Forecasting mFRR Prices and Volumes using Machine Learning

Study of mFRR CM and EAM balancing markets in the Swedish SE3 bidding zone using LSTM and MLP neural networks

Master's thesis in Complex Adaptive Systems (MPCAS)

PHILIP LAURI

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Forecasting mFRR Prices and Volumes using Machine Learning

Study of mFRR CM and EAM balancing markets in the Swedish SE3 bidding zone using LSTM and MLP neural networks

PHILIP LAURI

Department of Electrical Engineering
Division of Electric Power Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Forecasting mFRR Prices and Volumes using Machine Learning
Study of mFRR CM and EAM balancing markets in the Swedish SE3 bidding zone using LSTM and MLP neural networks
PHILIP LAURI

Supervisor: Kimmy Jakobsson, Göteborg Energi
Examiner: Torbjörn Thiringer, Chalmers, Department of Electrical Engineering

Master's Thesis 2026
Department of Electrical Engineering
Division of Electric Power Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover image: Wind turbines. Source: Göteborg Energi [1].

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Forecasting mFRR Prices and Volumes using Machine Learning
Study of mFRR CM and EAM balancing markets in the Swedish SE3 bidding zone using LSTM and MLP neural networks
PHILIP LAURI
Department of Electrical Engineering
Chalmers University of Technology

Abstract

The transition towards a more renewable and decentralized power system is increasing variability and operational complexity, leading to a growing need for balancing services. This thesis focuses on forecasting mFRR prices and volumes in the Swedish SE3 bidding zone for both the Capacity Market (CM) and Energy Activation Market (EAM). The objective is to identify relevant explanatory variables and to develop and evaluate data-driven forecasting models. Machine learning methods, including Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks, are compared against naive benchmark models. Different feature set configurations, including univariate, multivariate, and selected subsets, are evaluated using early stopping and a holdout test set.

Feature selection results reveal clear differences between the CM and EAM markets. For CM, the most informative variables are autoregressive signals, demand forecasts, renewable generation forecasts, spot market indicators, and seasonal temporal patterns, suggesting that market outcomes are largely driven by relatively stable system fundamentals and expected supply-demand conditions. In contrast, EAM forecasting is dominated by short-term market signals, particularly spot prices, autoregressive signals, and dependencies between EAM prices and volumes. Wind-related variables and system-state indicators such as inertia and frequency deviation also provide additional predictive value, while temporal variables show less relevance. This reflects the more dynamic and reactive nature of the EAM market.

Forecasting performance varies considerably across markets and targets, but no model architecture or feature configuration consistently outperforms the others. In the CM market, simple statistical baseline models often outperform machine learning models, suggesting that CM behavior is largely driven by recurring temporal patterns, stable system conditions, and expected future supply-demand conditions. In contrast, machine learning models outperform naive benchmarks in EAM, indicating more complex and reactive dynamics that are harder to capture with simple baselines, reflecting a distinction between longer-term planning in CM and near real-time balancing operations in EAM.

Overall, mFRR forecasting performance is highly target- and market-dependent, and increasing model or feature complexity does not systematically improve predictive performance.

Keywords: mFRR, Capacity Market, Energy Activation Market, balancing market, ancillary services, SE3, forecasting, machine learning, neural networks, LSTM, MLP, feature selection

Acknowledgements

I want to thank my supervisor and examiner at Chalmers, Torbjörn Thiringer, for helpful guidance and feedback throughout the whole thesis period.

I also want to thank Göteborg Energi for this interesting and rewarding opportunity, and special thanks to my supervisor Kimmy Jakobsson, and the rest of the Physical Trading team at Göteborg Energi, for being supportive and welcoming.

Philip Lauri, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms used throughout this thesis, grouped by category and ordered by relevance:

	Markets & Systems
mFRR	Manual Frequency Restoration Reserve
CM	Capacity Market
EAM	Energy Activation Market
aFRR	Automatic Frequency Restoration Reserve
FCR	Frequency Containment Reserve
FFR	Fast Frequency Reserve
SE3	Swedish Bidding Zone 3
	Machine Learning Models
ARIMA	Autoregressive Integrated Moving Average
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron
RF	Random Forest
	Evaluation Metrics
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
	Feature Set Configurations
fs	Forward Selection
mv	Multivariate
uv	Univariate

Contents

1
Introduction

1.1 Background

In today's society, electricity is fundamental to almost every aspect of daily life and is essential for a modern, functioning society. Homes, businesses, industries, transportation, communication systems, and digital services all depend on a reliable supply of electricity. To ensure a secure and efficient power system, many actors, systems, technologies, and markets must continuously work together.

The Swedish power system and electricity market are part of the integrated European electricity system. As a result, the Swedish electricity market is closely connected to the European electricity market and operates under conditions of free competition. Electricity prices are determined by the interaction of supply and demand across several market segments, including the day-ahead, intraday, and balancing markets. Price formation follows a marginal pricing principle, where the last produced or traded unit sets the market price, making the system highly sensitive to variations in both supply and demand. Price dynamics are influenced by both physical and market-related factors, such as weather conditions, electricity demand patterns, available generation and transmission capacity, system and grid constraints, as well as the bidding strategies and behaviors of market participants [2, 3].

At the same time, the Swedish, Nordic, and European electricity systems are undergoing a rapid transition towards a more renewable and decentralized energy system. An increasing share of variable renewable generation, primarily wind and solar power, introduces greater variability and uncertainty in electricity production. As a result, the need for flexibility and balancing services is increasing to maintain reliable and secure system operation and support a higher share of renewable energy sources [2, 4].
For the power system to function securely, electricity production and consumption must remain balanced at all times. This responsibility lies with the Swedish transmission system operator (TSO), *Svenska kraftnät*, which monitors and controls system operation in real time [5]. If this balance is not maintained, frequency deviations occur, which may compromise system stability. To prevent and correct such imbalances, *Svenska kraftnät* procures and activates balancing reserves through the balancing market [6].

The balancing market consists of several reserve products, including Fast Frequency Reserve (FFR), Frequency Containment Reserve (FCR), automatic Frequency Restoration Reserve (aFRR), and manual Frequency Restoration Reserve (mFRR).These reserves are activated at different timescales, under different system conditions, and for different operational purposes to maintain system frequency and restore balance following disturbances and forecast errors. As the share of variable renewable generation increases, and electricity and balancing markets evolve and open up to more participants, competition and market complexity increase, and the importance of balancing services and reserve markets continues to grow [6, 4].

Göteborg Energi is currently in a development phase where the need for more advanced, data-driven forecasting and analytical tools for ancillary service markets is increasing. At the same time, the company is expanding its participation in these markets, with ambitions to become a Balance Responsible Party (BRP) in the near future. In addition, parts of the company's wind generation portfolio are planned to participate in the mFRR market for the first time, making mFRR down-regulation particularly relevant. In this environment, the ability to forecast mFRR prices and activated volumes in order to support bidding and trading decisions becomes increasingly valuable, as accurate forecasts and well-founded decisions can improve both profitability and risk management.

This master's thesis is therefore carried out in close collaboration with the Physical Trading department within the Electricity and Gas Products business area at *Göteborg Energi*. The aim of this thesis is twofold: (1) to identify and analyze the variables that are most important for explaining mFRR prices and volumes, and (2) to develop and evaluate forecasting models that can support forecast-based bidding and trading decisions in the balancing market. This is achieved through a combination of data analysis and machine learning-based forecasting.

1.2 Purpose and Objectives

The purpose of this master's thesis is to analyse the factors influencing mFRR Capacity Market (CM) and Energy Activation Market (EAM) prices and activated volumes in SE3, and to develop and evaluate data-driven forecasting models for these markets.

The study focuses on short-term forecasting of mFRR CM and EAM prices and volumes for both upward and downward regulation, using historical observations and available forecasts of market, system, and environmental variables. A key objective is to investigate which system, market, environmental, and temporal variables are most important for explaining and predicting mFRR prices and volumes. Furthermore, the objective is to evaluate whether machine learning models improve forecasting performance compared to simple naive baseline models based on historical averages and seasonal repetition, and to investigate if forecasting performance improves by explicitly modelling temporal dependencies.

Finally, it evaluates how different input feature set configurations affect forecasting performance, and which setup performs best in terms of input dimensionality and feature selection strategy.

1.3 Scope and Limitations

The scope of this thesis is limited to short-term forecasting of mFRR Capacity Market (CM) and Energy Activation Market (EAM) prices and volumes in SE3, including both upward and downward regulation. The study focuses on SE3 due to its operational relevance for *Göteborg Energi* and to provide a focused case study of a single bidding zone. Particular focus is placed on downward regulation, as it is the natural balancing direction for wind generation and is therefore highly relevant to *Göteborg Energi*'s planned participation in the mFRR market.

While the target variables are restricted to SE3, the input variables are primarily derived from SE3 and are complemented by market, system, and environmental variables covering Sweden as a whole and the wider Nordic power system and electricity market. The study considers forecasting on both day-ahead (for CM) and intraday (for EAM) timescales, with all models formulated as supervised learning problems based on historical observations and available forecasts.

The study is limited to data from 2025. This is because the mFRR market is evolving rapidly, with market rules, participant behavior, and market dynamics changing significantly over time. As a result, data from previous years are not fully comparable to current market conditions, while future market conditions may differ from those observed in this study. This limits the ability to capture longer-term patterns and reduces the amount of training data available for the forecasting models.

The study does not include real-time forecasting, bidding, or operational decision-making. All models are trained and evaluated using historical data only, and are not tested or deployed in a live market environment.

1.4 Research Questions

To achieve the purpose and objectives of this thesis, the following main research questions are formulated:

1. Which system, market, environmental, and temporal variables are most important for explaining and forecasting mFRR prices and volumes?

2. To what extent, if any, do machine learning models improve forecasting performance compared to naive baseline models based on historical averages and recurring patterns?

3. To what extent, if any, do sequential time-variant models (LSTMs) improve forecasting performance compared to static time-invariant models (MLPs)?

4. To what extent do different input feature set configurations (univariate, multivariate, and selected feature subsets) affect forecast performance, and which configuration performs best?

1.5 Related Work

Forecasting electricity spot market prices and volumes has been extensively studied in the literature using both classical statistical methods and machine learning approaches. Traditional models such as ARIMA and SARIMAX have been widely applied due to their interpretability and strong theoretical foundation. However, modern electricity markets are characterized by nonlinear dynamics, changing conditions, and a large number of influencing factors, which can limit the performance of purely statistical approaches [7].

As a result, machine learning methods have received increasing attention in electricity market forecasting. Recent studies have shown that neural network-based models can improve forecasting accuracy by capturing complex nonlinear relationships and temporal dependencies in market data. In particular, recurrent architectures such as LSTM networks have become widely used for electricity price and energy forecasting tasks [7].

Compared to the extensive literature on day-ahead and intraday electricity price and volume forecasting, relatively fewer studies address balancing markets and ancillary service market forecasting, such as for mFRR prices and volumes. It is therefore not clearly established whether classical or machine learning-based approaches are superior in this area, as both remain less explored.

This thesis contributes to this research area by analysing the factors influencing mFRR prices and volumes, and by comparing naive benchmark models, MLP networks, and LSTM networks for forecasting mFRR Capacity Market and Energy Activation Market outcomes in SE3.

2

Theory

2.1 The Swedish Power System and Electricity Market

This section describes the Swedish power system and electricity market from a system-level perspective. It first outlines the European market framework and the physical structure of electricity production, consumption, and exchange, followed by the transmission and distribution network. It then introduces the main actors and their roles, before presenting the structure of electricity markets and balancing markets.

2.1.1 The European Electricity System and Market Framework

The European electricity system is an integrated cross-border market that enables electricity to be traded and physically exchanged between countries. It allows national electricity systems to operate within a shared European framework, where prices and flows are determined by supply, demand, and available transmission capacity [8].

The system is coordinated through cooperation between national Transmission System Operators (TSOs) within ENTSO-E, which supports the operational coordination and development of the European power system [9]. In addition, ACER provides regulatory oversight and supports the development and alignment of common rules and practices in the European electricity market [10].

2.1.2 Electricity Production, Consumption, and Exchange

The power system continuously balances electricity production, consumption, and exchange. Since large-scale electricity storage remains limited, supply and demand must be balanced in real-time to maintain stable system operation.

Electricity production in Sweden is based on a diversified generation mix consisting primarily of hydropower, nuclear power, wind power, and thermal generation. These are commonly divided into dispatchable (hydro, nuclear, thermal) generation, which can be adjusted after demand, and variable renewable (wind and solar) generation, whose output is weather-dependent [4].

Electricity consumption varies over time and is influenced by factors such as weather conditions, economic activity, and human behavior. Demand typically follows recurring daily, weekly, and seasonal patterns, with higher consumption during morning and evening peak hours, and during colder periods of the year.

Electricity is continuously exchanged through the interconnected Nordic and European power systems. Cross-zonal and cross-border exchanges enable electricity to flow between regions and countries, allowing electricity to be imported and exported as part of the integrated European electricity market [11, 12].

2.1.3 Transmission System and Power Grid

The Swedish power system is structured into a high-voltage transmission grid and multiple regional and local distribution grids.

The transmission grid operates at a national level and forms the backbone of the Swedish power system. It enables bulk power transfer over long distances between major generation and consumption centers, as well as cross-border interconnections. The transmission grid is owned and operated by *Svenska kraftnät*, the Swedish Transmission System Operator (TSO)[13].

The regional and local distribution grids are operated by different Distribution System Operators (DSOs), which are responsible for delivering electricity from the transmission grid to end-users within their respective areas.

In addition to operating the transmission network, *Svenska kraftnät* is responsible for maintaining system balance, frequency control, and secure real-time system operation, as well as ensuring sufficient transmission capacity within the network[5].

2.1.4 Actors and Responsibilities

The operation of the Swedish power system and electricity market involves several actors with different responsibilities [14].

At the system level, *Svenska kraftnät* acts as the Transmission System Operator (TSO) and is responsible for operating the transmission grid, maintaining system balance, and ensuring secure system operation. Distribution System Operators (DSOs) are responsible for operating and maintaining regional and local distribution grids [5].

Electricity producers generate and inject electricity into the system, while consumers withdraw and use electricity. Electricity traders and suppliers buy and sell electricity across different electricity markets and provide electricity supply contracts to end-users.

Within the balancing framework, Balance Responsible Parties (BRPs) are financially responsible for maintaining balance between planned production and consumption, while Balance Service Providers (BSPs) provide balancing services and reserve capacity that can be activated by the TSO when needed [14].

2.1.5 Electricity Markets

Electricity trading in Sweden takes place through a sequence of interconnected markets with different time horizons, shown in Figure 2.1. These markets enable participants to manage risk, schedule production and consumption, and maintain system balance. illustrates the overall market structure [3].

2.1.5.1 Bidding Zones

Sweden is divided into four bidding zones, or electricity areas, shown in Figure 2.2, denoted SE1 (Luleå), SE2 (Sundsvall), SE3 (Stockholm), and SE4 (Malmö), stretching from north to south. This division reflects structural limitations in the transmission network and enables secure and efficient operation of the power system. Northern Sweden generally has a surplus of electricity production, while most electricity consumption is concentrated in

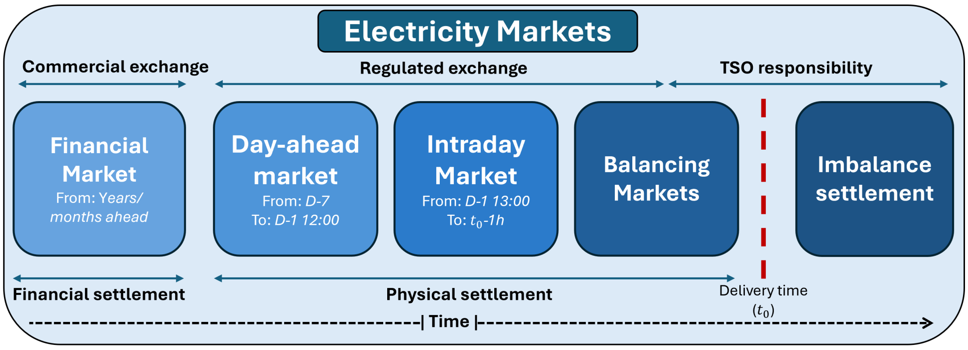

Figure 2.1: Overview of the electricity market structure and the sequence of its submarkets, spanning from long-term financial markets to real-time physical balancing markets. Adapted from *Acer* [15].

southern regions, particularly around major urban and industrial areas, where production is in deficit [16].

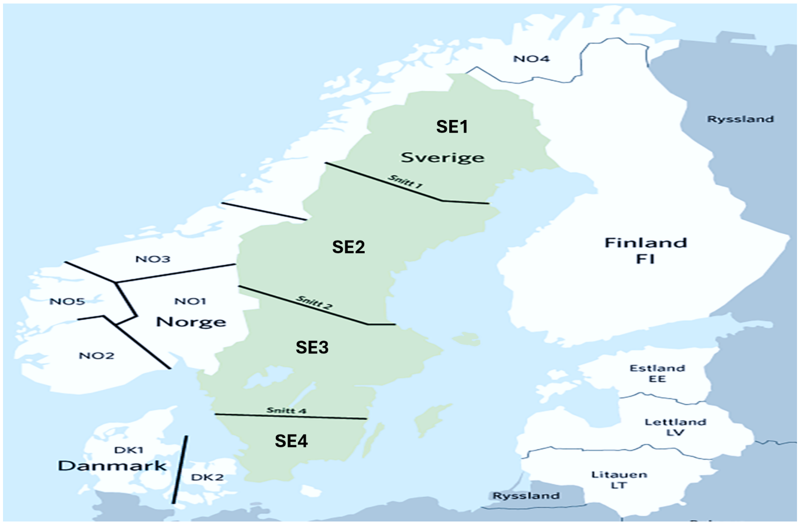

Figure 2.2: Swedish bidding zones: SE1, SE2, SE3, and SE4. Adapted from *Svenska kraftnät* [16].

The boundaries between bidding zones correspond to physical bottlenecks in the transmission grid, where limited transfer capacity restricts how much power can be transported between regions without compromising system stability. As long as sufficient transmission capacity is available, which is ensured and managed by the TSOs, electricity can flow freely between bidding zones and the prices equalize across zones. However, when demand exceeds the available capacity, congestion occurs and prices begin to diverge, with higher prices in deficit regions and lower prices in surplus regions. By embedding these network constraints directly into market pricing, the bidding zone system ensures a secure and efficient power allocation [17].

2.1.5.2 Electricity Price Formation

Electricity prices are generally formed through a marginal mechanism based on the merit order principle, where available resources are activated in order of increasing cost. As demand increases, progressively more expensive resources are required, and the price is ultimately determined by the marginal resource needed to balance supply and demand [3].

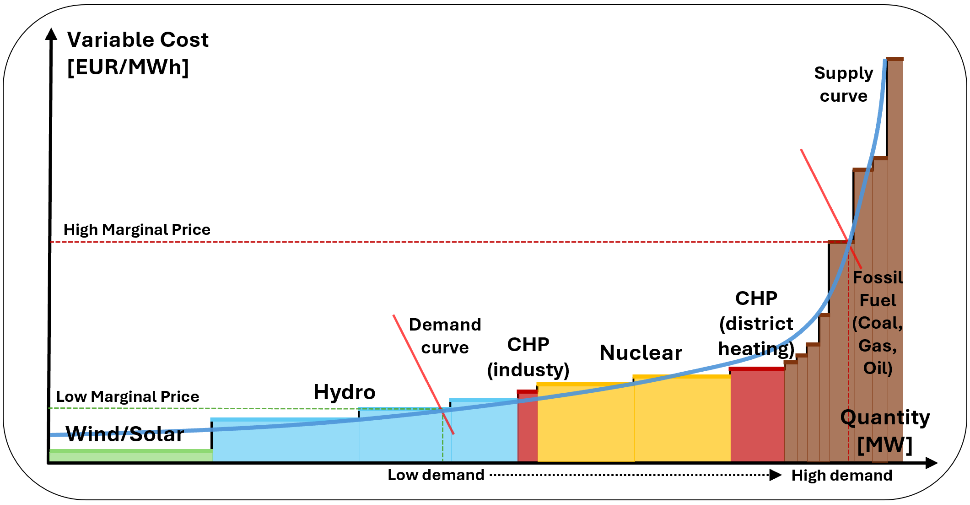

Figure 2.3: Marginal pricing according to the merit order curve, based on the intersection of supply and demand. Adapted from *Energimarknadsinspektionen* [18]
.

This is illustrated in Figure 2.3, which shows the interaction between supply and demand through the merit order curve. Marginal pricing is widely used in electricity markets, particularly in day-ahead markets and several balancing-related markets, although alternative pricing mechanisms such as pay-as-bid and capacity-based payments are also used depending on the market design.

2.1.5.3 Forward Market and Futures

The forward market allows market participants to hedge against future electricity price fluctuations and reduce uncertainty in future revenues and costs. This is commonly done through futures contracts, which specify a predetermined electricity price for a future delivery period. In the Nordic market, such contracts are available with monthly, quarterly, and yearly maturities, ranging from the following month up to several years ahead. In addition to contracts linked to the Nordic system price, market participants can also use EPAD contracts to hedge price differences between individual bidding zones [19].

2.1.5.4 Day-Ahead Market

The majority of electricity is traded on the day-ahead spot market, through organized electricity exchanges [20]. The main participants are electricity producers, electricity traders, and large energy consumers. In Sweden, trading takes place primarily on two exchange platforms, *Nord Pool* [21] and *EPEX SPOT* [22], where market actors submit bids for physical electricity delivery within specific electricity bidding zones. Despite trading on different exchanges, market coupling across Europe via the Single Day-Ahead Coupling (SDAC) ensures a single equilibrium price for each bidding zone [23].

Each day at 12:00 CET, an auction takes place for delivery during the following day. Buyers and sellers submit the volumes they wish to trade for each 15-minute interval from hour 00 to 24, together with the prices they are willing to pay or accept and the bidding zone in which they are located. All submitted bids are aggregated into supply and demand curves, and the resulting spot price, i.e. the equilibrium price, is set at their intersection. The market applies a marginal pricing principle, meaning that the cost of the last unit of electricity required to meet demand determines the price for all traded electricity within the same bidding zone and delivery interval. Consequently, all participants in a given area get the same price regardless of power source [20, 23, 24].

2.1.5.5 Intraday Market

The intraday spot market allows market participants to adjust their electricity positions after the closure of the day-ahead market and closer to physical delivery. It is mainly used to manage forecast errors related to demand, renewable generation, and unforeseen system events. The primary participants are balance responsible parties, who trade to reduce imbalances and thereby balancing costs [25].

In Sweden, intraday trading takes place primarily on *Nord Pool* [21] and *EPEX SPOT* [22], which are integrated into the European Single Intraday Coupling (SIDC) system. This market coupling enables cross-border trading as long as transmission capacity is available, resulting in common price formation across connected bidding zones regardless of the exchange used.

Intraday trading in Sweden consists of continuous trading and three intraday auctions. The continuous trading allows transactions to be executed immediately on a pay-as-bid basis, with contracts available in 15-minute, 30-minute, and hourly intervals, and trading is typically open until one hour before the delivery hour. In addition, three intraday auctions; IDA1, IDA2, and IDA3, are held at the specific times; 15:00 CET and 22:00 CET the day before, and 10:00 CET during the delivery day, to pool liquidity and establish reference prices for upcoming delivery periods. By enabling adjustments closer to delivery, the intraday market supports system balance and reduces the need for balancing actions such as mFRR activation [25, 26, 24].

2.1.6 Balancing Markets and Ancillary Services

To ensure system stability, maintain the real-time balance between electricity production, consumption, and cross-border exchanges, and maintain operational security of the power system, imbalances in the Nordic power system are managed through the balancing market. In this market, ancillary service products are procured and activated to provide both upward regulation, by increasing generation or reducing consumption, and downward regulation, by decreasing generation or increasing consumption [27]. The different types of reserves are shown in Table 2.1.

The reserves differ mainly in their functions and response speeds. Fast Frequency Reserve (FFR) reacts almost instantly to sudden frequency drops, while Frequency Containment Reserves automatically stabilize frequency during normal operating deviations (FCR-N) and larger frequency disturbances (FCR-D). Frequency Restoration Reserves restore the system frequency to its nominal value of 50 Hz, with aFRR activated automatically and mFRR activated manually (or automatically), to relieve the automatic reserves and maintain balance over longer periods [28].

Together, these reserve products form a layered balancing framework that maintains system stability and real-time balance in the Nordic power system. To manage the increasing shares of variable renewable generation, *Svenska kraftnät* continuously develops these markets and integrates new balancing resources [29].

Table 2.1: Overview of balancing market reserve products and their main technical characteristics. Adapted from *Svenska kraftnät* [28].

	FFR	FCR-N	FCR-D up	FCR-D down	aFRR	mFRR CM	mFRR EAM
Regulation direction	Up	Symmetric up/down	Up	Down	Up/down	Up/down	Up/down
Minimum bid size	0.5 MW	0.1 MW	0.1 MW	0.1 MW	1 MW	1 MW	1 MW
Market gate closure	Seasonal procurement	D-1 00:30 / 18:00	D-1 00:30 / 18:00	D-1 00:30 / 18:00	Opens 7 days ahead, closes D-1 07:30	D-1 09:00	Continuous until real time
Activation principle	Automatic response to rapid frequency drops at low inertia	Automatic linear frequency response	Automatic linear frequency response	Automatic linear frequency response	Automatic AGC-based frequency restoration	Capacity reservation for later activation	Manual TSO activation based on real-time imbalance
Activation range	Below frequency threshold	49.90–50.10 Hz	49.50–49.90 Hz	50.10–50.50 Hz	Continuous control signal	Activated through mFRR EAM	TSO dispatch signal
Full activation time	100% within 0.7–1.3 s	According to technical requirements	According to technical requirements	According to technical requirements	100% within 5 min	Available within 12.5 min	100% within 12.5 min
Minimum endurance	30 s (or 5 s), reactivation within 15 min	1 h at 100%	Min. 20 min	Min. 20 min	1 h	Contracted availability period	Typically 15–60 min

2.1.6.1 Manual Frequency Restoration Reserve (mFRR) Market

Manual Frequency Restoration Reserve (mFRR) is a balancing service used by transmission system operators (TSOs) to secure available capacity and activate reserves as needed to restore the system frequency to its nominal value of 50 Hz following imbalances between production, consumption, and cross-border exchanges. It operates after faster automatic reserves have stabilized the initial disturbance, providing additional upward or downward regulation to maintain system stability [30].

The mFRR market consists of two interconnected components: the mFRR Capacity Market (CM) and the mFRR Energy Activation Market (EAM). The CM procures balancing capacity ahead of delivery, while the EAM activates balancing resources during system operation. When imbalances occur, the TSO activates the most cost-efficient bids from the available mFRR resources, considering network constraints and system needs, to restore real-time balance in the power system [30, 31].

