Aalto-yliopisto
Aalto-universitetet
Aalto University

Master's Programme in Mathematics and Operations Research

Evaluation of Day-Ahead Electricity Price Forecasts for Multi-Market Optimization

Oskar Heikkilä

Master's Thesis
2026

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Oskar Heikkilä

Title Evaluation of Day-Ahead Electricity Price Forecasts for Multi-Market Optimization

Degree programme Mathematics and Operations Research

Major Systems and Operations Research

Supervisor University Lecturer Pekka Pere

Advisor Prof. Pauliina Ilmonen, MSc Arttu Hämäläinen

Collaborative partner NSC EnergyOpti Oy

Date 30 April 2026 **Number of pages** 46+1 **Language** English

Abstract

The increasing volatility of day-ahead electricity markets and the growing share of renewable energy production have made accurate day-ahead forecasting more important than ever. The transition towards multi-market optimization, where flexible assets operate simultaneously in multiple electricity markets, has introduced new requirements for a good forecast. In this context, classical forecast evaluation metrics may not fully capture the practical usefulness of forecasts for decision-making.

In this thesis, multiple day-ahead electricity price forecasts from different forecast providers are evaluated in the context of multi-market optimization. The analysis is conducted by using Finnish day-ahead electricity price forecasts from five different forecast providers and the realised day-ahead prices. Forecast performance is evaluated using classical error metrics, ranking-based metrics, and association and economic metrics. In addition, forecast performance is analysed under different market conditions.

The results show that while classical prediction error metrics provide useful information on average forecast accuracy, they alone are not sufficient for evaluating forecast quality in the context of multi-market optimization. Ranking-based and association metrics provide better insight how a forecast captures the timing of high-price and low-price periods, and economic metrics capture the economic value of timing the peaks and bottoms, which both are critical for operational decision-making.

The findings highlight the importance of selecting the evaluation metrics that align with the intended use. In multi-market optimization, emphasis should be placed on metrics that reflect the temporal structure and economic impact of forecast errors rather than solely minimizing average deviations. Through these comprehensive analyses, the most suitable forecast can be reliably selected for real-world trading and optimization applications.

Keywords electricity price forecasting, day-ahead market, forecast evaluation, multi-market optimization

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Oskar Heikkilä

Työn nimi Vuorokausimarkkinan sähkönhintaennusteiden vertailua monimarkkinaoptimointiin

Koulutusohjelma Matematiikka ja operaatiotutkimus

Pääaine Systeemianalyysi ja operaatiotutkimus

Työn valvoja Yliopistonlehtori Pekka Pere

Työn ohjaaja Prof. Pauliina Ilmonen, DI Arttu Hämäläinen

Yhteistyötaho NSC EnergyOpti Oy

Päivämäärä 30 April 2026 **Sivumäärä** 46+1 **Kieli** englanti

Tiivistelmä

Sähkön vuorokausimarkkinoiden sähkönhintojen kasvava vaihtelu sekä uusiutuvan energiatuotannon määrän lisääntyminen ovat tehneet tarkasta sähkönhinnan ennustamisesta entistä tärkeämpää. Yhä useammat sähköresurssit hyödyntävät monimarkkinaoptimointia, jossa joustavat resurssit osallistuvat samanaikaisesti useille sähkömarkkinoille. Tämä on asettanut uusia vaatimuksia laadukkaalle ennusteelle ja perinteiset ennusteiden arviointimetriikat eivät välttämättä täysin kuvaa ennusteiden käytännön hyödyllisyyttä päätöksenteossa.

Tässä diplomityössä arvioidaan useita vuorokausimarkkinoiden sähkönhintaennusteita monimarkkinaoptimoinnin näkökulmasta. Analyysi toteutetaan käyttäen viiden eri ennustetoimittajan Suomen vuorokausimarkkinoiden sähkönhintaennusteita sekä toteutuneita sähkönhintoja. Ennusteiden suorituskykyä arvioidaan käyttämällä klassisia virhemittareita, järjestysperusteisia mittareita sekä korrelaatio- ja taloudellisia mittareita. Lisäksi ennusteiden suorituskykyä tarkastellaan erilaisissa markkinatilanteissa.

Tulokset osoittavat, että vaikka klassiset virhemittarit kertovat ennusteiden keskimääräisestä ennustetarkkuudesta, ne eivät yksinomaan riitä kuvaamaan ennusteiden hyödyllisyyttä monimarkkinaoptimointiin. Järjestysperusteiset sekä korrelaatio- ja taloudelliset mittarit kuvaavat paremmin ennusteiden kykyä tunnistaa korkean ja matalan hinnan ajankohdat, mikä on keskeistä operatiivisen päätöksenteon kannalta.

Tulokset korostavat käyttötarkoitukseen soveltuvien arviointimittareiden valinnan merkitystä. Monimarkkinaoptimoinnin näkökulmasta tulisi painottaa mittareita, jotka kuvaavat ennusteiden kykyä ajoittaa taloudellisesti merkittävät hintapiikit ja matalan hinnan jaksot pelkän keskimääräisen virheen minimoinnin sijaan. Näiden kattavien analyysien avulla voidaan perustellusti valita käytännön kaupankäyntiin ja optimointiin soveltuvin ennuste.

Avainsanat sähkönhinnan ennustaminen, vuorokausimarkkina, ennusteiden arviointi, monimarkkinaoptimointi

Contents

Abbreviations

aFRR	Automatic Frequency Restoration Reserve
BESS	Battery Energy Storage System
BSP	Balancing Service Provider
DA	Day-Ahead
EL-EX	an early Nordic electricity power exchange
FCR	Frequency Containment Reserve
FFR	Fast Frequency Reserve
ID	Intraday
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
MaxAE	Maximum Absolute Error
MedAE	Median Absolute Error
ME	Mean Error
MPD	Min–Max Price Deviation
mFRR	Manual Frequency Restoration Reserve
MSE	Mean Squared Error
NEMO	Nominated Electricity Market Operator
RMSE	Root Mean Squared Error
SDAC	Single Day-Ahead Coupling
SOC	State of Charge
TSO	Transmission System Operator

1 Introduction

1.1 Background

The electricity market has undergone significant structural changes during recent years due to the increasing share of renewable energy production, growing electrification of society, and transition towards more flexible energy systems. Particularly, rapid expansion of wind and solar power has increased the volatility and uncertainty of electricity prices as renewable production is highly dependent on weather conditions. For example, during cold winter periods with high electricity consumption, low wind generation, and limited solar production, electricity prices may increase significantly due to tighter supply conditions.

However, flexible electricity assets, such as battery energy storage systems (BESS), flexible wind and solar farms, and a new-generation consumption resource, may now participate simultaneously in multiple electricity markets, including day-ahead, intraday, and reserve markets. This multi-market optimization improves risk management capabilities and enables revenue streams to be diversified across several markets.

In the context of multi-market optimization, operational decisions are strongly dependent on price forecasts. In particular, day-ahead electricity price forecasts are often used as a basis for operational planning and optimization decisions, as it is the primary market for electricity procurement and sales.

1.2 Research Problem

Traditionally, electricity price forecasts have been evaluated using classical prediction error metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). These metrics provide useful information on the average performance of the forecast. However, in a context of multi-market optimization, minimizing the average forecasting errors alone may not give enough information for selecting the best forecast, as the timing of price peaks and low-price periods are more important than predicting the exact price levels. Therefore, additional evaluation metrics are required to better assess the practical usefulness of forecasts.

1.3 Objectives and Research Questions

The objective of this thesis is to evaluate and compare different day-ahead electricity price forecasts. The study aims to assess the forecast quality not only through classical prediction error metrics, but also through metrics that evaluate the ability of forecasts to identify relevant price periods, such as daily price peaks and low-price periods. The forecasts are also evaluated under different market conditions and market resolutions, in order to assess their robustness.

1.4 Scope and Limitations

The scope of this thesis is limited to the evaluation of day-ahead electricity price forecasts in the Finnish electricity market. The analysis is based on five anonymised forecasts from five different forecast providers.

1.5 Thesis Structure

This thesis is structured in a following way. In Chapter 2, the theoretical background and literature review related to electricity markets, reserve markets, multi-market optimization and electricity price forecasting are studied. Chapter 3 introduces the dataset, preprocessing methods, and forecast evaluation metrics which are used in the study. Chapter 4 presents the results and analysis of the forecast comparison across different evaluation metrics, which are studied under different market resolution periods and market conditions. Finally, Chapter 5 summarizes the main conclusions of the thesis, limitations and presents suggestions for future research.

2 Literature Review: Electricity and Reserve Markets

2.1 Electricity Markets

2.1.1 Finnish Market Model and Structure

The Finnish electricity market is part of the European single electricity market. However, it has several national characteristics that affect its functioning such as operating as a single bidding zone, a strong dependence on cross-border transmission, and a production structure which is increasingly influenced by weather-dependent renewable generation. [1]

The liberalisation of Finnish electricity markets began in the mid-1990s. Finland was one of the first countries to start deregulation. As a result, Electricity Market Act of 1995 was initiated by Finnish Ministry of Trade [2]. This established a legal framework for competitive electricity production and retail supply. According to Pineau and Hämäläinen [3], the primary objective of the Electricity Market Act reform was to improve the efficiency of the Finnish electricity markets. Therefore, gradual opening of the electricity market was introduced in order to ensure a controlled transition from a regulated monopoly structure to a competitive market-based system. This reform was not only to modernize the domestic markets, but also to prepare Finnish electricity markets for international electricity system, particularly within the Nordic electricity markets.

As a result of the Electricity Market Act, many institutional, regulatory, and structural changes were implemented in Finland during the second half of the 1990s in order to implement the liberalised market. One of the first acts was to create the Electricity Market Authority under the Ministry of Trade in 1995. Its role was to supervise transmission pricing and issue licenses for transmission operations. Secondly, network was gradually opened during 1995 to 1997, meaning that an open access was given to lines over 500 kW. In addition, an independent power exchange was needed. Therefore, Nordic power exchange (EL-EX) was created to enable electricity trading by offering standard one-hour spot contracts. In 1996, unbundling of tariffs was introduced resulting in completely itemized components of electricity, such as delivery, namely energy, transmission and measurement. Additionally, unbundling of bookkeeping in 1996 meant that companies in both generation and distribution must have separate accounts for each activity. The reform also extended to the taxation. In 1997, electricity taxation was restructured in harmony with other Nordic countries. In the same year, the national transmission system operator (TSO) Fingrid was established. Finally, the markets were opened in 1997 and customers were able to choose their suppliers. [3]

Since the liberalisation during the late 1990s, Finnish Electricity System has become an integrated part of the Nordic Electricity system and also part of the European-wide Electricity system. [4]

In contrast to other energy markets, the electricity markets have a fundamental difference: the production and consumption must be balanced in real time. Consequently, the structure of electricity markets has been constructed in a unique manner to correspond to this condition. In order to address the requirement of continuous balance, the electricity market structure is organised into multiple interconnected market layers that operate across different time horizons. These layers include financial markets, wholesale electricity markets, reserve markets and imbalance settlement. [5]

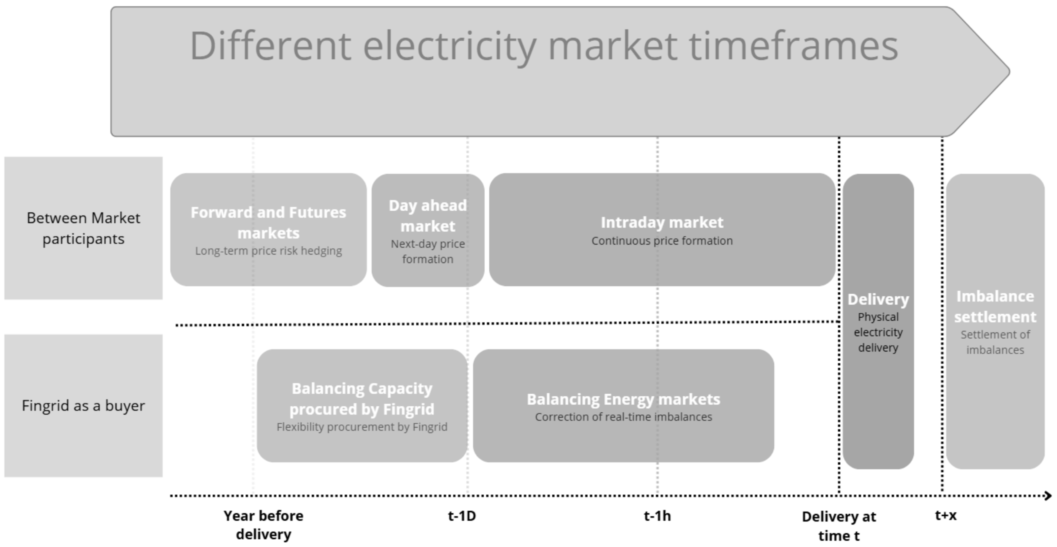

Figure 1: Different electricity market timeframes. Similar figure can be found in [6]

