M.Sc in Energy Systems
Hellenic Mediterranean University



Data-Driven Forecasting of Price Formation in the Greek Day-Ahead Electricity Market

Marios Nikologiannis

Supervisor: Prof. Emmanuel Karapidakis

Hellenic Mediterranean University

Heraklion, 2026

CONTENTS

List of figures 1

List of tables 1

Abstract 1

1. Introduction 2

2. Electricity Market in Greece 2

2.1. Market Structure 3

2.1.1. Nominated Energy Market Operator and Clearing House 3

2.1.2. Transmission System Operator 4

2.2. Market Participants 4

2.2.1. Direct and General Clearing Members 4

2.2.2. Types of Participants 4

2.3. Bidding Order Types 5

2.3.1. Priority Price-Taking Orders 5

2.3.2. Hourly Hybrid Orders 7

2.3.3. Block Orders 9

2.3.4. Import and Export Orders 10

3. Time-Series Forecasting 12

3.1. Methodology for Block Orders 12

3.1.1. Problem Formulation 12

3.1.2. Ridge Regression 13

3.1.3. Deterministic Regression Design 15

3.1.4. Residual Modelling 16

3.1.5. Final Forecast and Evaluation 17

3.2. Methodology for Hybrid Orders 18

3.2.1. Data Preparation 18

3.2.2. Hybrid Model Formulation 19

3.2.3. Functional Principal Component Analysis (FPCA) 19

3.3. Methodology for Implicit Orders 25

3.3.1. Overview of Implicit Import Orders 25

3.3.2. Markov Model 26
3.3.3. Logit Model 27
3.3.4. Magnitude Model 29
3.3.5. Holdout Method 29
4. Results 29
4.1. Block Orders Forecasts 29
4.1.1. Forecasting Setup for August 29
4.1.2. Forecasting Setup for September 32
4.2. Hybrid Orders Forecasts 35
4.2.1. Forecasted Hybrid Orders for August 35
4.2.2. Forecasted Hybrid Orders for September 37
4.2.3. Overview of FPCA Forecast Performance 40
4.3. Implicit Import Orders Forecasts 40
4.3.1. Optimal Configuration for August 40
4.3.2. Optimal Configuration for September 42
4.3.3. Conclusions and Future Work 44
4.3.4. Comparison with the Fourier-ARIMA Hybrid Framework 45
4.4. MCP Forecasts 45
4.4.1. MCP Forecasts of August 45
4.4.2. MCP Forecasts of September 47
5. Conclusion 48
Bibliography 50

LIST OF FIGURES

Figure 1. Greek Energy Spot Market Structure. 3
Figure 2. Flow of Forward Contracts into DAM Priority Orders. 6
Figure 3. Hourly Hybrid Order curve. 7
Figure 4. Acceptance Cases for a Hybrid Sell Order. 8
Figure 5. Supply and demand aggregated hourly hybrid curves. 8
Figure 6. Daily aggregated supply hybrid curves. 9
Figure 7. Example of a Sell Block Order. 9
Figure 8. Impact of Block Orders on DAM Clearing on 28/09 00:00. 10
Figure 9. Map of Greece's Interconnections [14]. 11
Figure 10. Accepted Block Orders from January to September 2025. 13
Figure 11. Original vs Fitted aggregated hybrid order curve. 18
Figure 12. Fitted Price Grid time-series for (a) January and (b) September 2025. 19
Figure 13. Implicit Import Orders from January to September 2025. 25
Figure 14. Accepted Block Orders Forecasting for August 31
Figure 15. Accepted Block Orders Metrics for August. 31
Figure 16. Accepted Block Orders Forecasting – September. 33
Figure 17. Accepted Block Orders Metrics for September. 34
Figure 18. Hybrid Forecast Heatmap Error for August. 35
Figure 19. Indicative Hybrid Curve Forecasts in August. 36
Figure 20. Hybrid Forecast Heatmap Error for September. 37
Figure 21. Indicative Hybrid Curve Forecasts in September. 38
Figure 22. Implicit Import Orders Forecasting for August. 42
Figure 23. Implicit Import Orders Forecasted for September. 44
Figure 24. Forecasted and actual MCP in August. 46
Figure 25. Forecasted and actual MCP in September. 47