2.1.6.1.1 mFRR Capacity Market (CM)

The mFRR Capacity Market (CM) is a day-ahead market where TSOs procure balancing capacity to ensure that sufficient reserves are available for the following day. Market participants submit bids specifying reserve volume and capacity price, and accepted bids receive a capacity payment for keeping resources available for potential activation. Participation in CM also requires providers to submit corresponding bids to the mFRR Energy Activation Market (EAM) [30].

The market opens seven days before delivery and closes at 07:30 on the day before delivery (D1). Bids are submitted per bidding zone (e.g. SE1–SE4) with a minimum bid size of 1 MW. The market clears on an hourly basis, producing marginal capacity prices and procured reserve volumes for the following day. The CM therefore forms the foundation for subsequent balancing activations in the EAM [30].

2.1.6.1.2 mFRR Energy Activation Market (EAM)

The mFRR Energy Activation Market (EAM) determines the real-time activation of balancing energy during system operation, compared to CM which procures availability of balancing reserves ahead of delivery. Activation is initiated by the TSOs based on forecasts of system imbalances in each bidding zone. Within the framework of the Nordic Balancing Model (NBM), this process is increasingly automated through a common activation optimization function (AOF), which selects bids based on cost efficiency while considering transmission constraints [30, 31].

In the EAM, bids can be submitted from 13:00 on the day before delivery (D-1) until 45 minutes before the relevant delivery period. Bids are submitted per regulating object with a minimum bid size of 1 MW. When activated, resources must deliver 100% of the requested power within 15 minutes, and they must be able to maintain the required effect for 15 minutes for scheduled activations and for 30 minutes for direct activation. Activated bids receive energy payments, typically based on marginal pricing. The market operates with a 15-minute temporal resolution and enables coordinated balancing across the Nordic power system [30].

2.2 Time-Series Analysis and Forecasting

This section introduces the main characteristics of the underlying dynamics of time-series, as well as their forecast-related key properties.

2.2.1 Time-Series Characteristics

A time-series is a sequence of observations recorded over time, usually at regular intervals such as hourly, daily, or monthly [32]. Unlike cross-sectional data, time series observations are ordered in time, and past values often influence future values and behaviors. As a result, time-series analysis focuses on identifying patterns that describe how data evolve over time.

Common time-series patterns include trends, seasonality, and cycles. A trend represents long-term upward and/or downward movements in the data. Seasonal patterns occur when regular and repeating fluctuations appear with a fixed frequency, such as daily, weekly, or yearly. Cycles describe longer-term fluctuations that do not have a fixed period and are often related to economic or structural conditions. Real-world time-series often contain a combination of these components, and understanding these characteristics is a fundamental step in selecting appropriate forecasting methods [32].

2.2.2 Stationarity and Seasonality

Stationarity refers to whether the statistical properties of a time-series, such as mean and variance, remain constant over time. Many forecasting models assume that the underlying process is stationary, making transformations such as differencing or detrending necessary when strong trends or changing variability are present [32].

Seasonality represents systematic and predictable patterns that repeat over fixed time intervals. While seasonality introduces non-stationarity, it can often be modeled explicitly or removed through seasonal adjustment or decomposition methods. Identifying both stationarity and seasonal effects is important for building reliable forecasting models [32].

2.2.3 Forecast Horizons and Temporal Resolution

The forecast horizon describes how far into the future predictions are made, such as short-term (minutes to hours), medium-term (days to weeks), or long-term (months to years). Different horizons require different modeling strategies; short-term forecasts rely more on recent observations and dynamics, while long-term forecasts are more dependent on trends and seasonal behaviors [32].

Temporal resolution refers to the frequency at which data are recorded, such as hourly, daily, or monthly intervals. Higher resolution provides more detailed information but may introduce more noise, while lower resolution smooths short-term fluctuations but can hide important details and dynamics [32].

The selection of suitable forecast horizons and time resolutions is important and should reflect and align with the practical settings and goals of the forecasting task.

2.3 Forecasting Models

This section presents the forecasting models used in this study, covering both classical statistical approaches and modern machine learning methods. The aim is to provide the theoretical background for the models used in this study and to describe their main assumptions and characteristics in a time-series forecasting context.

2.3.1 Classical Statistical Models

Classical statistical forecasting models are based on explicit assumptions about the underlying structure of time-series data, such as linear relationships, temporal dependence, and trend or seasonal components. These models are typically interpretable and rely on relatively strong assumptions about stationarity and the structure of relationships between variables.

Common approaches in this category include linear regression, exponential smoothing, and autoregressive models such as ARIMA and its extensions. These methods have historically been widely used in time-series forecasting and often serve as strong and interpretable baseline models in many applied forecasting problems [32].

2.3.1.1 ARIMA

Autoregressive Integrated Moving Average (ARIMA) models are a classical and widely used approach for time-series forecasting since their introduction in 1970 [33]. ARIMA models assume stationarity and capture temporal dependencies (autocorrelation) by relating current observations to past values and past forecast errors. In practice, real-world time-series often need to be transformed to achieve stationarity, and techniques such as detrending or differencing are commonly used to remove trends and seasonal effects [32].

Different ARIMA variants are applied depending on the structure of the data and the forecasting task. Seasonal extensions such as SARIMA are used to model repeating patterns, while SARIMAX includes external explanatory variables. For multivariate time-series, vector autoregression (VAR) models are commonly used to capture interactions between multiple variables. These model classes are widely used as classical benchmarks in time-series forecasting tasks [32, 34].

2.3.2 Machine Learning Models

Machine learning models are a broad class of data-driven methods that learn patterns and relationships directly from data. Depending on their underlying structure, they range from relatively interpretable models, such as linear regression and decision tree-based methods, to more complex black-box models, such as neural networks. The following subsections introduce machine learning models relevant to time-series forecasting and provide the theoretical background for the models considered in this study.

2.3.2.1 Multi-Layer Perceptron (MLP)

Multi-Layer Perceptrons (MLPs) are standard fully connected feed-forward neural networks composed of an input layer, one or more hidden layers, and an output layer. The model architecture is shown in Figure 2.4.

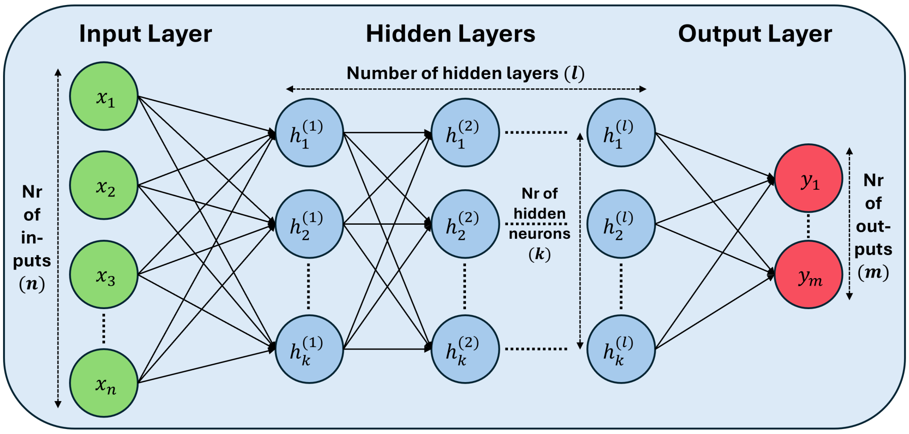

Figure 2.4: General architecture of a Multi-Layer Perceptron (MLP) consisting of an input layer, one or more hidden layers, and an output layer.

Each layer consists of a number of neurons (perceptrons), shown in Figure 2.5, where each neuron has its own parameters: weights (one for each input connection) and a bias term. Each neuron computes a weighted sum of its inputs, adds the bias, and applies a nonlinear activation function. This allows the network to learn complex nonlinear relationships between inputs and outputs.

MLPs are trained using the backpropagation algorithm, which adjusts weights based on the gradient of a loss function with respect to network parameters [35, 36]. By propagating error signals backward through the network, MLPs can learn representations that capture patterns in the input data, making them effective for a wide range of supervised learning tasks.

In practice, MLPs are best suited for fixed-size input data and tasks with no temporal dependencies, or with minimal temporal dependencies that can be encoded through feature

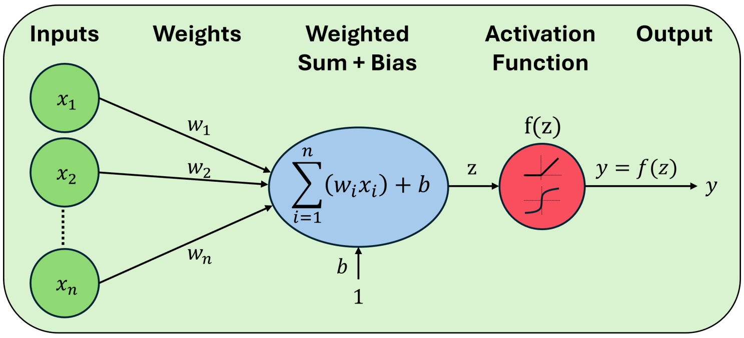

Figure 2.5: Structure of a perceptron, showing the weighted inputs, bias term, and activation function.

engineering, such as lagged values or rolling statistics. Unlike recurrent architectures, MLPs do not inherently model sequential dependencies since they do not maintain memory of past inputs, which limits their ability to model sequential or time-dependent patterns. MLPs therefore struggle to capture temporal dynamics, especially long-term ones. Despite this limitation, MLPs are computationally efficient, straightforward to implement, and often serve as a strong baseline for more complex models [36].

2.3.2.2 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) are a class of neural networks designed to handle sequential data by maintaining a hidden state that evolves over time. Through recurrent connections, the output of hidden units is fed back into the network as input for the next time step, allowing the model to form dynamic internal representations that capture temporal dependencies and provide a form of memory [37]. By incorporating past inputs and hidden states into current computations, RNNs can capture how previous events influence future outcomes, making them well-suited for tasks where context and sequence order matter, such as time-series forecasting.

In practice, standard RNNs are limited by vanishing and exploding gradient problems during training. As gradients are back-propagated through many time steps, they may decay or grow exponentially, causing the network to forget long-term dependencies, become numerically unstable, and ultimately limiting the network's ability to learn effectively [38]. As a result, RNNs are primarily effective for modelling short- to medium-term temporal dependencies. Despite these limitations, RNNs provide a useful baseline for sequential modelling and form the foundation for more advanced recurrent architectures, such as LSTM networks, which are designed to capture long-term dependencies more effectively.

2.3.2.3 Long-Short-Term-Memory (LSTM)

Long Short-Term Memory (LSTM) networks extend the RNN architecture by introducing memory cells and gating mechanisms that regulate the flow of information and preserve gradient propagation over time. By truncating gradients where it does not harm learning, LSTMs can bridge time lags spanning thousands of discrete steps through specialized memory units that enforce constant error flow. The gates determine which information is stored, updated, or forgotten within the memory cells, allowing LSTMs to overcome the vanishing gradient problem and retain relevant historical context over long sequences. This enables the modeling of both short-term fluctuations and long-term trends in sequential data [39].

In their original experiments, LSTM networks outperformed earlier recurrent models, including standard RNNs, backpropagation through time, and Elman networks, particularly on tasks involving long time delays and noisy data. They learned faster and produced more reliable results on complex sequence problems that traditional RNNs struggled to solve. These properties make LSTMs especially suitable for time-series with delayed effects and strong temporal structure [39].

2.3.2.4 Random Forest (RF)

Random Forest (RF) is an ensemble learning method that combines a large number of decision trees to improve predictive performance and robustness [40].

A decision tree is a hierarchical model consisting of nodes and branches. Each node represents a binary split of the data based on an input feature and a threshold, chosen to reduce the error or improve separation of the target variable. At each level, the data are split into two subsets according to this rule, and the process is repeated recursively until reaching the leaf nodes, where the final prediction is produced. While decision trees can capture complex patterns, they are highly flexible and prone to overfitting when used individually.

Each tree in a Random Forest is trained on a random subset of the training data and input features, introducing diversity among the trees. By aggregating many trees, Random Forest reduces variance and overfitting while retaining the ability to model non-linear relationships and feature interactions. This typically leads to improved predictive performance and generalization.

For regression tasks, predictions are averaged across trees, while for classification tasks a majority vote is used. Random Forest also provides an intrinsic measure of feature importance, computed during training, making it useful for both prediction and feature selection [40].

2.4 Data Preprocessing and Feature Engineering

This section presents the main data preprocessing steps and feature engineering techniques used to transform raw data into meaningful and suitable inputs for machine learning models.

2.4.1 Data Cleaning and Temporal Alignment

Data cleaning and temporal alignment are fundamental preprocessing steps in time-series analysis that ensure the data is consistent, complete, and properly structured over time. This typically involves handling missing values, outliers, noise, and irregular sampling. Common methods like resampling, aggregation, interpolation, forward- and backward-filling.

These steps help ensure a temporally aligned and coherent dataset suitable for subsequent data data processing and engineering steps [41].

2.4.2 Feature Construction

Appropriate data representations are essential for machine learning and strongly influence the achievable predictive performance of trained models. Feature construction refers to the process of creating new input variables derived from the original data. It provides an opportunity to incorporate domain knowledge into the feature set, making the resulting dataset highly application-specific. Using constructed features rather than relying only on raw input variables often leads to improved model accuracy, robustness, and interpretability [41].

Feature construction can involve mathematical transformations, combinations of variables, or more advanced techniques such as linear projections and cluster-based representations. In time-series and forecasting problems, features are commonly engineered from historical data to reflect temporal dynamics. Typical examples include lagged variables, moving averages, rolling variances, growth rates, and seasonal indicators. These derived features allow models to learn trends, volatility, and delayed effects that are not explicitly visible in the original observations.

2.4.3 Feature Scaling

Feature scaling is a fundamental data preprocessing step that transforms input variables with different units, magnitudes, and distributions into a common numerical scale, ensuring that no feature dominates the representation due to its scale alone [41]. The most commonly used approaches are standardization and normalization, which differ in how they transform the data, particularly in terms of centering, scaling, and bounding.

Standardization transforms features to have zero mean and unit variance:

$$X' = \frac{X - \mu}{\sigma} \tag{2.1}$$

where μ is the mean and σ is the standard deviation of the feature. This transformation preserves the underlying distributional structure while centering and rescaling the data.

Normalization (min-max scaling) rescales features to a fixed interval, typically $[0, 1]$:

$$X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}. \tag{2.2}$$

This approach preserves the relative ordering of values while constraining them to a bounded range, which can be useful when a fixed scale is required.

Overall, feature scaling improves the consistency of the data representation and can contribute to more stable and efficient model training.

2.4.4 Feature Selection

Feature selection focuses on identifying a subset of input variables that are most informative for a given prediction task while keeping the original features unchanged. This is achieved by removing irrelevant, redundant, or noisy variables, which helps improve model interpretability, generalization performance, and reduce computational complexity. An important idea in feature selection is that variables should not only be useful on their own but also work well together. Some features that seem weak individually may become valuable when combined, while highly correlated features may carry overlapping information and can be removed with little loss in performance. Feature selection methods are commonly grouped into filter methods based on simple statistical measures, embedded methods that perform selection during model training, and wrapper methods that evaluate feature subsets using model performance.

2.4.4.1 Statistical Filtering Methods

Statistical filtering methods select features as a preprocessing step, independently of the chosen predictive model. Each variable is evaluated based on simple statistical measures that describe its relationship with the target variable, such as correlation coefficients, t-tests, and F-tests [41].

These methods are computationally efficient and are often used to reduce the dimensionality of the input space before applying more complex models. A key limitation is that features are assessed individually, meaning that interactions and redundancy between variables are not considered. Nevertheless, filtering provides a practical first step for identifying potentially informative variables in large datasets [41].

2.4.4.2 Embedded Methods

Embedded methods perform feature selection as part of the model training process. Instead of selecting features as a separate preprocessing step, the model itself learns which variables are most relevant for the prediction task during training. This is typically achieved by assigning importance to features based on their contribution to reducing prediction error.

Compared to filtering methods, embedded approaches can capture nonlinear relationships and feature interactions, while being more computationally efficient than wrapper methods since feature selection and model training are performed simultaneously. Common examples of embedded methods include regularization-based models such as Ridge (L2), Lasso (L1), and Elastic Net, as well as tree-based models such as Random Forests [41].

These methods are particularly useful in high-dimensional settings, where they provide a balance between model complexity and predictive performance by automatically reducing the influence of less informative variables during training.

2.4.4.3 Wrapper Methods

Wrapper methods evaluate subsets of variables based on the predictive performance of a chosen learning model, treating the model as a black box [41]. Instead of assessing each feature individually, different combinations of variables are tested and compared according to their predictive power, typically using validation data or cross-validation. Compared to filtering methods, wrappers often provide improved predictive performance since they capture interactions between variables and are optimized directly for the selected model.

Common greedy search strategies include forward selection and backward elimination. Forward selection starts from an empty feature set and progressively adds the variable that yields the largest performance improvement, while backward elimination begins with all variables and iteratively removes the least useful one. These approaches produce nested subsets of variables and are computationally efficient compared to exhaustive search. However, wrapper methods remain more computationally demanding than filtering and embedded approaches, since the learning model must be retrained repeatedly [41].

2.5 Model Training, Validation, and Evaluation

This section describes the training procedure, validation strategy, hyperparameter selection, and evaluation methodology used to assess and compare the forecasting models.

2.5.1 Model Training and Optimization

Training a neural network consists of learning model parameters that minimize a predefined loss function. This is achieved through an iterative process involving forward propagation, loss evaluation, back-propagation, and parameter updates.

In the forward pass, input data are propagated through the network layer by layer to produce predictions. The magnitude of the deviation between predicted and true values is then quantified using a loss function, such as mean squared error (MSE). During back-propagation, gradients of the loss with respect to the model parameters are computed using the chain rule, allowing the model to identify how each parameter contributes to the prediction error [35].

The parameters are subsequently updated using gradient-based optimization algorithms, such as stochastic gradient descent (SGD) or adaptive methods like Adam [42]. These methods iteratively adjust the parameters in the direction that reduces the loss the most, with the learning rate controlling the step size of each update.

In practice, training is performed using mini-batches of data, which improves computational efficiency and introduces stochasticity that can improve generalization [43]. Several hyperparameters influence the training dynamics and convergence, including the learning rate, batch size, number of training epochs, and optimization algorithm. In addition, regularization techniques such as dropout are commonly applied to reduce overfitting by randomly deactivating neurons during training.

Finally, the architecture of the network itself, including the number of layers and neurons, determines the model capacity. Shallow models (traditional machine learning) typically have limited representational power, while deeper networks can capture more complex patterns but require careful training to avoid instability and overfitting.

2.5.2 Hyperparameters

The architecture and learning behavior of a neural network are determined by its hyperparameters, which control both model capacity and the training dynamics.

- **Number of hidden layers** determines network depth. Deeper models can learn more complex patterns, but increase the risk of overfitting.
- **Hidden layer size** defines the number of neurons per hidden layer. Larger sizes increase model capacity, but also computational cost and overfitting risk.
- **Learning rate** controls update step size during gradient-based optimization. Higher values may destabilize training, while lower values slow convergence.
- **Dropout rate** randomly deactivates neurons during training, reducing overfitting and improving generalization.
- **Epochs** determine the number of full passes of the training data through the model during training. Too few epochs may lead to underfitting, while too many may result in overfitting.

- **Batch size** determines the number of training samples processed before the model parameters are updated. Larger batch sizes improve computational efficiency and convergence speed, while smaller batches may improve generalization but with more noisy updates.

2.5.3 Model Complexity and Generalization

A key challenge in machine learning is achieving good generalization, meaning that the model performs well not only on the training data but also on unseen data [43]. This is closely related to the concept of model complexity and the bias–variance trade-off.

Models with low complexity may be unable to capture the underlying structure of the data, leading to underfitting and high bias. Conversely, highly complex models with many parameters, such as deep neural networks with multiple layers and large numbers of neurons, can fit the training data very closely but risk overfitting, where the model captures noise and fails to generalize to new data.

The bias–variance trade-off describes this balance: increasing model complexity typically reduces bias but increases variance. The goal is therefore to find an appropriate level of complexity that captures the essential patterns in the data without fitting noise.

Several factors influence this balance. Architectural choices, such as the number of layers and neurons, directly affect model capacity. Training-related hyperparameters, including learning rate and batch size, impact convergence behavior and stability. Regularization techniques, such as dropout, early stopping, and data splitting strategies, further help control overfitting by encouraging the model to learn more robust and generalizable patterns.

Careful tuning of both model architecture and training hyperparameters is therefore essential to achieve good predictive performance and reliable generalization.

2.5.4 Validation Methods

Model validation is used to estimate how well a forecasting model generalizes to unseen data. A common approach is hold-out validation, where the available data are divided into separate training, validation, and test sets. The model is trained on the training set, tuned using the validation set, and finally evaluated on the independent test set.

For time-series forecasting, validation must preserve the chronological order of observations to avoid data leakage from future to past. A static (chronological hold-out) split is therefore used for final evaluation, while rolling (walk-forward) validation is commonly applied during model tuning. In rolling validation, training and validation windows are shifted forward through time while maintaining temporal order. Standard random cross-validation is not suitable for time-series data, as it mixes observations across folds and can allow future information to influence model training, leading to overly optimistic performance estimates [32].

2.5.5 Evaluation Metrics

Forecasting models are commonly evaluated using error metrics that quantify the difference between predicted and observed values. Two of the most widely used metrics are Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

Root Mean Squared Error (RMSE) is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2.3}$$

and Mean Absolute Error (MAE) is defined as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{2.4}$$

where N denotes the number of observations, $\hat{y}_i$ the predicted value, and y_i the corresponding observed value.

RMSE penalizes large errors more strongly due to the squared error term and is therefore particularly useful when large forecasting errors are unwanted.

MAE measures the average magnitude of prediction errors and is easily interpretable since it is expressed in the same units as the target variable [32].

3

Methodology

3.1 Overall Methodological Framework

Forecasting electricity market variables requires methods that can handle dynamic, nonlinear, and multivariate time-series data. In particular, mFRR CM and EAM prices and volumes are influenced by a combination of market, system, environmental, and temporal factors, creating complex dependencies across both time and variables. To address these challenges, this study adopts a data-driven forecasting framework consisting of data collection, preprocessing, feature engineering, feature filtering and selection, model development, optimization, and evaluation.

3.1.1 Methodological Workflow

The methodology follows a strictly time-ordered workflow to preserve the temporal structure of the data and avoid data leakage. First, market, system, environmental, and temporal variables are collected, aligned, and preprocessed. Feature filtering and selection are then applied to reduce dimensionality and identify the most informative predictors. Next, multiple forecasting models are implemented, ranging from simple statistical baselines to deep learning architectures. Finally, all models are evaluated under a consistent train-validation-test framework to ensure a fair comparison of predictive performance.

The workflow is designed to support both predictive performance and interpretability, enabling analysis of which variables are most important for explaining and forecasting mFRR prices and volumes, as well as a structured comparison of forecasting performance across different modelling approaches and configurations.

3.1.2 Forecasting Problem Formulation

This study treats mFRR price and volume forecasting as a supervised learning problem based on historical observations and available forecasts of market, system, and environmental variables. Although the forecasting targets are defined for SE3, cross-zonal and system-wide effects are partially captured through the included Swedish and Nordic input variables.

The modelling framework consists of three model categories: naive baseline models, Multilayer Perceptrons (MLP), and Long Short-Term Memory (LSTM) networks. Baseline models provide reference performance levels, while MLP and LSTM models represent static and sequential machine learning approaches, respectively. Different feature set configurations are evaluated to assess the effect of input dimensionality and feature selection on forecasting performance.

This setup enables a structured comparison across several important dimensions of the study. It addresses the key research questions of (1) which variables are most important for explaining and forecasting mFRR prices and volumes, (2) whether machine learning models outperform simple historical baseline models, (3) whether sequential time-variant models provide advantages over time-invariant models, and (4) whether additional input features and feature selection improve forecasting performance, robustness, and generalization.

Overall, this structured approach makes it easier to relate differences in forecasting performance to specific modelling choices. This provides a clear and interpretable foundation for the analysis and discussion of the results.

3.2 Data Collection

This section describes the data used in this study, including its sources, temporal and spatial coverage, and the selection of input features and target variables. Accurate modelling of mFRR prices and volumes requires reliable data that capture system operation conditions and electricity market dynamics at high temporal resolution over a long time horizon. The dataset is constructed to represent key aspects of the power system and electricity market relevant for forecasting mFRR prices and volumes in the SE3 bidding zone.

To achieve this, time-series data from multiple sources, including transmission system operators (TSOs) and professional market analytics platforms, were collected and merged into a unified dataset. The resulting dataset consists of both historical observations and forecasts of system, market, and environmental variables such as weather and temporal features, with sufficiently high temporal resolution to reflect short-term dynamics in the power system and electricity markets. Together, these data provide a comprehensive and consistent foundation for the subsequent data preprocessing, feature engineering, modelling, training, and evaluation.

3.2.1 Data Sources

This study combines data from transmission system operators (TSOs) and professional electricity market analytics platforms, including both historical observations and forecasted time-series data covering system, market, and environmental variables.

The majority of the system, market, and environmental time-series data used in this study were obtained from *Montel Syspower* [44] (formerly *SKM Market Predictor*), a professional power market analytics platform focused on the Nordic and Baltic electricity markets. Access was provided through a subscription by *Göteborg Energi*.

Additional system-level data were obtained directly from the transmission system operators. Reserve market time-series data for Frequency Containment Reserves (FCR), automatic Frequency Restoration Reserves (aFRR), and manual Frequency Restoration Reserve capacity market data (mFRR CM) were downloaded from *Svenska kraftnät's* database *Mimer* [45]. Data for the Nordic power system frequency, time deviation, and inertia were obtained from *Fingrid's* open data services [46].

3.2.2 Temporal Coverage and Resolution

All time-series in the collected dataset have a temporal coverage of exactly one year, stretching from 1 January 2025 to 31 December 2025. Most time-series have a temporal resolution of 15 minutes or 1 hour, meaning they are sampled either four times every hour at minutes 0, 15, 30, and 45, or just once each full hour at minute 0. There are also a few time-series with other temporal resolutions (e.g. 1 minute, 3 minutes, 3 hours, and 1 day) in the dataset.

3.2.3 Spatial Scope

Since the forecasting targets are the mFRR CM and EAM prices and volumes for SE3, the feature set is primarily focused on SE3 and variables directly affecting balancing market dynamics in this bidding zone. The original dataset included several hundred candidate features across Swedish bidding zones (SE1–SE4), Nordic bidding zones (NO1–NO5, DK1–DK2, FI), and Germany (DE), covering market prices, production, consumption, weather, and system-state variables at zonal, national, and Nordic system levels.