Figure 1 illustrates how the layered electricity markets operate across different time horizons, ranging from long-term financial markets to real-time balancing and post-delivery imbalance settlement. The financial electricity markets mainly consist of forward and futures markets, through which market participants can manage long-term exposure to electricity price uncertainty i.e. hedge long-term price risks without physical delivery. Wholesale markets involve day-ahead market and intraday market. The day-ahead market is primary price formation mechanism for the next day. This is done through centralised auction, where supply and demand are matched. Once the day-ahead market price has formed, the intraday market allows market participants to adjust their positions closer to delivery as production or consumption forecasts are updated. The intraday market operates primarily through continuous trading, although intraday auctions have been introduced as a complementary mechanism.[6, 7]

In parallel with trading between market participants, there are reserve markets in which Fingrid, the Finnish transmission system operator, acts as a counterparty. Through these markets, Fingrid secures the functioning of the electricity grid by procuring balancing capacity in advance. This capacity ensures that sufficient upward and downward flexibility is available to respond to unexpected deviations in production or consumption. [6]

When activated, balancing energy is used to restore the instantaneous balance between production and consumption. This means that the frequency of the electricity system is maintained within acceptable limits. After the delivery period, imbalance settlement allocates the costs of balancing actions to market participants whose deviations in production or consumption contributed to the system imbalance. [6]

2.1.2 Wholesale Electricity Markets in Finland

The Finnish wholesale electricity markets primarily consist of day-ahead and intraday markets. These markets are operated through Nord Pool, which acts as the Nominated Electricity Market Operator (NEMO) in Finland. Nord Pool was established as a power exchange in the 1990s to facilitate electricity trading within the Nordics. The Finnish electricity wholesale market is constructed in a unique way as there is only one bidding zone in the whole country. In many countries, the electricity market is divided into several bidding zones due to transmission capacity constraints and geographical factors. For example, Sweden is divided into four and Denmark into two bidding zones [8].

Recent trading figures show that the day-ahead market remains dominant electricity energy market. According to Nord Pool's Markets Yearly Volumes report [9], approximately 101 TWh was traded in the Finnish day-ahead market area in 2025, whereas in the intraday continuous markets only less than 6 TWh was traded. In addition, intraday auctions accounted for approximately 0.12 TWh of traded volume. This highlights the central role of the day-ahead market in Finnish wholesale electricity trading.

2.1.3 Electricity Price Formation

As shown in Figure 1, the wholesale electricity markets are organised sequentially. The day-ahead auction takes place on a previous day at 14:00 Finnish time and establishes the initial price levels for the next day. Subsequently, the intraday market allows market participants to continuously adjust their positions. Together, these markets form the foundation of wholesale price formation in the Finnish electricity system. [10]

Electricity prices in the day-ahead market are determined through a centralised auction mechanism. Market participants submit supply offers and demand bids for each delivery period of the following day. Supply bids are arranged in ascending order according to marginal production costs, while demand bids are ordered according to consumers' willingness to pay. This can be seen in the illustrative Figure 2. Typically, renewable energy generation has the lowest marginal costs and is therefore located on the left-hand side of the supply curve. Hydropower and nuclear power are also generally positioned on the lower-cost side of the supply curve. In contrast, thermal power plants using fossil fuels are usually located on the right-hand

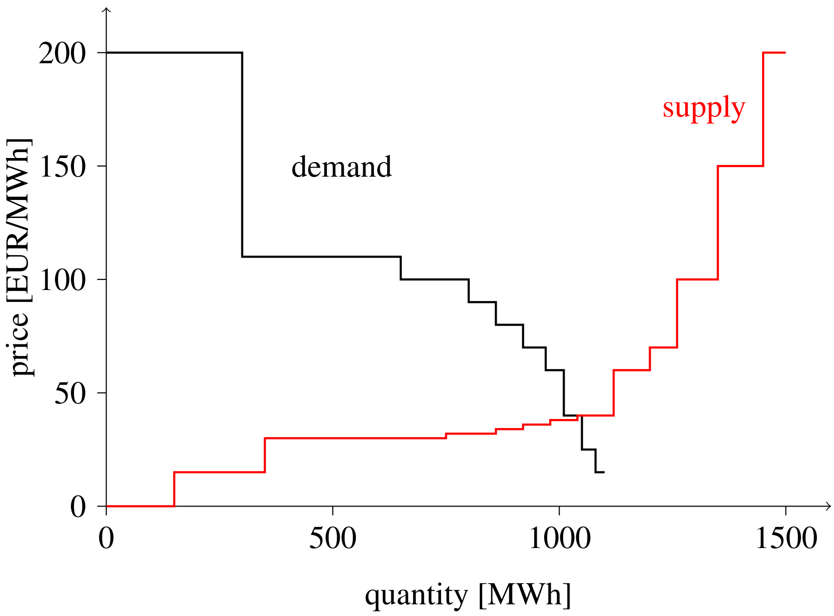

Figure 2: Illustrative aggregated demand and supply curves in a uniform-price auction. Similar Figure can be found in [11].

side, as their marginal costs are based on fuel prices and emission costs. The intersection between aggregated supply and demand curves determines the clearing price and the quantity for each bidding zone. If there were only one bidding zone or unlimited capacity between bidding zones, the market price would be the clearing price.

However, the European day-ahead markets are coupled through the Single Day-Ahead Coupling (SDAC), the price formation is determined jointly across all bidding zones rather than within a single national market. The coupling is conducted by using the Euphemia algorithm, which optimizes cross-border flows between bidding zones to maximize overall social welfare. Therefore, if there is sufficient amount of transmission capacity, prices converge between zones. On the other hand, if transmission capacity is limited, price differences may emerge. [12]

2.1.4 Challenges in Forecasting

Forecasting day-ahead electricity prices has become increasingly important as the price volatility has risen. However, the day-ahead price forecasting has many challenges due to the unique structure of electricity markets. Unlike most other commodity or financial markets, electricity is difficult to store economically at large scale and a power system requires a constant balance between production and consumption. As a result, electricity prices are highly sensitive to fluctuations in demand as well as in supply.

Electricity demand has very strong seasonality factors as it depends on economic activity, calendar effects, and weather conditions. Moreover, electricity prices have multiple layers of seasonality, including within a day (on-peak and off-peak hours), weekly, and annual cycles. These seasonal patterns interact with supply-side factors such as generation availability and fuel costs, creating complex price dynamics. In addition, electricity markets are very vulnerable to unexpected production-side outages, extreme weather events and limitations in cross-border transmission capacity. These characteristics lead to nonlinear behavior, which complicates the day-ahead forecasting. [13]

The increasing amount of renewable energy sources further complicates forecasting. Renewable production, such as wind and solar, has near zero marginal costs and high dependence on weather conditions. These characteristics increase supply-side uncertainty, and may lead to negative and extreme prices.

All in all, electricity price distributions are often heavily tailed and skewed distributions. Such characteristics violate many standard assumptions and therefore more flexible and sophisticated modeling is needed.

2.2 Reserve Markets

The electricity grid requires a constant balance between production and consumption in order to maintain the system frequency within acceptable limits and thus ensure functioning. The nominal system frequency in the European power system is 50 Hz. Wholesale markets define the scheduled production and consumption levels in advance and form the basis for market participants' planned positions. However, the deviation between planned and actual production, as well as consumption, is inevitable. These deviations may result from forecast errors, unexpected generation outages, or changes in consumption. [14]

In recent years, the need for balancing reserves has increased significantly. This development is primarily driven by the growing share of renewable energy sources, such as wind and solar power, which introduce higher variability and uncertainty into the power system. As a result, transmission system operators must procure increasing amounts of reserve capacity to ensure system stability. This trend highlights the growing importance of reserve markets in modern electricity systems. [15]

In Europe, the reserve markets are operated by transmission system operators. In Finland, Fingrid acts as the TSO and its responsibility is to ensure the reliable operation of the power system. To fulfil this responsibility, Fingrid procures different types of balancing reserves through reserve markets.

2.2.1 Reserve Market Products

Reserve markets consist of several different products which differ in activation speed, duration and operational purpose in order to allow the power system to respond to disturbances of varying magnitude. These products have a hierarchical structure and allow TSO to balance the power system in an intended manner. [16]

The fastest reserve products are frequency containment reserves (FCR). These include FCR-D UP, FCR-D DOWN and symmetrical FCR-N. These products activate automatically when the system frequency deviates enough and respond within seconds to stabilize the frequency. FCR products are primarily designed for short activations and serve as the first response to frequency deviation. In the Nordic power system, there is also an additional fast reserve product, Fast Frequency Reserve (FFR). FFR provides rapid activation in a low-inertia situation and therefore prevents large frequency drops. [16]

Whereas FCR products are designed as the first response, frequency restoration reserves are responsible for restoring the system frequency to its nominal value. Automatic Frequency Restoration Reserve (aFRR) and Manual Frequency Restoration Reserve (mFRR) provide secondary responses for maintaining system balance. Both aFRR and mFRR include upward and downward regulation and therefore system frequency can be adjusted in both directions. Operating in aFRR and mFRR markets involves two different time frames. On the previous day, market participants offer reserve capacity and commit to keeping this capacity available for activation by the transmission system operator if required. During the delivery day, market participants offer balancing energy bids that define the price at which the reserved capacity can be activated. If reserve capacity has been procured in advance, the corresponding balancing energy bids are mandatory. However, a market participant may also participate only in the energy market without offering reserve capacity. aFRR is activated through automatic activation signals by the transmission system operator and the required response time is in five minutes. mFRR are activated manually by the TSO and its response time is 7.5 minutes. mFRR is mainly used to manage larger system imbalances. [16]

2.2.2 Market Mechanism of Reserve Markets

Electricity assets may participate in reserve markets if they are capable of adjusting their power output and/or consumption when requested by the transmission system operator. These assets are for example production units, such as wind power plants or solar power plants, flexible electricity consumption, battery energy storage systems (BESS) or hybrid systems [17]. Participation in reserve markets is technology-neutral as long as the technical requirements of the specific reserve product are fulfilled. Reserve market participants are referred to as balancing service providers (BSPs).

Participation in reserve markets requires the market participant to sign a reserve agreement with Fingrid. Before an electricity asset can participate in the markets, the reserve unit must pass a prequalification process in which its technical and operational capabilities are verified. For different reserve products there are different prequalification tests to ensure that the unit is able to provide the required activation response within acceptable time limits. In addition, market participants must have a balance settlement agreement which is handled through the imbalance settlement organisation eSett. [16]

2.3 Multi-Market Optimization

2.3.1 Principles of Optimization

In the context of flexible electricity assets, multi-market optimization refers to an optimization problem where the available capacity of a flexible asset, such as battery energy storage system (BESS), is allocated across multiple electricity markets in order to maximize economic value while respecting technical and operational constraints [18]. These markets include day-ahead energy, intraday energy and various reserve markets.

The fundamental objective is to maximize the profit from different market revenue streams while also minimizing the degradation cost of an asset and operational risks. In practice, this means that the asset operator decides on the previous day, before market clearing, how much capacity to allocate to each market at each time step. Consequently, the optimization is strongly dependent on price forecasts for each market. Additionally, surplus or uncommitted capacity can be allocated during the delivery day to more dynamic energy markets such as intraday, aFRR or mFRR energy markets.

