on 01/29/2025).

[40] Robert Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996). Publisher: [Royal Statistical Society, Oxford University Press], pp. 267–288. ISSN: 0035-9246. URL: https://www.jstor.org/stable/2346178 (visited on 01/23/2025).

[41] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: KDD '16 (Aug. 2016), pp. 785–794. DOI: 10.1145/2939672.2939785. URL: https://dl.acm.org/doi/10.1145/2939672.2939785 (visited on 01/24/2025).

[42] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural Networks* 61 (Jan. 2015), pp. 85–117. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2014.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0893608014002135 (visited on 04/15/2025).

[43] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

[44] Wouter Bles. *wouterbles/tennet-afrr-archive*. original-date: 2025-01-12T22:27:10Z. Apr. 2025. URL: https://github.com/wouterbles/tennet-afrr-archive (visited on 04/25/2025).

[45] Essent. *Essent Internal data*. 2025. URL: https://essent.nl/ (visited on 01/30/2025).

[46] EnAppSys. *EnAppSys*. 2025. URL: https://app.enappsys.com/ (visited on 01/30/2025).

[47] Tennet API. *TenneT API*. 2025. URL: https://tennet.nl/ (visited on 01/30/2025).

[48] Ksenia Poplavskaya et al. "Making the most of short-term flexibility in the balancing market: Opportunities and challenges of voluntary bids in the new balancing market design". In: *Energy Policy* 158 (Nov. 2021), p. 112522. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2021.112522. URL: https://www.sciencedirect.com/science/article/pii/S030142152100392X (visited on 06/03/2025).

[49] Mohammad Taye. “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions”. In: *Computers* 12 (Apr. 2023), p. 91. DOI: 10.3390/computers12050091.

[50] Zehang Li, Andrés M. Alonso, and Lorenzo Pascual. “Predicting electricity supply and demand curves with functional data techniques”. In: *International Journal of Electrical Power & Energy Systems* 166 (May 2025), p. 110561. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2025.110561. URL: https://www.sciencedirect.com/science/article/pii/S0142061525001127 (visited on 04/28/2025).

[51] Rasmus Bro and Age K. Smilde. “Principal component analysis”. en. In: *Analytical Methods* 6.9 (2014). Publisher: Royal Society of Chemistry, pp. 2812–2831. DOI: 10.1039/C3AY41907J. URL: https://pubs.rsc.org/en/content/articlelanding/2014/ay/c3ay41907j (visited on 04/11/2025).

[52] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD ’19. New York, NY, USA: Association for Computing Machinery, July 2019, pp. 2623–2631. ISBN: 978-1-4503-6201-6. DOI: 10.1145/3292500.3330701. URL: https://dl.acm.org/doi/10.1145/3292500.3330701 (visited on 04/02/2025).

[53] Glenn Shafer and Vladimir Vovk. *A tutorial on conformal prediction*. arXiv:0706.3188 [cs]. June 2007. DOI: 10.48550/arXiv.0706.3188. URL: http://arxiv.org/abs/0706.3188 (visited on 04/10/2025).

[54] Anastasios N. Angelopoulos and Stephen Bates. *A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification*. arXiv:2107.07511 [cs]. Dec. 2022. DOI: 10.48550/arXiv.2107.07511. URL: http://arxiv.org/abs/2107.07511 (visited on 04/10/2025).

[55] Jakub Nowotarski and Rafał Weron. “Recent advances in electricity price forecasting: A review of probabilistic forecasting”. In: *Renewable and Sustainable Energy Reviews* 81 (Jan. 2018), pp. 1548–1568. ISSN: 1364-0321. DOI: 10.1016/j.rser.2017.05.234. URL: https://www.sciencedirect.com/science/article/pii/S1364032117308808 (visited on 03/23/2025).

[56] Francis X. Diebold and Roberto S. Mariano. “Comparing Predictive Accuracy”. In: *Journal of Business & Economic Statistics* 20.1 (Jan. 2002). Publisher: American Statistical Association, pp. 134–144. ISSN: 0735-0015. DOI: 10.1198/073500102753410444. URL: https://doi.org/10.1198/073500102753410444 (visited on 04/18/2025).

[57] Amand F. Schmidt and Chris Finan. “Linear regression and the normality assumption”. In: *Journal of Clinical Epidemiology* 98 (June 2018), pp. 146–151. ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2017.12.006. URL: https://www.sciencedirect.com/science/article/pii/S0895435617304857 (visited on 05/16/2025).

[58] *Lasso*. en. URL: https://scikit-learn/stable/modules/generated/sklearn.linear_model.Lasso.html (visited on 04/17/2025).

[59] *XGBoost Documentation — xgboost 3.0.0 documentation*. URL: https://xgboost.readthedocs.io/en/stable/index.html (visited on 04/17/2025).

[60] *tf.keras.layers.LSTM | TensorFlow v2.16.1*. en. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/LSTM (visited on 04/17/2025).

[61] David Schaurecker et al. *Maximizing Battery Storage Profits via High-Frequency Intraday Trading*. en. arXiv:2504.06932 [q-fin]. Apr. 2025. DOI: 10.48550/arXiv.2504.06932. URL: http://arxiv.org/abs/2504.06932 (visited on 06/07/2025).

[62] TenneT TSO B.V. *BSP register*. nl. URL: https://www.tennet.eu/nl/de-elektriciteitsmarkt/ondersteunende-diensten-bspcsp/bsp-register (visited on 04/21/2025).

A

BSP's in the Netherlands

Table A.1: List of all BSPs in the Netherlands [62].

Organization	**FCR**	**aFRR**		**mFRRda**	
		Capacity	Energy	Up	Down
DNO Energie BE B.V.		✓	✓	✓	✓
Eneco Energy Trade B.V.	✓	✓	✓	✓	✓
Energy Pool Développement SAS			✓	✓	✓
ENGIE Energie Nederland N.V.	✓	✓	✓	✓	✓
Enova Grid Management B.V.		✓	✓	✓	✓
Essent Sales Portfolio Management B.V.	✓		✓		
Flexcity Netherlands B.V.		✓	✓	✓	✓
GIGA Storage B.V.	✓		✓		
Greenchoice B.V.	✓		✓		
Groendus Energy Services B.V.			✓		
Mestral B.V.	✓		✓		
Next Kraftwerke Benelux B.V.	✓		✓	✓	✓
Peak Power UG			✓		
PZEM Energy Company B.V.	✓	✓	✓	✓	✓
Recoy B.V.			✓		
RWE Supply & Trading GmbH	✓	✓	✓		
Sappi Maastricht B.V.		✓	✓		
Scholt Energy Control B.V.	✓		✓		
Trafigura Denmark ApS			✓		
Vandebron Energie B.V.	✓		✓		
Vattenfall Energy Trading Netherlands N.V.	✓	✓	✓	✓	✓
AES Energy Storage Zeeland B.V.	✓				
Centrica Business Solutions B.V.	✓				
ECW Elektra B.V.					
E.D.Mij B.V.				✓	✓
HVC Energie B.V.					✓
Odura advies en projecten B.V.	✓				
Powerhouse B.V.				✓	✓
Pure Energie Levering B.V.	✓				
Repowered B.V.	✓				
Sympower Nederland B.V.					✓
The Mobility House GmbH	✓				
Uniper Benelux N.V.	✓				

B

Literature History

Day-ahead Literature

Day-ahead electricity price forecasting literature focuses on predicting market-clearing prices one day in advance. A wide range of forecasting approaches has been investigated that include statistical models, machine learning techniques, deep learning methodologies

Price Forecasting

Two review papers set a solid foundation for EPF in the day-ahead domain [28] [29]. The first provides a broader historical perspective and discusses the complexities, strengths, weaknesses and future directions of various solutions over the past 15 years, while the second focuses more recently on the recent statistical and deep learning methods for DAMs.

Weron (2014) set a set a fundament for EPF spot forecasting [27]. The paper reviews different methods categorised into: multi-agent, fundamental, reduce-form, statistical, and computational intelligence. It concludes that certain models, particularly those incorporating real-time market data and machine learning algorithms, show improved accuracy and reliability. These models outperform traditional methods by adapting to market volatilities and the integration of multiple data sources. The study emphasises that continuous model refinement and the incorporation of emerging data streams are crucial elements for improving forecast precision in the evolving energy markets.

Lago et al. (2021) did a comprehensive study and proposed a rigorous framework for spot price forecasting [29]. This study focused multiple international markets. The study's most important contribution is a comprehensive framework for price forecasting, accompanied by an open-source toolbox that implements two forecasting methods: the Lasso Estimated AutoRegressive (LEAR) model and Deep Neural Networks (DNN). The study is limited to open-access data, which can be improved by the incorporation of a broader dataset.