To reduce model complexity and workload, clear spatial delimitations were applied. The final feature set therefore focuses mainly on SE3-specific zonal-level variables, national-level variables for Sweden (e.g. aggregated production, consumption, and weather forecasts), cross-zonal variables (e.g. spot-price spread variables and scheduled intraday flows), and Nordic system-state variables (e.g. inertia, frequency, and time deviation). This reduces dimensional complexity and redundancy, making feature selection and model interpretation more manageable. This approach also allows the inclusion of a broader and more relevant set of variables for SE3, rather than fewer distinct variables repeated across multiple zones, which may introduce redundancy and reduce relevance.

3.2.4 Target Variables and Input Features

The dataset is structured into three main input feature categories: *system variables*, *market variables*, and *environmental variables*, as summarized in Tables 3.2, 3.3, 3.4, respectively. In addition, a separate set of *target variables* (see Table 3.1) is used as the ground truth, i.e. the future values that the models aim to predict, and serves as targets for model training and evaluation.

The tables provide an overview of all the variables used in the study and their main properties. Each row specifies the variable type and a brief description, followed by the value type, indicating whether the data represent historical observations (statistics) and/or forecasted values. The unit specifies the physical measurement, or currency, for the quantity of each variable, while the spatial scope defines the geographical level of aggregation. The temporal resolution indicates the sampling interval of the time-series, and the source identifies where the data were downloaded from.

3.2.4.1 Target Variables

The target variables are the mFRR upward (Up) and downward (Down) regulation prices and volumes for both the capacity market (CM) and the energy activation market (EAM), as summarized in Table 3.1. Only historical observed values are used for these targets, as forecasts for these variables are not available from any publicly accessible or subscription-based sources available in this study, and forecasting them is the objective of this work. These historical observations are therefore included as input features, allowing the models to capture temporal dependencies and autoregressive patterns between historical and future mFRR prices and volumes. Their future observed values are unknown to the model at the time of prediction and serve as the true target values (ground truth) against which forecasts are evaluated.

Table 3.1: Overview of the eight target variables used in this study, consisting of mFRR CM and EAM upward and downward regulation prices and volumes for SE3, and their main properties.

Variable	Description	Value Type	Unit	Spatial Scope	Temp. Res.	Source
mFRR CM						
Up Price	Marginal price for accepted up-regulation capacity in the mFRR capacity market for SE3.	Stat.	[€/MW]	SE3	1 h	Syspower
Up Volume	Total accepted up-regulation capacity in the mFRR capacity market for SE3.	Stat.	[MW]	SE3	1 h	Syspower
Down Price	Marginal price for accepted down-regulation capacity in the mFRR capacity market for SE3.	Stat.	[€/MW]	SE3	1 h	Syspower
Down Volume	Total accepted down-regulation capacity in the mFRR capacity market for SE3.	Stat.	[MW]	SE3	1 h	Syspower
mFRR EAM						
Up Price	Marginal price for activated up-regulation energy in the mFRR activation market for SE3.	Stat.	[€/MW]	SE3	15 min	Syspower
Up Volume	Total activated up-regulation energy in the mFRR activation market for SE3.	Stat.	[MW]	SE3	15 min	Syspower
Down Price	Marginal price for activated down-regulation energy in the mFRR activation market for SE3.	Stat.	[€/MW]	SE3	15 min	Syspower
Down Volume	Total activated down-regulation energy in the mFRR activation market for SE3.	Stat.	[MW]	SE3	15 min	Syspower

3.2.4.2 System Variables

The system variables describe the physical and operational state of the power system that directly drives imbalances and balancing needs. As summarized in Table 3.2, these include production (total and by generation type), consumption, and exchange flows for SE3 and Sweden, represented by historical observations and forecasts where available. Nordic power system-state variables are also included to capture system-wide operating conditions that influence local balancing needs and price formations.

3.2.4.3 Market Variables

The market variables represent prices and traded volumes across different electricity market stages and sub-markets, capturing market conditions and expectations that influence bidding behavior and price formation in balancing markets. As shown in Table 3.3, they include both realized and forecasted values and cover day-ahead spot markets, scheduled intraday flows, and balancing markets, primarily for SE3, but also for Sweden as a whole and the Nord Pool system price.

3.2.4.4 Environmental Variables

The environmental variables capture external factors, including weather and temporal (seasonal) patterns, that influence system conditions, market behavior, and the supply-demand balance. As summarized in Table 3.4, these variables are primarily forecast-based. Weather variables represent meteorological conditions affecting renewable generation and electricity consumption, while temporal features capture systematic patterns, such as hourly, daily, and weekly demand cycles, as well as seasonal variations.

Table 3.2: Overview of the system variables, and their main properties, used as input features for forecasting mFRR CM and EAM upward and downward prices and volumes for SE3.

Variable	Description	Value Type	Unit	Spatial Scope	Temp. Res.	Source
Frequency	System frequency measures the instantaneous power balance (supply vs. demand) in the Nordic system (nominal 50 Hz in equilibrium).	Stat.	[Hz]	Nordics	3 min	Fingrid
Time Deviation	Integrated deviation of the system frequency over time, indicating accumulated frequency imbalance in the Nordic power system.	Stat.	[s]	Nordics	3 min	Fingrid
Inertia	Inertia, or kinetic energy, of rotating masses in the Nordic power system	Stat.	[GWs]	Nordics	1 min	Fingrid
Production	Total electricity production statistics and forecasts for Sweden and SE3.	Stat./ Frcst.	[GW]	SE3/ Sweden	15 min	Syspower
Consumption	Total electricity consumption statistics and forecasts for Sweden and SE3.	Stat./ Frcst.	[GW]	SE3/ Sweden	15 min	Syspower
Exchange	Net exchange (import–export) statistics and forecasts for SE3 and Sweden with adjacent Swedish and Nordic bidding zones.	Stat./ Frcst.	[GW]	SE3/ Sweden	15 min	Syspower
Wind Production	Wind power generation statistics and forecasts for Sweden and SE3.	Stat./ Frcst.	[MW]	SE3/ Sweden	15 min	Syspower
Solar Production	Solar power generation statistics and forecasts for Sweden and SE3.	Stat./ Frcst.	[MW]	SE3/ Sweden	15 min	Syspower
Hydro Production	Total regulated hydro power production statistics for Sweden.	Stat.	[MW]	SE3/ Sweden	15 min	Syspower
Nuclear Production	Total nuclear power generation statistics and forecast for Sweden.	Stat.	[MW]	SE3/ Sweden	15 min	Syspower

3.3 Data Processing and Feature Engineering

Accurate and realistic forecasting requires careful handling of time-dependent data. Models must only use information that would have been available at the time when the forecast is made, and the temporal structure of the data must be preserved throughout preprocessing to avoid data leakage and overly optimistic performance estimates.

This section describes the preprocessing and feature engineering steps applied before feature selection and subsequent model development, training, and evaluation. First, a consistent and temporally aligned time-series dataset was constructed. Next, all features were shifted according to their real-world availability. Finally, the data were transformed into model-ready input–output pairs through a data pipeline, including splitting, scaling, sequence construction, and data loader creation. Together, these steps ensure temporally consistent and realistic model inputs, reduce the risk of data leakage, and provide a robust foundation for subsequent modelling and analysis.

3.3.1 Preprocessing – Data Cleaning and Temporal Alignment

To construct a consistent dataset suitable for time-series modelling, all raw data were first cleaned and temporally aligned. Since the data originate from multiple sources with different temporal resolutions and reporting frequencies, all timestamps were first converted to a common time zone (CET/CEST) and mapped onto a common 15-minute time grid covering the full year of 2025. This ensured temporal consistency across all time-series regardless of their source.

Table 3.3: Overview of the market variables, and their main properties, used as input features for forecasting mFRR CM and EAM upward and downward prices and volumes for SE3.

Variable	Description	Value Type	Unit	Spatial Scope	Temp. Res.	Source
Spot Markets						
Day-ahead Prices	Marginal prices on the day-ahead market for Swedish bidding zones, and the Nordic system price.	Stat./ Frcst.	[€/ MWh]	SE1234/ Nordics	15 min	Syspower
Day-ahead Volumes	Buy and sell volumes traded in the day-ahead market for Sweden in total and for SE3.	Stat./ Frcst.	[MW]	SE3/ Sweden	15 min	Syspower
Scheduled Intraday Flow	Total scheduled intraday flow (inflow, outflow, and net-flow) reflecting updated market nominations and physical schedule adjustments in the intraday market for SE3.	Stat.	[MWh]	SE3	15 min	Syspower
Balancing Markets						
FCR-N Price & Volume	Marginal price and procured volume in the Frequency Containment Reserve for Normal operation (FCR-N) for SE3.	Stat.	[€/MW] & [MW]	SE3	1 h	Mimer
FCR-D Up/Down Price & Volume	Marginal prices and procured volumes for Frequency Containment Reserve for Disturbances (FCR-D), both upward and downward regulation for SE3.	Stat.	[€/MW] & [MW]	SE3	1 h	Mimer
aFRR Up/Down Price & Volume	Marginal prices and procured volumes on the automatic Frequency Restoration Reserve capacity market (aFRR CM), both upward and downward regulation for SE3.	Stat.	[€/MW] & [MW]	SE3	1 h	Mimer
mFRR CM Up/Down Price & Volume	Marginal prices and accepted capacity volumes in the manual Frequency Restoration Reserve Capacity Market (mFRR CM), where balancing capacity is procured by the TSO one or several days ahead of real-time operation.	Stat.	[€/MW] & [MW]	SE3	1 h	Syspower
mFRR EAM Up/Down Price & Volume	Marginal activation prices and activated balancing energy volumes in the manual Frequency Restoration Reserve Energy Activation Market (mFRR EAM), reflecting real-time balancing actions by the TSO.	Stat.	[€/MW] & [MW]	SE3	15 min	Syspower

Table 3.4: Overview of the environmental (weather and temporal) variables, and their main properties, used as input features for forecasting mFRR CM and EAM upward and downward prices and volumes for SE3.

Variable	Description	Value Type	Unit	Spatial Scope	Temp. Res.	Source
Temperature	Mean temperature for Sweden and temperature for Gothenburg.	Frcst.	[Cel]	Sweden	1 h	Syspower
Precipitation Energy	Total precipitation energy for Sweden and for the SE3 area, and precipitation for Sweden and Gothenburg.	Frcst.	[GWh] & [mm/h]	Sweden	1 h	Syspower
Wind Velocity	Mean wind velocity in Sweden and wind velocity for Gothenburg.	Frcst.	[m/s]	Sweden	1 h	Syspower
Cloud Cover	Total mean cloud cover in Sweden and cloud cover for Gothenburg.	Frcst.	[%]	Sweden	1 h	Syspower
Date & Time	Cyclical representations of temporal variables to capture seasonal and temporal patterns for Hour of day, Day of Week and Month, and Week and Month of Year	Stat.	[-]	CET/ CEST	1 h	Syspower

The alignment was performed through resampling. Variables with higher temporal resolution (e.g. 1- or 3-minute measurements such as system inertia and frequency) were aggregated to 15-minute intervals using mean value of the observations within that 15-minute period, preserving short-term variations while reducing noise. Variables with lower temporal resolution were up-sampled using either forward-filling or interpolation, depending on the variable type. Forward-filling was applied when values were assumed constant within an interval, while interpolation was used for gradually varying variables. Instantaneous variables (e.g. prices and power levels) were retained directly after up-sampling, while aggregated variables (e.g. energy volumes or precipitation) were distributed evenly across the corresponding 15-minute steps to preserve the total sum over the original interval.

Missing values caused by reporting gaps or daylight saving time transitions were handled using time-aware interpolation applied individually to each time-series. Internal gaps were filled using linear interpolation, while leading and trailing gaps were handled using backward- and forward-filling, respectively. Additional manual inspection and sanity checks were performed to verify temporal consistency and identify obvious misalignments or clearly erroneous values.

This preprocessing resulted in a fully synchronized and temporally consistent multivariate time-series dataset, providing a reliable foundation for subsequent feature engineering, feature selection, and modelling.

3.3.2 Feature Construction – Engineered Features

To improve the predictive capability of the models, a range of engineered features was constructed in addition to the raw variables. These features were designed to capture deviations between expected and realized outcomes, rolling statistical properties such as trends and variability, structural relationships within the power system and electricity markets, and recurring temporal patterns. In this study, the engineered features were grouped into four main categories: forecast error features, derived system and market indicators, rolling statistical features, and temporal encodings. Table 3.5 provides an overview of the engineered feature groups and the specific features included in the original feature set.

Table 3.5: Overview of the engineered feature groups and corresponding variables included in the original feature set prior to feature filtering and feature selection

Feature Group	Engineered Features
Forecast Error Features	Production Forecast Error (SE3), Consumption Forecast Error (SE3), Exchange Forecast Error (SE3), Wind Production Forecast Error (SE3), Solar Production Forecast Error (SE3)
Derived System and Market Indicators	Power Balance Forecast (SE3), Renewable Share Forecast (Sweden) & Spot Price Spread (SE1234 Max–Min), Spot Price Spread (SE3–SE2), Spot Price Spread (SE3–SE4), Spot Price Spread (SE3–Nordic System Price)
Rolling Statistical Features	Spot Price RZD-1d (SE3), Spot Volume Buy RZD-1d (SE3), Spot Volume Sell RZD-1d (SE3)
Temporal Encodings	Quarter-Hour of Day (sin, cos), Hour of Day (sin, cos), Day of Week (sin), Week of Year (sin), Month of Year (sin)

3.3.2.1 Forecast Error Features

Forecast error features were constructed as the difference between forecasted and realized values for a given variable x_t:

$$e_t = x_t^{\text{forecast}} - x_t^{\text{outcome}} \tag{3.1}$$

These features were computed for core system and market variables, including production, consumption, exchange, and spot market prices and volumes. They capture deviations between expected and realized system conditions, providing information about recent imbalances that may influence balancing market prices and volumes.

3.3.2.2 Derived System and Market Indicators

A set of derived features was constructed to capture key structural relationships in the power system and electricity markets, providing more informative representations of system-wide conditions relevant for balancing market dynamics.

First, the overall power balance in SE3 is defined as:

$$\text{Power Balance}_t = \text{Production}_t - \text{Consumption}_t + \text{Exchange}_t \tag{3.2}$$

This variable serves as a direct proxy for system surplus or deficit, which is a key driver of balancing needs and price formation.

Second, the generation mix for Sweden was represented by the renewable (wind and solar) share, defined as wind and solar production relative to total production.

$$\text{Renewable Share}_t = \frac{(\text{Wind} + \text{Solar})\ \text{Production}_t}{\text{Total Production}_t} \tag{3.3}$$

These features provide insight into system flexibility, as a higher share of intermittent renewable generation typically increases the need for balancing resources.

Third, cross-zonal price differences are captured through spot price spreads, including SE3 relative to adjacent bidding zones (SE2, SE4, NO1, DK1, FI), the Nordic system price, and the maximum spread across Swedish bidding zones. These features reflect regional congestion, where transmission constraints lead to price differences between bidding zones, which are closely linked to balancing market dynamics.

Fourth, intraday market activity is represented through scheduled intraday flow variables, including total inflow, outflow, and net flow (inflow–outflow) for SE3. These were calculated by aggregating scheduled flows across all adjacent bidding zones involving SE3. These variables capture adjustments to day-ahead positions and cross-border balancing actions closer to delivery, and are therefore closely linked to balancing market dynamics.

Finally, weather conditions were represented using both national averages for Sweden and local forecasts for Gothenburg for temperature, wind velocity, precipitation, and cloud cover. The national averages were computed across multiple geographically distributed cities in Sweden, capturing system-wide conditions, while the Gothenburg forecasts provide local weather information. Together, these features capture external drivers affecting both demand (e.g. temperature) and supply (e.g. wind, solar, and hydro generation).

3.3.2.3 Rolling Statistical Features

To capture temporal dynamics such as trends, variability, and short-term momentum, rolling statistical features were constructed. These features summarize recent behavior of time-series over a moving (rolling) window and allow the model to incorporate information about local level, volatility, and deviations.

The moving average (MA) represents the local mean over a rolling window of size w, smoothing short-term fluctuations and highlighting underlying trends:

$$\mathrm{MA}_t = \frac{1}{w} \sum_{i=0}^{w-1} y_{t-i} \tag{3.4}$$

The moving standard deviation (MSD) measures local variability within the same window and captures short-term volatility:

$$\mathrm{MSD}_t = \sqrt{\frac{1}{w} \sum_{i=0}^{w-1} (y_{t-i} - \mathrm{MA}_t)^2} \tag{3.5}$$

The rolling Z-score deviation (RZD) normalizes the current observation relative to its recent history by scaling its deviation from the moving average by the moving standard deviation. This produces a dimensionless measure of how extreme the current value is compared to typical recent values, making it useful for detecting sudden shifts and spikes:

$$\mathrm{RZD}_t = \frac{y_t - \mathrm{MA}_t}{\mathrm{MSD}_t} \tag{3.6}$$

These features were computed for selected market, system, and weather-related variables using window sizes adapted to each variable's temporal characteristics, typically ranging from one day to one week.

3.3.2.4 Temporal Encoding

Finally, to capture recurring temporal patterns, calendar variables were encoded using cyclical sine and cosine transformations to preserve the periodic nature of these variables. In this study, the following cyclical encodings were applied: quarter-hour of day (0-95), hour of day (0–23), day of week (1–7), week of year (1–52), and month of year (1–12), enabling the model to effectively learn daily, weekly, and yearly seasonal patterns.

Let x denote a temporal variable (e.g., hour of day, day of week). The cyclical encoding was defined as:

$$x_{\sin} = \sin\left(2\pi \frac{x}{T}\right), \quad x_{\cos} = \cos\left(2\pi \frac{x}{T}\right) \tag{3.7}$$

where T is the period of the variable (e.g. $T = 24$ for hours of the day, and $T = 7$ for days of the week).

This transformation maps discrete temporal variables to the continuous interval $[-1, 1]$, while preserving their periodic structure. As a result, values that are close in time (e.g. hour 23 and hour 0) are also close in the transformed feature space, avoiding artificial discontinuities.

3.3.3 Information Availability and Temporal Shifting

In real-world forecasting settings, different variables become available at different times and with varying reporting delays. If this information structure is not accounted for, models may implicitly use future information during training that would not be known in a real-world scenario, resulting in data leakage and unrealistically optimistic performance estimates. To ensure a realistic forecasting setup, all features were temporally shifted to reflect their true availability at the *forecast origin*, i.e. the time at which a prediction is generated. For each feature i, the shifted value is defined as

$$x_t^{(i,\text{shifted})} = x_{t-\text{shift}_i}^{(i)} \tag{3.8}$$

where shift_i denotes the feature-specific temporal offset. Positive shifts correspond to delayed historical observations, while negative shifts represent available forecast information.

The forecasting setup differs between the two market settings considered in this study in terms of prediction frequency, the time horizon between the forecast origin and the start of the prediction window, and the length of the prediction window. Table 3.6 summarizes the information availability and temporal shifts applied to all raw variables for both the CM and EAM setup. In general, historical variables involve relatively short reporting delays (typically 1–2 hours), while forecast variables may be available several days ahead of delivery.

For the *mFRR Capacity Market (CM)*, forecasts are generated once per day at 06:00 on day D. The first prediction starts 72 quarter-hour steps (18 hours) after the forecast origin, corresponding to the start of day $D+1$, and the full forecast covers 192 quarter-hour steps (48 hours), i.e. all timesteps for day $D+1$ and day $D+2$. The CM setup uses a lookback window of 264 quarter-hour steps, aligned with the full forecast horizon, meaning that forecasted input features are shifted by up to -264 so that future values available at the forecast origin are correctly aligned with their corresponding prediction targets.

For the *mFRR Energy Activation Market (EAM)*, forecasts are generated every third hour ($00{:}00, 03{:}00, \ldots, 21{:}00$). Each forecast begins 4 quarter-hour steps (1 hour) after the forecast origin and covers the following 24 quarter-hour steps (6 hours). The EAM setup uses a lookback window of 48 quarter-hour steps, while the prediction setup covers 28 quarter-hour steps in total (forecast gap plus prediction window). Forecasted input features are shifted by up to -48, depending on their availability, ensuring that all inputs represent information available at the forecast origin, including both historical observations and available future forecasts.

Historical variables (realized observations, referred to as *statistics*) were shifted positively to account for reporting delays. Forecast variables were shifted negatively according to their forecast availability horizon. Market and balancing-market variables were shifted according to their publication times, which differ between the CM and EAM forecasting settings, while intraday scheduled flow is based on day-ahead nominations and continuously updated through intraday market adjustments. Target variables were not shifted, as they represent the future values to be predicted.

Engineered features inherited the temporal shift of the variables from which they were derived. Forecast error features were shifted according to the availability of the corresponding historical observations, while rolling statistical features inherited the shift of their underlying variables. Derived system and market indicators were shifted consistently with the inputs used in their construction. Temporal encoding features were not shifted, as calendar information is known exactly at the forecast origin.

Applying these shifts introduces boundary effects at the beginning and end of the dataset. These were handled by trimming invalid time ranges and applying interpolation or forward- and backward-filling where appropriate. This procedure ensures that all model

Table 3.6: Information availability and temporal alignment of features used in the forecasting models. Positive shifts indicate delayed reporting of realized data, while negative shifts represent forecast information available ahead of delivery.

Variable	**Information availability**	**CM shift**	**EAM shift**
System variables			
Frequency	$t_0 - 1h$	+4	+4
Inertia	$t_0 - 1h$	+4	+4
Time deviation	$t_0 - 1h$	+4	+4
Total production statistics	$t_0 - 2h$	+8	+8
Total production forecast	$t_0 + 1d$	-96	-48
Consumption statistics	$t_0 - 2h$	+8	+8
Consumption forecast	$t_0 + 7d$	-264	-48
Exchange statistics	$t_0 - 1h$	+4	+4
Exchange forecast	$t_0 + 7d$	-264	-48
Wind production statistics	$t_0 - 2h$	+8	+8
Wind production forecast	$t_0 + 7d$	-264	-48
Solar production statistics	$t_0 - 2h$	+8	+8
Solar production forecast	$t_0 + 7d$	-264	-48
Nuclear production statistics	$t_0 - 2.5h$	+10	+10
Hydro production statistics (SE)	$t_0 - 2.5h$	+10	+10
Market variables			
Spot market variables			
Spot (day-ahead) prices	D (and D+1 from ≈12:40)	-72	-46
Spot prices forecast	$t_0 + 7d$	-264	-48
Spot volumes buy/sell	D (and D+1 from ≈12:40)	-72	-46
Spot volumes buy/sell forecast	$t_0 + 7d$	-264	-48
Intraday scheduled flow	D (intraday updates)	-72	-24
Balancing market variables			
FCR prices/volumes	D (and D+1)	-72	-24
aFRR prices/volumes	D (and D+1)	-72	-24
mFRR CM prices/volumes	D (and D+1)	-72	-48
mFRR EAM prices/volumes	$t_0 - 1h$	+4	+4
Environmental variables			
Temperature forecast	$t_0 + 3d$	-264	-48
Wind velocity forecast	$t_0 + 3d$	-264	-48
Precipitation forecast	$t_0 + 3d$	-264	-48
Cloud cover forecast	$t_0 + 3d$	-264	-48

inputs remain temporally consistent and that only information available at the forecast origin is used for prediction, ensuring a realistic forecasting setup.

3.3.4 Data Pipeline

This section outlines the data pipeline used to transform the time-series data into a format suitable for model training and evaluation in a realistic, data leakage-free, and temporally consistent way. The pipeline consists of four main steps:

1. **Train–Validation–Test Splitting** – chronological split of the dataset.

2. **Feature Scaling** – standardization of input features and target using training data statistics.

3. **Sequence Construction** – transformation into supervised learning samples using a sliding window approach.

4. **Data Loader Creation** – conversion into PyTorch dataloaders for model training.

Overall, the pipeline ensured a consistent and reproducible preprocessing workflow, while preserving the temporal structure of the data and preventing data leakage between training, validation, and test sets.

3.3.4.1 Step 1: Train–Validation–Test Splitting

To ensure a realistic, fair, and unbiased training and evaluation of the predictive performance of the models, a holdout validation approach was used to split the full dataset into training, validation, and test sets in a strictly chronological order. In this way, the models were always trained on past data and evaluated on future, unseen data. This is particularly important for time-series problems, where observations are temporally dependent, and any use of future information during training would lead to data leakage.

In this study, the full dataset covered the exact 1-year period from 1 January 2025 to 31 December 2025. The data was split chronologically into three consecutive periods, following a 58.3/16.7/25 split ratio, as shown in Figure 3.1. The purpose of each split is summarized below:

- **Training set**: used for training the model and fitting its parameters.

- **Validation set**: used for evaluating the model during training and development, for hyperparameter tuning and optimization, feature selection, and model selection.

- **Test set**: used for final unbiased evaluation of model performance and generalizability on unseen data.

Overall, this splitting strategy ensured a realistic forecasting setup, where models were trained exclusively on past observations and evaluated on future data. It provided a reliable and unbiased estimate of model generalization and predictive performance by maintaining a strict separation between training, validation, and test sets.

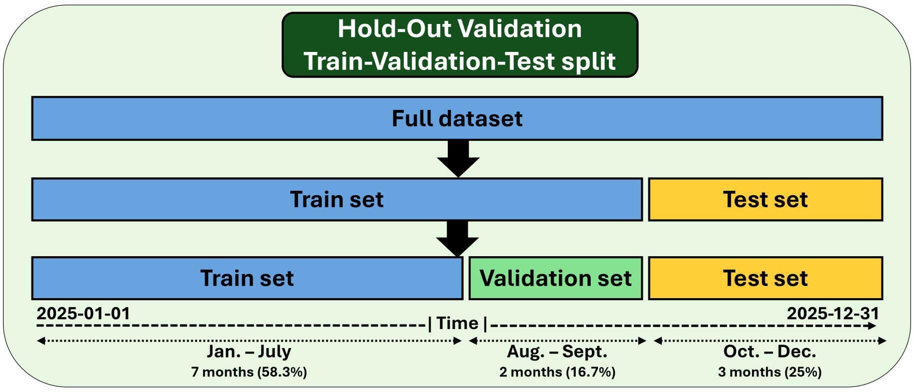

Figure 3.1: Temporal train–validation–test split used in the study, illustrating the chronological hold-out validation approach. The one-year dataset is divided into a training period (January–July), validation period (August–September), and test period (October–December), following a 58.3/16.7/25 split ratio.

3.3.4.2 Step 2: Feature Scaling

All input features and, in this study, also the target variable were scaled to ensure stable and efficient model training. Standardization was used, meaning that each feature was transformed to have zero mean and unit variance using `StandardScaler` from `sklearn.preprocessing`. This is important for neural networks, as it ensures comparable feature scales, improves numerical stability, and supports faster convergence.