The multi-market optimization for battery energy storage systems is in many ways more complex as the energy resource is limited and therefore the optimization is inherently time-dependent. Decisions made at time step t_y affect the future time steps $t_{y+1}...t_{y+n}$. Therefore, state-of-charge (SOC) dynamics play a central role in the optimization. Charging the battery during low-price periods reduces the available capacity for future charging and downward regulation reserve activations, whereas discharging limits the ability to respond to future upward regulation or price spikes.

Another important aspect of multi-market optimization is uncertainty. Market prices and reserve activations are not known at the time of decision-making. Especially, mFRR and aFRR activations are binary: either activated or not. Therefore, their impact on the state-of-charge (SOC) is really difficult to predict in advance. As a result, continuous SOC management during the delivery day is needed in order to respond the reserve obligations. This SOC management can be done by continuously trading in the intraday market.

2.3.2 Role of Day-Ahead Price Forecasting in Optimization

The role of the day-ahead price forecasting in multi-market optimization is central, because bidding and capacity allocation decisions are made before the delivery day and therefore rely on forecasted market prices. However, accurately predicting the exact price level is not the only relevant objective. It is also important that the forecast captures the timing of low- and high-price periods within the day. For flexible assets, this timing information is essential for identifying economically favourable charging, discharging, and market allocation opportunities.

Trading the asset in the day-ahead market has several functions in the context of operating in a multiple markets. On the one hand, it determines how capacity is allocated between energy and reserve markets in order to maximize arbitrage revenues. On the other hand, it enables maintaining feasibility under reserve commitments by shaping the state-of-charge trajectory in advance.

2.3.3 Characteristics of a Good Day-Ahead Price Forecast

A good day-ahead price forecast has several unique characteristics in a use for multi-market optimization. Traditionally, classical error metrics such as mean absolute error (MAE) and root mean squared error (RMSE) are used to evaluate forecast performance. These metrics measure the magnitude of the forecast errors and therefore every time step has the same weight.

However, in the context of multi-market optimization, only minimizing the absolute forecast errors is not sufficient. The more interesting feature of a forecast is to know how it identifies the timing of price peaks and low-price periods, rather than predicting exact price levels. Therefore, the forecast can be systematically biased meaning poor MAE and RMSE but the forecast is still good if it captures the intraday price shape correctly. [19]

3 Research Data and Methods

3.1 Data

3.1.1 Day-Ahead Price Forecasts

The study uses day-ahead price forecasts from five different forecast providers together with realised Finnish day-ahead electricity prices. The forecasts are anonymised and referred to as Forecast 1, Forecast 2, Forecast 3, Forecast 4, and Forecast 5 throughout the thesis. The realised Finnish day-ahead prices are fetched from Nord Pool API [20]. All forecasts are point forecasts for the Finnish day-ahead market price for a specific delivery period. The prices are expressed in euros per megawatt-hour (€/MWh).

3.1.2 Test Period

The evaluation period is divided into two separate test periods as the day-ahead market resolution changed from 1 hour to 15-minute resolution on 1 October 2025. The first test period covers the hourly resolution from 1 October 2024 to 30 September 2025. During this period, each delivery day consists of 24 hourly price observations. The second test period covers 15-minute resolution period from 1 October 2025 to 1 February 2026. During this period, each delivery day consists of 96 quarter-hourly price observations.

The two test periods are not comparable in all aspects. The hourly test period covers a full year, whereas the 15-minute test period covers only approximately four months and mainly represents winter season. However, these differences are taken into account when interpreting the results.

3.1.3 Data Preprocessing

The data preprocessing mainly consisted of transforming the timestamps into a common timestamp format Coordinated Universal Time (UTC). After that, the forecasted prices were matched with the corresponding realised day-ahead prices based on the delivery timestamp. For daily metrics, such as epsilon hit rates, Corr-f, and Min–Max Price Deviation, the data were grouped by delivery day, as they are based on identifying daily maximum and minimum prices and the relative ordering of prices within a day. In addition, the data were divided into normal and volatile market conditions by filtering based on the realised daily price range. Days belonging to the upper 25% of the daily price range were classified as volatile market conditions, while the remaining days were classified as normal market conditions. This classification was performed separately for the hourly and 15-minute periods.

3.2 Forecast Evaluation Metrics

3.2.1 Classical Error Metrics

Prediction error metrics are widely used to evaluate the accuracy of point forecasts. They quantify forecast performance by calculating the difference between predicted values and realised outcomes. By aggregating these residuals into a single numerical value, these metrics enable a consistent comparison across different point forecasting models.

Let x_t denote the realised price at the time t and $\hat{x}_t$ denote the corresponding forecasted price at the time t, and n denote sample size i.e. length of the time series. Definition of residual e_t at the time t is given as follows:

$$e_t = x_t - \hat{x}_t.$$

Mean Squared Error (MSE) is widely used prediction error metric, which measures the average deviation between predicted values and observations. MSE is defined as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2,$$

where n is sample size [21].

The Root Mean Squared Error (RMSE) is defined as a square root of MSE,

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}.$$

While MSE describes the average of the squared errors, RMSE is its square root meaning that the error is in a same scale as the original sample. Both metrics are sensitive to outliers as the residuals are squared and therefore large outliers have a large weight. [21]

While squared error metrics have a quadratic weight on error terms, another commonly used class of prediction error metrics have a linear weight on residuals. This means that the metric is more robust to large outliers. For instance the Mean Absolute Error (MAE) measures the average of the absolute values of errors. MAE is defined as follows [22]:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t|.$$

If a scale-independent metric is required, the Mean Absolute Scaled Error (MASE) can be applied. It is the mean absolute error of the forecast values, divided by the mean absolute error of a naive forecast. Here the naive forecast is the price observed at the same time of the previous day. This property makes it particularly suitable for comparing forecast performance across different time series and market conditions, as many forecasts have heuristic tendencies. The MASE is defined as follows [25]:

$$\mathrm{MASE} = \frac{\frac{1}{n}\sum_{t=1}^{n} |e_t|}{\frac{1}{n-m}\sum_{t=m+1}^{n} |x_t - x_{t-m}|},$$

where m denotes the seasonal lag. For hourly electricity prices, $m = 24$, corresponding to a seasonal naive forecast defined as $\hat{x}_t = x_{t-24}$, i.e., the price observed at the same hour of the previous day. For 15-minute resolution data, $m = 96$ is used. The MASE value smaller than one indicates that the forecasting model outperforms the naive benchmark, whereas a value greater than one implies weaker performance relative to the benchmark.

From a mathematical perspective, prediction error metrics can be interpreted through vector norms L_1 and L_2 applied to the residual vector. The Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) are derived from the L_2 norm of the residual vector e_t. In contrast, the Mean Absolute Error (MAE) corresponds to the normalised L_1 distance between predicted values $\hat{x}_t$ and observations x_t. [23]

While many classical prediction error metrics are based on L_1 and L_2 norms of the residuals, there are multiple alternative, more system specific, prediction error measures. These measures include for example maximum, median-based, weighted and asymmetrical loss functions which are designed to capture robustness and heterogeneous importance of forecast errors.

In addition to average-based error measures, it is often informative to examine the extreme deviations of the residual vector. The Maximum Absolute Error (MaxAE) measures the largest individual forecast deviation within the sample and is defined as follows [24]:

$$\text{MaxAE} = \max_{t=1,\ldots,n} |e_t|.$$

This metric highlights the worst-case forecast error and therefore can be used in applications where extreme deviations may have significant consequences.

The Median Absolute Error (MedAE) is defined as follows [25]:

$$\text{MedAE} = \text{median}(|e_1|, |e_2|, \ldots, |e_n|).$$

Unlike the MAE, the median-based MedAE metric is very robust to extreme outliers.

In addition to absolute and squared error measures, it is often useful to quantify systematic forecast bias. The Mean Error (ME), also called bias, is defined as the average of the signed residuals [26]:

$$\text{ME} = \frac{1}{n}\sum_{t=1}^{n} e_t.$$

A positive ME indicates that forecasts are on average lower than the realised values meaning that they are under-forecasting, whereas a negative ME indicates over-forecasting.

Overall, there are multiple different prediction error metrics. Each of them has its own characteristics and, therefore emphasizes different aspects of a forecast performance. This means that no single prediction error measure is capable of capturing all dimensions of prediction accuracy. Consequently, multiple prediction error metrics are often used jointly in order to provide a more comprehensive assessment of a point forecast performance.

3.2.2 Ranking and Order-Based Metrics

In addition to classical prediction error metrics, it is useful to evaluate whether a forecast identifies the relative ordering of prices within a day. In the context of multi-market optimization, this is more useful than classical error metrics, as the objective is often to find the time periods that are close enough from peaks and lows relative to daily profile. Many operational and trading decisions depend not only on the absolute forecast accuracy but also how accurately the model predicts the timing of daily peaks and low-price periods. Therefore, ranking- and order-based metrics are introduced to complement the classical error measures.

The order-based metric used in this study is called epsilon hit rate. The purpose of this metric is to evaluate whether the forecast correctly identifies time periods that are sufficiently close to the realised daily maximum or minimum prices.

For each day d, let us denote $x_{d,t}$ as the realised price and $\hat{x}_{d,t}$ as the forecasted price at the time t. The predicted peak time is defined as

$$\hat{t}^{\max} = \arg\max_t \hat{x}_{d,t}.$$

A hit is recorded if the realised price at the predicted peak is close enough to the realised daily maximum. The hit rate can be divided into relative and absolute hit rate. In relative formulation, a hit occurs if the predicted peak time is at least a certain percentage of the actual daily maximum. The relative hit is defined as

$$x_{d,\hat{t}^{\max}} \geq (1 - \varepsilon) \max_t x_{d,t},$$

where $\epsilon \in [0, 1]$ is the allowed proportional deviation from the realised daily maximum.

In contrast, the absolute hit rate defines a hit based on an absolute tolerance level. An absolute hit is defined as

$$x_{d,\hat{t}^{\max}} \geq \max_t x_{d,t} - \gamma,$$

where $\gamma > 0$ is a tolerance allowed from the realised daily maximum.

Now, the epsilon hit rate is calculated as the proportion of days for which hit is recorded. It is defined as

$$\text{Hit Rate}_{\text{peak}} = \frac{1}{D} \sum_{d=1}^{D} \text{Hit}_d,$$

where D is the number of valid days and $\text{Hit}_d \in \{0, 1\}$ indicates whether a hit is recorded for day d.

In addition to peak detection, the hit rate can also be applied to evaluate the detection of low-price periods (bottoms). The predicted bottom time is defined as

$$\hat{t}^{\min} = \arg\min_{t} \hat{x}_{d,t}.$$

Similar to peak detection, a hit is recorded if the realised price at the predicted bottom is close enough to the daily minimum. In the relative formulation, a hit occurs if

$$x_{d,\hat{t}^{\min}} \leq (1+\varepsilon) \min_{t} x_{d,t},$$

where $\varepsilon \in [0, 1]$ defines the allowed proportional deviation from the realised daily minimum.

The absolute hit for bottom detection is defined as

$$x_{d,\hat{t}^{\min}} \leq \min_{t} x_{d,t} + \gamma,$$

where $\gamma > 0$ is the allowed absolute tolerance.

Similarly, the bottom hit rate is calculated analogously as

$$\text{Hit Rate}_{\text{bottom}} = \frac{1}{D} \sum_{d=1}^{D} \text{Hit}_d^{\min},$$

where $\text{Hit}_d^{\min} \in \{0, 1\}$ indicates whether a bottom hit is recorded for day d.