Probalistic Forecasting

The growing volatility of energy markets has increased the need to account for uncertainty in forecasting. This is done by probabilistic forecasting.

Nowotarski and Weron (2018) provided a comprehensive review of probabilistic forecasting methods in this context [28]. The study emphasises the importance of these methods in managing forecast uncertainty and supporting risk-aware decisions. However, applying probabilistic forecasting to day-ahead electricity prices is more complex than point forecasting, especially in model development evaluation.

Intra-day Literature

As stated, the literature on intra-day markets is smaller than day-ahead. Because this market is 'pay-as bid', no supply curves exists in this domain. Therefore therefore studies focus more on market characteristics [30, 31, 34] and price derivatives such as Volume-Weighted Average Price indices (VWAP) [32].

Shinde and Amelin (2019) wrote a review paper addressing low liquidity, high volatility and high trading activity [30]. The research focuses on Spanish, German, and Nordic markets. Later Shinde, Kouveliotis-Lysikatos, and Amelin (2021) added another review paper that analyses trading trends in continuous intra-day markets [31]. The study addresses the impact of renewable energy sources and impact of forecast errors on intra-day prices.

Narajewski and Ziel (2020) did an econometric analysis on the ID3 price, which is the weighted average price 3 hours before delivery [32]. The research focuses on the German market. The study did an econometric time series analysis using LASSO and Elastic Net techniques. This research showed the potential of linear models such as LASSO in combination with orderbook data.

Kotsias (2022) did an exploratory comparison of different statistical and state-of-art models [33]. This work forecasted the VWAP of the last three hours of intra-day trading in Nordic markets. Linear regression, ARX, SARIMAX models are compared to DeepAR and a Temporal Fusion Transformer. Traditional methods resulted to be more accurate, but also more understandable.

Birkeland and AlSkaif (2024) did a literature review on 132 primary studies in European intra-day electricity markets [34]. The paper highlights the need for trading strategies and forecasting methods that consider the stochastic nature of renewables. The research also notes a lack of research in the Dutch and French markets. It also adds that intra-day data can give insights into price volatility and price movements, which could be useful to predict aFRR supply curves.

Imbalance Literature

Due to volatility and complex market structures, imbalance literature lacks attention compared to day-ahead and intra-day markets. Imbalance market literature focuses on price and probabilistic forecasts Browell and Gilbert (2022) discuss the need for forecasting electricity imbalance prices and volumes, they underscore their importance in short-term markets [10]. They advocate for increased research efforts to leverage diverse data sources, which could improve the accuracy of imbalance price and volume forecasts.

Price Forecasting

Price forecasting in balancing and ancillary energy markets has been explored using various statistical, machine learning, and deep learning models, with studies consistently highlighting the challenges of high uncertainty and dynamic market behavior, where no single method proves universally superior, but approaches like XGBoost and LSTM show promising performance under specific conditions.

Multiple time-series models are researched by Klæboe, Eriksrud, and Fleten (2015) to forecast balancing market prices for Nord Pool price zone NO2 in Norway [12]. Forecasts are conducted for 1-hour and 1 day ahead using statistical models such as ARMA, ARX, ARM, ARIMAX, and Markov models. It is addressed by the study that reliable forecasts could not be made. The complexity of predicting balancing prices, due to last-minute market adjustments which are difficult to account for, is emphasised. Nevertheless, the value of imbalance prices is still vouched for by the study. Forecasts incorporating the balancing state tend to forecast better. However, this research, conducted in 2015, does not take into account new market dynamics that have since arisen.

Lucas et al. (2020) employs decision tree-based models, Random Forrest (RF), Gradient Boosting (GB), and eXtreme Gradient Boosting (XGBoost) in Germany [35]. It highlights a small likelihood to forecast prices with very high precision. However, XGBoost seemed to outperform Random Forrest and Gradient Boosting methods.

Merten et al. (2020) did a review paper on forecast methods for probability for bid acceptance, mixed marginal price and percentiles prices in Germany [36]. It used an Exponential Smoothing (ES), SARIMA, Neural Network (NN) and Recurrent Neural Network model (RNN). No overall winner is identified between these models, results are strongly influenced by previous outcomes. Statistical models generally outperformed machine learning models in terms of stability and performance.

Cardo-Miota, Pérez, and Beltran (2023) did a study on deep-learning based predictions in the Iberian electricity market. The study compared a Feed-Forward Neural Network (FFNN), Convolutional Neural Network (CNN), and LSTM model to predict the marginal price [37]. The LSTM had the best performance due to its accuracy and ability to capture sequential dependencies. The input data is split into known (historical) and delayed (forecasts) inputs. It underscores the growing importance of accurate forecasting methods due to the increasing share of intermittent renewable energy sources and the dynamic nature of ancillary service markets.

Probalistic Forecasting

Multiple studies have been done on probabilistic forecasting [38, 39, 23]. They highlight the evolving complexity of electricity markets and the growing necessity for advanced predictive models that can handle the uncertainties of renewable integration and market volatility.

Dumas et al. (2019) did a probabilistic forecast in the Belgian context [38]. The research involved a two-step approach: it first computes the Net Regulation Volume (NRV) probability and afterwards matches it to the prices based on a price table uploaded a day before by the Belgian TSO Elia. This is compared to a deterministic, Multi-Layer Perceptron (MLP), and Gaussian Process model.

Later Bottieau et al. (2020) did a study on an LSTM-based encoder-decoder model in the Belgian and German market [39]. It compared it to ARIMA, Quantile Regression Forrest (QRF), MLP, GRU, and XGBoost models. The LSTM and XGBoost showed the best results, but the research is done on a relatively short training window (2 months).

Afterwards, Narajewski (2022) did a study on multiple forecasting probabilistic methods in the German imbalance market 30 minutes before delivery [23]. Methods used include LASSO with bootstrap, Generalised Additive Models for Location Scale and Shape (GAMLSS), and probabilistic neural networks. No method is able to outperform the benchmark of the ID1 price.

Bid Curve Prediction

Although the majority of literature focuses on price or probabilistic forecasts, a few articles covered bid curve predictions [24], [25], [26], all in the day-ahead domain.

Ziel and Steinert (2016) developed the X-model. This is the intersection of the sale and purchase curve to predict the day-ahead price in Germany and Austria [24]. The research employs a combination of dimension reduction and a LASSO-based estimation to handle all auction data. The model makes probabilistic forecasts that are able to capture non-linear behaviour. This model's focus on auction data and its ability to forecast price probabilities can be particularly relevant for forecasting aFRR bid ladders and the associated risk, as it provides a detailed insight into market behaviours.

Guo et al. (2021) developed a Long Short-Term Memory (LSTM) based model for forecasting Aggregated Supply Curves (ASC) in the Midcontinent Independent System Operator (MISO) DAM [25]. Their approach integrates a novel data dimensionality reduction technique, leveraging Principal Component Analysis (PCA) to transform complex, high-dimensional ASC data into a more manageable form. The model demonstrates robust performance in ASC curve prediction. Whole bid curve predictions can reveal additional market dynamics, which can provide insights to price formation.

Pinhão, Fonseca, and Covas (2022) introduced a method for forecasting market offers for DAM curves in the Iberian market [26]. The bid curve is fit through the prediction of a sixth-degree polynomial to predict the market curve. This is done with a Vector AutoRegressive (VAR) and XGBoost benchmark model. A key contribution of their work is highlighting the importance of analysing underlying market structures. Focusing on supply and demand curves, rather than solely on price, is emphasised as a way to improve forecasting performance. The study highlights that incorporating these curves directly into price forecasts enhances accuracy. However, it also acknowledges that forecasting full market curves remains a complex task with significant room for improvement.

Li, Alonso, and Pascual (2025) proposed a methodology for predicting electricity supply and demand curves in the Spanish DAM using functional data techniques combined with machine learning approaches [50]. In their framework, supply and demand curves are forecasted by modelling volume as a function of price, with price positioned on the x-axis and volume on the y-axis, allowing the method to naturally accommodate curves of varying lengths. Functional Principal Component Analysis (FPCA) and functional regression techniques are employed to reduce the high dimensionality of the data while preserving key structural characteristics. Dimension reduction techniques and meteorological covariates significantly improve the forecast accuracy of both the curves and the resulting market prices.

C

Data Analysis

Bidladder Statistics

Table C.1: Bid ladder statistics presented for different volume bin intervals.