The scaling parameters (mean and standard deviation) were computed using only the training data. The fitted scaler was then applied to both the validation and test sets. The target variable was also standardized using a scaler fitted on the training targets, and the inverse transformation was applied to model outputs to restore real-world units for evaluation.

This procedure of restricting scaling to the training data prevents data leakage, as fitting the scaler on the full dataset would allow information from validation and test data to influence training, and supports a realistic forecasting setting and an unbiased evaluation.

3.3.4.3 Step 3: Sequence Construction

To apply machine learning models to time-series data, continuous observations must be transformed into a supervised learning format with fixed input and output shapes.

In this study, a sliding window approach was used to construct input-output pairs, where each pair (sample) represents a forecasting event. Each sample consists of a lookback window (input sequence X) and a prediction window (output sequence y), defined as:

$$X = \{x_{t-\text{seq\_len}+1}, \ldots, x_t\} \tag{3.9}$$

$$y = \{y_{t+\text{horizon}}, \ldots, y_{t+\text{horizon}+\text{nr\_pred\_steps}-1}\} \tag{3.10}$$

where t denotes the forecast origin time. The input sequence X contains past observations of length `seq_len`, providing historical context including both realized statistics and available forecast information. The prediction window y is defined by the parameters `horizon`, which determines the forecasting gap, and `nr_pred_steps`, which specifies the number of future time steps to predict.

Together, these parameters allow for flexible forecasting setups with varying lookback windows and prediction horizons, as illustrated in Figure 3.2.

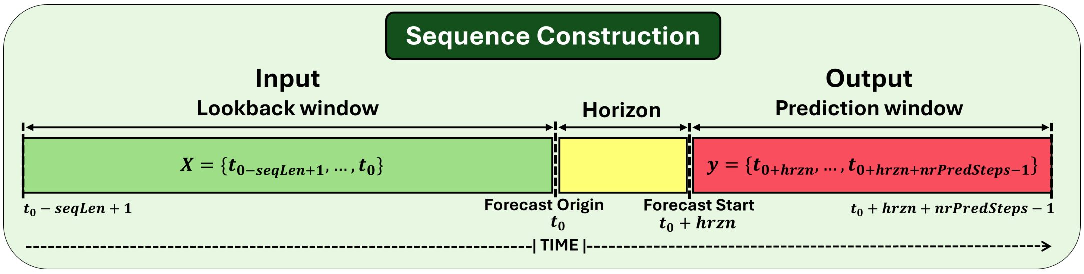

Figure 3.2: Sliding window-based sequence construction for supervised time-series forecasting. Each sample consists of a lookback window (input sequence), a forecast gap (horizon) between the forecast origin and forecast start, and a fixed-length prediction window (output sequence).

In practice, different configurations were used depending on the characteristics of the market being modelled. For mFRR CM, longer input sequences (`seq_len=264`), a larger horizon (`horizon=72`), and a two-day prediction window (`nr_pred_steps=192`) were used to capture longer-term dynamics. For mFRR EAM, shorter sequences (`seq_len=48`), a shorter horizon (`horizon=4`), and a six-hour prediction window (`nr_pred_steps=24`) were applied to enable more short-term forecasting.

In addition, the sequence construction process allowed control over when and how often forecasts were generated by setting the forecast origin time to specific times of the day or to regular intervals, depending on the application. For mFRR CM, forecasts were generated once daily at 06:00 for the following two days. For mFRR EAM, forecasts were generated every third hour (00:00–21:00), each covering a 6-hour prediction window starting one hour ahead.

To reduce the risk of data leakage, a small temporal gap of 4 days (corresponding to the maximum combined length of the lookback and prediction windows) was introduced between the train, validation, and test splits to prevent overlap between input sequences and prediction targets across different sets.

Overall, this sequence construction step enabled the models to learn temporal dependencies between past input features and future target values across different forecasting horizons, while maintaining a flexible and realistic forecasting framework.

3.3.4.4 Step 4: Dataset and Data Loader Creation

The generated sequences were converted into PyTorch datasets consisting of input-output pairs (X, y), and corresponding dataloaders were created for efficient batching and iteration during training and evaluation. A separate set of train, validation, and test dataloaders was created for each target variable and feature configuration (uv, fs5, fs10, and mv), resulting in four dataloader sets per target.

The batch size is an important training hyperparameter, as it affects computational efficiency and convergence stability. In this study, batch sizes were chosen to approximately correspond to four days of forecasting samples per update step, providing a balance between efficient computation and stable gradient updates. For the mFRR CM setting, generating one forecast per day, a batch size of 4 was used. For the mFRR EAM setting, a batch size of 32 was used, since forecasts are generated eight times per day. This resulted in approximately 52 training batches, 14–15 validation batches, and 22 test batches for both market settings.

Shuffling is commonly used for many machine learning datasets to improve generalization by exposing the model to samples in different orders. However, for time-dependent forecasting tasks it is typically avoided, since preserving chronological order is important for maintaining temporal dependencies and preventing unrealistic information leakage. Therefore, shuffling was disabled (`shuffle=False`) for all dataloaders in this study.

Overall, this structured pipeline ensured that the models were trained and evaluated in a realistic and reproducible way while fully respecting the temporal nature of the data.

3.3.5 Feature Filtering and Selection

Feature selection is an important step in time-series forecasting, particularly when working with high-dimensional datasets. Including too many input features increases computational cost in terms of time and model complexity, introduces noise, and can lead to overfitting, especially for flexible models such as LSTMs. Therefore, the goal of this step was to progressively refine the feature space by removing irrelevant, noisy, and redundant features in order to obtain a more compact subset of informative and stable features that improve predictive performance while preserving interpretability and transparency.

In this project, a multi-stage feature selection approach was applied, combining:

1. **Correlation-based filtering** of features with respect to the target (filter method).
2. **Feature importance ranking** based on Random Forest models (embedded method).
3. **Forward selection** of features using LSTM models (wrapper method).

This multi-stage feature selection strategy follows established feature selection methodology by combining filter, embedded, and wrapper-based approaches [41], which supports both improved predictive performance and interpretability.

The complete feature filtering and selection pipeline used is shown in Figure 3.3.

First, correlation-based filtering is used to remove weak and noisy features, followed by Random Forest importance ranking to capture nonlinear relationships and feature interactions. Finally, LSTM-based forward selection evaluates candidate feature subsets directly based on forecasting performance.

This pipeline reduces dimensionality and computational complexity by systematically filtering and refining the feature space, resulting in a compact and informative feature subset for explaining and forecasting mFRR prices and volumes.

3.3.5.1 Step 1: Correlation-based filtering (filtering method)

In this study, correlation-based filtering was used as an initial feature selection step to quantify the strength and direction of the relationship between each input feature and the target variables. This provides a simple, interpretable, and computationally efficient way to identify weakly informative features and reduce dimensionality and noise in the input data before applying more advanced selection methods.

To capture monotonic relationships between features and target variables, the Spearman rank correlation coefficient was computed for each feature–target pair. Spearman correlation was chosen over the standard Pearson correlation because it is robust to outliers and non-normal distributions, which are common characteristics of electricity market time series.

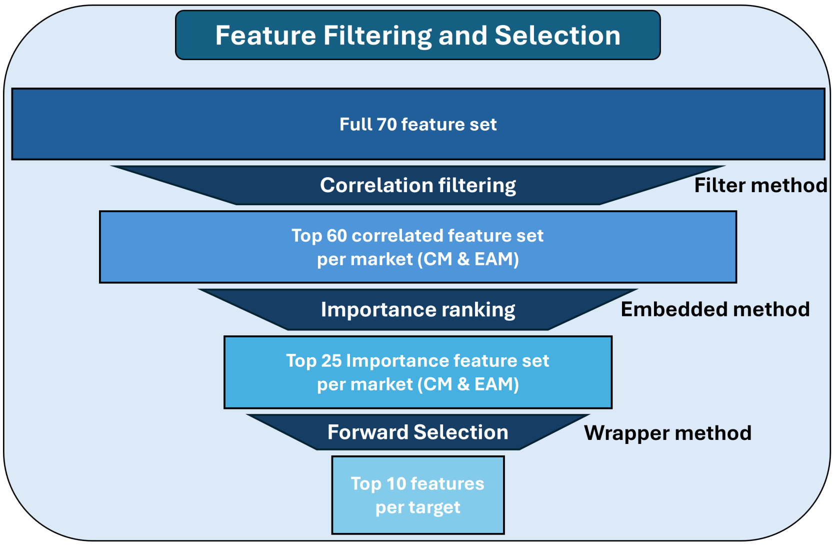

Figure 3.3: Overview of the feature filtering and selection pipeline. The feature space is gradually reduced through correlation-based filtering, Random Forest importance ranking, and LSTM-based forward selection, resulting in target-specific top-10 feature subsets.

All correlations were computed using temporally aligned observations, i.e. real-time correlations between feature x_t and target y_t. Features were first ranked for each target based on the absolute value of their correlation coefficients, ensuring that both positive and negative relationships were treated equally. For each market setting (CM and EAM), mean absolute correlation scores were then aggregated across the four target variables (up/down price and volume). In addition, features were required to exhibit non-negligible correlation for at least two of the four targets. This aggregation ensures that selected features are not only strongly related to a single target but are consistently informative across multiple related prediction tasks. The resulting output is a ranked subset of robust and relevant candidate features for subsequent feature selection and modelling.

Due to its computational efficiency and simplicity, correlation-based filtering serves as a practical preprocessing step before applying more advanced feature selection methods. However, it is limited to monotonic relationships and does not account for more complex nonlinear dependencies or feature interactions. Therefore, model-based feature selection methods, such as feature importance ranking using Random Forest (Section 3.3.5.2) and LSTM-based forward selection (Section 3.3.5.3), were applied to further refine the feature set.

3.3.5.2 Step 2: Feature Importance Ranking using Random Forest (embedded method)

Random Forest (RF) is an ensemble-based machine learning method that combines multiple decision trees to capture nonlinear relationships and feature interactions. In addition to its predictive capability, RF provides an embedded measure of feature importance based on each feature's contribution to reducing prediction error. In this study, RF-based ranking was used as an intermediate feature selection step following correlation-based filtering to further reduce dimensionality while retaining the most informative variables for downstream LSTM modelling.

Table 3.7: Random Forest configuration used for feature importance estimation.

Parameter	Value
n_estimators	100
max_depth	15
criterion	squared_error
n_runs	10

For each target variable, RF importance scores were computed using the correlation-filtered feature subset, with a time-based split of 7 months for training, 2 months for validation, and 3 months for testing to avoid information leakage.

To improve robustness, the RF model was trained over $n_{\text{runs}} = 10$ runs with different random seeds. For each feature, the mean importance μ_i and standard deviation σ_i were computed across runs. Stability was quantified using the coefficient of variation,

$$s_i = \frac{\sigma_i}{\mu_i}, \tag{3.11}$$

and a stability-adjusted importance score was defined as

$$\text{score}_i = \frac{\mu_i}{1 + s_i}. \tag{3.12}$$

This formulation prioritizes features that are both highly important and stable across runs while penalizing unstable estimates. Features were ranked by this score, and the top 50 were retained for each target.

To improve consistency across related prediction tasks within each market, similarly to the correlation-based filtering step, feature importance results were further aggregated at the market level (CM and EAM) by combining the top-ranked features across the four corresponding targets (up/down price and volume). Features were ranked according to their mean importance across targets, while selection frequency was used as an indicator of consistency.

This produced a compact feature subset that was both informative and consistently relevant across forecasting tasks within the same market, providing a stable and representative input set for the subsequent LSTM-based feature selection stage.

3.3.5.3 Step 3: Forward Selection with LSTM (wrapper method)

Finally, to complement the filter and embedded feature selection steps, a wrapper-based forward selection method was applied using the LSTM model. Unlike the previous methods, which rely on statistical relationships or model-specific importance measures, this approach evaluates feature subsets directly based on forecasting performance. This makes it well-suited for capturing nonlinear temporal dependencies and feature interactions, which are central to LSTM-based time-series modelling.

A simple LSTM architecture was used during feature selection to ensure computational efficiency and stable feature ranking. A single-layer LSTM with a moderate hidden size (25), learning rate 0.001, and no dropout was sufficient to capture relevant temporal patterns while avoiding overfitting and excessive training time. To improve robustness, candidate feature sets were evaluated over multiple training durations (20, 25, and 30 epochs).

The forward selection was implemented as an iterative process. Starting from an empty feature set, features were added one at a time based on their contribution to model performance. At each step, all remaining candidate features were evaluated by temporarily adding them to the current feature set, training the LSTM model, and assessing the resulting forecasting performance. The feature yielding the largest improvement was then permanently included. This procedure was repeated until a predefined number of features (10) was reached.

To ensure realistic and robust evaluation, a time-aware validation strategy was applied. Rolling time-series cross-validation with four folds was used to preserve the temporal structure of the data and avoid information leakage. Each fold used 60 days for training and 30 days for validation, and three independent training runs were performed per fold to account for the stochastic nature of neural network training.

For each candidate feature set, the mean and standard deviation of the root mean squared error (RMSE) were computed across all folds and runs. To balance predictive accuracy and stability, a combined scoring metric was defined as:

$$\text{score} = \text{RMSE}_{\text{mean}} + \lambda \cdot \text{RMSE}_{\text{std}} \tag{3.13}$$

where λ ($\lambda = 0.33$ in this work) controls the trade-off between accuracy and stability. This formulation penalizes feature sets that perform inconsistently across different time periods or training runs.

The forward selection procedure was performed separately for each target variable, resulting in a final subset of the top 10 features for each target. These feature subsets were used as inputs to the final LSTM and MLP models, ensuring that the selected variables are both relevant and robust for forecasting.

3.4 Forecasting Model Implementation

This section describes the forecasting setup, including the definition of the two market-specific forecasting scenarios (mFRR CM and EAM) and the benchmark models used as performance references for the machine learning models.

3.4.1 Forecast Settings

The forecasting models were evaluated under two different market settings: the mFRR Capacity Market (CM) and the mFRR Energy Activation Market (EAM). Although the same forecasting models were applied in both settings, the sequence construction and forecasting horizons differ due to the different operational characteristics of the two markets.

For the CM setting, forecasts are generated once per day at 06:00 on day D, predicting all quarter-hour timesteps for day $D+1$ and day $D+2$. This corresponds to a forecasting gap of `horizon = 72` quarter-hour steps and a prediction window of `nr_pred_steps = 192`. The input sequence length was set to `seq_len = 264`.

For the EAM setting, forecasts are generated every third hour (00:00, 03:00, . . . , 21:00). Each forecast starts one hour ahead, corresponding to `horizon = 4` quarter-hour steps, and predicts the following six hours, corresponding to `nr_pred_steps = 24`. The input sequence length was set to `seq_len = 48`.

Apart from these differences in sequence construction and forecasting horizons, the same baseline and machine learning models were implemented and evaluated under both market settings.

3.4.2 Benchmark and Reference Models (Naive Baselines)

To evaluate the forecasting performance of the machine learning models, a set of naive benchmark models was implemented. These models provide simple yet important forecasting approaches that capture fundamental characteristics of the data, such as persistence, seasonality, and average system behavior. They serve both as practical forecasting baselines and as reference models for assessing whether more advanced machine learning methods provide meaningful predictive improvements.

3.4.2.1 Seasonal Naive Model

The seasonal naive model assumes that future values follow the same pattern as a previous period. Forecasts are constructed by repeating observations from the same timesteps, either from the previous day ($d = 1$), or from the same day of the previous week ($d = 7$):

$$\hat{y}_{t+h} = y_{t+h-d} \qquad h = 1, \ldots, H \tag{3.14}$$

In the CM setting, forecasts are based on the previous day and are repeated across the full prediction horizon. In the EAM setting, each timestep is aligned with the corresponding value from the previous day, preserving the intraday structure.

This baseline model captures strong daily and weekly seasonal patterns, but does not account for changes in overall level, volatility, or noise.

3.4.2.2 Mean Naive Model

The mean naive model predicts all future values as a constant equal to the average over a historical lookback window:

$$\hat{y}_{t+h} = \bar{y}_t \qquad h = 1, \ldots, H \tag{3.15}$$

where

$$\bar{y}_t = \frac{1}{N} \sum_{i=1}^{N} y_{t-i} \qquad N = 96 \cdot d \tag{3.16}$$

For CM, the mean is computed over the previous day and applied to the full forecast horizon. For EAM, a rolling 24-hour window is used, ending one hour before the forecast origin to avoid data leakage.

This baseline captures the overall level of the time-series and is robust to noise, but ignores temporal structure such as seasonality.

3.4.2.3 Seasonal Mean Naive Model

The seasonal mean naive model combines seasonality and averaging by predicting each timestep as the mean of that timestep over the past d days:

$$\hat{y}_{t+h} = \frac{1}{d} \sum_{i=1}^{d} y_{t+h-i\cdot 96} \qquad h = 1, \ldots, H \tag{3.17}$$

This is applied consistently across CM and EAM, capturing recurring intraday and weekly patterns while smoothing short-term fluctuations.

This baseline produces a time-varying forecast that preserves intraday patterns while reducing variance through averaging, making it more stable than the pure seasonal naive model.

3.4.2.4 Role of Benchmark Models

Together, these baselines capture complementary aspects of the data: recurring seasonal patterns, overall level, and a combination of both.

The benchmark models serve two main purposes:

- Provide simple, yet potentially competitive, reference forecasting models that capture fundamental temporal patterns in the data, including level, seasonality, and short-term dependencies.

- Serve as a reference for evaluating the predictive performance of the LSTM and MLP models, enabling quantification of performance gains, if any, of machine learning models relative to rule-based forecasting models.

They provide strong and interpretable reference models, allowing for a clear assessment of whether machine learning models offer meaningful improvements in forecasting performance beyond simple baseline models based on repetition, averaging, and seasonal persistence patterns.

3.4.3 Machine Learning Models

This section presents the machine learning models used to forecast mFRR prices and volumes. The selected architectures allow comparison between static feed-forward models and recurrent models that explicitly account for temporal dependencies.

3.4.3.1 Multi-Layer Perceptron (MLP)

The MLP consists of a fixed input layer, a variable number of fully connected hidden layers, and a fixed output layer. The number of hidden layers, hidden layer size, and dropout rate were treated as tunable hyperparameters.

The input sequence is flattened into a single feature vector of size $seq\_len \times n_{\text{features}}$ and processed simultaneously by the network. The output layer predicts all future timesteps in the prediction window.

For CM, the input size is $264 \times n_{\text{features}}$ and the output size is 192 timesteps. For EAM, the input size is $48 \times n_{\text{features}}$ and the output size is 24 timesteps.

3.4.3.2 Long Short-Term Memory (LSTM)

The LSTM consists of a fixed input layer, one or more stacked LSTM layers, and a fixed output layer. The number of LSTM layers, hidden state size, and dropout rate were treated as tunable hyperparameters.

Unlike the MLP, the LSTM processes the input sequence sequentially. At each timestep, a feature vector of size n_{features} is provided to the network, which updates an internal hidden state of size $hidden\_size$. This process is repeated for seq_len timesteps. The hidden representation from the final timestep is then used to predict all future timesteps in the prediction window.

For CM, the LSTM processes 264 sequential timesteps, each containing n_{features} input features, and predicts 192 future timesteps. For EAM, it processes 48 sequential timesteps, each containing n_{features} input features, and predicts 24 future timesteps.

3.5 Model Training and Validation

To ensure robust and unbiased model performance, a structured training and validation framework was implemented. The objective of this framework was to obtain models that generalize well to unseen data, while preventing overfitting and ensuring fair comparison across different model configurations.

3.5.1 Training Procedure

The models were trained using mini-batch gradient descent with the Adaptive Moment Estimation (Adam) optimizer and mean squared error (MSE) loss function. For each training epoch, the data were processed in sequential mini-batches, preserving the temporal dependencies required for LSTM-based modelling. For each batch, a forward pass was first performed to generate predictions, followed by loss calculation as the MSE between predicted and true target values.

The overall training loss for the epoch is computed by first averaging the MSE across samples within each batch, and then averaging the batch-wise losses across all batches to obtain the final epoch loss.

$$\mathcal{L}_{\text{train}} = \frac{1}{B}\sum_{j=1}^{B}\left(\frac{1}{N_j}\sum_{i=1}^{N_j}(\hat{y}_{j,i} - y_{j,i})^2\right) \tag{3.18}$$

where B denotes the total number of batches, N_j is the number of samples in batch j, $\hat{y}_{j,i}$ is the predicted value, and $y_{j,i}$ is the corresponding true target value for sample i in batch j.

The gradients of this loss were then computed through backpropagation, after which the optimizer updated the model parameters to minimize the training loss, with the parameter update step expressed as:

$$\theta^{(k+1)} = \theta^{(k)} - \eta \cdot \nabla_\theta \mathcal{L}_{\text{train}}, \tag{3.19}$$

where $\theta^{(k)}$ denotes the model parameters at iteration k, η is the learning rate, and $\nabla_\theta \mathcal{L}_{\text{train}}$ is the gradient of the training loss with respect to the parameters, computed via backpropagation. Adam further improves this update by adapting the learning rates using estimates of the first and second moments of the gradients, improving convergence speed and stability through adaptive learning rates.

This process was repeated across all batches, iteratively updating the model parameters to minimize the training loss over the full dataset.

3.5.2 Validation Strategy

During training, following each training epoch, after the parameter optimization step, the models were evaluated on a separate validation dataset to provide an unbiased estimate of the model's generalization performance on unseen data and to enable the use of an early stopping mechanism (see Section 3.5.3).

Predictions were generated for the entire validation set, and since standardization scaling was applied in the preprocessing step, predictions were transformed back to the original scale using the inverse transform operation of the corresponding scaler. Performance was then evaluated using two complementary metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

RMSE was selected as the primary metric due to its quadratic sensitivity to large errors, which is particularly relevant in electricity markets where extreme price and volume spikes are common. MAE is additionally reported to provide a more interpretable measure of average deviation.

3.5.3 Early Stopping and Model Selection

To prevent overfitting and determine an appropriate stopping point during training, an early stopping strategy based on the validation root mean squared error (RMSE) was employed.

A patience-based early stopping strategy was applied. At each training epoch, the validation RMSE was computed and compared to the best observed value across all previous epochs. If the validation RMSE improved, the current model parameters were stored as the best-performing parameters. Otherwise, the patience counter was incremented. Training was terminated once no improvement in validation RMSE was observed for a predefined number of consecutive epochs, defined by the patience threshold. The optimal model parameters, denoted by θ^*, were selected as:

$$\theta^* = \arg\min_{\theta} \ \mathrm{RMSE}_{\mathrm{val}}(\theta), \tag{3.20}$$

where θ represents the model parameters after each completed training epoch.

In addition to RMSE, the corresponding mean absolute error (MAE) at the selected epoch was recorded for reporting purposes. After training, the model parameters were restored to those corresponding to the lowest validation RMSE, i.e. the best validation performance.

Throughout training, both training loss and validation performance were tracked at each epoch. This enabled monitoring and analysis of the learning dynamics, including convergence behavior and potential under- or overfitting. Figure 3.4 shows representative learning curves for LSTM models trained using different feature sets.

Specifically, different patterns in the learning curves were used as diagnostic indicators:

- Convergence: Indicated when both training and validation errors stabilize with minimal improvement across successive epochs.
- Overfitting: Characterized by a continued decrease in training loss while the validation error begins to increase, indicating that the model is fitting noise in the training data rather than generalizable patterns.
- Underfitting: Observed when both training and validation errors remain relatively high and continue to decrease at the final training epoch, suggesting that the model has not yet fully captured the underlying data structure.

These observations helped assess model behavior and complemented the early stopping procedure by making it possible to detect convergence, overfitting, and underfitting.

3.5.4 Training and Validation Framework Summary

The described training and validation framework was designed to ensure reliable model evaluation under realistic forecasting conditions. Strict temporal separation between training, validation, and test sets prevented data leakage, while all model configurations were

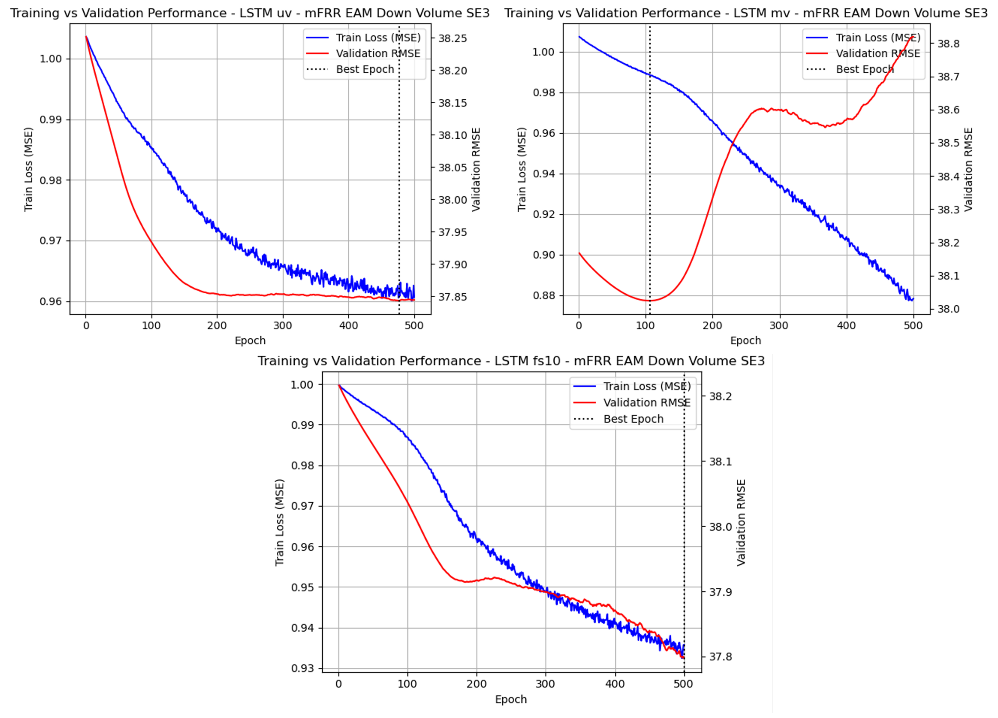

Figure 3.4: Training (blue) and validation (red) loss curves for representative ML model training behaviors: *convergence* (top left), *overfitting* (top right), and *underfitting* (bottom), with model selection based on the lowest validation error, represented by the dashed vertical line.

evaluated under identical conditions to ensure fair comparison. Validation-based early stopping improved robustness by reducing overfitting, and the selection of models based on validation out-of-sample performance ensured that optimization prioritized performance on unseen data and generalization, rather than in-sample fit.

Overall, this framework ensures that the reported results reflect realistic forecasting performance and provide a solid basis for model evaluation and comparison.