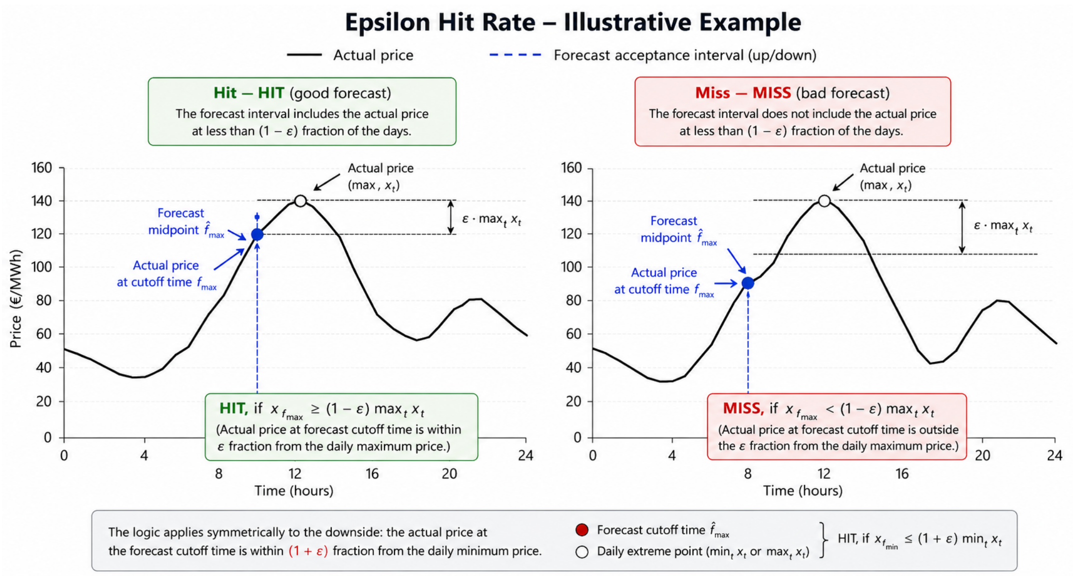

Figure 3: Illustrative example of the epsilon hit rate logic

Figure 3 illustrates the logic behind the epsilon hit rate. The metric evaluates whether the peak selected by the forecast is sufficiently close to the actual daily extrema or bottom. The left side of the figure shows a successful hit and the right side demonstrates a miss as the forecast-selected peak deviates more than epsilon from the actual daily maximum price.

3.2.3 Association and Economic Metrics

In addition to classical prediction error and hit rate metrics, it is important to know how the forecast captures the structure of the price profile and its economic value. The epsilon hit rate focuses on identifying timing of the extreme price events within a tolerance. However, it lacks an understanding of the overall price profile during the day. Therefore, the association metric Corr-f and the Min–Max Price Deviation (MPD) are introduced.

The association metric Corr-f evaluates how well the forecast preserves the relative ordering of prices within a day, independently of the absolute price level. It is explained as follows. For each day d, let $x_{d,t}$ denote the realised price and $\hat{x}_{d,t}$ the forecasted price at time t and the daily association is computed using the Spearman rank correlation coefficient:

$$\rho_d = \mathrm{corr}_{\mathrm{Spearman}}\left(\{x_{d,t}\}_{t=1}^{T}, \{\hat{x}_{d,t}\}_{t=1}^{T}\right).$$

The overall Corr-f metric is defined as the average of daily correlations:

$$\text{Corr-f} = \frac{1}{D}\sum_{d=1}^{D}\rho_d,$$

where D is the total number of days. [19]

The MPD is used to evaluate the economic impact of errors in identifying extreme price periods. It measures the deviation between the realised extreme prices and the prices observed at the forecasted extreme time steps.

For each day d, the realised and predicted extreme times are defined as

$$t_d^{\max} = \arg\max_t x_{d,t}, \quad \hat{t}_d^{\max} = \arg\max_t \hat{x}_{d,t},$$

$$t_d^{\min} = \arg\min_t x_{d,t}, \quad \hat{t}_d^{\min} = \arg\min_t \hat{x}_{d,t}.$$

The daily deviation is defined as

$$\mathrm{MPD}_d = \left|x_{d,t_d^{\min}} - x_{d,\hat{t}_d^{\min}}\right| + \left|x_{d,t_d^{\max}} - x_{d,\hat{t}_d^{\max}}\right|.$$

The overall MPD is computed as

$$\text{MPD} = \frac{1}{D}\sum_{d=1}^{D}\text{MPD}_d,$$

where D is the number of valid days. [19]

3.3 The Use of Technical Tools and Artificial Intelligence

Version control and software development were conducted using Visual Studio Code and a GitLab repository. AI-assisted coding tools were used to support the implementation of data processing, analysis, and visualization scripts.

In addition, artificial intelligence tools were utilized to support linguistic refinement, grammar checking, and improving the clarity of the written text. AI tools were not used to generate scientific conclusions, interpretations, or research findings independently.

The author of this thesis has verified all produced code, analyses, results, and written content and therefore takes full responsibility for the work.

4 Results and Discussion

4.1 Visualisations

In this section a visual comparison between five forecast providers is introduced anonymously. The comparison is divided into two time frames that reflect the structural change in the day-ahead market as the resolution changed from hourly to 15-minute frequency. The first period is 1 October 2024 to 30 September 2025 and is evaluated using hourly price data and the second time frame is 1 October 2025 to 1 February 2026 when the new 15-minute frequency is valid. Not all forecast providers cover the whole time interval. However, the selected time interval is chosen such that the test period is expected to capture as much information as possible.

4.1.1 Hourly Resolution Period (1 Oct 2024 – 30 Sep 2025)

The comparison between anonymised forecasts and realised day-ahead prices is shown in Figure 4. In the hourly resolution period, four different forecast providers are tested in which data from Forecast 2 is only partially available. The evaluation period covers over 8000 data points, resulting in the visual inspection can be challenging and alone does not provide all the information.

Nevertheless, several structural observations can be drawn from the figure. During the

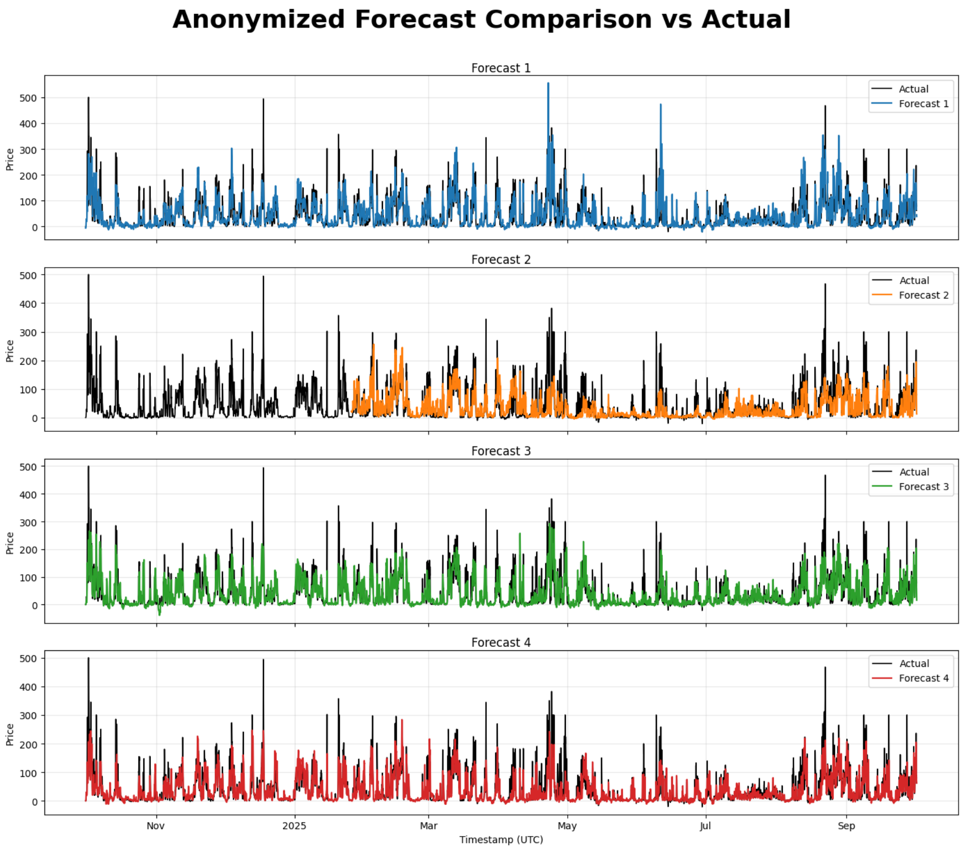

Figure 4: Comparison of anonymised forecasts and realised day-ahead prices (hourly resolution, 1 Oct 2024 – 30 Sep 2025)

year, seasonal price variation is clear. In particular, the summer period can be seen, which is low price period, as there is increased renewable generation, especially solar production, and typically lower electricity demand.

Overall, all the forecast providers follow the price level reasonably well. However, differences appear during periods of higher price levels and increased volatility. For example, Forecast 1 appears to have a tendency to capture extreme values, whereas Forecast 2 seems to have a more conservative approach and tends to under-forecast during peak hours.

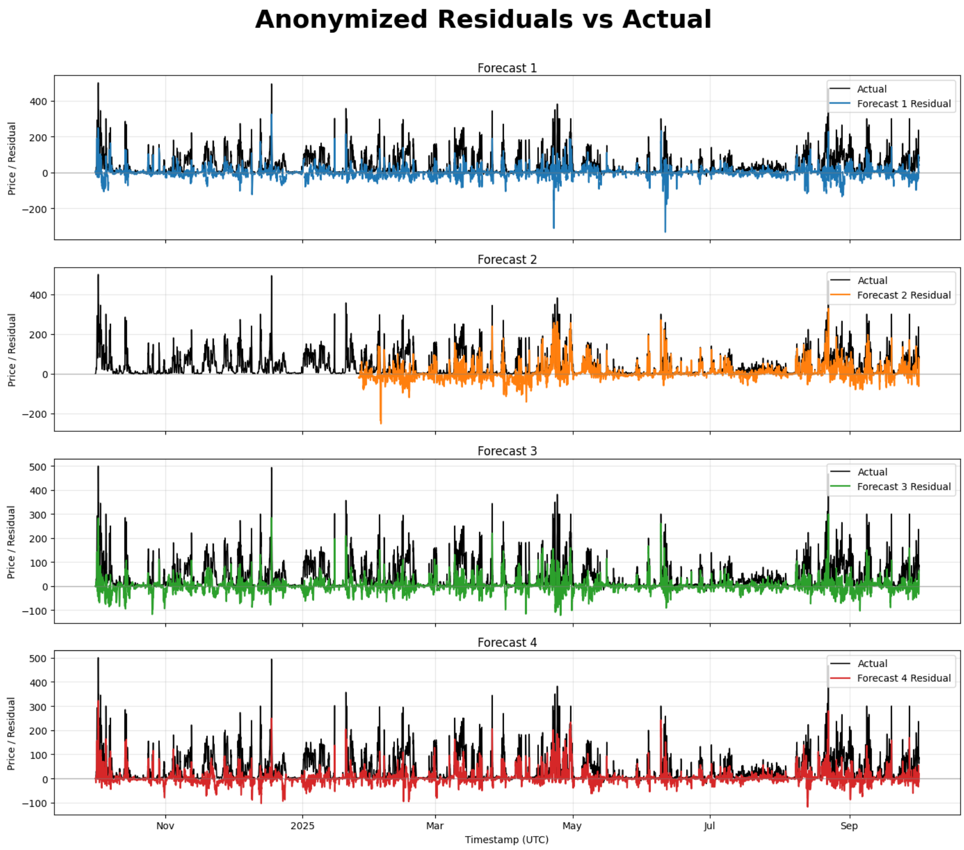

Figure 5: Residuals of anonymised forecasts relative to realised day-ahead prices (hourly resolution)

The Figure 5 illustrates how residuals behave relative to day-ahead prices. The residual plot provides a better view to analyse how forecasts have performed during the test period.

Several conclusions can be drawn from the figure. First, the residuals i.e. forecast errors increase during periods of the high price volatility. In contrast, all models perform relatively well during low day-ahead prices as the residuals remain clearly smaller. This structural pattern suggests that the forecast accuracy is strongly dependent on market conditions. When the electricity system is in a stable condition, this is often during low day-ahead price periods, all forecasts perform relatively well. In contrast, during stressed system conditions, such as tight supply-demand balance or unexpected disturbances, day-ahead prices are often high and forecast errors increase significantly.