LIST OF TABLES

Table 1. Categories of Market Participants. 4
Table 2. Interconnections between Countries that submit Import Orders. 11
Table 3. Table Format of Fitted Hybrid Orders. 18
Table 4. Forecast Parameters for August Block Orders. 30
Table 5. Forecasting Metrics for August Block Orders. 32
Table 6. Forecast Parameters for September Block Orders. 32
Table 7. Forecasting Metrics for September Block Orders. 34
Table 8. Optimal number of AIC parameters for August. 36
Table 9. Hybrid Order Metrics for August. 37
Table 10. Optimal number of AIC parameters for September. 39
Table 11. Hybrid Order Metrics for September. 39
Table 12. Markov-Logit Parameters for August 40
Table 13. Markov-Logit Parameters for September 42

Marios Nikologiannis
Student ID No.: MM176

ABSTRACT

This thesis presents a data-driven framework for modelling and forecasting price formation in the Greek Day-Ahead Electricity Market through the structural reconstruction of historical data from accepted key energy generation order types. Instead of directly predicting the Market Clearing Price (MCP), the proposed approach decomposes the auction mechanism into its fundamental components, granting valuable insights regarding market behaviour. The orders' energy generation categories are Block Orders, Hourly Hybrid Orders and Implicit Import Orders, and each category is forecasted through different techniques, based on their time-series properties. Ridge regression and Autoregressive Integrated Moving Average (ARIMA) are used for Block Orders, hybrid order curves are forecasted through Functional Principal Component Analysis (FPCA) that capture their high-dimensional temporal structure, while Implicit Import Orders are forecasted through a Markov-LASSO/Logit model that is useful in capturing the time-series activation-like behaviour and traded energy magnitudes. The individual forecasts are recombined to produce the MCP forecast, which is compared with the recorded data for the time interval under study.

Keywords: Day-Ahead Electricity Market, Greek Electricity Market, Electricity Market Design, Price Formation, Electricity Price Forecasting, Hourly Hybrid Orders, Block Orders, Implicit Orders, Time-Series Forecasting, Data-Driven Modelling

1. INTRODUCTION

The transformation of electricity systems under increasing renewable energy penetration has fundamentally altered the dynamics of wholesale electricity systems. Liberalised electricity markets, particularly those operating under the European Target Model, rely on auction-based mechanisms to balance supply and demand while ensuring economic efficiency and adequate grid capacity allocation. In this context, accurate modelling and forecasting of price formation has become essential for market participants, system operators, and policymakers.

The DAM constitutes the cornerstone of wholesale electricity trading in Europe. Even though the vast majority of energy volumes that are traded by market participants are in the form of forwards, bilateral contracts, etc., the DAM price serves as the main metric of risk management in energy markets [1]. Prices in the DAM are determined through a price auction, where supply and demand orders are accepted and matched within network and regulatory constraints [2]. This auction process is influenced by each participant's bidding strategy, which integrates the existing bidding mechanisms, in the form of Long-Term Physical Transmission Rights (LT-PTRs), Priority Price-Taking Orders (PPTs), Block Orders, Hybrid Orders and Implicit or Explicit Import Orders.

In the Greek Electricity Market, the implementation of the European Target Model introduced significant structural changes, including market coupling and cross-zonal capacity allocation. Beside direct price forecasting ([3], [4]), structural approaches explicitly modelling supply stacks and demand interactions have gained prominence in recent years, particularly in capacity-constrained electricity systems them [5].

This thesis proposes a structural, data-driven methodology for forecasting price formation in the Greek DAM. The methodology focuses on the forecasting of energy volumes submitted and accepted under separate order types that are available to participants responsible for electricity generation. The combination of the separate forecasts constitutes the overall forecasted hourly energy generation, and its intersection with the recorded hourly energy demand leads to the forecasted hourly MCP. The forecasting period is August and September of 2025, the last months of hourly price formation before the implementation of the 15-minute Market Time Unit (MTU) [6].