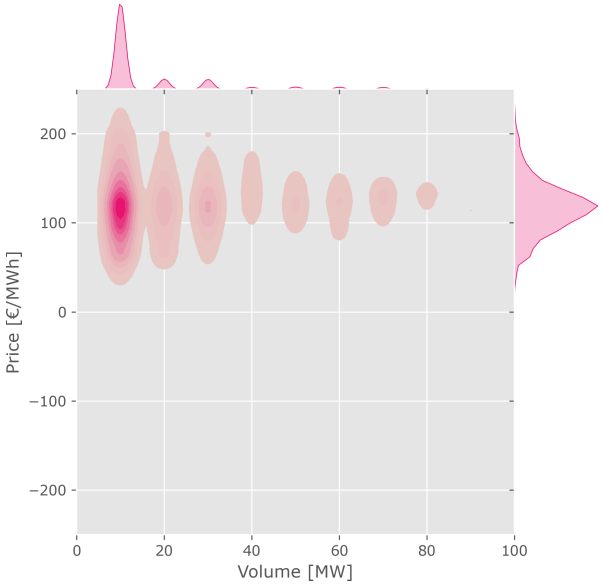
(a) Upward bidding direction

	10	100	250	500	750	1000	1250	1500	1750	2000
count	9692	9692	9692	9692	8725	5289	2747	1824	1192	693
mean	96.17	65.41	36.59	-76.94	-296.56	-555.11	-555.30	-559.16	-557.88	-610.60
std	22.40	39.01	56.07	193.87	330.75	312.49	247.84	247.25	247.02	258.48
min	6.65	-69.38	-531.01	-1000	-1456.42	-1453.89	-1538.65	-1400	-1538.65	-1481.42
25%	83.98	54.57	-5.95	-66.98	-513.37	-895	-593.08	-660.82	-657	-673.79
50%	93.69	76.09	55.98	-29.80	-132.99	-524.23	-531.57	-531.57	-542.31	-562.76
75%	107.68	91.73	78	34.98	-50	-316.54	-514.40	-514.40	-514.40	-559.14
max	338.51	131.90	108.57	87.90	54.48	-29.01	-32.11	-33.30	-35	-57.19

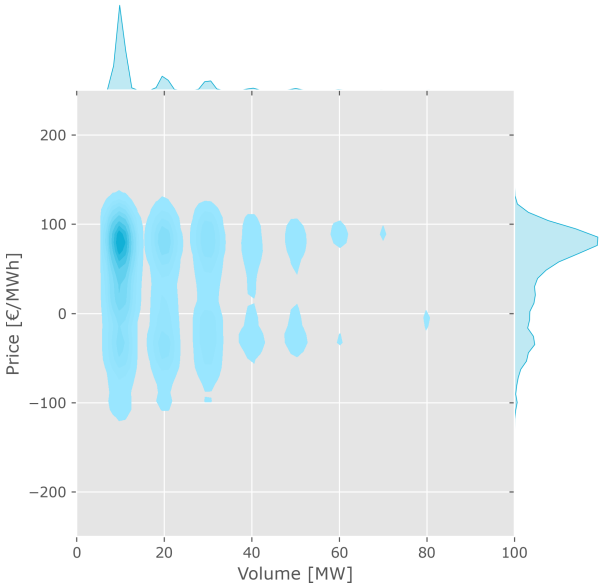
(b) Downward bidding direction

	10	100	250	500	750	1000	1250	1500	1750	2000
count	9692	9692	9692	9692	8725	5289	2747	1824	1192	693
mean	96.17	65.41	36.59	-76.94	-296.56	-555.11	-555.30	-559.16	-557.88	-610.60
std	22.40	39.01	56.07	193.87	330.75	312.49	247.84	247.25	247.02	258.48
min	6.65	-69.38	-531.01	-1000	-1456.42	-1453.89	-1538.65	-1400	-1538.65	-1481.42
25%	83.98	54.57	-5.95	-66.98	-513.37	-895	-593.08	-660.82	-657	-673.79
50%	93.69	76.09	55.98	-29.80	-132.99	-524.23	-531.57	-531.57	-542.31	-562.76
75%	107.68	91.73	78	34.98	-50	-316.54	-514.40	-514.40	-514.40	-559.14
max	338.51	131.90	108.57	87.90	54.48	-29.01	-32.11	-33.30	-35	-57.19

2D KDE



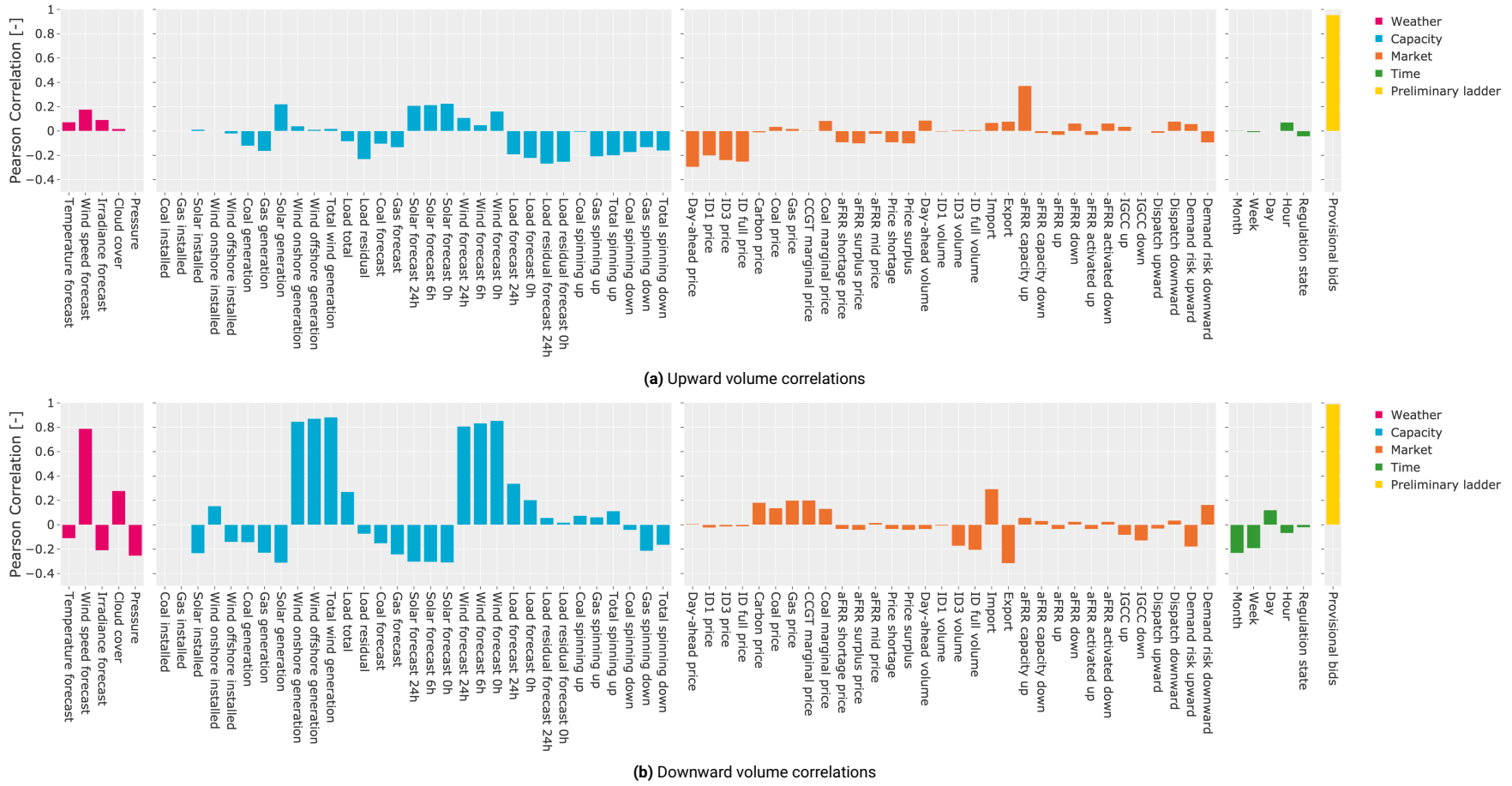
(a) Downward bids distribution



(b) Downward bids distribution

Figure C.1: Joint 2 KDE of all bids with 10 MW steps, where blue (downward) and pink (upward) indicate the direction.