4.1.2 15-Minute Resolution Period (1 Oct 2025 – 1 Feb 2026)

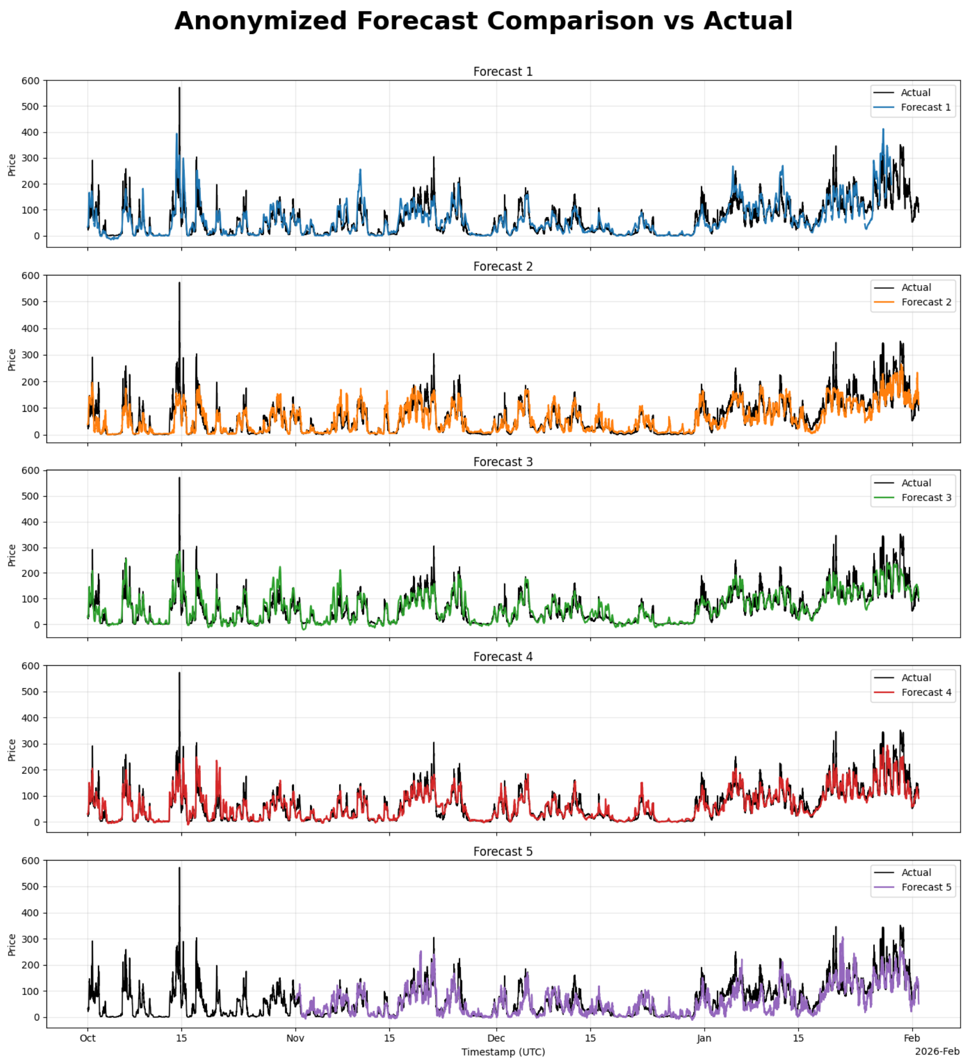

Figure 6: Comparison of anonymised forecasts and realised day-ahead prices (15-minute resolution, 1 Oct 2025 – 1 Feb 2026)

The comparison between anonymised forecasts and day-ahead prices for the 15-minute test period is presented in Figure 6 and spans from 1 Oct 2025 to 1 Feb 2026. The test period has now five different forecast providers, of which the first four correspond to those analysed in the hourly resolution period. Unlike the previous hourly test period, this interval reflects the new market structure where the day-ahead market operates with a 15-minute resolution. This means that the number of observations is four times higher compared to the hourly frequency.

Visually, all forecast providers seem to continue the general price development well. A structural pattern similar to the hourly period can be observed across the test period. When day-ahead price is low or close to zero, all forecast providers seem to perform reasonably well. In contrast, during the high price periods, forecasting becomes more challenging.

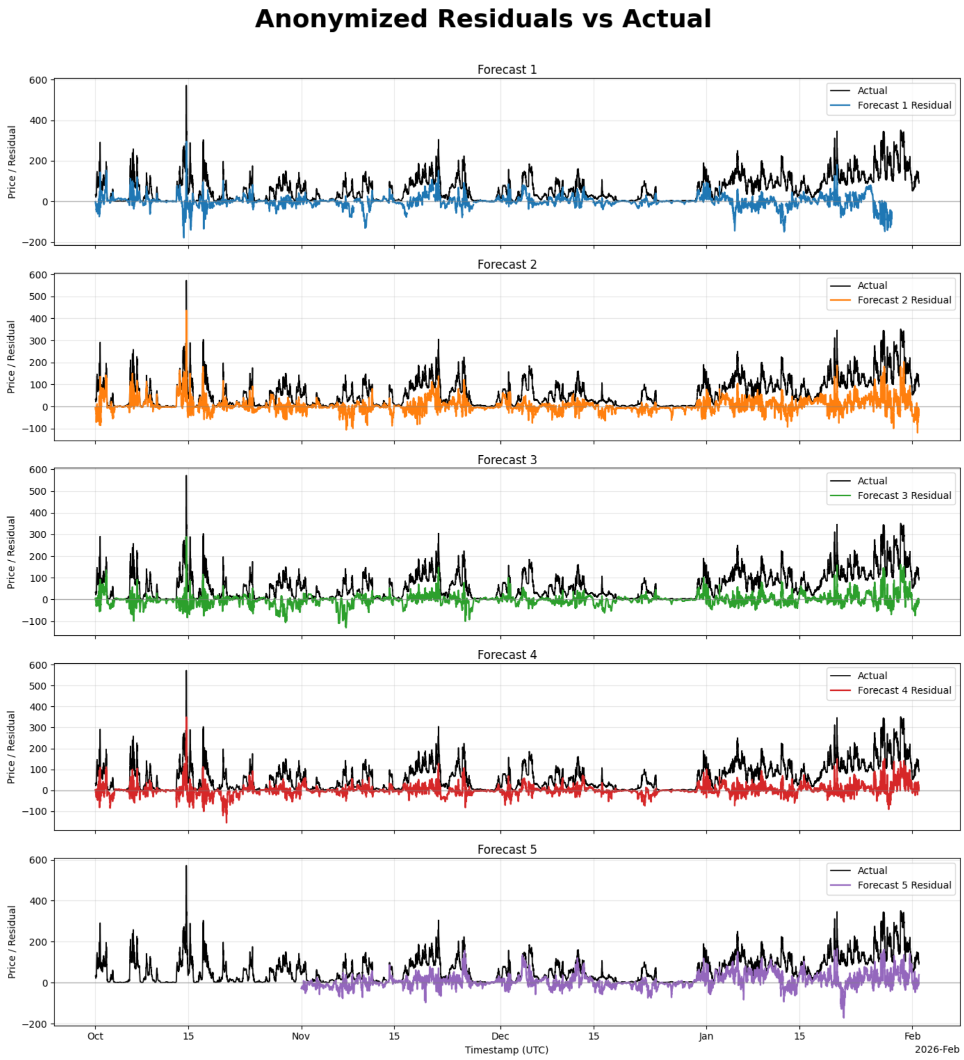

Figure 7: Residuals of anonymised forecasts relative to realised day-ahead prices (15-min resolution)

The behaviour of the residuals for the 15-minute resolution period is shown in Figure 7. Compared to hourly period Figure 5, the deviations seem larger. All models seem to have reasonably good accuracy during low-price periods. However, as soon as the daily price profile is more volatile, the residuals seem to have larger absolute values. From the Figure 7 and Figure 6, it can be observed that Forecast 1 seems to follow price spikes relatively closely but also cause residual spikes during high price periods. Forecast 3 and Forecast 4 seem to have the most balanced behavior, appearing to capture the general trend. Forecast 5 shows the highest residuals, especially during high price periods.

Overall, the residuals seem to follow general pattern where forecasting during high-price periods is not that accurate compared to low-price periods. This is consistent with the usual functioning of electricity system. It is often more stable during low price periods and therefore easier to predict. In contrast, high-price periods are typical when there is a tight supply-demand conditions or unexpected outages, making accurate forecasting more challenging.

4.2 Classical Error Metrics

Tables 1 and 2 present the classical error metrics for the day-ahead forecasts from the forecast providers. The evaluated metrics are MSE, RMSE, MAE, MedAE, MaxAE, ME, and MASE. For each metric, the best forecast provider is bolded. Each of these metrics captures different aspects of forecasting accuracy, such as the magnitude of errors, robustness to extreme values, and potential systematic bias. Therefore, together these results enable proper validation of forecast performance.

Table 1: Classical prediction error metrics (Hourly resolution, 1 Oct 2024 – 30 Sep 2025)

Provider	MSE	RMSE	MAE	MedAE	MaxAE	ME	MASE
Forecast 1	893.26	29.89	17.03	8.04	333.12	1.10	0.454
Forecast 2	1514.81	38.92	22.76	11.34	328.28	-4.01	0.651
Forecast 3	**762.75**	**27.62**	15.84	7.90	**299.38**	**0.08**	0.431
Forecast 4	830.47	28.82	**15.13**	**6.23**	321.91	-3.83	**0.411**

Table 1 presents the forecasting performance during the hourly resolution period from 1 October 2024 to 30 September 2025. Forecast 3 provides the best results considering MSE and RMSE. This indicates that Forecast 3 is particularly effective in avoiding large deviations as squared metrics penalize large errors more heavily. However, Forecast 1 and Forecast 4 are reasonably close in terms of these metrics.

When comparing typical forecasting errors, MAE and MedAE are considered. MAE is robust for outliers and MedAE is very robust for outliers and therefore highlights which one has the smallest typical forecasting errors. In this comparison Forecast 4 performs the best, followed by Forecast 3 and Forecast 1. MaxAE indicates which forecast has the largest individual outlier. Forecast 3 has the smallest and therefore the best value. Forecast 3 has the best ME, only deviating 0.08 on average. In comparison, Forecast 4 tends to under-forecast -3.77 and Forecast 1 tends to over-forecast 1.10 on average. Forecast 4 has the best MASE, indicating the strongest performance relative to the naive benchmark model.

Overall, the classical error metrics for hourly period present the deviation clear differences in performance. Forecast 2 consistently performs worse than other providers across most metrics. Particularly, Forecast 2 performs poorly in terms of MSE and RMSE, indicating larger deviations between forecasts and realised prices especially during high-price periods. Among the top three performing models, the differences are relatively small. However, the top three can be ranked as Forecast 3, Forecast 4, and Forecast 1 in relatively small margins.

Table 2: Classical prediction error metrics (15-minute resolution, 1 Oct 2025 – 1 Feb 2026)

Provider	MSE	RMSE	MAE	MedAE	MaxAE	ME	MASE
Forecast 1	870.35	29.50	18.72	11.24	295.39	**1.19**	0.493
Forecast 2	960.40	30.99	19.93	11.57	436.20	-4.58	0.523
Forecast 3	670.21	25.89	17.02	10.82	289.50	2.07	0.447
Forecast 4	**580.33**	**24.10**	**14.64**	**7.70**	342.02	**-1.25**	**0.385**
Forecast 5	1178.75	34.33	24.00	16.48	**171.89**	-8.52	0.715

Table 2 presents the forecasting performance during the 15-minute resolution period from 1 October 2025 to 1 February 2026. Compared to hourly resolution period 1, the results only reflect certain winter season period and therefore are not generalisable. Nevertheless, it is the best guess how the forecasting accuracy has evolved among providers when new 15-minute resolution has been introduced.

It can easily be seen that Forecast 4 achieves the best overall performance across classical error metrics. It has the best metrics values except MaxAE. This indicates that Forecast 4 performs well in typical error while also having an ability to avoid larger forecasting errors.