The remainder of this thesis is followed by Chapter 2, where the regulatory framework of the Greek electricity market is presented, and the order types that are available in the DAM Spot Market are explained. Chapter 3 describes in detail the methodology for the forecasting process of each time-series under study, while Chapter 4 compiles the methodology's results, including the final MCP forecasts. Finally, Chapter 5 concludes with the thesis' key findings and directions for future research.

2. ELECTRICITY MARKET IN GREECE

The architecture of the Greek DAM is designed in accordance with the European Target Model, particularly under Regulation (EU) 2015/1222 [7], which establishes the guideline on Capacity Allocation and Congestion Management. This regulatory framework harmonises Day-Ahead and Intra-Day market coupling across European Union member states and defines the responsibilities of Nominated Electricity Market Operators (NEMOs), Transmission System Operators (TSOs) and clearing entities.

2.1. MARKET STRUCTURE

The Greek market clearing of submitted energy volume orders in the DAM is not an isolated national optimisation process, but part of a multiregional price coupling process under Single Day-Ahead Coupling. Articles 36-39 of the Regulation (EU) 2015/1222 introduce the Market Coupling Operator, who is responsible for developing and operating the Single Day-Ahead and Intra-Day Coupling algorithms. Order clearing determines the price setting of the energy commodities that are traded in each Bidding Zone, in full cooperation between the NEMOs, the Clearing House, the TSOs, and the market participants, as organised in Figure 1.

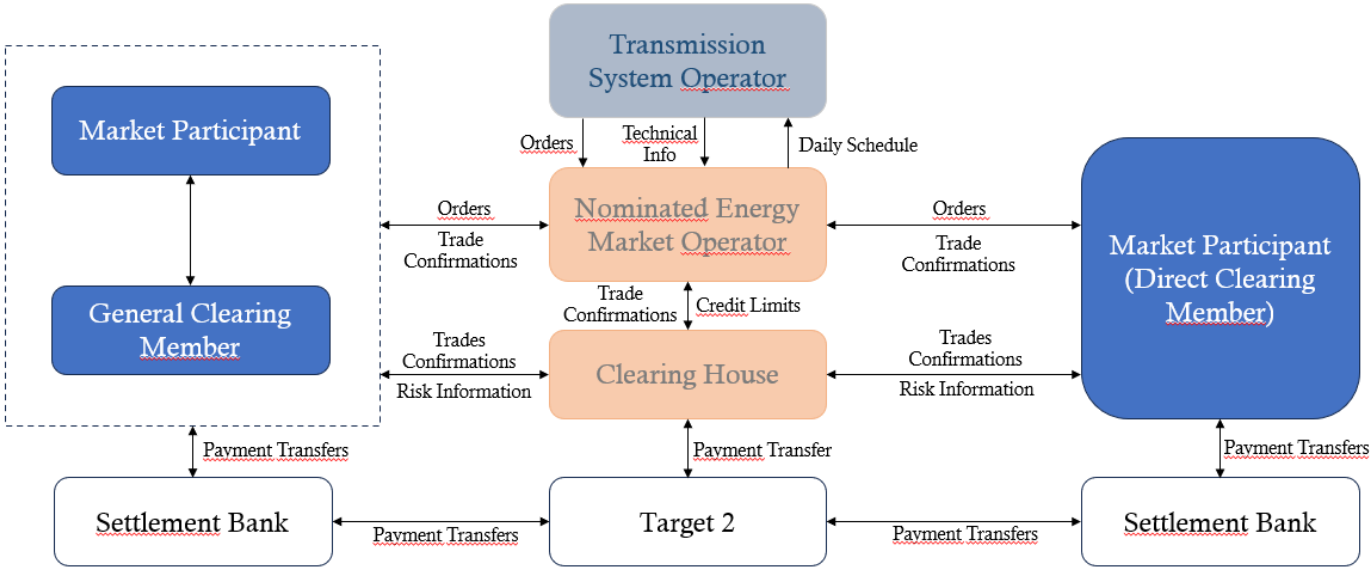


Figure 1. Greek Energy Spot Market Structure.