All Correlations
Volume Correlations

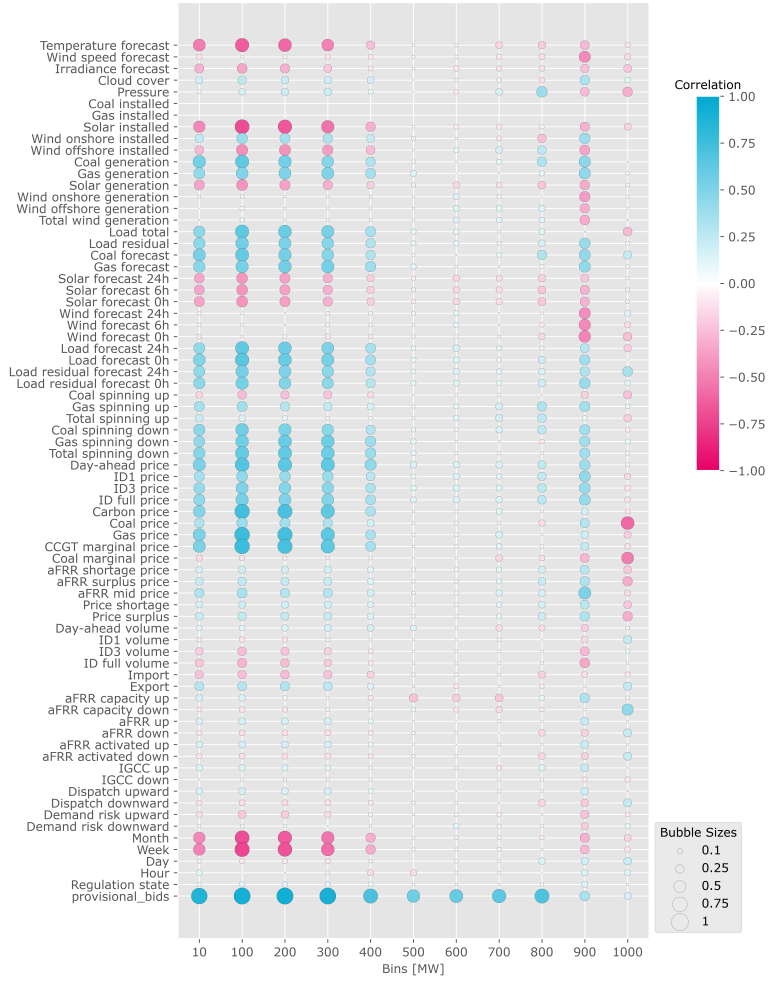


(a) Upward volume correlations

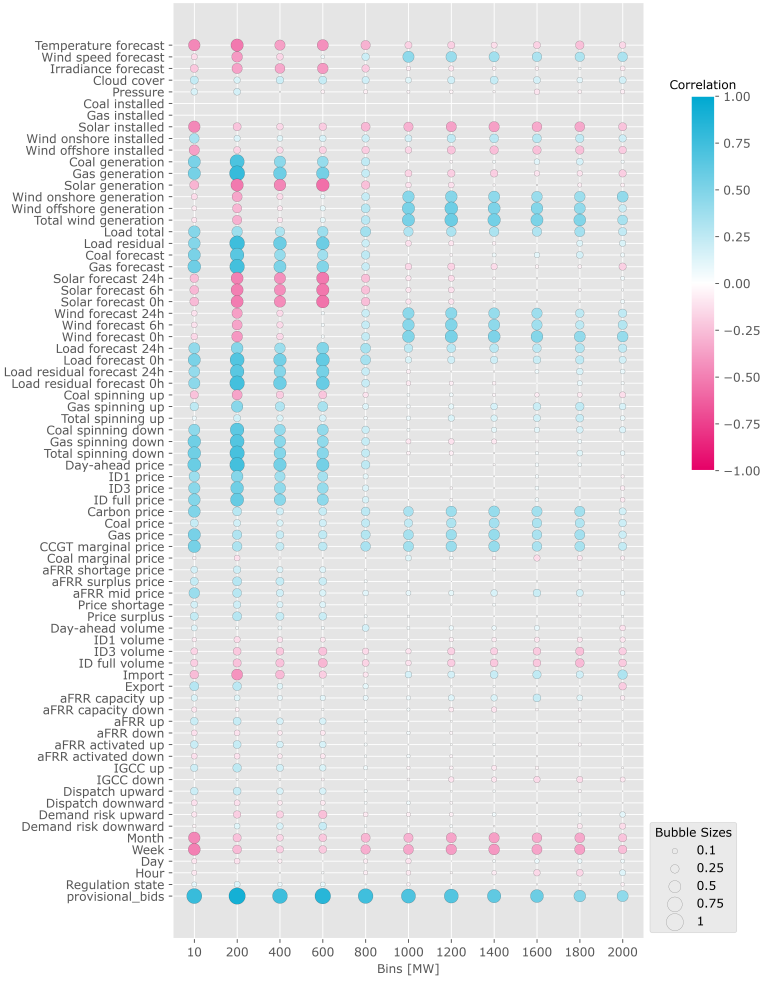
(b) Downward volume correlations

Figure C.2: Correlations between explanatory features and binned aFRR volumes. Bubble size shows magnitude, color indicates direction (blue: positive, pink: negative)

Price Correlations



(a) Downward price correlations



(b) Downward price correlations

Figure C.3: Pearson correlation between all analysed features and total bid volume. Features are grouped by category: weather (pink), capacity (blue), market (orange), time (green), and preliminary bidladders (yellow).

Predictive Value of Preliminary Bidladders

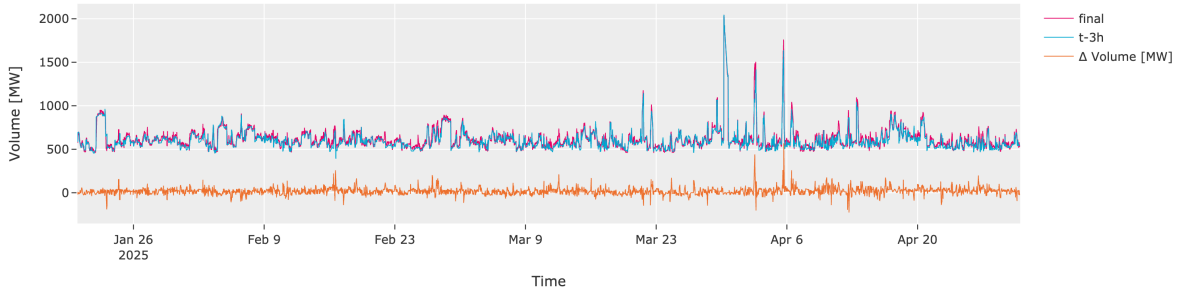
To evaluate the preliminary bids, this section examines the change in bid volume and bid prices three hours before delivery (t-3).

Delta Volume

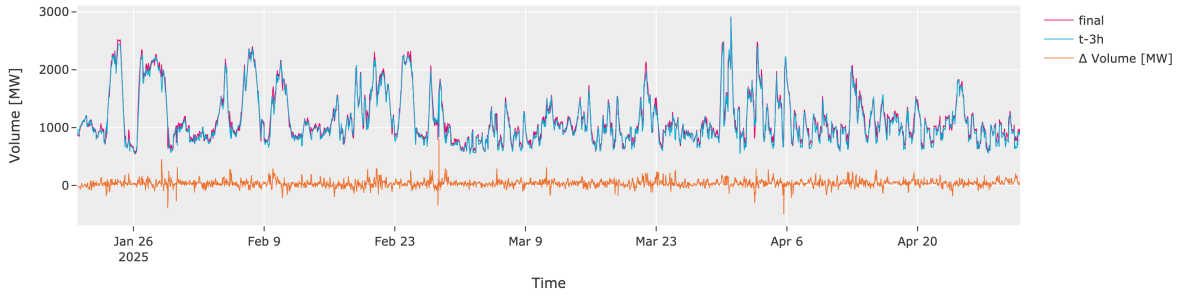
Figure C.4 shows the time series of bid volumes at t−3h and final delivery for upward and downward regulation, respectively. The pink and blue lines represent the total bid volume at final delivery and at t−3h, while the orange line tracks the difference between the two, denoted as ΔVolume (final − t−3h).

For upward regulation, the lines align closely, and ΔVolume remains narrow around zero, indicating stable volumes and minimal adjustments near delivery. Small positive spikes suggest occasional volume additions.

Downward bids exhibit more variation, with larger and more frequent ΔVolume spikes. This reflects a more flexible market where volumes are often adjusted closer to real-time in response to changing conditions.