Forecast 3 appears to be the second-best forecast according to most classical error metrics, while Forecast 1 has slightly higher error levels. However, Forecast 1 has the best mean error (1.19), meaning small over-forecasting tendencies. Forecast 2 performs worse than these top three providers, nevertheless achieving reasonably comparable results.

Forecast 5 performs clearly the weakest overall, with the highest MSE, RMSE, MAE, MedAE, ME and MASE. This indicates substantial difficulties to capture price dynamics during the test period. However, Forecast 5 achieves the lowest MaxAE value, meaning that even though the average performance is poor, it avoids extremely large individual errors.

4.3 Ranking and Order-Based Metrics

Tables 3, 4, 5 and 6 present the results for epsilon hit rates. The results are shown differently for peak and low-price (bottom) detection and for both hourly resolution period (1 October 2024 – 30 September 2025) and 15-minute resolution period (1 October 2025 – 1 February 2026). These results evaluate how each forecast provider performs in identifying daily highs and lows rather than the magnitude of forecasting errors.

Table 3: Peak detection performance using epsilon hit rates (hourly resolution)

Provider	Rel (ϵ=0.05)	Rel (ϵ=0.10)	Rel (ϵ=0.20)	Abs (ϵ=5)	Abs (ϵ=10)	Abs (ϵ=50)
Forecast 1	0.414	0.536	0.675	0.558	0.653	0.922
Forecast 2	0.259	0.344	0.457	0.429	0.530	0.757
Forecast 3	0.420	0.536	0.684	0.588	0.692	0.923
Forecast 4	**0.486**	**0.574**	**0.745**	**0.640**	**0.717**	**0.945**

In the hourly resolution Table 3, it can easily be seen the differences in peak detection performances. Forecast 4 constantly outperforms the other providers in all epsilon rates. However, the differences between top three Forecast 4, Forecast 3 and Forecast 1 are narrow. The absolute tolerance level of 50 EUR/MWh is particularly notable, as all top three providers have over 90% hit rate and Forecast 2 only has approximately 76% .

Table 4: Low-price (bottom) detection performance using epsilon hit rates (hourly resolution)

Provider	Rel (ϵ=0.05)	Rel (ϵ=0.10)	Rel (ϵ=0.20)	Abs (ϵ=5)	Abs (ϵ=10)	Abs (ϵ=50)
Forecast 1	0.419	0.486	0.561	0.875	0.906	**0.997**
Forecast 2	0.296	0.336	0.360	0.773	0.810	0.967
Forecast 3	0.396	0.429	0.489	**0.890**	**0.923**	**0.997**
Forecast 4	**0.500**	**0.527**	**0.607**	0.868	0.920	0.995

The results for low-price (bottom) detection in the hourly resolution period can be seen in Table 4. It shows generally higher hit rates across all models compared to peak detection. This can be explained as in typical daily day-ahead price profile is such that there are at least two low price periods, for example during early mornings and late nights, and high price period is typically seen only once per day. The low-price periods are perhaps also easier to predict as they are associated with more stable system conditions. Forecast 4 performs best in the relative metrics, while Forecast 3 achieves better results in strict absolute tolerance levels. All in all, the differences between forecast providers are smaller than in peak detection, suggesting they are more reliable identifying low-price periods.

Next, the 15-minute period is analysed. As the frequency increases from 24 to 96, the decrease in performance can be expected.

Table 5: Peak detection performance using epsilon hit rates (15-minute resolution)

Provider	Rel (ϵ=0.05)	Rel (ϵ=0.10)	Rel (ϵ=0.20)	Abs (ϵ=5)	Abs (ϵ=10)	Abs (ϵ=50)
Forecast 1	**0.280**	0.381	**0.644**	**0.390**	**0.517**	0.856
Forecast 2	0.137	0.266	0.516	0.274	0.379	0.831
Forecast 3	0.242	0.395	0.629	0.363	0.484	0.879
Forecast 4	0.260	**0.398**	0.602	0.350	0.480	0.855
Forecast 5	0.100	0.200	0.444	0.222	0.344	**0.943**

Table 5 illustrates the peak detection performance during 15-minute resolution period. Compared to the hourly resolution, the hit rates are generally lower across all models. This could be due to the increase in frequency but also due to the change in system dynamics. Market participants may not yet have fully adapted to the new resolution and therefore making the forecasting more challenging.

Forecast 1 achieves the best performance in most of the evaluated metrics, including in relative $\epsilon = 0.05$ and $\epsilon = 0.20$, as well as in absolute tolerance levels of $\epsilon = 5$ and $\epsilon = 10$. Forecast 3 has the second best performance, followed by Forecast 4, Forecast 2 and Forecast 5.

In contrast to the hourly resolution period, Forecast 4 is no longer the best performing model. This suggests that either Forecast 4 has not been able to adapt the new market dynamics or it is only less responsive to short-term price fluctuations and captures broader daily trends. If the results are compared with Table 2, the second option is more probable, as the classical error metrics suggest that Forecast 4 performs very well on average. Forecast 5 clearly performs again the poorest across all metrics.

Table 6: Low-price (bottom) detection performance using epsilon hit rates (15-minute resolution)

Provider	Rel (ϵ=0.05)	Rel (ϵ=0.10)	Rel (ϵ=0.20)	Abs (ϵ=5)	Abs (ϵ=10)	Abs (ϵ=50)
Forecast 1	0.347	**0.449**	0.602	0.771	**0.898**	0.992
Forecast 2	0.323	0.427	0.548	**0.782**	0.887	0.976
Forecast 3	0.306	0.387	0.548	**0.782**	0.871	**1.000**
Forecast 4	**0.398**	0.447	**0.634**	0.780	0.894	**1.000**
Forecast 5	0.222	0.289	0.433	0.578	0.789	0.944

The results for low-price detection in the 15-minute resolution period are presented in Table 6. Similar to the hourly period, the hit rates are generally higher than for peak detection.

Forecast 1 performs the best in all relative tolerance levels. This indicates that Forecast 1 has a good capture rate relative to the daily profile. However, Forecast 2 and Forecast 3 are slightly better in absolute tolerance levels.

Forecast 4 shows again slightly weaker performance compared to its hourly results. This strengthens the assumption that Forecast 4 model is maybe not that responsive to short-term fluctuations in higher-frequency data. Forecast 5 performs again the weakest in all metrics.

These results indicate that low-price periods are generally easier to detect than peak periods across all forecasts.

4.4 Association and Economic Metrics

Table 7 and Table 8 present the results for association metric Corr-f and economic metric MPD.

Table 7: Association and economic performance metrics (hourly resolution)

Provider	Corr-f	MPD
Forecast 1	0.80	17.62
Forecast 2	0.55	37.41
Forecast 3	0.75	17.63
Forecast 4	**0.82**	**14.18**

In the hourly resolution period, Forecast 4 shows the strongest overall performance, achieving 0.82 Corr-f and 14.18 MPD. Forecast 1 and Forecast 3 also perform relatively well whereas Forecast 2 performs clearly the worst. This suggests that, at least in this framework, models that better capture the price profile structure (higher Corr-f) also tend to achieve lower economic losses (lower MPD).

Table 8: Association and economic performance metrics (15-minute resolution)

Provider	Corr-f	MPD
Forecast 1	0.80	27.04
Forecast 2	0.71	33.64
Forecast 3	0.79	24.68
Forecast 4	**0.82**	**22.45**
Forecast 5	0.66	38.94

In the 15-minute resolution period, similar patterns can be observed. Forecast 4 dominates in both metrics, while Forecast 1 and Forecast 3 follow with comparable performance. Forecast 5 performs clearly the worst. It is worth noting that the Corr-f has similar scale independent from resolution whereas MPD is significantly higher in the 15-minute period.

4.5 Forecast Performance under Different Market Conditions

From an economic perspective, higher revenues are typically associated with periods of increased market volatility. Although day-ahead arbitrage may represent only a small share of total revenue in the multi-market optimization, it is important that the optimization is able to detect the low- and high-price periods during volatile day-ahead market conditions. Similarly, the potential losses are greater when market is in volatile conditions. Therefore, models are additionally evaluated through different market conditions. Two different conditions, referred to as volatile and normal, defined based on the realised daily price range. Specifically, days belonging to the upper quartile (top 25%) of the daily price range distribution are classified as volatile, while the remaining days are classified as normal market conditions. The 75th percentile threshold value was examined both in hourly and a 15-minute period differently. For the hourly period, the daily volatile threshold value is 132.185 and for 15-minute period it is 130.690. In the following tables, each metric cell contains two values: first one is the value during hourly period and the second one during the 15-minute period.

Table 9: Forecast performance under volatile market conditions

Provider	MSE	RMSE	MAE	MedAE	MaxAE	ME	MASE
Forecast 1	2381.09 / 2207.47	48.80 / 46.98	32.07 / 33.40	20.37 / 22.72	333.12 / 295.39	-3.31 / 5.79	0.47 / 0.49
Forecast 2	4011.41 / 2350.08	63.34 / 48.48	42.91 / 33.18	27.41 / 22.50	328.28 / 436.20	-19.78 / -19.15	0.66 / 0.50
Forecast 3	**2173.44 / 1352.63**	**46.62 / 36.78**	**30.57** / 26.01	19.39 / 19.00	**299.38 / 289.50**	-8.31 / **-4.43**	**0.45** / 0.39
Forecast 4	2561.01 / 1451.63	50.61 / 38.10	32.02 / **25.55**	**18.44 / 16.00**	321.91 / 342.02	-19.36 / -8.17	0.47 / **0.38**
Forecast 5	– / 2184.23	– / 46.74	– / 34.40	– / 24.92	– / 165.68	– / -24.93	– / 0.67

Table 9 presents the forecasting accuracy under volatile market conditions for both hourly and 15-minute period. The results seem a little counterintuitive as the classical error metrics indicate that the forecasting accuracy is better in 15-minute resolution. First, the comparison is not fully beyond criticism. The hourly results cover a year, whereas the 15-minute results only cover 4 months period during recent winter. There are different seasonality aspects for sure and therefore comparison is not that straightforward. Second, the number of observations during 15-minute period is significantly smaller and therefore individual days have larger weight on the results. However, it is yet possible that the results are indeed better and the forecast providers' models have improved as time period is more recent. This would need more sophisticated examinations. Nevertheless, the results together are suited for comparison with performance under normal market condition which are presented in the Table 10.

Table 10: Forecast performance under normal market conditions

Provider	MSE	RMSE	MAE	MedAE	MaxAE	ME	MASE
Forecast 1	353.85 / 426.83	18.81 / 20.66	11.64 / 13.84	5.95 / 9.24	146.58 / 107.57	3.02 / **-0.20**	0.63 / 0.56
Forecast 2	559.05 / 479.57	23.64 / 21.90	15.04 / 15.35	7.89 / 9.17	239.44 / 119.39	2.03 / 0.46	0.90 / 0.62
Forecast 3	288.51 / 434.09	16.99 / 20.84	10.89 / 13.91	6.12 / 8.89	**103.81 / 130.40**	2.91 / 4.32	0.59 / 0.56
Forecast 4	**248.69 / 287.54**	**15.77 / 16.96**	**9.45 / 10.98**	**4.46 / 6.19**	106.41 / 134.76	**1.39** / 1.08	**0.51 / 0.45**
Forecast 5	– / 903.53	– / 30.06	– / 21.15	– / 14.62	– / 171.89	– / -4.03	– / 0.83

The results under normal market conditions are more consistent compared to the volatile conditions. Forecast 4 outperforms the other models in nearly all classical error metrics in both time intervals. It achieves the best results in MSE, RMSE, MAE, and MASE values. This indicates that under stable price dynamics, the model is able to find daily price patterns well. Forecast 3 also performs competitively, particularly in the hourly setting followed by Forecast 1 and Forecast 2. Again, Forecast 5 performs the weakest among the 15-minute resolution.