3.6 Hyperparameter Tuning

The predictive performance of LSTM models is highly sensitive to the choice of hyperparameters, i.e. the parameters that define both the model architecture and the training process. Selecting suitable hyperparameter values is therefore essential for achieving good predictive accuracy, stable convergence, and robust generalization. To ensure a systematic and computationally efficient tuning procedure, a structured grid search strategy was applied.

3.6.1 Two-Stage Hyperparameter Grid Search Approach

Instead of performing a full grid search over all hyperparameters simultaneously, which is computationally expensive, the tuning process was divided into two stages:

- **Model architecture:** (1) hidden layer size and (2) number of hidden layers for the LSTM model

- **Training parameters:** (1) learning rate, (2) dropout rate, and (3) number of training epochs

In the first stage, the model architecture was fixed to a representative configuration (i.e. a reasonable hidden size and number of layers), while the training-related hyperparameters were systematically explored. In the second stage, the best-performing training parameters identified in the first stage were fixed, and the architectural hyperparameters were varied to determine a suitable model complexity under these optimized training conditions.

This separation reduces the computational cost of the grid search while maintaining an effective exploration of the hyperparameter space. The approach is motivated by the different roles these parameter groups play in model training. Training parameters, such as learning rate, dropout rate, and number of epochs, primarily influence the optimization process, including convergence speed, stability, and generalization during training. In contrast, architectural parameters, such as hidden size and number of hidden layers, determine the model's capacity to learn patterns in the data, controlling the balance between underfitting and overfitting. By tuning these groups sequentially, the search process becomes both more efficient and more interpretable.

3.6.1.1 Step 1: Architecture Grid Search (Hidden Size and Number of Layers)

In the first stage, the training parameters, i.e. the learning rate and the dropout rate were fixed to reasonable mid-sized values (`lr=[0.0001]`, `dropout=0.05`), and the LSTM models were trained for 200 epochs, evaluated every 10:th epoch (`epochs=[10, 20, 30, ..., 190, 200]`). The following architectural hyperparameters were explored:

- `hidden_size = [25, 50, 75, 100, 150, 250]`
- `nr_layers = [1, 2, 3]`

After running the architecture grid search for each target variable, the results were aggregated within the two market settings (mFRR CM and EAM). Hyperparameter configurations were evaluated based on their average predictive performance and stability across all four targets within each market.

For the CM setting, the optimal architecture was obtained with `hidden_size = 100` and `nr_layers = 2`. For the EAM setting, the best-performing configuration was `hidden_size = 25` and `nr_layers = 3`.

3.6.1.2 Step 2: Training Parameter Grid Search (Learning Rate, Dropout rate, Epochs)

In the second stage, for each market (mFRR CM and EAM) setting, the best-performing architecture hyperparameters from the previous grid search were fixed, and the following hyperparameters were explored:

- `learning_rate = [0.000003, 0.00001, 0.00003, 0.0001, 0.0003, 0.001, 0.003, 0.01]`
- `dropout_rate = [0, 0.05, 0.10, 0.20, 0.30, 0.50, 0.75]`
- `nr_epochs = [25, 50, 75, 100, 125, 150, 175, 200, 225, 250]`

The resulting training configurations were evaluated in the same manner, based on average predictive performance and stability across all targets within each market.

The optimal training parameters were for the CM setting, `learning_rate = 0.00003` and `dropout_rate = 0.5`, and for the EAM setting, `learning_rate = 0.00003` and `dropout_rate = 0.3`.

3.6.2 Final Hyperparameter Selection

To improve robustness, the tuning procedure was repeated across all target variables and aggregated separately for the two market settings considered in this study: mFRR Capacity Market (CM) SE3 and mFRR Energy Activation Market (EAM) SE3. For each market, hyperparameter configurations were evaluated based on their average performance across the four corresponding targets (up/down price and volume). The final selected LSTM hyperparameters for the mFRR CM and EAM models are summarized in Table 3.8.

Table 3.8: Final selected LSTM hyperparameter configurations obtained through grid search for the mFRR CM and EAM settings.

Hyperparameter	**mFRR CM (SE3)**	**mFRR EAM (SE3)**
hidden_size	100	25
nr_layers	2	3
learning_rate	0.00003	0.00003
dropout_rate	0.5	0.3

Model performance was evaluated using Root Mean Squared Error (RMSE). To enable fair comparison across targets with different magnitudes, RMSE values were normalized relative to the best-performing configuration within each target group. For each configuration, the mean and standard deviation of the normalized RMSE were computed, allowing both predictive performance and stability to be considered during selection.

3.7 Model Evaluation

This section describes how the forecasting models are evaluated and compared. The goal is to assess model performance in a consistent setting across all models, feature configurations, and target variables. All evaluations are based on a strict separation between training, validation, and test data to ensure that the results reflect true out-of-sample forecasting performance.

3.7.1 Evaluation Procedure

For each target variable, LSTM and MLP models were evaluated using four feature configurations: univariate (uv), top-5 forward-selected features (fs5), top-10 forward-selected features (fs10), and the full multivariate feature set (mv), resulting in eight machine learning model configurations per target.

Each model was trained on the training dataset and selected based on the validation RMSE using the early stopping and model selection procedure described in Section 3.5.3. In addition, five naive benchmark models were included for comparison.

Consequently, a total of thirteen forecasting models were evaluated for each target variable: four LSTM models, four MLP models, and five naive benchmark models. The selected machine learning models and all benchmark models were subsequently evaluated on the independent test dataset using RMSE and MAE metrics.

3.7.2 Model Comparison

For each target variable, based on the evaluated model performance on the test set, all thirteen forecasting models were ranked based on their RMSE, with lower values indicating better predictive performance. Mean Absolute Error (MAE) was additionally reported as a complementary measure of average absolute deviation.

The evaluation was performed separately for all eight forecasting targets. This setup ensures that model performance is assessed consistently across different prediction tasks, including both markets (CM and EAM), as well as price and volume forecasting for upward and downward regulation.

This enables a direct and fair comparison between LSTM, MLP, and naive benchmark models, as well as a systematic evaluation of different feature set configurations within the machine learning models.

3.8 Implementation and Reproducibility

This section summarizes the software tools used to implement the proposed methodology and discusses considerations related to the reproducibility of the experimental workflow.

3.8.1 Software Tools and Libraries

All implementations were carried out in `Python`, using Visual Studio Code as the main development environment. Jupyter-notebooks were used extensively, as they provide a structured and efficient workflow for iterative development, exploratory analysis, and visualization.

Data processing and numerical computations were performed using `NumPy` and `Pandas`. Data preprocessing, feature scaling (StandardScaler, MinMaxScaler, RobustScaler), dataset splitting, and performance metrics such as RMSE and MAE were implemented using `Scikit-learn`. Classical machine learning methods, specifically Random Forest regression, were also implemented using `Scikit-learn`.

Machine learning models and training pipelines were implemented using `PyTorch`, including `torch.nn` for model architectures, `torch.optim` for optimization, and `torch.utils.data` for data loading and batching. Model architectures were inspected using `torchinfo`.

Visualization, plotting, and exploratory data analysis were performed using `Matplotlib` and `Plotly`, while interactive analysis was supported using `ipywidgets`. Additional utilities for time handling and experiment tracking were implemented using Python's standard libraries, including `datetime`, `time`, and `copy`.

3.8.2 Reproducibility Considerations

Several components of the machine learning pipeline are stochastic, meaning they involve randomness and will produce different results between runs even when using the same configuration. This includes model training (such as weight initialization and batch sampling), feature selection methods, and hyperparameter tuning through grid search, where different runs can lead to different selected features, model parameters, and final results.

To improve reproducibility, fixed random seeds were used in NumPy and PyTorch where possible, and the full workflow from preprocessing to evaluation was implemented as a consistent pipeline across all experiments.

Despite this, the overall procedure is fully specified, allowing the same experimental setup to be reproduced and comparable results to be obtained.

4

Results & Discussion

4.1 Feature Selection

4.1.1 Correlation-Based Feature Filtering

Correlation-based feature filtering provides an initial indication of which features are most strongly associated with each target (upward and downward regulation prices and volumes) in both CM and EAM markets. As it relies on pairwise real-time feature-target correlations without accounting for lag structure or feature interactions, it should be interpreted as an initial screening of potentially informative variables rather than a measure of predictive importance.

4.1.1.1 mFRR CM (SE3)

The correlation-based feature filtering results for the CM market are shown in Table 4.1 (and presented in more detail in Table A.1), providing an initial overview of the most strongly correlated features with CM upward and downward regulation prices and volumes.

Table 4.1: Top 15 correlated features for mFRR CM Up/Down Price/Volume SE3, sorted by mean absolute correlation across all CM targets.

Feature	mFRR CM Mean Abs. Corr.	mFRR CM Up Price (corr)	mFRR CM Up Volume (corr)	mFRR CM Down Price (corr)	mFRR CM Down Volume (corr)	count
Month of Year (sin) (E.F) [-]	**0.370**	0.185	-0.533	0.542	0.218	4
Week of Year (sin) (E.F) [-]	**0.313**	0.098	-0.558	0.395	0.199	4
mFRR CM Down Price SE3 [EUR/MW]	**0.266**	0.325	-0.303	NaN	0.436	3
Production Forecast Error SE3 (E.F) [MW]	**0.246**	0.196	-0.317	0.203	0.269	4
Total Mean Temperature Forecast Sweden (E.F) [CEL]	**0.237**	0.219	0.224	0.412	0.091	4
mFRR CM Down Volume SE3 [MW]	**0.234**	0.218	-0.282	0.436	NaN	3
Total Wind Production Forecast Sweden [MW]	**0.222**	-0.372	-0.203	-0.125	-0.188	4
Temperature Forecast for Göteborg [CEL]	**0.221**	0.206	0.206	0.394	0.079	4
Total Production Forecast Sweden [MW]	**0.218**	-0.198	-0.204	-0.419	-0.052	4
Spot Volume Buy Forecast for SE3 [MW]	**0.208**	-0.092	-0.153	-0.489	-0.098	4
Total Spot Volume Sell Forecast Sweden [MW]	**0.208**	-0.176	-0.165	-0.429	-0.061	4
Total Spot Volume Buy Forecast Sweden [MW]	**0.204**	-0.091	-0.152	-0.488	-0.087	4
Spot Price Forecast SE3 [EUR/MWh]	**0.199**	0.336	0.109	-0.312	0.039	4
Production Forecast SE3 [MW]	**0.195**	-0.249	-0.244	-0.286	0.001	4
NP Area Spot Systemprice Forecast [EUR/MWh]	**0.194**	0.264	0.130	-0.365	0.016	4

The strongest patterns are observed for long-term temporal variables. *Month of*

Year (sin) (0.370) and *Week of Year (sin)* (0.313) show the highest correlations across CM targets, indicating clear seasonal structure in capacity needs, likely reflecting recurring monthly and weekly patterns in electricity demand and supply throughout the year.

Several system variables also appear consistently across all CM targets. In particular, *Production Forecast Error SE3*, with a mean absolute correlation of 0.246, *Total Wind Production Forecast Sweden* (0.222), and *Total Production Forecast Sweden* (0.218) show stable relevance across both upward and downward CM markets, for both volume and price forecasting tasks. The full results in Appendix A.1 further show that solar forecasts, consumption forecasts for both SE3 and Sweden, and the renewable share forecast for Sweden also have relatively high correlations across the CM target variables. This suggests that CM price formations and volume procurements are strongly influenced by expected production and consumption levels, production forecast errors, and renewable generation availability, indicating that CM outcomes are closely linked to expected overall system balance conditions.

CM balancing market variables, including *mFRR CM Down Price* (0.266) and *mFRR CM Down Volume SE3* (0.234), further indicate intra-market dependencies within CM, suggesting interactions between balancing prices and volumes across regulation states.

Weather-related variables, particularly temperature and renewable generation forecasts, also show moderate correlations, especially for CM Down Price. Their influence is likely indirect through their impact on electricity demand, renewable share production, and overall system balance.

Market variables, including spot prices, spot volumes, and CM balancing variables, also appear among the most correlated features. This indicates that CM outcomes are linked both to expected day-ahead market conditions and to internal dependencies within the balancing market itself.

Overall, CM correlations are dominated by temporal and system variables, with market and weather variables providing additional explanatory value. This suggests that CM behavior is primarily driven by relatively stable and predictable system conditions.

4.1.1.2 mFRR EAM (SE3)

The correlation-based feature filtering results for the EAM market, shown in Table 4.2 (and presented in more detail in Table A.2), provide an overview of the most strongly correlated features with EAM upward and downward regulation prices and volumes.

In particular, *Spot Price SE3* shows the strongest relationship across all EAM targets, with a mean absolute correlation of 0.411, with a very strong correlation with EAM Up Price (0.872), a moderately high correlation with EAM Down Price (0.668), but substantially weaker relationships with both EAM Up Volume and EAM Down Volume. Similar patterns are observed for *Spot Price Forecast SE3* (0.372) and *NP Area Spot Systemprice Forecast* (0.363), both showing strong positive correlations with EAM balancing prices. This suggests that EAM pricing is closely linked to expected spot market conditions within the same bidding zone and the broader Nordic power market.

All EAM balancing market variables, i.e. *mFRR EAM Down Price* (with mean absolute correlation 0.411), *mFRR EAM Up Price* (0.332), *mFRR EAM Up Volume* (0.316), and *mFRR EAM Down Volume* (0.298), rank among the most strongly correlated features for all other EAM balancing targets. In particular, strong relationships are observed between upward and downward price variables, as well as between volume variables. This indicates strong internal dependencies within the EAM balancing market, where price

Table 4.2: Top 15 correlated features for mFRR EAM Up/Down Price/Volume SE3, sorted by mean absolute correlation across all EAM targets.

Feature	**mFRR EAM Mean Abs. Corr.**	mFRR EAM Up Price (corr)	mFRR EAM Up Volume (corr)	mFRR EAM Down Price (corr)	mFRR EAM Down Volume (corr)	count
Spot Price SE3 [EUR/MWh]	**0.411**	0.872	0.082	0.668	-0.024	4
mFRR EAM Down Price SE3 [EUR/MW]	**0.411**	0.723	0.378	NaN	0.545	3
Spot Price Forecast SE3 [EUR/MWh]	**0.372**	0.782	0.073	0.608	-0.026	4
NP Area Spot Systemprice Forecast [EUR/MWh]	**0.363**	0.728	0.111	0.610	0.005	4
Total Scheduled Intraday In Flow Statistics SE3 (E.F.) [MWh]	**0.335**	0.600	0.120	0.554	0.066	4
mFRR EAM Up Price SE3 [EUR/MW]	**0.332**	NaN	0.423	0.723	0.182	3
mFRR EAM Up Volume SE3 [MW]	**0.316**	0.423	NaN	0.378	0.464	3
Inertia (kinetic energy) - Nordic power system [GWs]	**0.303**	0.519	0.119	0.528	0.046	4
Total Netto Scheduled Intraday Flow Statistics SE3 (E.F.) [MWh]	**0.300**	0.514	0.114	0.468	0.103	4
mFRR EAM Down Volume SE3 [MW]	**0.298**	0.182	0.464	0.545	NaN	3
Renewable (Wind/Solar) Share Forecast Sweden (E.F.) [%]	**0.293**	-0.515	-0.077	-0.479	-0.101	4
Exchange Forecast SE3 [MW]	**0.291**	0.513	0.073	0.496	0.082	4
Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	**0.257**	0.544	0.052	0.369	-0.062	4
Spot Price SE3 RZD-1d (E.F.) [-]	**0.241**	0.520	0.046	0.377	-0.023	4
Spot Price Forecast Spread SE1234 (Max-Min) (E.F.) [EUR/MWh]	**0.235**	0.505	0.051	0.336	-0.048	4

formation and activation needs across upward and downward regulation states are closely interconnected.

System-state variables, including intraday flow statistics, exchange forecasts, inertia, and time deviation, also show relatively strong correlations, particularly for EAM prices. This suggests that EAM balancing is strongly influenced by short-term operational conditions and transmission-flow dynamics within SE3, Sweden, and the broader Nordic power system.

Weather variables also remain relevant, though not to the same degree as for the CM market, particularly renewable-related forecasts such as renewable share forecasts, wind velocity forecasts, and wind production forecasts. These variables generally show negative correlations with EAM balancing prices, indicating that higher renewable availability is associated with lower balancing price levels, likely reflecting reduced system stress and greater availability of balancing resources.

Overall, the EAM results reveal a significantly different structure compared to the CM market. While temporal, system, and weather variables remain relevant, market and balancing market variables are among the most strongly correlated features, reflecting the more dynamic and near real-time nature of the balancing energy activation market. EAM correlations are dominated by day-ahead and balancing market variables together with short-term system-state conditions, whereas weather variables provide additional signal through their impact on renewable generation and balancing resource availability, and temporal variables are less relevant. This indicates that EAM behavior is primarily reactive, driven by short-term system imbalances and market interactions rather than the more predictable seasonal and demand–supply patterns observed for CM.

4.1.2 Feature Importance Filtering - Random Forest

Random Forest feature importance was used as an intermediate feature-selection step between correlation filtering and forward selection. Its primary purpose was to capture potential nonlinear feature-target relationships and further reduce the feature space before the computationally more expensive forward-selection procedure.

This step primarily serves to reduce the number of candidate features before forward selection rather than to provide feature interpretation. Therefore, the results for feature importance filtering are not discussed in this study, but the full rankings are included for CM and EAM for completeness in Appendix A, Section A.1.2.

4.1.3 Forward Feature Selection – LSTM

Forward feature selection was applied using an LSTM estimator with rolling time-series cross-validation to evaluate different combinations of input features in terms of predictive performance for each target. Complete forward-selection rankings for all CM and EAM targets are provided in Appendix A, Section A.1.3 in Tables A.5–A.12. Since downward regulation forecasting is of particular relevance for wind-based balancing participation, the results presented and discussed in this section are focused on downward regulation prices and volumes for the CM and EAM markets.

4.1.3.1 Forward Selection – mFRR CM (SE3)

The forward selection results for the CM market are shown in full in Section A.1.3.1 in Tables A.5–A.8, and show a relatively structured and interpretable feature selection process for both CM downward and upward price and volume forecasting.

4.1.3.1.1 mFRR CM Down Price SE3

For CM Down Price (see Table 4.3), the historical target value is selected first and achieves a mean RMSE of 19.42, confirming strong autoregressive behavior and showing that recent CM down prices provide the strongest baseline signal for predicting future prices. This is reasonable, since balancing prices often evolve relatively continuously over time and are influenced by recent price levels and short-term market conditions. However, in periods of system imbalance and high volatility, these dynamics break, resulting in abrupt price spikes.

The largest improvement occurs after adding Total Wind Production Forecast Sweden in Step 2, reducing RMSE to 18.85, highlighting the relevance of renewable generation for downward balancing needs and price formation. In Step 3, Total Consumption Forecast Sweden further reduces RMSE to 18.40, indicating that expected demand is also an important driver of CM down prices.

After the first few steps, improvements become marginal and performance largely stabilises, reaching a minimum RMSE of 18.29 at Step 9. Later-selected features include system inertia, exchange forecast error, spot prices, and cross-zonal price spreads. Their limited impact suggests that most predictive information is already captured by historical prices, wind generation, and demand forecasts, while additional variables mainly add redundancy and noise once the main effects are included.

Table 4.3: mFRR CM Down Price SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR CM Down Price SE3 [EUR/MW]	19.42	8.53	22.24
2	Total Wind Production Forecast Sweden [MW]	18.85	8.34	21.60
3	Total Consumption Forecast Sweden [MW]	18.40	8.24	21.12
4	Inertia (kinetic energy) - Nordic power system [GWs]	18.43	8.12	21.11
5	Exchange Forecast Error SE3 (E.F.) [MW]	18.34	8.24	21.06
6	Spot Price SE3 [EUR/MWh]	18.38	8.21	21.09
7	Spot Price Forecast Spread (SE3-SE4) (E.F.) [EUR/MWh]	18.30	8.39	21.07
8	Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	18.36	8.31	21.10
9	Spot Price Forecast Spread SE1234 (Max-Min) (E.F.) [EUR/MWh]	18.29	8.42	21.07
10	Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	18.31	8.44	21.10

4.1.3.1.2 mFRR CM Down Volume SE3

For CM Down Volume (see Table 4.4), a similar pattern appears. The first selected feature is the CM down volume historical gives an RMSE of 49.62, while adding Total Consumption Forecast Sweden immediately reduces it to 47.43, representing the largest improvement seen across all target variables when adding a feature to the set. This indicates that expected system load is particularly important for volume activation decisions. Spot Price SE3 further improved performance to 46.08, suggesting direct coupling between balancing needs and day-ahead market prices and conditions.

Table 4.4: mFRR CM Down Volume SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR CM Down Volume SE3 [MW]	49.62	8.41	52.40
2	Total Consumption Forecast Sweden [MW]	47.43	8.67	50.29
3	Spot Price SE3 [EUR/MWh]	46.08	9.09	49.08
4	Total Mean Cloud Cover Forecast Sweden (E.F.) [%]	46.80	8.46	49.59
5	Day of Week (sin) (E.F.) [-]	45.85	7.68	48.38
6	Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	45.90	7.62	48.42
7	Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	45.55	7.90	48.16
8	Total Wind Production Forecast Sweden [MW]	44.76	6.71	46.98
9	mFRR CM Up Price SE3 [EUR/MW]	44.74	7.28	47.14
10	NP Area Spot Systemprice Forecast [EUR/MWh]	45.00	7.94	47.62

Additional weather-related variables such as cloud cover, wind velocity, and total wind production are then selected, gradually reducing RMSE to 44.76, which reflects the operational influence of renewable forecast uncertainty on balancing volume requirements. Temporal seasonality (Day of Week) and cross-market coupling through mFRR CM Up Price also contribute, indicating recurring weekly activation patterns and interaction between upward and downward balancing states.

Overall, improvements remain small and performance is largely stable after the first few steps, suggesting that most additional variables provide overlapping or redundant information rather than new predictive signal.

4.1.3.1.3 Summary – mFRR CM

The full CM forward-selection rankings are presented in Appendix A, Section A.1.3.1, and show similar patterns across all four CM targets. Historical target values are consistently selected first, confirming strong autoregressive behavior throughout the CM market. These are typically followed by wind generation forecasts, consumption forecasts, spot-market signals, and cross-zonal price spread variables.

Overall, this indicates that CM outcomes are largely shaped by predictable structural system conditions, where expected production-demand balance dominates both price formation and reserve volume procurement, while additional market variables mainly provide secondary refinements.

4.1.3.2 Forward Selection – mFRR EAM (SE3)

In contrast to the CM market, the forward selection results for EAM (see Tables A.9–A.12 in Section A.1.3.2) show smaller and less stable marginal improvements. This reflects the more dynamic and close-to-real-time nature of activation energy markets, where prices and volumes are more directly influenced by rapidly changing system conditions and exhibit stronger short-term variability. At the same time, the results indicate tighter coupling between different EAM products and related market signals.

4.1.3.2.1 mFRR EAM Down Price SE3

For EAM Down Price (see Table 4.5), the historical target value is selected first and provides a baseline RMSE of 156.33, confirming autoregressive behavior, where recent EAM down prices already contain substantial predictive information for future prices.

Table 4.5: mFRR EAM Down Price SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR EAM Down Price SE3 [EUR/MW]	156.33	161.96	209.77
2	Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	155.60	161.91	209.03
3	Inertia (kinetic energy) - Nordic power system [GWs]	155.53	161.75	208.91
4	Total Wind Production Forecast Sweden [MW]	154.98	162.14	208.49
5	mFRR EAM Up Price SE3 [EUR/MW]	154.99	162.29	208.54
6	Wind Production Forecast SE3 [MW]	155.10	162.12	208.60
7	Wind Velocity Forecast for Göteborg [m/sec]	155.36	161.99	208.81
8	mFRR EAM Up Volume SE3 [MW]	155.57	161.95	209.01
9	Total Exchange Forecast Sweden [MW]	155.95	161.48	209.23
10	Spot Volume Sell for SE3 RZD-1d (E.F.) [-]	156.07	161.62	209.40

In Step 2, adding Total Mean Wind Velocity Forecast Sweden reduces RMSE to 155.60, indicating that short-term wind conditions contain useful information for EAM prices. In Step 3, Inertia (kinetic energy) further reduces RMSE to 155.53, suggesting that system stability also affects price formation. In Step 4, Total Wind Production Forecast Sweden lowers RMSE to 154.98, reinforcing the role of overall wind generation in downward activation pricing.

After these initial improvements, performance remains largely flat and slightly fluctuates, ending at 156.07 in Step 10. Later-selected features include EAM price and volume variables, wind-related forecasts, exchange forecasts, and spot market indicators. Overall, these features add little additional explanatory power beyond wind conditions and system stability, and may introduce redundancy once the main drivers are already captured.

4.1.3.2.2 mFRR EAM Down Volume SE3

For EAM Down Volume (see Table 4.6), the historical target value provides a baseline RMSE of 33.65, confirming autoregression and that recent EAM down volumes contain predictive information about future volumes.

Table 4.6: mFRR EAM Down Volume SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR EAM Down Volume SE3 [MW]	33.65	6.09	35.66
2	Spot Price SE3 [EUR/MWh]	33.35	5.87	35.28
3	Spot Price SE3 RZD-1d (E.F.) [-]	33.29	5.78	35.20
4	Wind Production Forecast SE3 [MW]	33.24	5.58	35.08
5	Spot Volume Sell for SE3 RZD-1d (E.F.) [-]	33.22	5.59	35.07
6	Wind Velocity Forecast for Göteborg [m/sec]	33.18	5.51	35.00
7	mFRR EAM Up Price SE3 [EUR/MW]	33.20	5.46	35.00
8	mFRR EAM Up Volume SE3 [MW]	33.18	5.37	34.95
9	mFRR EAM Down Price SE3 [EUR/MW]	33.21	5.46	35.01
10	Total Wind Production Forecast Sweden [MW]	33.26	5.46	35.06

Spot-market variables provide early improvements. In Step 2, Spot Price SE3 reduces RMSE to 33.35, followed by Spot Price SE3 RZD-1d in Step 3 (33.29). In Step 5, Spot Volume Sell for SE3 RZD-1d is selected, indicating that both price and volume information from the day-ahead market contribute to explaining EAM down volumes. Wind-related variables also appear consistently, including Wind Production Forecast SE3 (Step 4), Wind Velocity Forecast for Göteborg (Step 6), and Total Wind Production Forecast Sweden (Step 10). In addition, all four EAM variables (up/down prices and volumes) are selected within the top 10 steps, indicating strong interdependence within the EAM market.

Overall, improvements remain small and performance is largely stable after the first few steps, suggesting that most additional variables provide overlapping or redundant information rather than new predictive signal.