(a) Delta upward volume



(b) Delta downward volume

Figure C.4: Comparison between the final volumes (pink), the three-hour-ahead volumes (blue), and their differences (orange)

To complement the time series perspective, a distributional analysis helps quantify how often and by how much bid volumes are adjusted between t−3h and final delivery.

Figure C.5 shows the evolution of market volumes at t−3h and at final delivery, for both upward and downward bids. The KDE plot visualises the distribution of ΔVolume, defined as the difference between final and t−3h bid volumes. In both directions, the distribution is not centred at zero but exhibits a positive skew, indicating that additional volume is more often added than withdrawn closer to delivery.

For upward bids, the ΔVolume distribution is narrow and peaked, suggesting limited changes after t−3h. This reflects a more stable and predictable market, where volumes are largely fixed in advance and only minor adjustments occur. The slight positive bias confirms that some additional volume is added near delivery, but the magnitude of these changes remains relatively small.

Downward bids display a wider and more dispersed distribution, indicating frequent and larger volume changes near delivery. The stronger positive skew suggests that the addition of volume is more pronounced in the downward market. This points to a more flexible and reactive bidding behaviour, with participants actively adjusting bids in response to real-time system conditions.

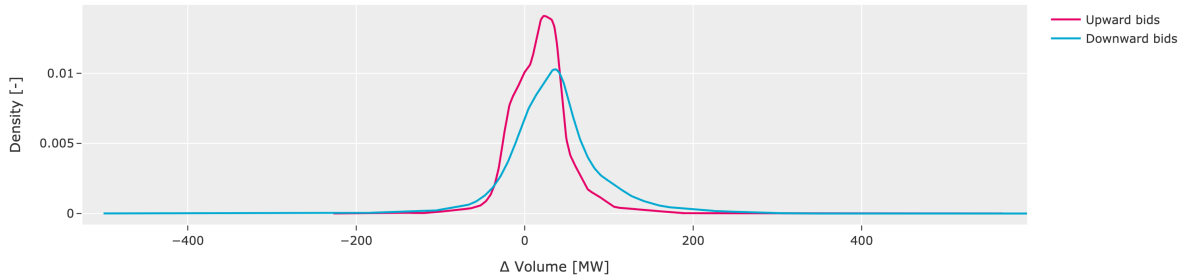


Figure C.5: KDE of volume changes three hours ahead for upward bids (pink) and downward bids (blue) is presented.

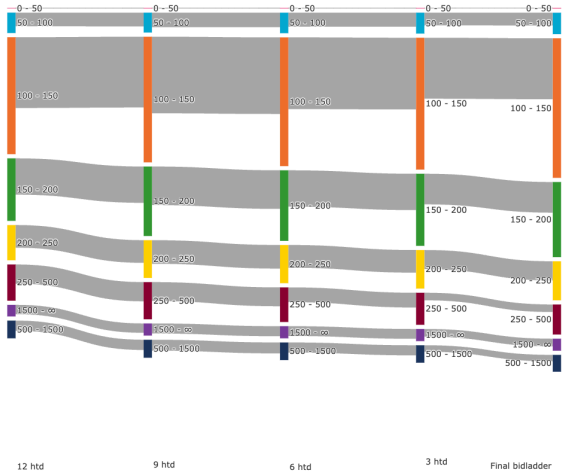
While volume adjustments reveal when market activity intensifies, the following analysis explores how bid prices evolve over time leading up to delivery.

Delta Price

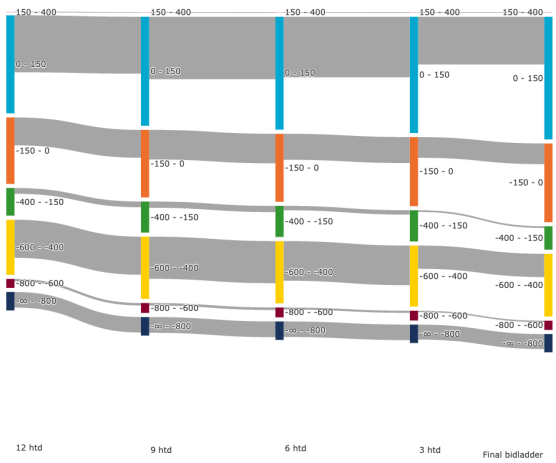
The Sankey diagrams in Figures C.6a, and C.6b, illustrate the evolution of upward and downward bid volumes across price intervals and time layers, from 12 hours before delivery until final submission. Each time layer represents a snapshot of the market state at a specific point before delivery, while each node corresponds to a specific price interval. The thickness of the flows indicates the volume of bids within each price bin, and grey flows represent bids that persist into the next time layer.

For upward bids, the majority of the volume is concentrated between 100 and 250 €/MWh. This concentration remains relatively stable over time, as shown by the consistent thickness of flows across successive time layers and the dominance of grey flows. Most upward bids are already established nine hours before delivery, with mostly price adjustments closer to real-time instead of the addition of extra volume.

For downward bids, the majority of the volume lies between 150 and −600 €/MWh. The persistence of grey flows across time layers indicates that participants establish the bulk of their downward bids well in advance of delivery, with only minor reallocations across adjacent price intervals observed closer to real-time. Stability remains most evident in mid-price ranges for both directions.



(a) Upward bid evolution



(b) Downward bid evolution

Figure C.6: Sankey diagram of upward bid volumes across price intervals and time layers, where nodes represent bid volumes in each interval and grey flows indicate bids persisting into the next time layer.

D

Model Specifications

Hyperparameters

Table D.1: Hyperparameters for each principal component model.

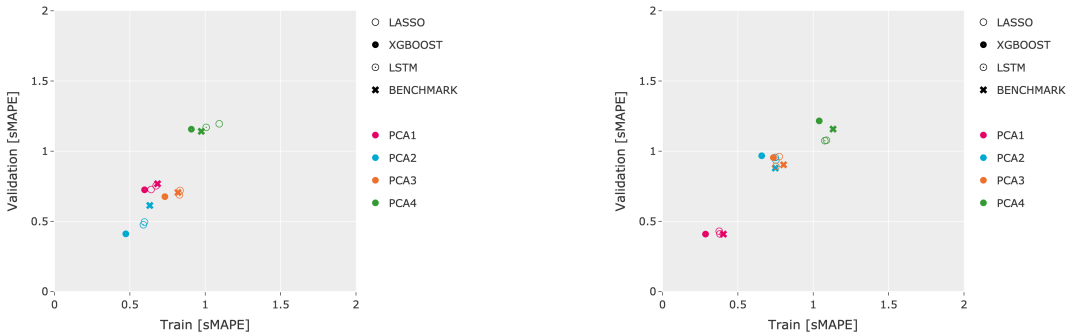
(a) Upward models

		PCA1	PCA2	PCA3	PCA4
LASSO	alpha	0.00010	0.00010	0.00201	0.01
XGBOOST	max depth	3	5	5	4
	learning rate	0.10	0.09	0.03	0.07
	n estimators	362	764	656	865
	subsample	0.83	0.88	0.74	0.60
	colsample bytree	0.83	0.76	0.85	0.66
	min child weight	45	49	38	19
	gamma	4.29	1.46	0.17	0.72
	reg alpha	5.15	8.94	9.05	2.80
	reg lambda	3.41	7.18	0.90	3.54
LSTM	timesteps	3	12	12	12
	lstm units	16	16	32	32
	dense units	32	32	32	64

(b) Downward models

		PCA1	PCA2	PCA3	PCA4
LASSO	alpha	0.00010	0.00192	0.02	0.00010
XGBOOST	max depth	6	6	3	3
	learning rate	0.08	0.10	0.05	0.07
	n estimators	236	596	534	168
	subsample	0.81	0.53	0.67	0.86
	colsample bytree	0.82	0.57	0.77	0.72
	min child weight	16	36	30	25
	gamma	1.15	2.80	1.85	0.44
	reg alpha	0.80	0.47	2.19	7.03
	reg lambda	4.77	6.72	0.84	5.98
LSTM	timesteps	288	3	3	12
	lstm units	16	16	16	16
	dense units	128	64	64	128

Train and Validation Errors

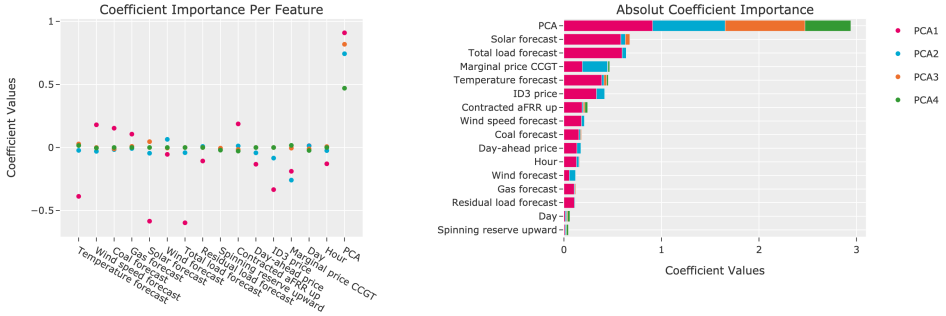


(a) Upward sMAPE for training and validation sets.