When comparing overall results between volatile and normal market conditions, there is a notable increase of forecasting errors across all providers during volatile periods. This is expected as more volatile market increases the difficulty of forecasting. The best RMSE and MAE are over tripled during the hourly resolution period and over doubled during the 15-minute resolution period between normal and volatile market conditions. However, MASE behaves more consistently across both market conditions. This behavior can be explained by its definition as MASE is scaled error metric, where the forecast error is normalised by the error of a naive benchmark i.e. the previous days error. During the volatile periods, both the forecast errors and the benchmark errors tend to increase and therefore their relative ratio keeps stable. As a result, MASE is less sensitive to changes in overall price levels and volatility compared to absolute error metrics such as RMSE and MAE.

A similar comparison between market conditions is implemented also for peak and low-price detection using selected epsilon hit rates. The results can be seen in the Table 11 and Table 12.

Table 11: Peak and low-price detection performance using selected epsilon hit rates in volatile market conditions

Provider	Peak Rel (ϵ=0.10)	Peak Abs (ϵ=10)	Bottom Rel (ϵ=0.10)	Bottom Abs (ϵ=10)
Forecast 1	0.49 / 0.32	0.40 / 0.21	0.57 / **0.57**	0.83 / **0.82**
Forecast 2	0.22 / 0.31	0.20 / 0.19	0.43 / 0.47	0.69 / 0.72
Forecast 3	0.51 / 0.28	0.44 / 0.13	0.53 / 0.47	**0.86** / 0.69
Forecast 4	**0.53 / 0.45**	**0.47 / 0.29**	**0.69** / 0.58	**0.86** / 0.71
Forecast 5	– / 0.26	– / 0.16	– / 0.37	– / 0.58

The results in Table 11 reveal the differences between forecast providers' performance for peak and low-price detection under volatile market conditions. Forecast 4 achieves the highest hit rates in peak detection and also bottom detection in hourly resolution. All forecast providers perform better in bottom detection which is expected as the level of the low-price is easier to forecast generally. The daily price profile has usually more tendency to have high peaks than deep bottoms. As a result, the daily minimum price is more likely to fall within the defined epsilon range, leading to higher hit rates in bottom detection.

Table 12: Peak and low-price detection performance under normal market conditions

Provider	Peak Rel (ϵ=0.10)	Peak Abs (ϵ=10)	Bottom Rel (ϵ=0.10)	Bottom Abs (ϵ=10)
Forecast 1	0.55 / 0.40	0.73 / **0.61**	0.46 / **0.41**	0.93 / 0.92
Forecast 2	0.39 / 0.25	0.65 / 0.45	0.30 / **0.41**	0.85 / 0.95
Forecast 3	0.54 / **0.44**	0.78 / **0.61**	0.39 / 0.36	**0.95** / 0.94
Forecast 4	**0.59** / 0.38	**0.80** / 0.54	**0.47** / 0.40	0.94 / **0.96**
Forecast 5	– / 0.18	– / 0.39	– / 0.27	– / 0.85

The results of normal market conditions, presented in Table 12, show more balanced performance in all forecast providers compared to volatile. Again Forecast 4 achieves the highest hit rates in hourly resolution except the bottom absolute hit rate. Forecast 3 and Forecast 1 also perform relatively well across all metrics.

The overall differences between volatile and normal market conditions are relatively closer than classical error metrics. This indicates that even though market conditions are different the models, especially the best performing models, are able to time the peaks and bottoms despite the market condition. Nevertheless, the timing accuracy is still easier for models under normal market conditions.

Table 13: Association and economic performance under volatile market conditions

Provider	Corr-f	MPD
Forecast 1	0.84 / 0.83	39.28 / 66.47
Forecast 2	0.61 / 0.73	90.05 / 74.89
Forecast 3	0.86 / 0.84	40.40 / 56.88
Forecast 4	**0.86 / 0.85**	**33.46 / 44.40**
Forecast 5	– / 0.74	– / 76.26

Table 14: Association and economic performance under normal market conditions

Provider	Corr-f	MPD
Forecast 1	0.79 / 0.79	10.51 / 14.78
Forecast 2	0.53 / 0.70	18.62 / 19.29
Forecast 3	0.72 / 0.77	9.94 / **13.48**
Forecast 4	**0.81** / **0.80**	**7.66** / 15.05
Forecast 5	– / 0.64	– / 28.95

Tables 13 and 14 present the association metric (Corr-f) and the economic metric (MPD) under volatile and normal market conditions. The results indicate that the Corr-f remains relatively high regardless of the market condition. In contrast, MPD decreases significantly in the normal market conditions for all models. Forecast 4 performs best across most metrics, with the exception of the 15-minute MPD under normal conditions, followed by Forecast 3 and Forecast 1.

The results indicate that while capturing the daily profile of the forecast (Corr-f) is relatively stable under both market conditions, the economic impact of forecast errors (MPD) is highly dependent on market volatility. Under volatile conditions, even small deviations in identifying the timing of extreme price periods can lead to substantially larger economic losses.

Interestingly, Corr-f values are slightly higher under volatile market conditions. This could be explained by the fact that stronger price variations make the relative ordering of prices more distinct, which is easier for models to capture even if the absolute errors are larger.

4.6 Comparison Across Forecasts

This section focuses on identifying the best overall forecast by combining the different evaluation metrics. Thus, an illustrative composite performance score was constructed in order to capture all metrics within a single value. The purpose of this score is not to replace the metric-level analysis presented in the previous sections, but to complement it by enabling a consistent view across multiple evaluation dimensions.

The composite score combines all three categories of metrics including classical error metrics, ranking-based metrics, and association and economic metrics. Each metric is first normalised to a scale from 0 to 1 across the forecast providers meaning that the best metric value is 1 and the worst is 0. The normalisation for metrics where higher value is preferred is defined as:

$$x^{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)},$$

whereas for metrics where lower values are preferred:

$$x^{\mathrm{norm}} = \frac{\max(x) - x}{\max(x) - \min(x)}.$$

After normalisation, category-level scores are calculated as the average of the normalised metrics within the category. Finally, the composite score is then found as an equally weighted average of the three category scores, meaning that each category of metrics contributes one third of the overall score.

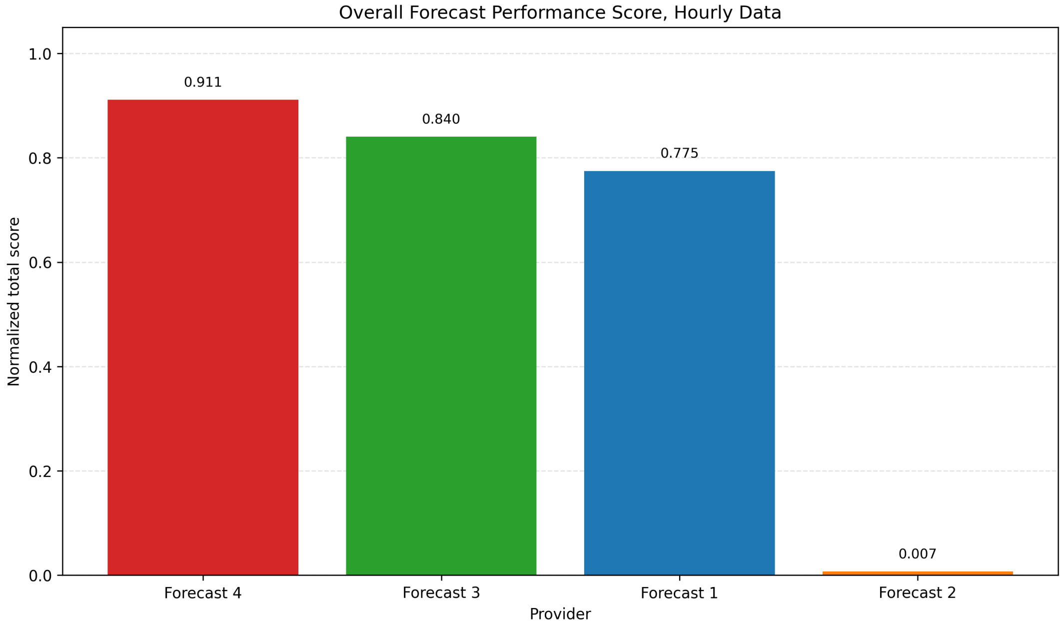

Figure 8: Illustrative composite forecast performance scores for the hourly resolution period.

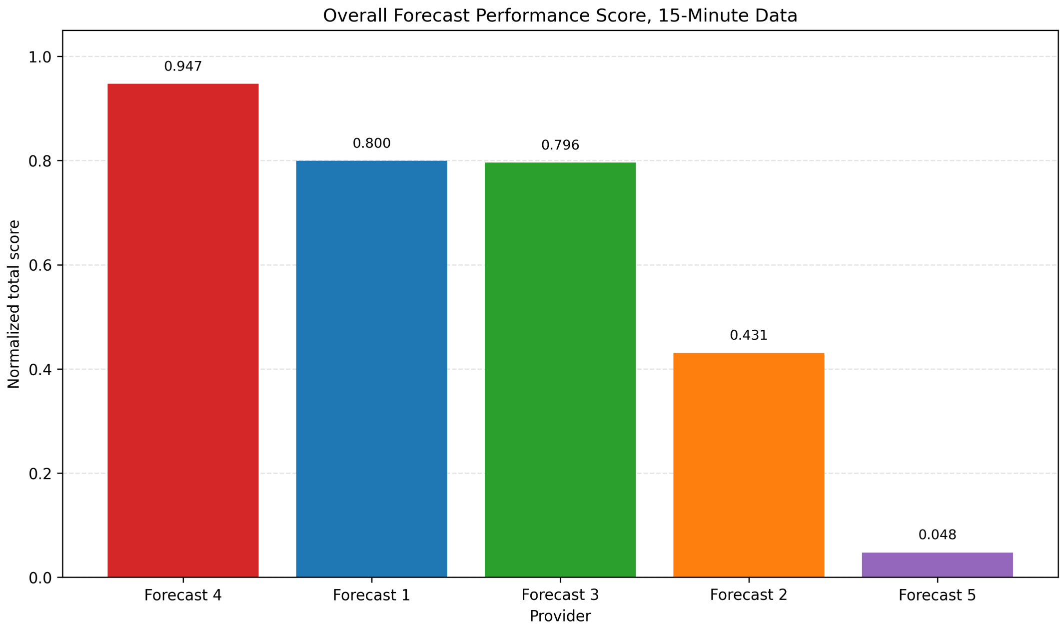

Figure 9: Illustrative composite forecast performance scores for the 15-minute resolution period.

Figures 8 and 9 show the composite performance scores for the hourly and 15-minute test periods. In both cases, Forecast 4 achieves the highest score. This indicates strong overall performance across multiple evaluation dimensions. Forecast 3 and 1 perform relatively well in both time frames. However, their ranking slightly changes between the two time resolutions, indicating that the relative ranking of these forecasts depends on the selected evaluation period and resolution. Forecast 2 presents clearly weaker performance, while Forecast 5 is evidently the weakest in the 15-minute comparison.