4.1.3.2.3 Summary – mFRR EAM SE3

The full EAM forward-selection rankings are presented in Appendix A, Section A.1.3.2, and show similar patterns across all four EAM targets. Feature selection consistently prioritizes the historical target value, confirming autoregressive behavior, followed by spot-market signals, wind-related forecasts, cross-market EAM variables, and system-state indicators such as inertia and time deviation.

Compared to CM, improvements from additional features are generally smaller and less stable, with predictive information distributed across more interacting variables rather than a few dominant structural drivers. While CM forecasting is mainly explained by broader physical system fundamentals such as demand and renewable generation levels, EAM forecasting depends more strongly on immediate market-state interactions and rapidly changing market, weather, and balancing conditions.

This suggests that EAM behavior is less persistent and more reactive to short-term operational fluctuations, making the feature selection process less stable and more tightly coupled across related EAM markets. Compared to CM, EAM appears less predictable and offers less benefit from adding large feature sets, reflecting the faster and more stochastic nature of balancing activation closer to real time.

4.2 Forecasting Performance Evaluation and Comparison

This section compares forecasting performance across LSTMs, MLPs, and naive baseline models for the CM and EAM markets in SE3, both at the market level and for each of the eight target variables (CM/EAM Up/Down Price/Volume). Performance was evaluated using univariate (uv), multivariate (mv), and top 5 and top 10 forward-selected feature subsets (fs5 and fs10). Full performance rankings for all target variables are provided in Appendix A in Section A.2.

4.2.1 Forecasting Performance – mFRR CM SE3

The CM results show clear differences in forecasting performance across targets, where simple baseline models are highly competitive and, on average, stronger and more stable than machine learning models, suggesting that CM outcomes are largely driven by persistent and recurring patterns.

4.2.1.1 mFRR CM Down Price SE3

The *CM Down Price SE3* train-validation-test time-series is shown in Figure 4.1. The forecasting performance on the test set across all models and feature sets is summarized in Table 4.7.

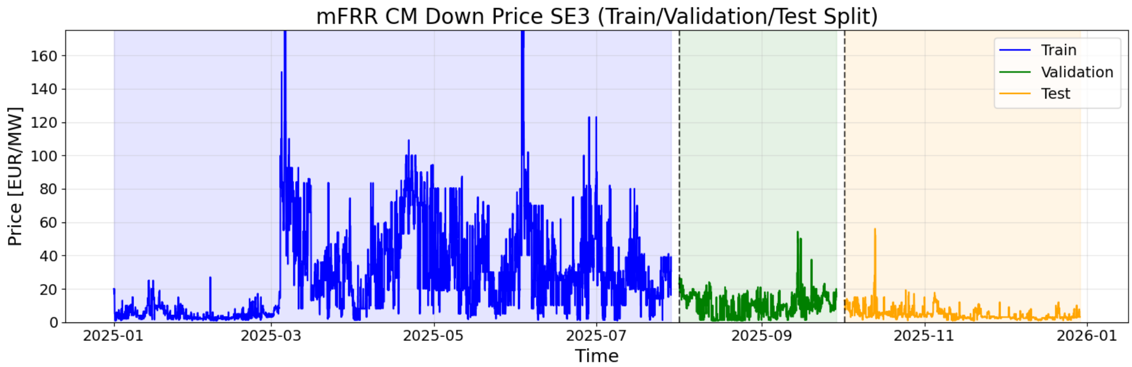

Figure 4.1: mFRR CM Down Price SE3 train-validation-test time-series.

A clear result is that all naive baseline models outperform all machine learning models, by a clear margin. The strongest overall performance was achieved by the Naive Mean-1w baseline, with a test RMSE of 3.87, closely followed by the Naive Seasonal-Mean-1w (3.96) and Naive Mean-1d (4.26). For the machine learning models, the best-performing model was MLP (uv) with an RMSE of 5.30, while most of the remaining MLP and LSTM models performed considerably worse.

This result suggests that the CM Down Price is largely driven by strong mean-level persistence at both daily and weekly scales, and seasonal effects are most effectively captured through weekly-averaged patterns rather than pure daily or weekly repetition.

The prediction comparison of the top models from each model class is shown in Figure 4.6. A likely explanation for the poor neural network performance, is the clear regime shift visible in Figure 4.1. During the training period, particularly from March

Table 4.7: Forecasting performance for mFRR CM Down Price SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, top 5 and top 10 forward-selected features, and multivariate), ranked by RMSE on the test set.

Rank	Model	RMSE	MAE
1	Naive Mean-1w	3.87	2.22
2	Naive Seasonal-Mean-1w	3.96	2.24
3	Naive Mean-1d	4.26	2.23
4	Naive Seasonal-1d	5.10	2.49
5	Naive Seasonal-1w	5.25	2.89
6	MLP (uv)	5.30	4.01
7	LSTM (mv)	6.79	5.05
8	LSTM (uv)	7.65	6.65
9	LSTM (fs10)	10.95	7.53
10	MLP (fs5)	11.19	9.49
11	MLP (fs10)	11.93	10.01
12	LSTM (fs5)	12.46	11.49
13	MLP (mv)	22.50	9.03

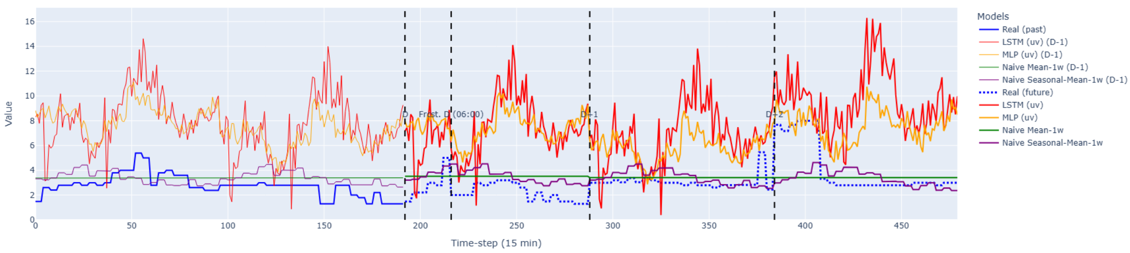

Figure 4.2: Forecast comparison for the top-performing models from each model-class for mFRR CM Down Price SE3: Naive Mean-1w (green), Naive Seasonal-Mean-1w (purple), MLP uv (yellow), LSTM mv (red), showing predicted vs. true target (blue) values.

to July, the price level is clearly higher and more volatile than during the validation and test periods, where prices return to a lower and more stable range. As a result, the machine learning models are trained primarily on dynamics that differ significantly from those observed during evaluation, reducing their ability to generalize. In contrast, the naive baseline models rely directly on recent historical observations and are therefore more robust to such distributional shifts.

4.2.1.2 mFRR CM Down Volume SE3

The *CM Down Volume SE3* train-validation-test time-series is shown in Figure 4.3. The forecasting performance on the test set across all models and feature sets is summarized in Table 4.8.

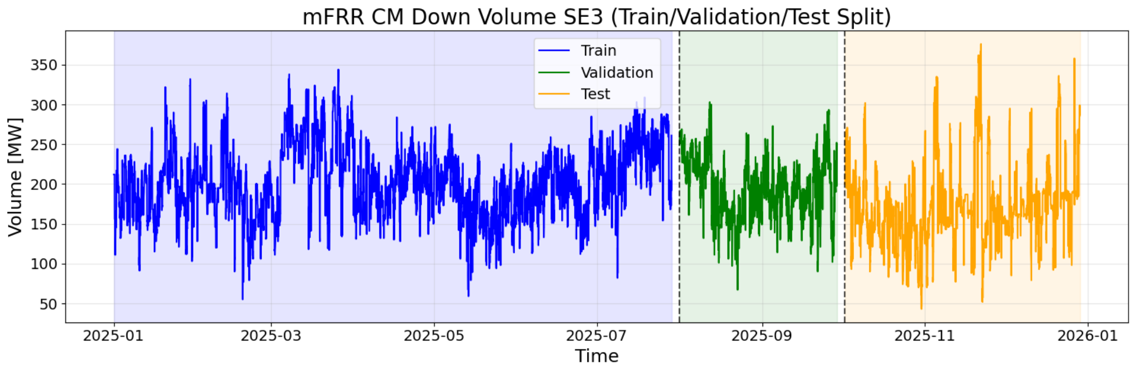

Figure 4.3: mFRR CM Down Volume SE3 train-validation-test time-series.

Overall, the results show a much more competitive relationship between machine learning models and the best-performing baseline models compared to the CM Down Price case. The best overall performance is achieved by the LSTM (fs10) model with an RMSE of 52.68, closely followed by Naive Mean-1w (53.33) and Naive Seasonal-Mean-1w (53.77). Several other LSTM and MLP models perform in a similar range, while the weakest models are the seasonal naive baselines which show noticeably higher errors.

Table 4.8: Forecasting performance for mFRR CM Down Volume SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, top 5 and top 10 forward-selected features, and multivariate), ranked by RMSE on the test set.

Rank	Model	RMSE	MAE
1	LSTM (fs10)	52.68	39.74
2	Naive Mean-1w	53.33	37.77
3	Naive Seasonal-Mean-1w	53.77	38.70
4	LSTM (fs5)	53.95	41.78
5	LSTM (uv)	54.20	41.97
6	MLP (mv)	54.77	40.27
7	MLP (uv)	55.25	41.65
8	MLP (fs5)	56.34	44.14
9	LSTM (mv)	56.39	44.83
10	MLP (fs10)	59.44	47.10
11	Naive Mean-1d	60.91	42.83
12	Naive Seasonal-1d	69.12	49.11
13	Naive Seasonal-1w	72.85	53.49

Among the machine learning models, no consistent benefit from increasing input dimensionality is observed. For LSTMs, the fs10, fs5, and univariate models perform best, while the multivariate model performs worse. MLP performance is more irregular across feature sets, although LSTMs are generally slightly more stable. At the same time, the poor performance of the daily and weekly seasonal baselines suggests that simple repetition of past values is insufficient to capture the dynamics of the series. Instead, the strong performance of the Mean-1w and Seasonal-Mean-1w baselines indicates that CM Down Volume is better characterized by averaged weekly behavior, such as underlying level and intraday patterns.

The results suggest that CM Down Volume is characterized by more complex dynamics than CM Down Price, including time-varying level shifts, stronger variability, and less stable short-term repetition. This is consistent with the observed time-series structure in Figure 4.3, where the volume fluctuates within a broad range and shows both gradual changes in level and frequent short-term spikes. These characteristics make pure repetition-based forecasting less effective, while baseline models capturing averaged historical behavior and machine learning models capturing nonlinear interactions and temporal dependencies perform relatively better, as illustrated in Figure 4.4, which compares predictions from the best-performing models.

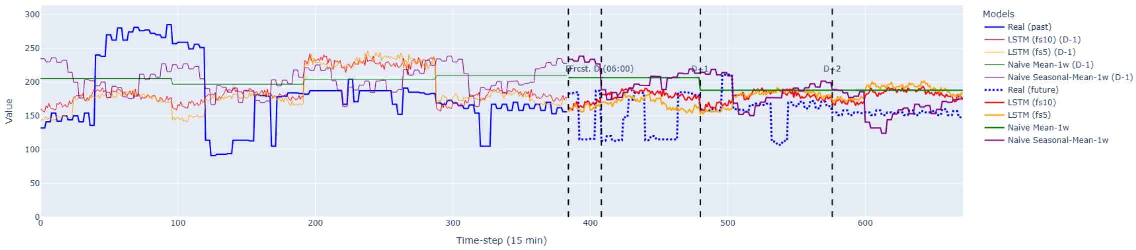

Figure 4.4: Forecast comparison for the top-performing models for mFRR CM Down Volume SE3: LSTM fs10 (red), Naive Mean-1w (green), Naive Seasonal-Mean-1w (purple), LSTM fs5 (yellow), showing predicted vs. true target (blue) values.

Both Table 4.8 and Figure 4.4 suggest that the top-performing models achieve very similar forecasting performance, both in terms of statistical accuracy and actual prediction behavior. While the models appear to capture different dynamics in the series, no single model consistently dominates the others.

4.2.1.3 mFRR CM Up Price & Volume SE3

Detailed results for CM Up Price and CM Up Volume are reported in Appendix A.2. In summary, CM Up Price benefits moderately from LSTM-based models, particularly when using selected feature sets, although the performance differences relative to strong averaging-based baselines remain small. In contrast, CM Up Volume is largely dominated by averaging-based baseline models, indicating highly persistent and repetitive dynamics, with most machine learning models performing substantially worse.

4.2.1.4 Summary – mFRR CM

Overall, CM forecasting is dominated by strong persistence and seasonal structure, indicating relatively stable and predictable dynamics. Naive averaging-based baselines are highly competitive and often outperform machine learning models. Feature selection provides limited benefits, and larger multivariate feature sets rarely improve performance. While some targets benefit from LSTM-based models, forecasting performance is generally explained well by simple statistical models.

4.2.2 Forecasting Performance – mFRR EAM SE3

The EAM results differ substantially from CM, showing that machine learning models generally outperform naive baselines, reflecting more complex and reactive short-term dynamics.

4.2.2.1 mFRR EAM Down Price SE3

The *EAM Down Price SE3* train-validation-test time-series is shown in Figure 4.1.

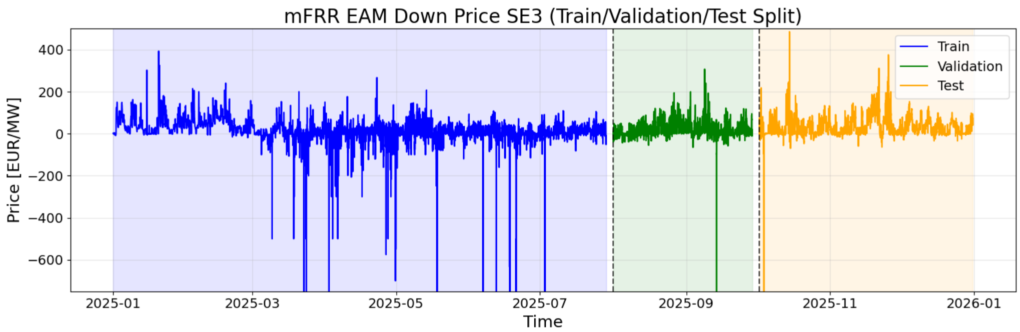

Figure 4.5: mFRR EAM Down Price SE3 train-validation-test time-series.

The forecasting performance on the test set across all models and feature sets is summarized in Table 4.9.

Overall, the results show that machine learning models clearly outperform the naive baseline methods. The best performance is achieved by the MLP (mv) model with an RMSE of 117.84, closely followed by LSTM (mv) (118.44). The best naive baseline is Naive Mean-1d with an RMSE of 120.14, while the seasonal naive models perform substantially worse. In general, MLP and LSTM models dominate the top ranks, indicating that nonlinear and multivariate learning approaches provide a clear advantage for this target.

Among machine learning models, multivariate (mv) inputs perform best for both MLP and LSTM, while univariate (uv) and feature-selected models (fs10 and fs5) are only slightly worse. This suggests that multiple explanatory variables are useful, but that gains from adding more inputs are limited, indicating diminishing returns beyond a moderate feature set.

For baselines, mean-based methods outperform seasonal variants, as they better capture short-term level dynamics, although with reduced responsiveness to sudden changes. Seasonal-Mean-1w captures recurring structure but introduces bias, while pure seasonal models perform poorly, indicating that strict repetition is insufficient for this target.

The time-series in Figure 4.5 shows clear regime changes with varying volatility and shifts in both level and spike intensity. This makes the series difficult for repetition-based models. While mean-based approaches capture average behavior reasonably well, they fail to react to abrupt changes, whereas machine learning models benefit from their ability to capture nonlinear relationships and interactions across variables.

Table 4.9: Forecasting performance for mFRR EAM Down Price SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, top 5 and top 10 forward-selected features, and multivariate), ranked by RMSE on the test set.

Rank	Model	RMSE	MAE
1	MLP (mv)	117.84	25.75
2	LSTM (mv)	118.44	27.06
3	MLP (uv)	119.38	28.21
4	MLP (fs10)	119.56	29.40
5	LSTM (uv)	119.69	28.91
6	Naive Mean-1d	120.14	28.96
7	LSTM (fs5)	120.84	31.29
8	LSTM (fs10)	121.04	31.78
9	Naive Mean-1w	121.20	34.01
10	MLP (fs5)	121.45	32.54
11	Naive Seasonal-Mean-1w	127.99	35.12
12	Naive Seasonal-1d	166.59	35.54
13	Naive Seasonal-1w	169.12	41.50

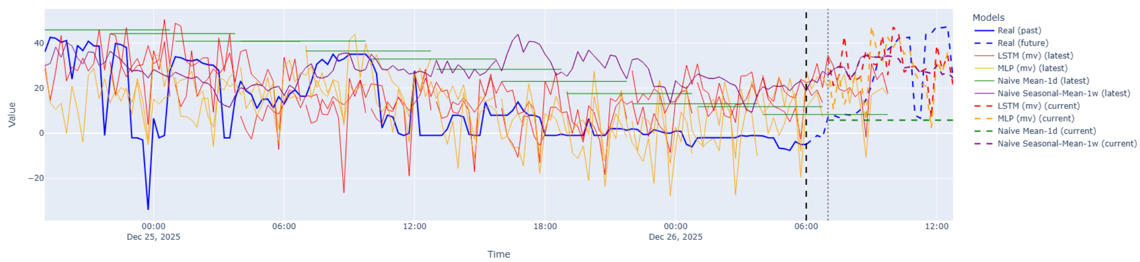

Figure 4.6: Forecast comparison for the top-performing models from each model-class for mFRR EAM Down Price SE3: 1) MLP mv (yellow), LSTM mv (red), Naive Seasonal-Mean-1w (purple), Naive Mean-1d (green), showing predicted vs. true target (blue) values.

Finally, the prediction comparison in Figure 4.6 shows that machine learning models capture the overall level and trend direction well, but their predictions are more volatile than the observed series, with frequent short-term oscillations. This suggests that the models are sensitive to short-term fluctuations in the training data. A potential improvement is to apply post-processing smoothing, such as a rolling average, to produce more stable and interpretable forecasts.

4.2.2.2 mFRR EAM Down Volume SE3

The *EAM Down Volume* train-validation-test time-series is shown in Figure 4.7. The forecasting performance on the test set across all models and feature sets is summarized in Table 4.10.

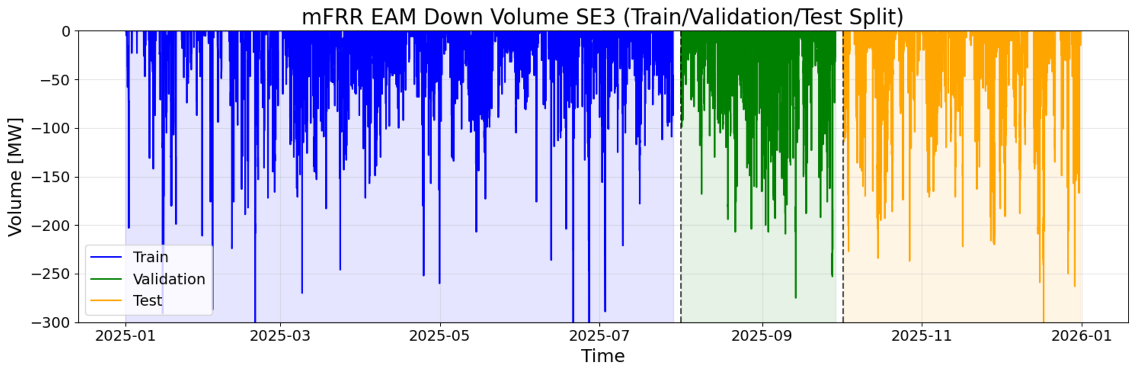

Figure 4.7: mFRR EAM Down Volume SE3 train-validation-test time-series.

Overall, the results show that machine learning models outperform the naive baseline methods. The best performance is achieved by LSTM (fs5) with an RMSE of 40.51, closely followed by MLP (fs5) (40.56), while the best-performing baseline model is Naive Mean-1w (42.24). Mean-based baselines (Naive Mean-1d and Naive Mean-1w) achieve intermediate performance, while Naive Seasonal-Mean-1w performs slightly worse. In contrast, the pure seasonal naive baselines perform clearly worse than all other models.

Table 4.10: Forecasting performance for mFRR EAM Down Volume SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, top 5 and top 10 forward-selected features, and multivariate), ranked by RMSE on the test set.

Rank	Model	RMSE	MAE
1	LSTM (fs5)	40.51	25.38
2	MLP (fs5)	40.56	24.99
3	LSTM (uv)	40.74	24.49
4	MLP (uv)	40.88	24.96
5	MLP (fs10)	41.48	24.85
6	MLP (mv)	41.65	26.94
7	LSTM (mv)	42.06	26.02
8	Naive Mean-1w	42.24	28.58
9	LSTM (fs10)	42.27	26.46
10	Naive Mean-1d	42.73	26.93
11	Naive Seasonal-Mean-1w	44.52	29.60
12	Naive Seasonal-1d	56.15	31.69
13	Naive Seasonal-1w	59.00	34.58

Among the machine learning models, LSTM and MLP perform very similarly. In terms of feature sets, fs5 performs best, followed by univariate (uv), while fs10 and multivariate (mv) perform slightly worse. This suggests that autoregressive information is highly informative, while a small number of selected external variables provides the strongest signal. Larger feature sets may introduce noise and reduce performance.

The time-series in Figure 4.7 shows a highly intermittent activation pattern dominated by zero values with occasional negative spikes (typically -50 to -150 MW, with extremes below -300 MW). This makes forecasting challenging, as the model must both detect activation events and estimate their magnitude. Mean-based models tend to predict near-zero values but fail to distinguish inactive periods from activations. Seasonal models also perform poorly due to their rigid structure.

Machine learning models better capture nonlinear dependencies, but tend to produce smoothed predictions close to zero or slightly negative values. This indicates that they learn the average activation level rather than accurately separating inactive periods from spike events. They do, however, partially detect activation periods, as seen by downward shifts in forecasts during likely activation windows, even if spike magnitudes are not well captured.

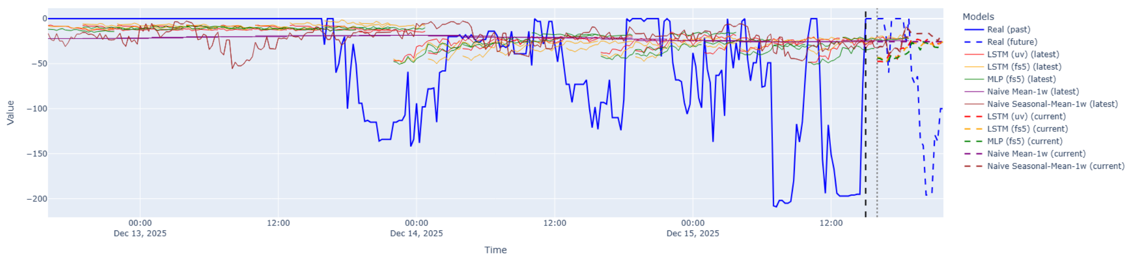

Figure 4.8: Forecast comparison for the top-performing models from each model class for mFRR EAM Down Volume SE3: LSTM fs5 (yellow), MLP fs5 (green), LSTM uv (red), Naive Mean-1w (purple), Naive Seasonal-Mean-1w (brown), showing predicted vs. true target (blue) values.

4.2.2.3 mFRR EAM Up Price & Volume SE3

Detailed results for EAM Up Price and EAM Up Volume are reported in Appendix A.2. In summary, machine learning models outperform the naive baselines for both targets. EAM Up Price benefits from external explanatory variables, with feature-selected and multivariate models performing best, indicating that incorporating additional variables improves predictive performance beyond purely autoregressive information. EAM Up Volume is best captured by autoregressive information, where univariate models perform strongest, while adding more features appears to introduce noise or poorly learned dynamics, leading to reduced performance.

4.2.2.4 Summary – mFRR EAM

Overall, EAM forecasting exhibits more complex dynamics than CM. Machine learning models generally outperform naive baselines, suggesting that forecasting performance depends more on nonlinear relationships, interactions between variables, and rapidly changing market conditions than on simple persistence and seasonal patterns. At the same time, univariate and forward-selected feature sets are often competitive with full multivariate models, indicating that most predictive information is captured by autoregressive behavior together with a small number of relevant external variables.

4.2.3 Overall Model Performance Comparison

This section summarizes forecasting performance across models, feature configurations, markets, and targets in the mFRR SE3 dataset. Figure 4.9 presents the average ranking of all evaluated models and model classes across the CM and EAM markets.

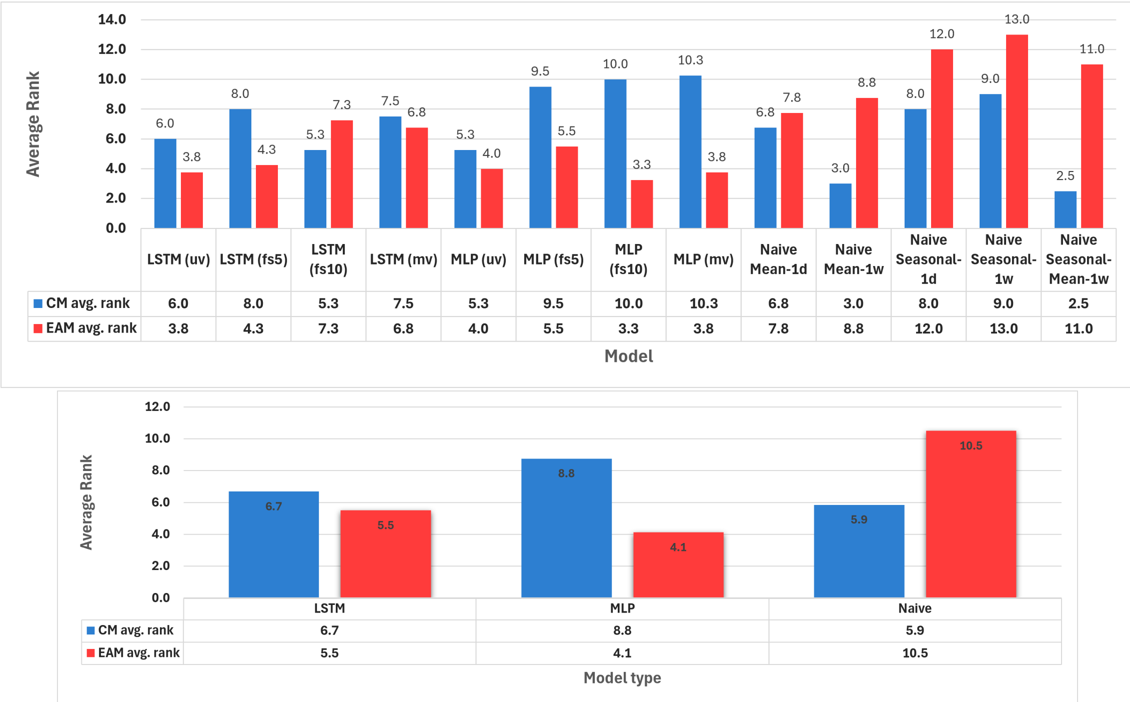

Figure 4.9: Model performance comparison using average model rankings for individual models (top) and model classes (bottom) in the CM and EAM markets. Rankings are computed across the four forecasting targets (up/down price and volume), where lower ranks indicate better performance.