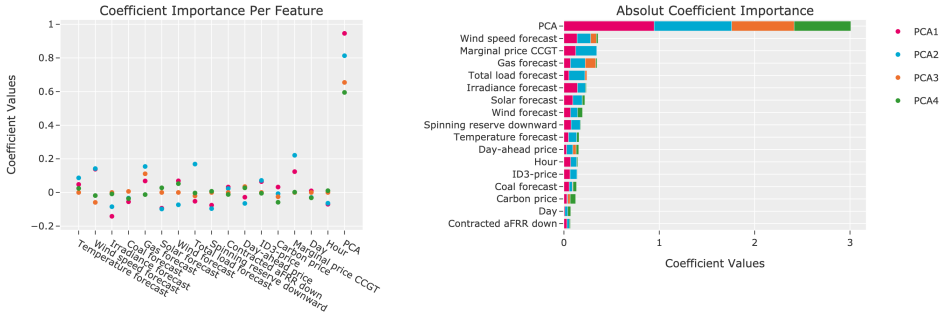
(b) Downward sMAPE for training and validation sets.

Figure D.1: Training and validation sMAPE across models (LASSO, XGBOOST, LSTM, BENCHMARK) and principal components (PCA1–PCA4). Lower values indicate a better fit.

Feature Importance

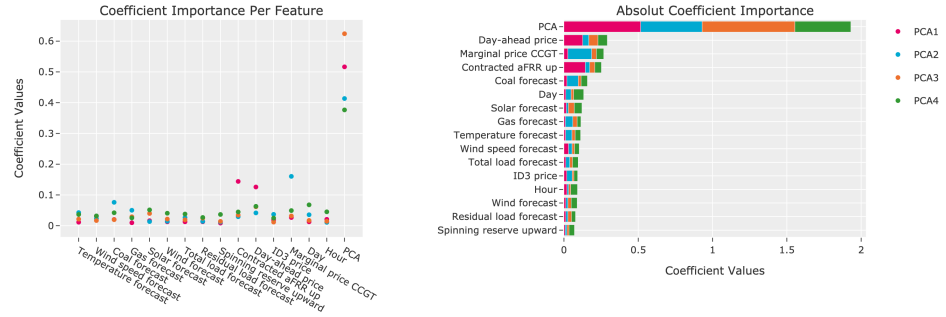


(a) Upward LASSO feature importances.

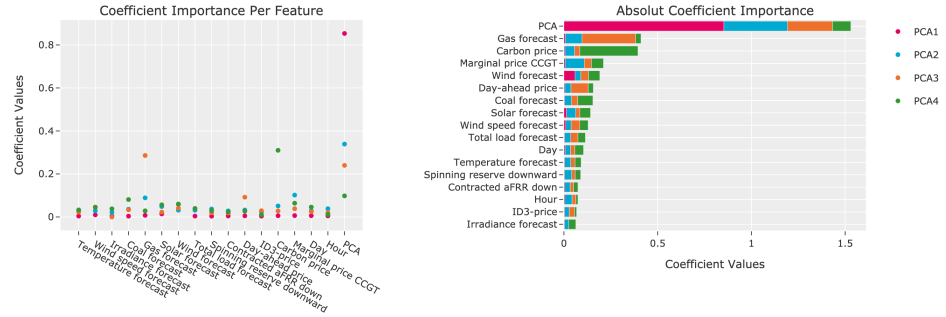


(b) Downward LASSO feature importances.

Figure D.2: Coefficient importance for LASSO models for (a) upward and (b) downward regulation: left shows feature coefficients (direction and magnitude), right shows absolute importance (PCA1: pink, PCA2: blue, PCA3: orange, PCA4: green).



(a) Upward XGBOOST feature importances.



(b) Downward XGBOOST feature importances.

Figure D.3: Coefficient importance for XGBOOST models for (a) upward and (b) downward regulation: left shows feature coefficients (direction and magnitude), right shows absolute importance (PCA1: pink, PCA2: blue, PCA3: orange, PCA4: green).

E

Additional Forecast Results

sMAPE over Time

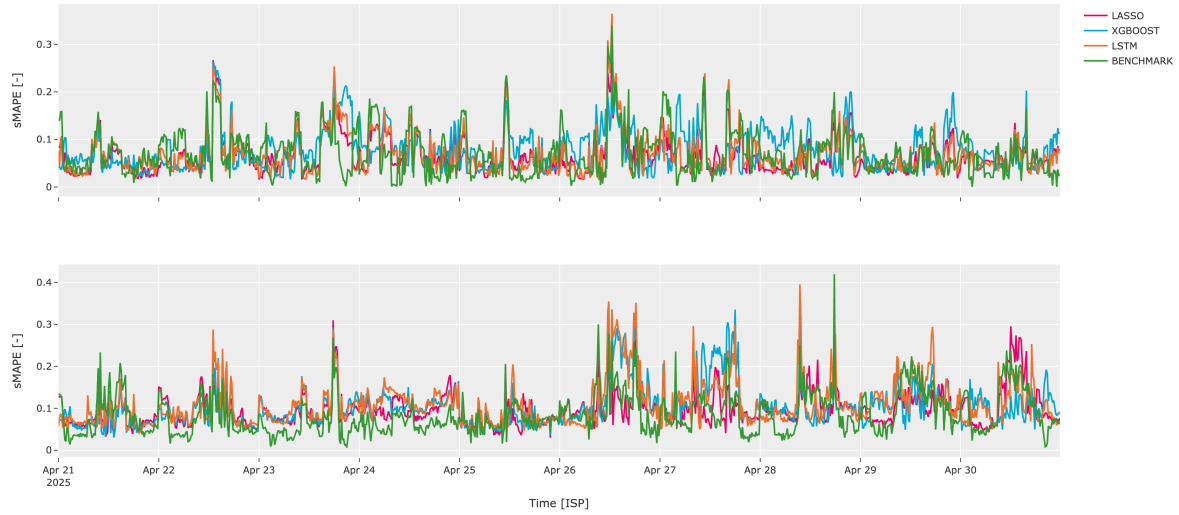


Figure E.1: Mean sMAPE over time under upward (top) and downward (bottom) trends for LASSO (pink), XGBOOST (blue), LSTM (orange), and BENCHMARK (green).