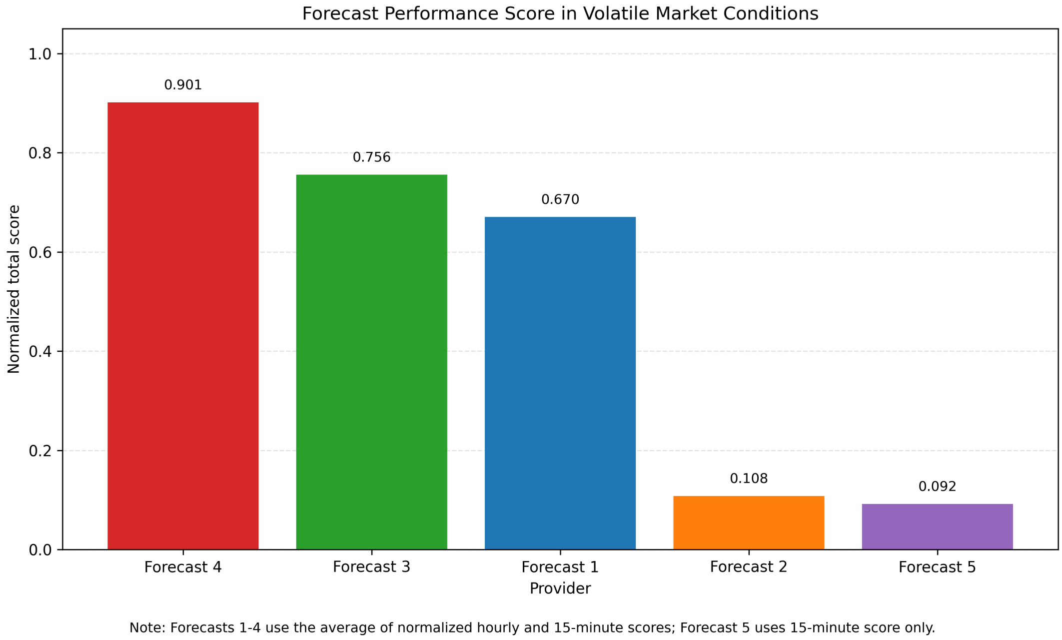

Figure 10: Illustrative composite forecast performance scores under volatile market conditions. Forecasts 1–4 use the average of normalised hourly and 15-minute scores, while Forecast 5 uses only the 15-minute score.

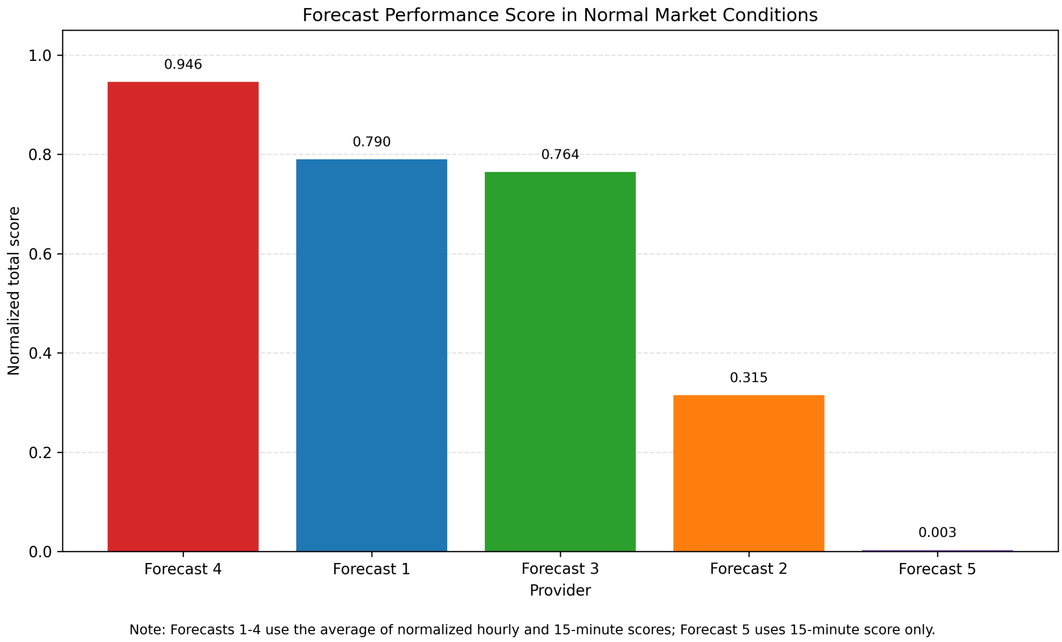

Figure 11: Illustrative composite forecast performance scores under normal market conditions. Forecasts 1–4 use the average of normalised hourly and 15-minute scores, while Forecast 5 uses only the 15-minute score.

The market condition -based comparison results shown in Figures 10 and 11 further support these findings. Forecast 4 maintains the highest score in both volatile and normal market conditions, indicating robustness for different market situations. Forecast 3 performs particularly well under volatile market conditions, while Forecast 1 achieves slightly better performance under normal market conditions. Forecasts 2 and 5 remain clearly weaker in both market conditions.

Overall, the composite score forecast analysis supports the conclusion that Forecast 4 provides the strongest and most consistent overall performance, followed by Forecasts 3 and 1. Nevertheless, the composite score should not be interpreted as an absolute ranking of forecast providers as the score depends on the normalisation method and weighting. Instead, the results strengthen the individual results that have been found in previous analysis.

5 Conclusions and Future Work

5.1 Key Findings

The objective of this thesis was to evaluate and compare different day-ahead electricity price forecasts from the perspective of multi-market optimization. The evaluation and analysis were conducted not only through traditional prediction error metrics, but also through ranking-based, association, and economic evaluation metrics.

The results indicated that Forecast 4 achieved the strongest and most consistent overall performance across multiple evaluation dimensions. Forecasts 3 and 1 also performed relatively well, even though their performance varied more depending on the evaluation settings and market conditions. Forecasts 5 and 2 showed clearly the weakest overall performance.

The results showed also that volatile market conditions increase the forecasting difficulty, especially in peak detection. In contrast, detecting the low-price periods was generally easier to identify. Furthermore, the transition to 15-minute resolution has had a clear impact on forecasting performance, as both peak detection and MPD results became weaker.

5.2 Limitations of the Study

There are some limitations and constraints related to this study, and these should be taken into account when interpreting the results. First, the 15-minute resolution time period covers a significantly shorter time period than the hourly resolution dataset. In addition, the 15-minute test period mainly covers the winter season, and therefore it may affect the comparability of the two temporal resolutions. Since the day-ahead electricity market has strong seasonal characteristics throughout the year, the study is not able to provide evidence on how the forecasts would perform, for example, during the summer season.

Secondly, the main goal of this study was to evaluate and compare different day-ahead electricity price forecasts from the perspective of multi-market optimization. However, the evaluation was conducted through selected evaluation metrics, and therefore the practical benefits for multi-market optimization were evaluated indirectly. Consequently, there is no certainty about how these metrics correlate with the actual optimization results.

5.3 Future Research Directions

The limitations identified previously also provide several directions for future research. In particular, longer evaluation periods, especially for 15-minute resolution periods, would improve the generalisability of the results across different seasonal and market conditions.

In addition, future research could evaluate forecast performance directly from multi-market optimization results. This could be done by backtesting the optimization algorithm with different day-ahead forecasts and then analysing the final economic values.

Future studies could also investigate whether combining multiple forecasts through ensemble forecasting methods would improve the performance compared to individual forecast providers.

Finally, the evaluation could be extended by developing additional forecast evaluation metrics that better capture the practical operational value of forecasts.

References

[1] Energiavirasto, *National Report on Electricity and Gas Markets in 2024 in Finland*. Available: https://energiavirasto.fi/documents/11120570/13026619/National+Report+on+electricity+and+gas+markets+in+2024+in+Finland.pdf (accessed on 16 January 2026).

[2] Ministry of Trade and Industry, Finland, *Electricity Market Act (386/1995)*. Available: https://www.finlex.fi/en/legislation/translations/1995/eng/386 (accessed on 16 January 2026).

[3] P.-O. Pineau and R. Hämäläinen, "A dynamic analysis of the Finnish electricity market," *Energy Policy*, vol. 27, no. 8, pp. 459–472, 1999. Available: https://sal.aalto.fi/publications/pdf-files/mpina.pdf (accessed on 16 January 2026).

[4] OECD, *Finland: Energy Policy Review 2023*. OECD Publishing, 2023. Available: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/05/finland-2023-energy-policy-review_68830904/d435fa51-en.pdf (accessed on 16 January 2026).

[5] Fingrid Oyj, *Johdanto sähkömarkkinoihin*. Available: https://www.fingrid.fi/sahkomarkkinat/markkinoiden-yhtenaisyys/johdanto-sahkomarkkinoihin/ (accessed on 22 January 2026).

[6] Agency for the Cooperation of Energy Regulators (ACER), *Market Rules for Different Electricity Market Timeframes*. Available: https://www.acer.europa.eu/electricity/market-rules/market-rules-different-electricity-market-timeframes (accessed on 22 January 2026).

[7] MCSC and SIDC, *Information Package About Intraday Auction (IDA)*. European intraday market design documentation. Available: https://www.nemo-committee.eu/assets/files/information-package-about-intraday-auction.pdf (accessed on 24 February 2026).

[8] Nord Pool Group, *Bidding Areas*. Nord Pool market information. Available: https://www.nordpoolgroup.com/en/the-power-market/Bidding-areas/ (accessed on 24 February 2026).

[9] Nord Pool Group, *Yearly Volumes 2025*. Nord Pool annual volume statistics. Available: https://www.nordpoolgroup.com/48f6fa/globalassets/trading-and-services/nord-pool-yearly-volumes-2025.pdf (accessed on 24 February 2026).

[10] Nord Pool ASA, *History of Nord Pool*. Available: https://www.nordpoolgroup.com/en/About-us/History/ (accessed on 23 February 2026).

[11] P. Pinson, *Renewables in Electricity Markets*, Module 2, Section 2.1. Technical University of Denmark (DTU), 2025. Available: https://pierrepinson.com/31761/Slides/31761lecture2p1.pdf (accessed on 12 September 2025).

[12] NEMOs Committee, *EUPHEMIA Public Description: Single Price Coupling Algorithm*, Version 10.3, 2023. Available: https://www.nemo-committee.eu/assets/files/euphemia-public-description.pdf (accessed on 24 February 2026).

[13] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014. doi: https://doi.org/10.1016/j.ijforecast.2014.08.008.

[14] ENTSO-E, *Electricity Balancing Guideline*. European Network of Transmission System Operators for Electricity, 2017.

[15] Fingrid Oyj, *Reserve Market Day 2025: English Summaries of the Presentations*. Available: https://www.fingrid.fi/globalassets/dokumentit/en/electricity-market/reserves/reservimarkkinapaiva-2025-eng.pdf (accessed on 24 February 2026).

[16] Fingrid Oyj, *Reserve Markets*. Fingrid, 2025. Available: https://www.fingrid.fi/en/electricity-market/reserves/reserve-markets/ (accessed on 20 March 2026).

[17] Fingrid Oyj, *Wind and Solar in the Nordic Reserve Markets*. Nordic TSOs, 2023. Available: https://www.fingrid.fi/globalassets/dokumentit/fi/tiedotteet/ajankohtaista/julkaisu-wind-and-solar-in-the-nordic-reserve-markets-id-431541.pdf (accessed on 20 March 2026).

[18] M. Aghamohammadloo et al., *Multi-market Optimization of Hybrid Energy Storage Systems*. CIRED Workshop, 2023. Available: https://lutpub.lut.fi/bitstream/handle/10024/170655/aghamohammadloo_et_al_multi-market_optimization_aam.pdf?sequence=3 (accessed on 2 April 2026).

[19] K. Maciejowska, A. Lipiecki, and B. Uniejewski, "Statistical and economic evaluation of forecasts in electricity markets: Beyond RMSE and MAE," *Energy Conversion and Management*, vol. 306, Art. no. 121408, 2026. doi: https://doi.org/10.1016/j.enconman.2026.121408.

[20] Nord Pool Group, *Day-Ahead Auctions Market Data*. Nord Pool power market data services. Available: https://www.nordpoolgroup.com/en/services/power-market-data-services/day-ahead-market-data/ (accessed on 10 May 2026).

[21] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013, ch. 5.

[22] G. Shmueli and K. C. Lichtendahl Jr., *Practical Time Series Forecasting with R: A Hands-On Guide*, 2nd ed. Axelrod Schnall Publishers, 2016.

[23] L. Beilina, E. Karchevskii, and M. Karchevskii, *Numerical Linear Algebra: Theory and Applications*. Springer, 2013.

[24] S. Zhan and A. Chong, "Data requirements and performance evaluation of model predictive control in buildings: A modeling perspective," *Energy and Buildings*, vol. 245, Art. no. 111064, 2021. doi: https://www.sciencedirect.com/science/article/pii/S1364032121001295.

[25] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006. doi: https://doi.org/10.1016/j.ijforecast.2006.03.001.

[26] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. OTexts, 2021. Available: https://otexts.com/fpp3/ (accessed on 10 May 2026).