The chart describes the data exchange between all parties of the Greek electricity market, as well as the currency exchange that takes place after the price settlement. Cooperation between the TSO, the NEMO, and the Clearing House is essential for the calculation of the MCP, the physical transmission of energy products through the national electricity grid, and the compensation of market participants.

2.1.1. NOMINATED ENERGY MARKET OPERATOR AND CLEARING HOUSE

Under Article 7(1) of Regulation 2015/1222, NEMOs are responsible for receiving and processing orders, matching them using a social welfare maximisation algorithm, EUPHEMIA [8], and publishing market coupling results. In Greece, HEnEx S.A. functions as the NEMO, which both manages and operates the DAM, the Intra-Day Markets (IDAs) and the Cross-Border Intra-Day Market (XBID), a continuous trading process across Bidding Zones within the EU. The HEnEx Spot Trading Rulebook aligns with this requirement by assigning HEnEx responsibility for market operation and the management of the Energy Trading Spot System (ETSS), a secure platform where the submission of orders by participants takes place.

According to Section 2.1.4 of the Trading Rulebook, the Clearing House is responsible for the clearing and financial settlement of the trades that are concluded in the Spot Markets, specifying that the Clearing House may act as a Shipping Agent within the context of Regulation (EU) 2015/1222 for transferring Net Positions in coupled markets, based on the trading records transmitted by HEnEx. The daily cash settlement of financial obligations and receivables is executed via TARGET-GR, the payment system of the European

Central Bank, with the involvement of EnExClear, the Clearing Members, and the designated Settlement Banks [9]. EnExClear serves as the Clearing House and is part of EnEx Group, along with HEnEx.

2.1.2. TRANSMISSION SYSTEM OPERATOR

The Hellenic Transmission System Operator (HTSO), commonly known as the Independent Power System Operator (IPTO), performs the physical settlement of accepted orders and system-level balancing actions through the Balancing Market. According to Section 2.1.4 of the Trading Rulebook [10], the clearing and settlement of Physical Deliveries that result from the Spot Markets is performed by the HTSO. The HTSO is also responsible for the provision of cross-zonal capacities and coordinating congestion management within the electricity grid and the capacity calculation, which is necessary for the trade and settlement of Long-Term Physical Transmission Rights (LT-PTRs) and Explicit Import-Export Orders.

2.2. MARKET PARTICIPANTS

2.2.1. DIRECT AND GENERAL CLEARING MEMBERS

According to Section 2.1.3 of the Trading Rulebook, access to HEnEx Markets is granted to registered Participants through the ETSS. In order to acquire the Participant capacity, Section 3.3 dictates that the approval by the Clearing House to act as a Direct Clearing Member is required, or the submission of a certificate of cooperation with a General Clearing Member. Under Section 2.7, a Participant acts as a Direct Clearing Member if they fulfil their obligation directly towards the Clearing House and maintain adequate Credit Limits that are subject to real-time monitoring through the ETSS. On the other hand, a General Clearing Member cooperates with Participants that are not Direct Clearing Members and clears electricity trades carried out by the contracted Participants on the Electricity Markets.

2.2.2. TYPES OF PARTICIPANTS

Both Energy Producers and Energy Suppliers can participate in the DAM. Different types of electricity generation technologies are represented in the Market, as well as various types of electricity buyers and suppliers. Table 1 summarises the available Participants and their respective trading products:

Table 1. Categories of Market Participants.

Participant	Asset
Electricity Producers	Generating Unit Pumping Unit
RES producers/ RES Aggregators Last Resort RES Aggregators	RES Portfolio (Dispatchable and Non-dispatchable)
Suppliers/ Self-supplying Consumers	Load Portfolio (LV, MV, HV)

Traders/ Suppliers/ Self-supplying Consumers	(Electricity Grid Interconnection)
IPTO	Mandatory Hydro Injections System Losses
RES and GO Operator	RES FiT Portfolio Rooftop Photovoltaics Dispatchable, High-Efficiency CHP Unit

Most of the Table's content is self-explanatory, except for IPTO's Mandatory Hydro Injections and RES and GO Operator. IPTO has to submit these orders, and they must be accepted, since they represent energy volumes produced by hydropower stations that need to lower their water reserve levels to prevent overflow and/or assume daily, necessary water services (irrigation, etc.).