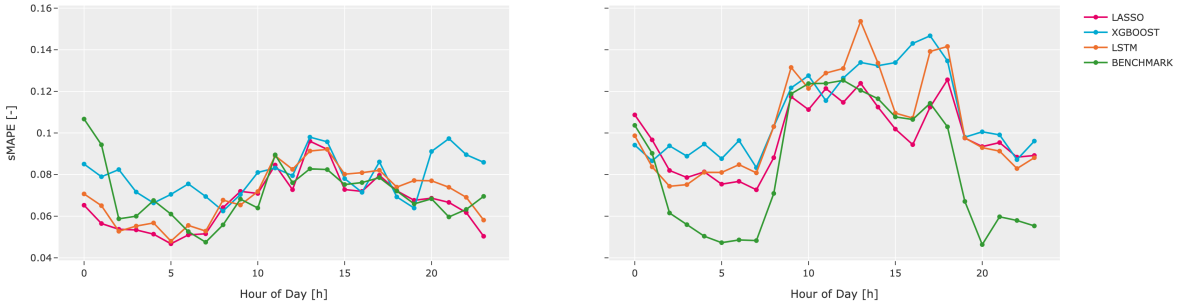
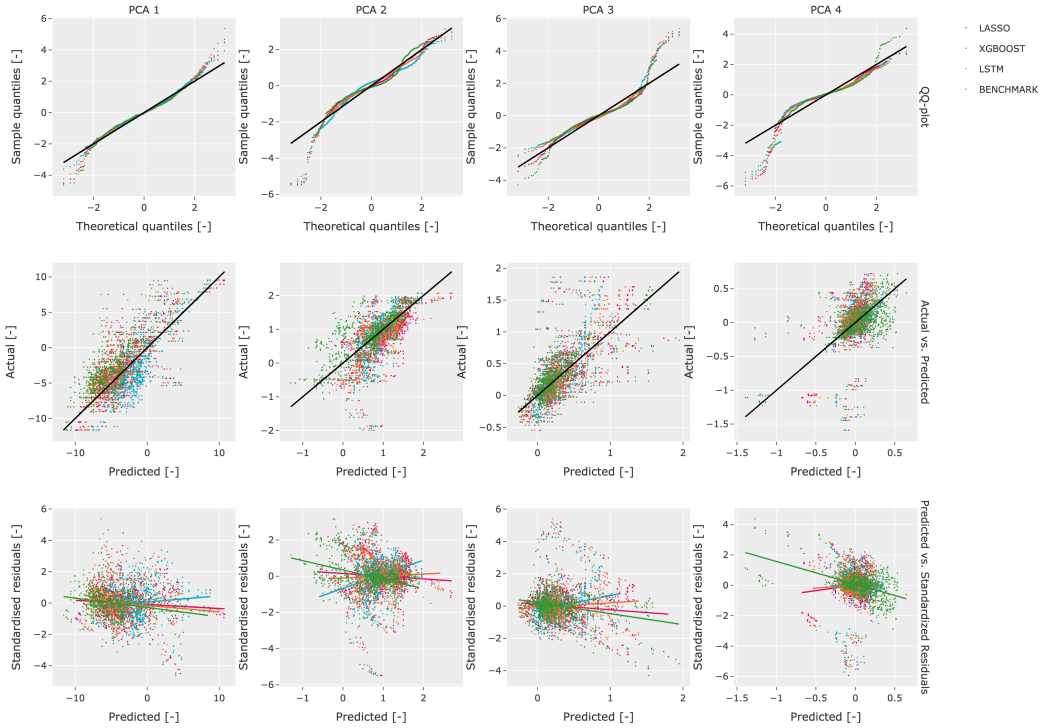
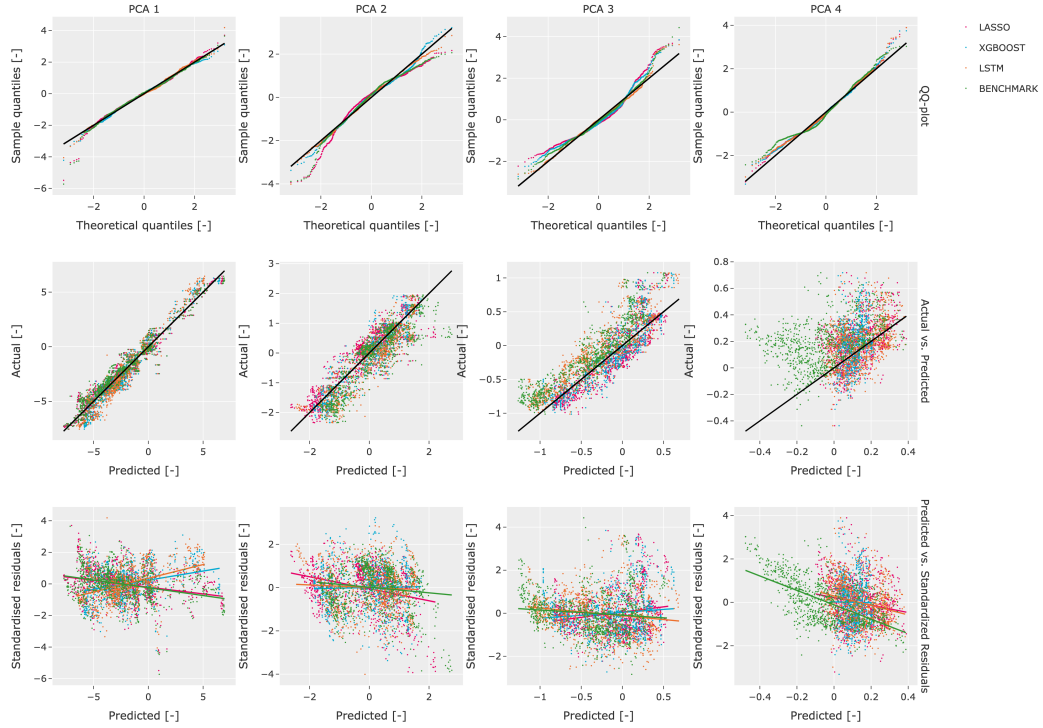


Figure E.2: Mean sMAPE over hour of the day upward (left) and downward (right) for LASSO (pink), XGBOOST (blue), LSTM (orange), and BENCHMARK (green).

Residual Diagnostics



(a) Upward residuals



(b) Downward residuals

Figure E.3: Residual diagnostics showing standardised residuals vs. predictions, actual vs. predicted values, and QQ-plots for LASSO (pink), XGBOOST (blue), LSTM (orange), and the BENCHMARK (green).

F

Battery Strategy

Battery Head and Footroom

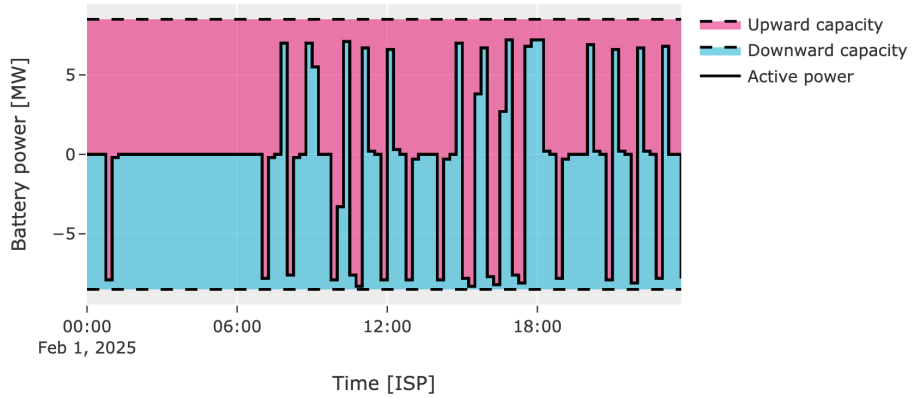


Figure F.1: Headroom (pink) and footroom (blue) indicate the battery's available upward and downward aFRR capacity, with the black line representing actual power usage.

Battery Bid Location

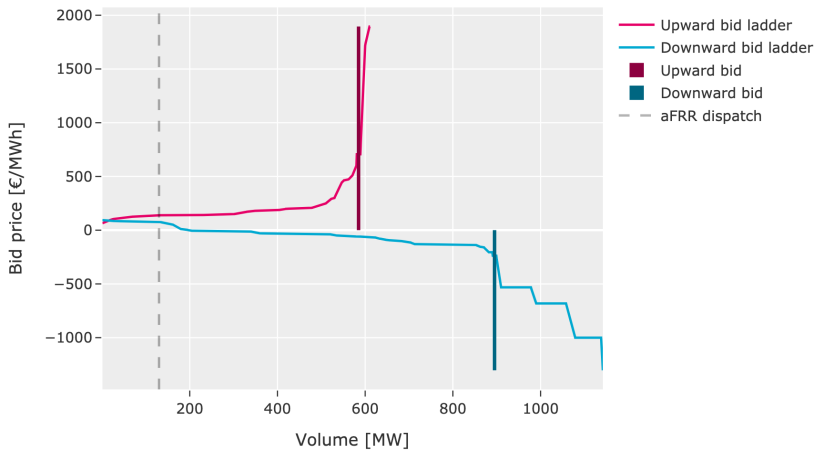
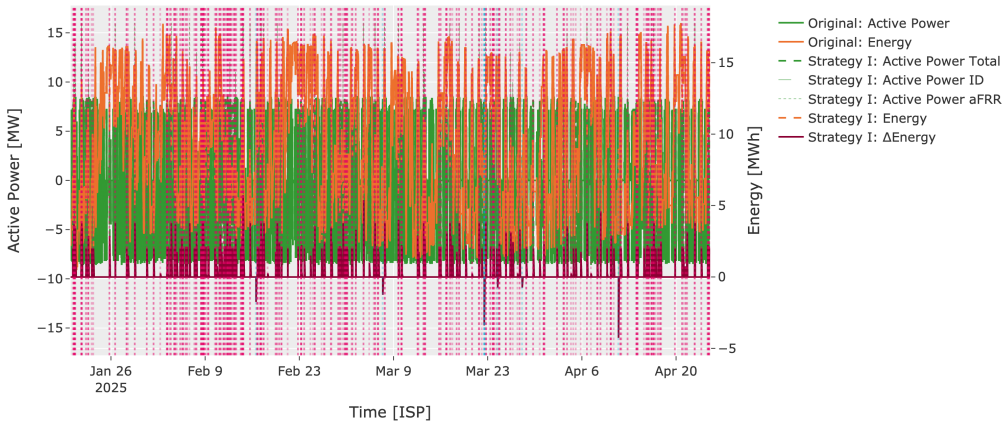
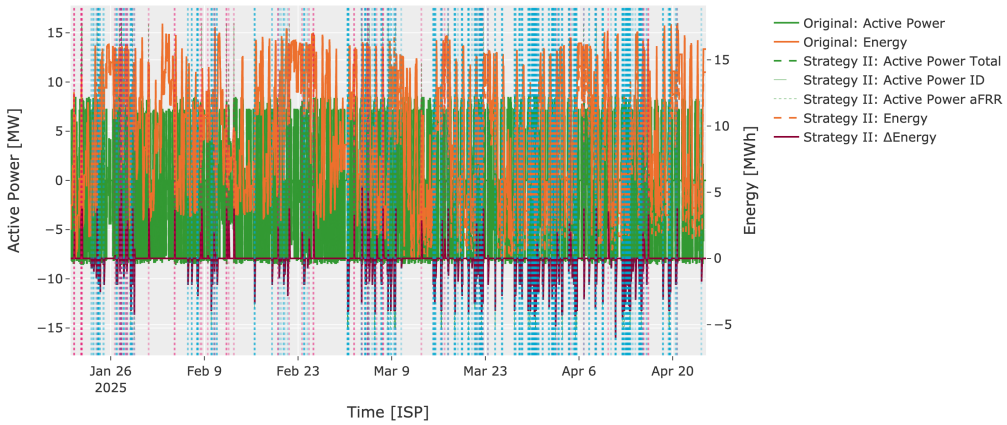