Overall, the results show clear differences between markets. For CM, the best-performing models are simple statistical baselines, especially weekly and seasonal means, which consistently achieve the lowest average ranks. This indicates that CM forecasting is mainly driven by persistent structure and recurring temporal patterns, which are well captured by simple historical averages. While some neural models perform well for specific targets, machine learning approaches do not provide consistent improvements over strong baselines.

In contrast, machine learning models generally perform better in the EAM market, where both MLP and LSTM models outperform naive benchmarks on average. This suggests that EAM dynamics contain more nonlinear and short-term dependencies that are not well captured by simple averaging. However, differences between architectures are small and inconsistent across targets.

Across feature configurations, no approach is consistently superior. Univariate models are often competitive and stable, while multivariate models show higher variability, sometimes improving but also degrading performance. Forward-selected feature sets typically provide a good trade-off between dimensionality and stability.

Overall, forecasting performance is driven more by market and target characteristics than by model complexity or feature set size. CM is generally well captured by simple statistical structure, whereas EAM benefits more from machine learning methods that can capture more complex and rapidly changing dynamics.

4.2.4 Practical Use of Forecasts

The forecasts can be used as decision-support tools for trading and bidding in the mFRR CM and EAM markets. For a given market and regulation direction, the forecasts provide estimates of expected price and volume.

The simplest use is to visualize the forecasted price and volume curves as graphs to support manual decision-making. A more advanced use is to use the forecasts as input to rule-based or expected-value-based trading strategies, where forecasts are converted into trading signals based on thresholds or expected profitability.

In practice, such forecast strategies support comparisons between expected revenue from mFRR activation with alternative spot-market revenues and potential imbalance costs, helping identify the market action with the highest expected profitability. This also contributes to improved risk management by quantifying expected outcomes and reducing uncertainty in trading decisions.

5
Conclusion

5.1 Overview of the study

This thesis investigated forecasting of mFRR capacity market (CM) and activation energy market (EAM) prices and volumes in the Swedish bidding zone SE3 using statistical and machine learning methods. The objective was to identify informative features and compare forecasting performance across markets, targets, model architectures, and feature configurations.

5.2 Answers to research questions

This section summarizes the main findings of the thesis by directly addressing the four research questions based on the results from the feature selection and the forecasting performance evaluation and comparison across mFRR CM and EAM markets.

5.2.1 RQ1 – Important explanatory variables and predictors for mFRR

One of the most central findings is that historical targets are almost always selected first during forward selection, indicating strong autoregressive structure in all mFRR prices and volumes. Beyond this common pattern, the most informative additional features differ clearly between CM and EAM. CM forecasting is primarily influenced by temporal patterns, demand and renewable generation forecasts, day-ahead spot prices, and weather forecasts. EAM forecasting depends more on intra-EAM balancing-market variables, spot prices, system-state variables, and wind-related variables. This suggests that CM is largely driven by expected future system and market conditions, while EAM is more strongly influenced by near real-time system and market conditions and balancing activity, reflecting a distinction between longer-term planning in CM and short-term reactions in EAM.

5.2.2 RQ2 – Machine learning vs baseline performance

The relative performance of machine learning models and naive baselines depends strongly on the market and target. In CM, simple averaging-based baselines often match or outperform machine learning models, indicating strong persistence and recurring temporal structure. In contrast, machine learning models consistently outperform the baselines in EAM, suggesting that EAM dynamics are more nonlinear, reactive, and less effectively captured by simple historical averaging.

5.2.3 RQ3 – LSTM vs MLP: sequential vs static forecasting models

No architecture consistently outperforms the other across all targets. LSTMs are slightly stronger on average in CM and MLPs slightly stronger in EAM, but both architectures achieve comparable performance and alternate as the best-performing model depending on the target. This suggests that explicitly modeling sequential temporal dependencies through LSTMs does not provide a consistent forecasting advantage over static MLP models, and that market- and target-specific dynamics determine whether sequential or static modeling approaches are more effective.

5.2.4 RQ4 – Impact of feature set configurations

No feature set configuration consistently outperforms the others. Univariate, feature-selected, and multivariate inputs achieve similar overall performance, suggesting that increasing feature dimensionality does not systematically improve accuracy and may introduce noise rather than additional predictive signal.

5.3 Key findings

The results demonstrate a clear distinction between the CM and EAM markets. CM forecasts are largely driven by recurring temporal patterns and stable system conditions, making simple statistical baselines highly competitive. In contrast, EAM forecasts depend more strongly on short-term market conditions, balancing activity, and system-state variables, resulting in greater benefits from machine learning approaches.

Across all targets, no single model architecture or feature configuration consistently outperformed the others. LSTM and MLP models achieved similar performance, while univariate, feature-selected, and multivariate inputs generally produced comparable results. This suggests that forecasting performance is driven more by market-specific characteristics than by increasing model or feature complexity.

Overall, the findings indicate that simple approaches are often sufficient for CM forecasting, while EAM forecasting benefits from nonlinear machine learning models capable of capturing more dynamic market behavior.

5.4 Limitations

This study is subject to several limitations. The available dataset is relatively small, covering only one year of data and resulting in limited training, validation, and test sets, which limits learning, forecasting performance, and generalization.

Electricity and balancing markets are also inherently noisy and highly uncertain, which introduces variability that is difficult to fully capture using the available features and models. This limits the achievable forecasting accuracy regardless of model choice.

In addition, computational constraints limited the extent of hyperparameter tuning and model experimentation, particularly for more complex neural network configurations.

Furthermore, while all eight forecasting targets (four per market) were modeled using a consistent framework for comparability, a more target-specific modeling approach could potentially improve performance, as different targets may benefit from different feature and model configurations.

Finally, neural network models are highly limited in interpretability due to their black-box nature, making their internal decision processes non-transparent and very difficult to explain.

5.5 Future work

A natural extension of this work is to use a longer historical dataset. This study is based on only one year of data, while machine learning models are data-driven and generally benefit from larger datasets, assuming sufficiently stable underlying dynamics. Future work should also consider structural changes in the Nordic balancing market, including Sweden's planned integration into the European mFRR balancing platform MARI in Q1 2027.

Another potential improvement is to reformulate EAM volume forecasting as a two-stage task, where activation is first predicted as a classification problem and the activated volume is then forecasted conditional on activation. This could improve forecasting performance by separating activation occurrence from activation magnitude.

Finally, future work could focus on developing and evaluating bidding or trading strategies based on the forecasting outputs, allowing the forecasts to be assessed in terms of practical profitability and decision support.

5.6 Final remarks

This thesis has shown that forecasting performance differs significantly between the mFRR capacity market (CM) and activation energy market (EAM). While simple statistical models are often sufficient for CM forecasting, EAM forecasting benefits more from machine learning methods capable of capturing nonlinear and rapidly changing market behavior. Overall, the results highlight that forecasting performance is primarily determined by market- and target-specific dynamics, with model and feature complexity playing a secondary role, emphasizing that forecasting approaches should be tailored to the specific dynamics of each market and target. As balancing markets become increasingly important for ensuring system balance, operational flexibility, and integration of renewable energy, accurate forecasting tools are expected to play an important role in supporting future market participation and decision-making.

Bibliography

[1] Göteborg Energi. *Vindkraft*. https://www.goteborgenergi.se/privat/elavtal/vindkraft. Accessed: 2026-06-18. 2026.

[2] Energimyndigheten. *Elsystemet och elmarknaden*. Accessed: 2026-01-19. URL: https://www.energimyndigheten.se/energisystem-och-analys/nulaget-i-energisystemet/om-det-svenska-elsystemet/.

[3] Svenska kraftnät. *Om elmarknaden*. Accessed: 2026-01-19. 2025. URL: https://www.svk.se/om-kraftsystemet/om-elmarknaden/.

[4] Svenska kraftnät. *Så påverkar olika kraftslag systemet*. Accessed: 2026-01-19. 2025. URL: https://www.svk.se/om-kraftsystemet/sa-paverkar-olika-kraftslag-systemet/.

[5] Svenska kraftnät. *Svenska kraftnäts ansvar i kraftsystemet*. Accessed: 2026-01-19. 2026. URL: https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/svenska-kraftnats-ansvar-i-kraftsystemet/.

[6] Svenska kraftnät. *Balansmarknaden*. Accessed: 2026-01-19. 2025. URL: https://www.svk.se/om-kraftsystemet/om-elmarknaden/balansmarknaden/.

[7] Ciaran O'Connor et al. "A Review of Electricity Price Forecasting Models in the Day-Ahead, Intra-Day, and Balancing Markets". In: *Energies* 18.12 (2025), p. 3097. DOI: 10.3390/en18123097. URL: https://www.mdpi.com/1996-1073/18/12/3097.

[8] European Commission. *Electricity market design*. Accessed: 2026-06-01. 2026. URL: https://energy.ec.europa.eu/topics/markets-and-consumers/electricity-market-design_en.

[9] ENTSO-E. *About ENTSO-E*. Accessed: 2026-06-01. 2026. URL: https://www.entsoe.eu/about/.

[10] ACER. *About ACER*. Accessed: 2026-06-01. 2026. URL: https://www.acer.europa.eu/the-agency/about-acer.

[11] Svenska kraftnät. *Om kraftsystemet*. Accessed: 2026-01-19. 2025. URL: https://www.svk.se/om-kraftsystemet/.

[12] Svenska kraftnät. *EU:s inre elmarknad*. Accessed: 2026-05-01. 2025. URL: https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/eus-inre-elmarknad/.

[13] Svenska kraftnät. *Om transmissionsnätet*. Accessed: 2026-01-19. 2024. URL: https://www.svk.se/om-kraftsystemet/om-transmissionsnatet/.

[14] Svenska kraftnät. *Roller i kraftsystemet*. Accessed: 2026-01-25. n.d. URL: https://www.svk.se/om-kraftsystemet/oversikt-av-kraftsystemet/roller-i-kraftsystemet/.

[15] Acer. *Market rules for different electricity market timeframes.* https://www.acer.europa.eu/electricity/market-rules/market-rules-different-electricity-market-timeframes. Accessed: 2026-05-01.

[16] Svenska Kraftnät. *Elområden.* https://www.svk.se/om-kraftsystemet/om-elmarknaden/elomraden/. Accessed: 2026-02-10. 2025.

[17] Svenska Kraftnät. *Elområden och prisskillnader.* https://www.svk.se/om-kraftsystemet/om-elmarknaden/elomraden/elomraden-och-prisskillnader/. Accessed: 2026-02-10. 2024.

[18] Energimarknadsinspektionen. *Elmarknaden.* https://ei.se/konsument/el/elmarknaden. Accessed: 2026-05-01.

[19] Svenska Kraftnät. *Förhandsmarknaden – prissäkringsmarknad.* Accessed: 2026-02-09. 2023. URL: https://www.svk.se/om-kraftsystemet/om-elmarknaden/forhandsmarknaden--prissakringsmarknad/.

[20] Svenska Kraftnät. *Dagen före-marknaden – fysisk handel med el.* https://www.svk.se/om-kraftsystemet/om-elmarknaden/dagen-fore-marknaden--fysisk-handel-med-el/. Accessed: 2026-02-09. 2025.

[21] Nord Pool. *Nord Pool (homepage).* https://www.nordpoolgroup.com/en/. Accessed: 2026-02-09. 2026.

[22] EPEX SPOT. *EPEX SPOT (homepage).* https://www.epexspot.com/en. Accessed: 2026-02-09. 2026.

[23] Nord Pool. *Day-ahead market.* Accessed: 2026-02-09. URL: https://www.nordpoolgroup.com/en/the-power-market/Day-ahead-market/.

[24] EPEX SPOT. *Basics of the Power Market.* https://www.epexspot.com/en/basicspowermarket. Accessed: 2026-02-09.

[25] Svenska kraftnät. *Intradagsmarknaden – justering av dagen före-handel.* https://www.svk.se/om-kraftsystemet/om-elmarknaden/intradagsmarknaden--justering-av-dagen-fore-handel/. Accessed: 2026-02-09. 2025.

[26] Nord Pool. *Intraday market.* https://www.nordpoolgroup.com/en/the-power-market/Intraday-market/. Accessed: 2026-02-09.

[27] Svenska kraftnät. *Balansmarknaden.* https://www.svk.se/om-kraftsystemet/om-elmarknaden/balansmarknaden/. Accessed: 2026-03-16. 2026.

[28] Svenska kraftnät. *Om olika reserver.* https://www.svk.se/aktorsportalen/bidra-med-reserver/om-olika-reserver/. Accessed: 2026-03-16. 2025.

[29] Svenska kraftnät. *Stödtjänster och avhjälpande åtgärder.* https://www.svk.se/om-kraftsystemet/om-systemansvaret/verktyg-for-systemdrift/stodtjanster-och-avhjalpande-atgarder/. Accessed: 2026-03-16.

[30] Svenska kraftnät. *Manuell frekvensåterställningsreserv (mFRR).* Accessed: 2026-01-19. 2025. URL: https://www.svk.se/aktorsportalen/bidra-med-reserver/om-olika-reserver/mfrr/.

[31] Svenska kraftnät. *Automatiserad nordisk energiaktiveringsmarknad för mFRR.* 2024. URL: https://www.svk.se/utveckling-av-kraftsystemet/systemansvar--elmarknad/ny-nordisk-balanseringsmodell-nbm/automatiserad-nordisk-energiaktiveringsmarknad-for-mfrr/.

[32] George Hyndman Rob J Athanasopoulos. *Forecasting: Principles and Practice.* English. 3rd. Australia: OTexts, 2021. URL: https://otexts.com/fpp3/.

[33] G.E.P. Box and G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. 1st ed. Holden-Day series in time series analysis and digital processing. Holden-Day, 1970. ISBN: 9780816210947. URL: https://books.google.se/books?id=5BVfnXaq03oC.

[34] G.E.P. Box et al. *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN: 9781118674925. URL: https://books.google.se/books?id=rNt5CgAAQBAJ.

[35] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323 (1986), pp. 533–536. DOI: 10.1038/323533a0.

[36] Geoffrey E Hinton. "Connectionist learning procedures". In: *Artificial intelligence* 40.1-3 (1989), pp. 185–234. DOI: 10.1016/0004-3702(89)90049-0. URL: https://api.semanticscholar.org/CorpusID:7840452.

[37] Jeffrey L. Elman. "Finding structure in time". In: *Cognitive Science* 14.2 (1990), pp. 179–211. DOI: 10.1016/0364-0213(90)90002-E.

[38] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult". In: *IEEE Transactions on Neural Networks* 5.2 (1994), pp. 157–166. DOI: 10.1109/72.279181. URL: https://ieeexplore.ieee.org/document/279181.

[39] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735. URL: https://ieeexplore.ieee.org/abstract/document/6795963.

[40] Leo Breiman. "Random forests". In: *Machine Learning* 45 (2001), pp. 5–32. DOI: 10.1023/A:1010933404324. URL: https://link.springer.com/article/10.1023/A:1010933404324.

[41] Isabelle M Guyon and André Elisseeff. "An Introduction to Variable and Feature Selection". In: *J. Mach. Learn. Res.* 3 (2003), pp. 1157–1182. URL: https://api.semanticscholar.org/CorpusID:379259.

[42] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv preprint arXiv:1412.6980* (2014). URL: https://arxiv.org/abs/1412.6980.

[43] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: https://www.deeplearningbook.org/.

[44] Montel. *Montel Syspower*. https://montel.energy/platforms/montel-syspower. Accessed: 2026-02-02.

[45] Svenska Kraftnät. *Mimer*. https://mimer.svk.se/Start/. Accessed: 2026-02-02.

[46] Fingrid. *Findgrid Open Data*. https://data.fingrid.fi/en. Accessed: 2026-02-02.

A

Appendix

A.1 Feature Selection (results)

A.1.1 1) Correlation Selection/Filtering – Top 25 features per market (CM and EAM)

A.1.1.1 mFRR CM SE3 – Top 25 correlation feature set

Table A.1: Top 25 correlated features for mFRR CM Up/Down Price/Volume SE3.

Feature	mFRR CM Mean Abs. Corr.	mFRR CM Up Price (corr)	mFRR CM Up Volume (corr)	mFRR CM Down Price (corr)	mFRR CM Down Volume (corr)	count
Month of Year (sin) (E.F) [-]	**0.370**	0.185	-0.533	0.542	0.218	4
Week of Year (sin) (E.F) [-]	**0.313**	0.098	-0.558	0.395	0.199	4
mFRR CM Down Price SE3 [EUR/MW]	**0.266**	0.325	-0.303	NaN	0.436	3
Production Forecast Error SE3 (E.F) [MW]	**0.246**	0.196	-0.317	0.203	0.269	4
Total Mean Temperature Forecast Sweden (E.F) [CEL]	**0.237**	0.219	0.224	0.412	0.091	4
mFRR CM Down Volume SE3 [MW]	**0.234**	0.218	-0.282	0.436	NaN	3
Total Wind Production Forecast Sweden [MW]	**0.222**	-0.372	-0.203	-0.125	-0.188	4
Temperature Forecast for Göteborg [CEL]	**0.221**	0.206	0.206	0.394	0.079	4
Total Production Forecast Sweden [MW]	**0.218**	-0.198	-0.204	-0.419	-0.052	4
Spot Volume Buy Forecast for SE3 [MW]	**0.208**	-0.092	-0.153	-0.489	-0.098	4
Total Spot Volume Sell Forecast Sweden [MW]	**0.208**	-0.176	-0.165	-0.429	-0.061	4
Total Spot Volume Buy Forecast Sweden [MW]	**0.204**	-0.091	-0.152	-0.488	-0.087	4
Spot Price Forecast SE3 [EUR/MWh]	**0.199**	0.336	0.109	-0.312	0.039	4
Production Forecast SE3 [MW]	**0.195**	-0.249	-0.244	-0.286	0.001	4
NP Area Spot Systemprice Forecast [EUR/MWh]	**0.194**	0.264	0.130	-0.365	0.016	4
Total Solar Forecast (long-term) Sweden [MW]	**0.193**	0.215	-0.072	0.350	0.136	4
Spot Price SE3 [EUR/MWh]	**0.191**	0.351	0.062	-0.295	0.057	4
Solar Forecast (long-term) for SE3 [MW]	**0.190**	0.212	-0.073	0.341	0.133	4
Spot Volume Sell Forecast for SE3 [MW]	**0.186**	-0.239	-0.215	-0.275	-0.015	4
Total Consumption Forecast Sweden [MW]	**0.182**	-0.056	-0.204	-0.409	-0.059	4
Renewable (Wind/Solar) Share Frcst. Sweden (E.F) [%]	**0.181**	-0.318	-0.165	0.083	-0.157	4
Consumption Forecast SE3 [MW]	**0.173**	-0.042	-0.197	-0.393	-0.060	4
mFRR EAM Up Price SE3 [EUR/MW]	**0.172**	0.326	0.072	-0.252	0.039	4
Inertia (kinetic energy) - Nordic power system [GWs]	**0.171**	-0.067	-0.118	-0.477	0.024	4
Wind Production Forecast SE3 [MW]	**0.170**	-0.345	-0.075	-0.085	-0.175	4

A.1.1.2 mFRR EAM SE3 – Top 25 correlation feature set

Table A.2: Top 25 correlated features for mFRR EAM Up/Down Price/Volume SE3.

Feature	**mFRR EAM Mean Abs. Corr.**	mFRR EAM Up Price (corr)	mFRR EAM Up Volume (corr)	mFRR EAM Down Price (corr)	mFRR EAM Down Volume (corr)	count
Spot Price SE3 [EUR/MWh]	**0.411**	0.872	0.082	0.668	-0.024	4
mFRR EAM Down Price SE3 [EUR/MW]	**0.411**	0.723	0.378	NaN	0.545	3
Spot Price Forecast SE3 [EUR/MWh]	**0.372**	0.782	0.073	0.608	-0.026	4
NP Area Spot Systemprice Forecast [EUR/MWh]	**0.363**	0.728	0.111	0.610	0.005	4
Total Scheduled Intraday In Flow Statistics SE3 (E.F.) [MWh]	**0.335**	0.600	0.120	0.554	0.066	4
mFRR EAM Up Price SE3 [EUR/MW]	**0.332**	NaN	0.423	0.723	0.182	3
mFRR EAM Up Volume SE3 [MW]	**0.316**	0.423	NaN	0.378	0.464	3
Inertia (kinetic energy) - Nordic power system [GWs]	**0.303**	0.519	0.119	0.528	0.046	4
Total Netto Scheduled Intraday Flow Statistics SE3 (E.F.) [MWh]	**0.300**	0.514	0.114	0.468	0.103	4
mFRR EAM Down Volume SE3 [MW]	**0.298**	0.182	0.464	0.545	NaN	3
Renewable (Wind/Solar) Share Forecast Sweden (E.F.) [%]	**0.293**	-0.515	-0.077	-0.479	-0.101	4
Exchange Forecast SE3 [MW]	**0.291**	0.513	0.073	0.496	0.082	4
Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	**0.257**	0.544	0.052	0.369	-0.062	4
Spot Price SE3 RZD-1d (E.F.) [-]	**0.241**	0.520	0.046	0.377	-0.023	4
Spot Price Forecast Spread SE1234 (Max-Min) (E.F.) [EUR/MWh]	**0.235**	0.505	0.051	0.336	-0.048	4
Total Spot Volume Buy Forecast Sweden [MW]	**0.225**	0.412	0.077	0.411	NaN	3
Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	**0.225**	-0.443	-0.047	-0.362	-0.048	4
Spot Volume Buy Forecast for SE3 [MW]	**0.222**	0.405	0.077	0.407	0.000	4
Time Deviation - Nordic power system [s]	**0.205**	-0.139	-0.238	-0.206	-0.236	4
Spot Volume Buy for SE3 RZD-1d (E.F.) [-]	**0.199**	0.404	0.089	0.296	0.007	4
Wind Velocity Forecast for Göteborg [m/sec]	**0.197**	-0.391	-0.049	-0.307	-0.040	4
mFRR CM Down Price SE3 [EUR/MW]	**0.196**	-0.252	-0.132	-0.337	-0.061	4
Total Consumption Forecast Sweden [MW]	**0.196**	0.357	0.060	0.361	0.005	4
Wind Production Forecast SE3 [MW]	**0.194**	-0.380	-0.014	-0.313	-0.068	4
Consumption Forecast SE3 [MW]	**0.193**	0.355	0.055	0.357	0.006	4

A.1.2 2) Feature Importance Filtering (Random Forest) – Top 25 features per market (CM and EAM)

While correlation analysis captures linear explanatory relationships, it does not account for nonlinear interactions or feature redundancy. To further assess predictive relevance under nonlinear model assumptions, Random Forest-based feature importance filtering was subsequently applied.

A.1.2.1 mFRR CM SE3 – Top 25 importance feature set

Table A.3: Top 25 RF-importance features for mFRR CM Up/Down Price/Volume SE3.

Feature	**Total Mean Importance for mFRR CM**	mFRR CM Up Price (mean imp.)	mFRR CM Up Volume (mean imp.)	mFRR CM Down Price (mean imp.)	mFRR CM Down Volume (mean imp.)	count
mFRR CM Up Price SE3 [EUR/MW]	**0.079**	NaN	0.211	0.057	0.050	3
Week of Year (sin) (E.F.) [-]	**0.079**	0.061	0.089	0.028	0.138	4
mFRR CM Down Price SE3 [EUR/MW]	**0.065**	0.063	0.039	NaN	0.159	3
Month of Year (sin) (E.F.) [-]	**0.050**	0.004	0.172	0.003	0.021	4
NP Area Spot Systemprice Forecast [EUR/MWh]	**0.039**	0.141	0.004	0.004	0.006	4
Exchange Forecast SE3 [MW]	**0.036**	0.006	0.119	0.012	0.008	4
mFRR CM Down Volume SE3 [MW]	**0.035**	0.020	0.017	0.102	NaN	3
Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	**0.032**	0.100	0.007	0.004	0.017	4
Inertia (kinetic energy) - Nordic power system [GWs]	**0.031**	0.015	0.006	0.038	0.065	4
Total Mean Cloud Cover Forecast Sweden (E.F.) [%]	**0.030**	0.011	0.011	0.088	0.010	4
Total Scheduled Intraday Out Flow Statistics SE3 (E.F.) [MWh]	**0.028**	0.057	0.007	0.037	0.012	4
Day of Week (sin) (E.F.) [-]	**0.025**	0.014	0.068	NaN	0.020	3
Production Forecast Error SE3 (E.F.) [MW]	**0.025**	0.058	0.008	0.011	0.024	4
Production Forecast SE3 [MW]	**0.022**	0.007	0.009	0.004	0.069	4
Spot Price Forecast Spread (SE3-SE4) (E.F.) [EUR/MWh]	**0.021**	0.068	0.004	0.005	0.007	4
Total Mean Temperature Forecast Sweden (E.F.) [CEL]	**0.021**	0.011	0.046	0.014	0.011	4
Spot Price Forecast Spread SE1234 (Max-Min) (E.F.) [EUR/MWh]	**0.020**	0.009	0.011	0.054	0.004	4
mFRR CM Up Volume SE3 [MW]	**0.019**	0.032	NaN	0.008	0.036	3
Exchange Forecast Error SE3 (E.F.) [MW]	**0.019**	0.011	0.006	0.050	0.008	4
Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	**0.017**	0.005	0.006	0.053	0.005	4
Spot Price SE3 [EUR/MWh]	**0.015**	0.026	0.003	0.020	0.011	4
Total Consumption Forecast Sweden [MW]	**0.015**	0.018	0.003	0.031	0.006	4
Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	**0.015**	0.011	0.005	0.033	0.010	4
Temperature Forecast for Göteborg [CEL]	**0.014**	0.006	0.008	0.024	0.020	4
Total Wind Production Forecast Sweden [MW]	**0.014**	0.010	0.009	0.012	0.026	4

A.1.2.2 mFRR EAM SE3 – Top 25 importance feature set

Table A.4: Top 25 RF-importance features for mFRR EAM Up/Down Price/Volume SE3.