Figure F.2: Bid locations of the battery (dark pink, dark blue) shown alongside the upward (pink) and downward (blue) bid ladders, with the actual activated upward dispatch indicated by a dashed black line.

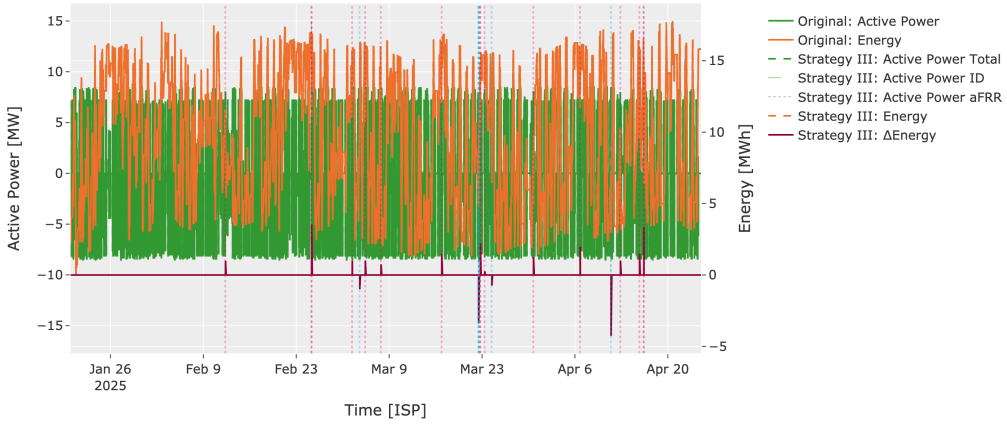
Battery Energy Levels



(a) Strategy I



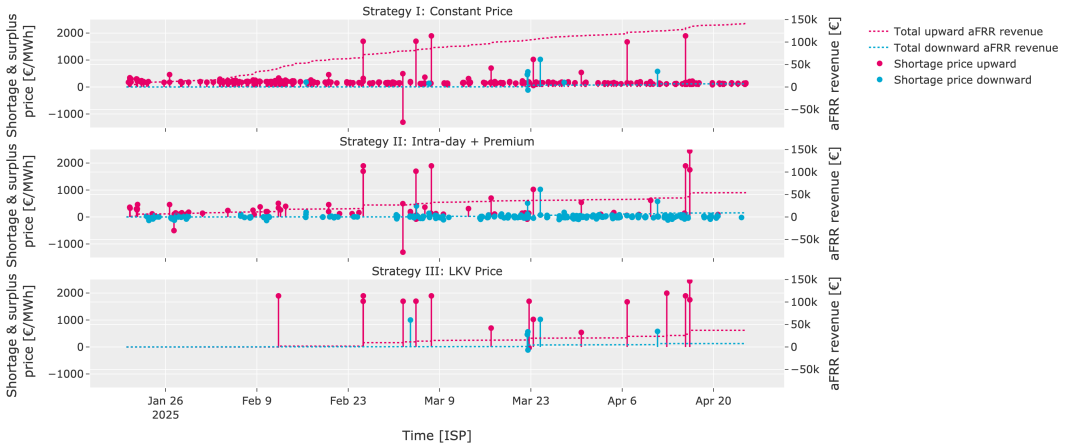
(b) Strategy II



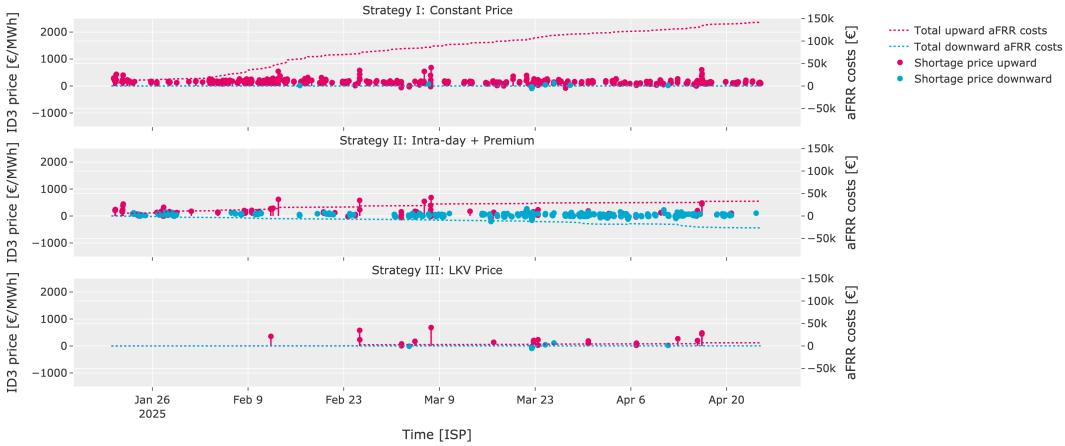
(c) Strategy III

Figure F.3: Battery energy levels under three pricing strategies: power output (green), battery energy (orange), and energy delta (red). Solid lines show the original strategy versus the dashed modified strategy.

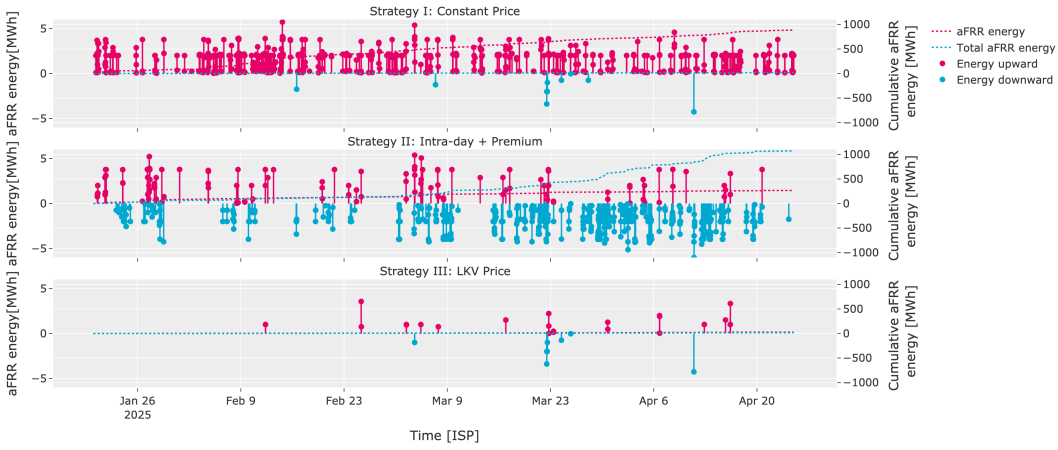
aFRR Activation



(a) aFRR revenues.



(b) aFRR costs.



(c) aFRR energy deployment.

Figure F.4: Comparison of aFRR revenue, cost, and energy deployment across three strategies, with upward (pink), downward (blue), and total (dotted black) values.