Feature	Total Mean Importance for mFRR EAM	mFRR EAM Up Price (mean importance)	mFRR EAM Up Volume (mean importance)	mFRR EAM Down Price (mean importance)	mFRR EAM Down Volume (mean importance)	count
mFRR EAM Up Price SE3 [EUR/MW]	**0.121**	NaN	0.460	0.013	0.010	3
mFRR EAM Up Volume SE3 [MW]	**0.119**	0.461	NaN	NaN	0.014	2
Inertia (kinetic energy) - Nordic power system [GWs]	**0.082**	0.018	0.032	0.155	0.121	4
mFRR EAM Down Price SE3 [EUR/MW]	**0.074**	0.018	0.006	NaN	0.272	3
mFRR EAM Down Volume SE3 [MW]	**0.059**	0.017	0.029	0.188	NaN	3
Spot Price SE3 [EUR/MWh]	**0.055**	0.070	0.072	0.014	0.063	4
Total Exchange Forecast Sweden [MW]	**0.040**	0.004	0.009	0.141	0.007	4
Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	**0.035**	0.005	0.007	0.120	0.008	4
Spot Price Forecast SE3 [EUR/MWh]	**0.021**	0.004	0.015	0.044	0.022	4
Week of Year (sin) (E.F.) [-]	**0.017**	0.044	0.010	0.003	0.012	4
mFRR CM Down Volume SE3 [MW]	**0.016**	0.010	0.018	0.022	0.015	4
Total Wind Production Forecast Sweden [MW]	**0.016**	0.008	0.041	0.006	0.007	4
Wind Velocity Forecast for Göteborg [m/sec]	**0.014**	0.005	0.005	0.037	0.010	4
Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	**0.012**	0.005	0.012	0.015	0.016	4
Exchange Forecast SE3 [MW]	**0.012**	0.005	0.008	0.002	0.032	4
Wind Production Forecast SE3 [MW]	**0.012**	0.010	0.010	0.017	0.009	4
Production Forecast Error SE3 (E.F.) [MW]	**0.011**	0.009	0.008	0.018	0.009	4
Spot Volume Sell for SE3 RZD-1d (E.F.) [-]	**0.011**	0.015	0.012	0.006	0.011	4
mFRR CM Down Price SE3 [EUR/MW]	**0.011**	0.016	0.004	0.003	0.021	4
Month of Year (sin) (E.F.) [-]	**0.010**	0.039	NaN	0.001	NaN	2
mFRR CM Up Volume SE3 [MW]	**0.010**	0.007	0.018	0.004	0.010	4
Spot Volume Sell Forecast for SE3 [MW]	**0.010**	0.016	0.004	0.010	0.008	4
Spot Price SE3 RZD-1d (E.F.) [-]	**0.009**	0.013	0.008	0.005	0.013	4
Precipitation Energy Forecast (AROME00) SE3 [GWh]	**0.009**	NaN	NaN	0.036	NaN	1
Time Deviation - Nordic power system [s]	**0.009**	0.006	0.006	0.011	0.013	4

A.1.3 3) Forward Selection (LSTM) – Top 10 forward-selected features per target

A.1.3.1 mFRR CM Up/Down Price/Volume SE3

Table A.5: mFRR CM Up Price SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR CM Up Price SE3 [EUR/MW]	50.45	23.53	58.22
2	Production Forecast SE3 [MW]	49.13	24.29	57.15
3	Spot Price Forecast Spread SE1234 (Max-Min) (E.F.) [EUR/MWh]	48.81	22.92	56.38
4	mFRR CM Down Volume SE3 [MW]	48.68	23.05	56.29
5	Total Mean Temperature Forecast Sweden (E.F.) [CEL]	48.21	22.61	55.67
6	Inertia (kinetic energy) - Nordic power system [GWs]	48.29	22.25	55.64
7	Spot Price SE3 [EUR/MWh]	47.73	22.28	55.08
8	Total Consumption Forecast Sweden [MW]	48.01	22.24	55.35
9	Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	47.93	21.65	55.07
10	Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	47.77	21.51	54.87

Table A.6: mFRR CM Up Volume SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR CM Up Volume SE3 [MW]	63.73	9.49	66.87
2	Exchange Forecast Error SE3 (E.F.) [MW]	63.23	8.95	66.18
3	NP Area Spot Systemprice Forecast [EUR/MWh]	62.65	9.41	65.75
4	mFRR CM Up Price SE3 [EUR/MW]	62.23	9.31	65.30
5	Total Scheduled Intraday Out Flow Statistics SE3 (E.F.) [MWh]	61.45	10.51	64.92
6	Spot Price SE3 [EUR/MWh]	61.59	10.17	64.95
7	Total Consumption Forecast Sweden [MW]	61.79	9.00	64.76
8	Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	63.29	9.48	66.42
9	Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	63.68	10.30	67.08
10	Production Forecast Error SE3 (E.F.) [MW]	64.48	10.57	67.97

Table A.7: mFRR CM Down Price SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR CM Down Price SE3 [EUR/MW]	19.42	8.53	22.24
2	Total Wind Production Forecast Sweden [MW]	18.85	8.34	21.60
3	Total Consumption Forecast Sweden [MW]	18.40	8.24	21.12
4	Inertia (kinetic energy) - Nordic power system [GWs]	18.43	8.12	21.11
5	Exchange Forecast Error SE3 (E.F.) [MW]	18.34	8.24	21.06
6	Spot Price SE3 [EUR/MWh]	18.38	8.21	21.09
7	Spot Price Forecast Spread (SE3-SE4) (E.F.) [EUR/MWh]	18.30	8.39	21.07
8	Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	18.36	8.31	21.10
9	Spot Price Forecast Spread SE1234 (Max-Min) (E.F.) [EUR/MWh]	18.29	8.42	21.07
10	Spot Price Forecast Spread (SE3-SE2) (E.F.) [EUR/MWh]	18.31	8.44	21.10

Table A.8: mFRR CM Down Volume SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR CM Down Volume SE3 [MW]	49.62	8.41	52.40
2	Total Consumption Forecast Sweden [MW]	47.43	8.67	50.29
3	Spot Price SE3 [EUR/MWh]	46.08	9.09	49.08
4	Total Mean Cloud Cover Forecast Sweden (E.F.) [%]	46.80	8.46	49.59
5	Day of Week (sin) (E.F.) [-]	45.85	7.68	48.38
6	Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	45.90	7.62	48.42
7	Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	45.55	7.90	48.16
8	Total Wind Production Forecast Sweden [MW]	44.76	6.71	46.98
9	mFRR CM Up Price SE3 [EUR/MW]	44.74	7.28	47.14
10	NP Area Spot Systemprice Forecast [EUR/MWh]	45.00	7.94	47.62

A.1.3.2 mFRR EAM Up/Down Price/Volume SE3

Table A.9: mFRR EAM Up Price SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	Spot Price SE3 RZD-1d (E.F.) [-]	171.76	62.71	192.45
2	mFRR EAM Up Price SE3 [EUR/MW]	171.94	62.85	192.68
3	mFRR CM Down Volume SE3 [MW]	172.03	62.60	192.69
4	Wind Velocity Forecast for Göteborg [m/sec]	172.47	62.43	193.08
5	mFRR CM Up Volume SE3 [MW]	172.45	62.09	192.94
6	Month of Year (sin) (E.F.) [-]	174.70	59.42	194.30
7	Time Deviation - Nordic power system [s]	174.75	61.55	195.06
8	Spot Price SE3 [EUR/MWh]	173.84	61.28	194.06
9	mFRR EAM Down Price SE3 [EUR/MW]	173.02	62.15	193.53
10	Spot Price Forecast SE3 [EUR/MWh]	174.39	60.78	194.45

Table A.10: mFRR EAM Up Volume SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	Spot Price SE3 RZD-1d (E.F.) [-]	35.23	3.80	36.49
2	mFRR EAM Down Price SE3 [EUR/MW]	35.20	3.91	36.49
3	Wind Velocity Forecast for Göteborg [m/sec]	35.15	4.03	36.48
4	Time Deviation - Nordic power system [s]	35.18	4.05	36.52
5	mFRR EAM Up Volume SE3 [MW]	35.25	3.88	36.54
6	mFRR EAM Up Price SE3 [EUR/MW]	35.12	3.98	36.43
7	Spot Price Forecast SE3 [EUR/MWh]	35.14	4.03	36.47
8	Spot Price SE3 [EUR/MWh]	35.14	4.04	36.48
9	Spot Price Forecast Spread (SE3-NP) (E.F.) [EUR/MWh]	35.19	3.98	36.51
10	Inertia (kinetic energy) - Nordic power system [GWs]	35.31	3.85	36.58

Table A.11: mFRR EAM Down Price SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR EAM Down Price SE3 [EUR/MW]	156.33	161.96	209.77
2	Total Mean Wind Velocity Forecast Sweden (E.F.) [m/sec]	155.60	161.91	209.03
3	Inertia (kinetic energy) - Nordic power system [GWs]	155.53	161.75	208.91
4	Total Wind Production Forecast Sweden [MW]	154.98	162.14	208.49
5	mFRR EAM Up Price SE3 [EUR/MW]	154.99	162.29	208.54
6	Wind Production Forecast SE3 [MW]	155.10	162.12	208.60
7	Wind Velocity Forecast for Göteborg [m/sec]	155.36	161.99	208.81
8	mFRR EAM Up Volume SE3 [MW]	155.57	161.95	209.01
9	Total Exchange Forecast Sweden [MW]	155.95	161.48	209.23
10	Spot Volume Sell for SE3 RZD-1d (E.F.) [-]	156.07	161.62	209.40

Table A.12: mFRR EAM Down Volume SE3 – top 10 forward-selected features using LSTM-based forward selection with rolling time-series cross-validation.

Step	Selected Feature	RMSE Mean	RMSE Std	Score
1	mFRR EAM Down Volume SE3 [MW]	33.65	6.09	35.66
2	Spot Price SE3 [EUR/MWh]	33.35	5.87	35.28
3	Spot Price SE3 RZD-1d (E.F.) [-]	33.29	5.78	35.20
4	Wind Production Forecast SE3 [MW]	33.24	5.58	35.08
5	Spot Volume Sell for SE3 RZD-1d (E.F.) [-]	33.22	5.59	35.07
6	Wind Velocity Forecast for Göteborg [m/sec]	33.18	5.51	35.00
7	mFRR EAM Up Price SE3 [EUR/MW]	33.20	5.46	35.00
8	mFRR EAM Up Volume SE3 [MW]	33.18	5.37	34.95
9	mFRR EAM Down Price SE3 [EUR/MW]	33.21	5.46	35.01
10	Total Wind Production Forecast Sweden [MW]	33.26	5.46	35.06

A.2 Model Evaluation and Comparison

A.2.1 mFRR CM Up Price SE3

For CM Up Price (see Table A.13), machine learning models generally outperform the naive baselines, although the differences between the top models are relatively small. LSTM models perform best overall, with LSTM (fs10) (RMSE=33.96), LSTM (uv) (34.19), and LSTM (mv) (34.43) occupying the top ranks, while MLP performance is more unstable across feature sets. The strong performance of Naive Seasonal-Mean-1w (RMSE=34.48), and the reasonable performance of Naive Mean-1w (36.13), indicates that the series contains clear weekly and intraday persistence patterns, while the poor performance of pure seasonal baselines suggests that averaged historical behavior is more informative than strict daily or weekly repetition. Overall, the relatively small performance differences between

Table A.13: Forecasting performance for mFRR CM Up Price SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	Model	RMSE	MAE
1	LSTM (fs10)	33.96	24.38
2	LSTM (uv)	34.19	25.78
3	LSTM (mv)	34.43	25.80
4	Naive Seasonal-Mean-1w	34.48	24.00
5	MLP (uv)	34.71	27.02
6	MLP (fs10)	35.10	25.80
7	Naive Mean-1w	36.13	26.55
8	LSTM (fs5)	37.25	25.38
9	Naive Mean-1d	38.82	25.73
10	MLP (mv)	39.39	28.26
11	Naive Seasonal-1d	42.53	25.20
12	Naive Seasonal-1w	46.07	29.72
13	MLP (fs5)	53.56	35.72

the top models suggest that CM Up Price contains both predictable seasonal structure and some nonlinear dynamics, allowing machine learning models to achieve modest improvements over the strongest baselines.

A.2.2 mFRR CM Up Volume SE3

For CM Up Volume (see Table A.14), the results are dominated by naive baseline models, with Naive Seasonal-Mean-1w (RMSE=72.90) and Naive Mean-1w (76.33) clearly outperforming most other models. Among the ML models, only MLP (uv) (82.28) remains moderately competitive, while more flexible models with larger feature sets and multivariate inputs perform substantially worse. In this case, the additional model flexibility appears to introduce overfitting rather than improved predictive performance. This suggests that CM Up Volume is highly stable and repetitive, with limited nonlinear structure for neural networks to exploit.

The prediction comparison of the top models from each model class is shown in Figure A.1.

Table A.14: Forecasting performance for mFRR CM Up Volume SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	Model	RMSE	MAE
1	Naive Seasonal-Mean-1w	72.90	56.85
2	Naive Mean-1w	76.33	59.12
3	MLP (uv)	82.28	67.30
4	Naive Mean-1d	84.16	65.07
5	Naive Seasonal-1d	87.90	67.95
6	Naive Seasonal-1w	94.06	73.96
7	MLP (fs5)	96.08	78.50
8	LSTM (fs5)	97.44	82.23
9	LSTM (uv)	101.46	85.39
10	LSTM (fs10)	126.76	106.64
11	LSTM (mv)	144.34	110.77
12	MLP (mv)	160.00	137.07
13	MLP (fs10)	189.17	149.33

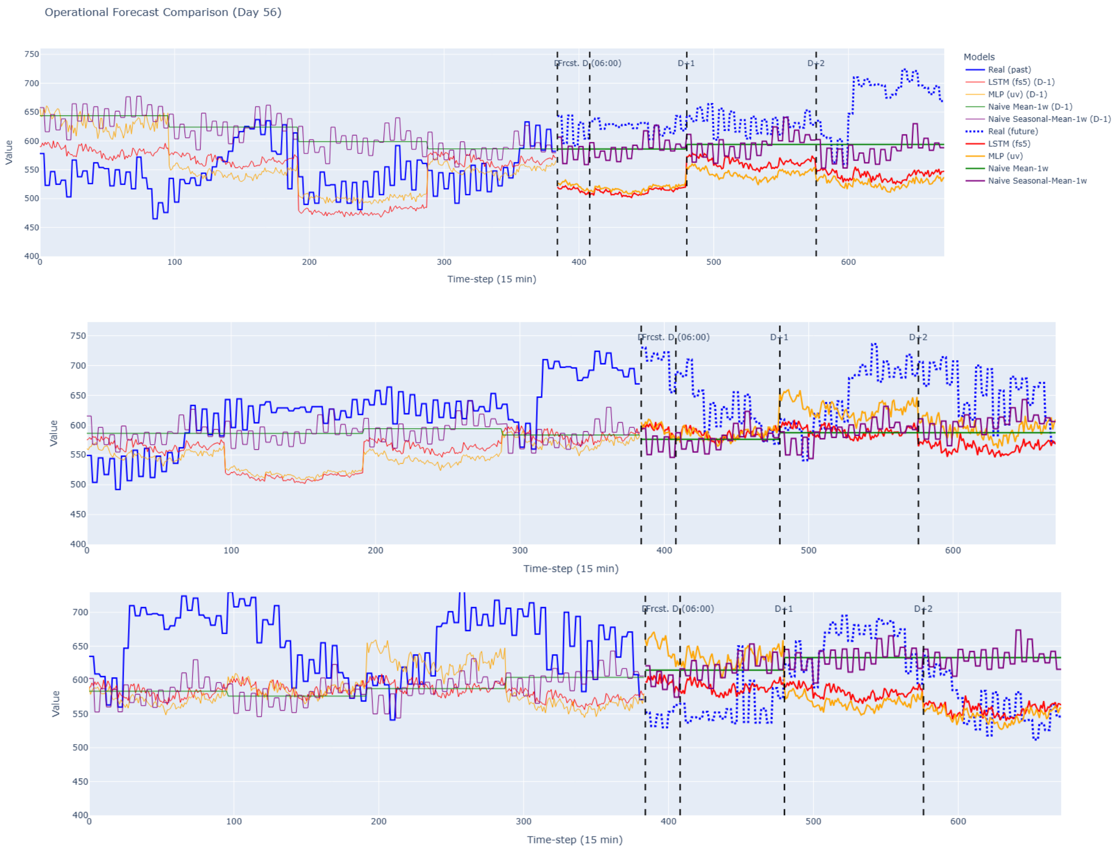

Figure A.1: Forecast comparison for the top-performing models from each model class for mFRR CM Up Volume SE3: 1) Naive Seasonal-Mean-1w (purple), Naive Mean-1w (green), 3) MLP uv (yellow), 4) LSTM fs5 (red), showing predicted vs. true target (blue) values, and differences in how the models capture variability, level shifts, and short-term dynamics in the series.

Overall, the results suggest that CM Up Volume is driven by stable recurring patterns that are captured effectively by simple averaging-based baselines, while more complex machine learning models tend to overfit and perform worse.

A.2.3 mFRR CM Down Price SE3

Table A.15: Forecasting performance for mFRR CM Down Price SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	Model	RMSE	MAE
1	Naive Mean-1w	3.87	2.22
2	Naive Seasonal-Mean-1w	3.96	2.24
3	Naive Mean-1d	4.26	2.23
4	Naive Seasonal-1d	5.10	2.49
5	Naive Seasonal-1w	5.25	2.89
6	MLP (uv)	5.30	4.01
7	LSTM (mv)	6.79	5.05
8	LSTM (uv)	7.65	6.65
9	LSTM (fs10)	10.95	7.53
10	MLP (fs5)	11.19	9.49
11	MLP (fs10)	11.93	10.01
12	LSTM (fs5)	12.46	11.49
13	MLP (mv)	22.50	9.03

A.2.4 mFRR CM Down Volume SE3

Table A.16: Forecasting performance for mFRR CM Down Volume SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	Model	RMSE	MAE
1	LSTM (fs10)	52.68	39.74
2	Naive Mean-1w	53.33	37.77
3	Naive Seasonal-Mean-1w	53.77	38.70
4	LSTM (fs5)	53.95	41.78
5	LSTM (uv)	54.20	41.97
6	MLP (mv)	54.77	40.27
7	MLP (uv)	55.25	41.65
8	MLP (fs5)	56.34	44.14
9	LSTM (mv)	56.39	44.83
10	MLP (fs10)	59.44	47.10
11	Naive Mean-1d	60.91	42.83
12	Naive Seasonal-1d	69.12	49.11
13	Naive Seasonal-1w	72.85	53.49

A.2.5 mFRR EAM Up Price SE3

For EAM Up Price (see Table A.17), machine learning models clearly outperform all naive baseline methods. The clearly best performance is achieved by MLP (fs10) with an RMSE of 108.75, followed by LSTM (fs10) (111.27) and MLP (mv) (111.35). Both LSTM and MLP models benefit from external input features, particularly forward-selected feature sets, while purely autoregressive (uv) models perform noticeably worse. Among the baselines, Naive Mean-1d (119.18) performs best, while pure seasonal baselines perform significantly worse.

Table A.17: Forecasting performance for mFRR EAM Up Price SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	Model	RMSE	MAE
1	MLP (fs10)	108.75	37.99
2	LSTM (fs10)	111.27	45.27
3	MLP (mv)	111.35	39.93
4	MLP (fs5)	113.31	44.77
5	LSTM (fs5)	113.64	47.02
6	LSTM (uv)	116.11	50.81
7	MLP (uv)	116.82	49.80
8	Naive Mean-1d	119.18	53.85
9	LSTM (mv)	119.61	51.92
10	Naive Mean-1w	120.61	58.47
11	Naive Seasonal-Mean-1w	123.85	59.84
12	Naive Seasonal-1d	158.46	63.76
13	Naive Seasonal-1w	159.53	70.01

The prediction comparison of the top models from each model class is shown in Figure A.2.

Overall, the results suggests that EAM Up Price contains complex nonlinear relationships and strong dependence on external explanatory variables that are not sufficiently captured by simple daily or weekly mean or seasonal baseline models.

A.2.6 mFRR EAM Up Volume SE3

For EAM Up Volume (see Table A.18), machine learning models also outperform all naive baseline methods. LSTM (uv) performs best with an RMSE of 40.89, closely followed by MLP (uv) (41.20). In contrast to the EAM price targets, autoregressive models consistently perform best, while fs5, fs10, and multivariate (mv) models perform relatively similarly at slightly worse levels. The relatively strong performance of Naive Mean-1d (RMSE=42.53) and Naive Mean-1w (42.63) indicates that daily and weekly average persistence remain highly informative for EAM Up Volume, while pure seasonal repetition models again perform clearly worse.

Similar to the EAM Down Volume case, the results suggest that the machine learning models primarily learn a smoothed average activation level rather than the true zero-versus-activation dynamics of the series. Since the data is dominated by zero activation periods, with positive activation spikes occurring less frequently, the models tend to converge toward local minima that minimize the training loss by predicting values close to the

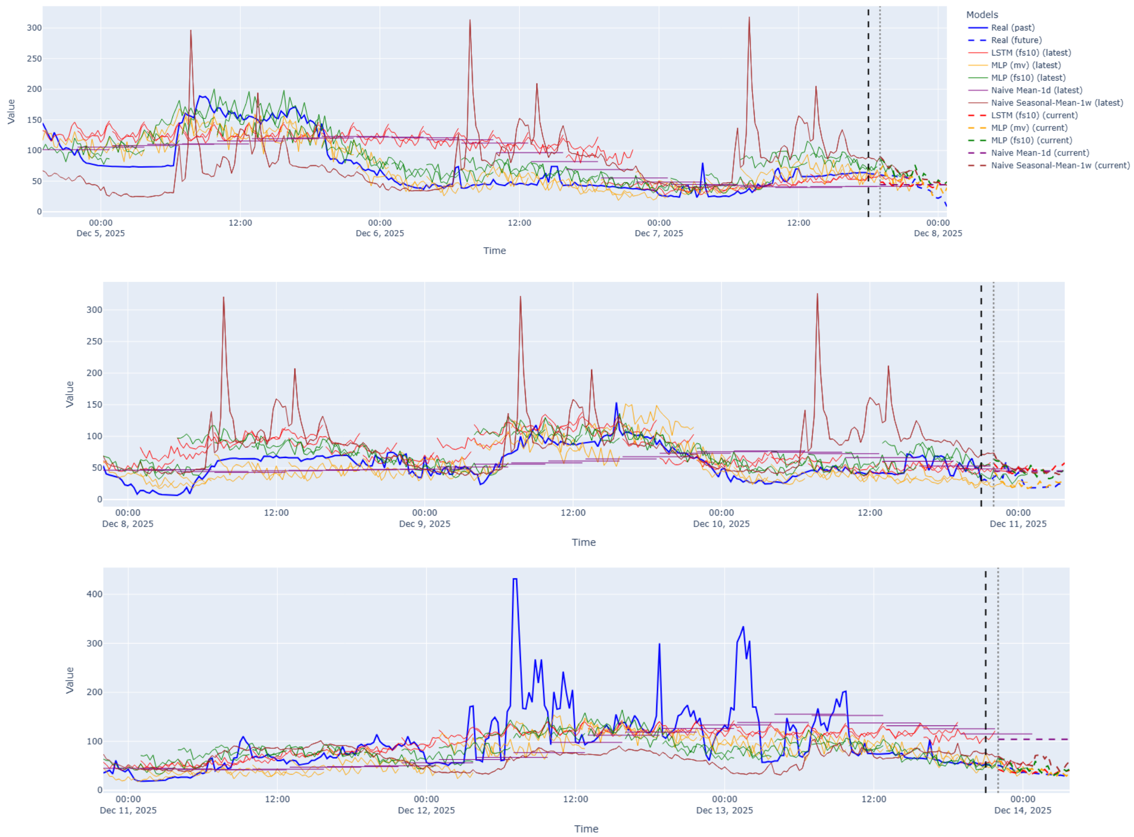

Figure A.2: Forecast comparison for the top-performing models from each model class for mFRR EAM Up Price SE3: 1) MLP fs10 (green), 2) LSTM fs10 (red), 3) MLP mv (yellow), 4) Naive Mean-1d (purple), 5) Naive Seasonal-Mean-1w (brown), showing predicted vs. true target (blue) values, and differences in how the models capture variability, level shifts, and short-term dynamics in the series.

moving average level of the series. As a result, the models struggle to correctly capture both inactive periods (0 MW) and the timing and magnitude of activation spikes.

Table A.18: Forecasting performance for mFRR EAM Up Volume SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	**Model**	**RMSE**	**MAE**
1	LSTM (uv)	40.89	22.04
2	MLP (uv)	41.20	21.73
3	MLP (fs10)	42.03	21.33
4	LSTM (fs5)	42.04	22.34
5	MLP (mv)	42.20	25.56
6	MLP (fs5)	42.39	22.14
7	Naive Mean-1d	42.53	23.57
8	Naive Mean-1w	42.63	25.80
9	LSTM (mv)	42.77	23.01
10	LSTM (fs10)	43.50	22.48
11	Naive Seasonal-Mean-1w	44.68	26.59
12	Naive Seasonal-1d	55.84	27.17
13	Naive Seasonal-1w	57.38	29.46

A.2.7 mFRR EAM Down Price SE3

Table A.19: Forecasting performance for mFRR EAM Down Price SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	**Model**	**RMSE**	**MAE**
1	MLP (mv)	117.84	25.75
2	LSTM (mv)	118.44	27.06
3	MLP (uv)	119.38	28.21
4	MLP (fs10)	119.56	29.40
5	LSTM (uv)	119.69	28.91
6	Naive Mean-1d	120.14	28.96
7	LSTM (fs5)	120.84	31.29
8	LSTM (fs10)	121.04	31.78
9	Naive Mean-1w	121.20	34.01
10	MLP (fs5)	121.45	32.54
11	Naive Seasonal-Mean-1w	127.99	35.12
12	Naive Seasonal-1d	166.59	35.54
13	Naive Seasonal-1w	169.12	41.50

A.2.8 mFRR EAM Down Volume SE3

Table A.20: Forecasting performance for mFRR EAM Down Volume SE3 evaluated across all models (LSTMs, MLPs, and baseline models) and feature sets (univariate, multivariate, top 10 forward-selected features, and top 5 forward-selected features).

Rank	**Model**	**RMSE**	**MAE**
1	LSTM (fs5)	40.51	25.38
2	MLP (fs5)	40.56	24.99
3	LSTM (uv)	40.74	24.49
4	MLP (uv)	40.88	24.96
5	MLP (fs10)	41.48	24.85
6	MLP (mv)	41.65	26.94
7	LSTM (mv)	42.06	26.02
8	Naive Mean-1w	42.24	28.58
9	LSTM (fs10)	42.27	26.46
10	Naive Mean-1d	42.73	26.93
11	Naive Seasonal-Mean-1w	44.52	29.60
12	Naive Seasonal-1d	56.15	31.69
13	Naive Seasonal-1w	59.00	34.58

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

AVANCEZ
1829

CHALMERS
UNIVERSITY OF TECHNOLOGY