The role of RES and GO (Renewable Energy Sources and Guarantee of Origin) Operator is to oversee and maintain a registry of renewable energy installations and the issuance, transfer, or cancellation of GOs. A GO is an electronic certificate which proves that a certain energy volume has been produced from a renewable source. This certificate is transferable and can be traded independently of the physical energy transmission itself. In Greece, DAPEEP S. A. is this Operator [11], and, in accordance with Section 2.2.3 of the Trading Rulebook, they provide to HEnEx the necessary information from this registry for settlement and scheduling processes and submit Priority Price-Taking Orders to the ETSS.

2.3. BIDDING ORDER TYPES

2.3.1. PRIORITY PRICE-TAKING ORDERS

Both forward and derivative contracts with physical delivery obligations must be nominated into the ETSS before the DAM gate closure. These are transformed into Priority Price-Taking (PPT) Orders by HEnEx, ensuring they are executed in the DAM order book. As such, they do not set the price directly, but they shift net positions (imports, exports, delivery vs offtake). As a result, they contribute to the supply/demand balance and can indirectly push the MCP by changing the energy volumes the algorithm clears.

According to Chapters 6.4-6.8 of the Rulebook, the Central Counter Party registers trades in Energy Financial Instruments (forwards, futures, OTC contracts) that are selected for Physical Settlement and are registered in the ETSS (for exchange-traded derivatives), or by the Participants themselves (for OTC contracts). They must nominate these quantities as Physical Delivery (Sell) or Physical Offtake (Buy) positions for each Market Time Unit.

Section 4.4.2.3 dictates that there are limits on how much of a supplier's portfolio can be covered by physically settled derivatives relative to DAM trades. If exceeded, non-compliance charges are imposed. This prevents suppliers from relying too heavily on forwards to circumvent competitive DAM bidding. It should be noted that Crete's priority orders are excluded from this calculation. Figure 2 below showcases the settlement process of the derivative and contracts.

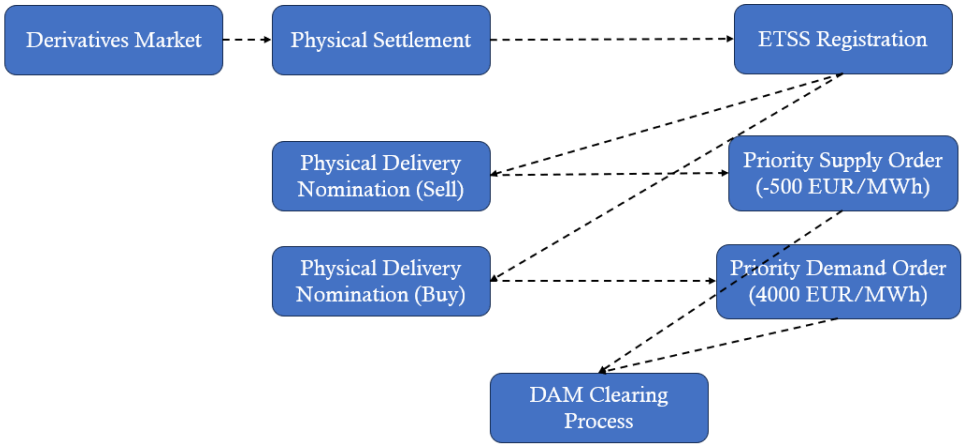


Figure 2. Flow of Forward Contracts into DAM Priority Orders.

The Derivatives trade is concluded on the Energy Derivatives Market (or OTC). During the Physical Settlement Choice, the Participant opts for physical delivery/offtake. For the ETSS registration, the contract is registered in the Energy Trading Spot System. The branches show that the Physical Delivery Nomination becomes a Priority Supply Order at the minimum price, while the Physical Offtake Nomination (Buy) becomes a Priority Demand Order at the maximum price. Both feed into the DAM Clearing Algorithm (EUPHEMIA), which shifts net supply/demand but does not set the MCP directly.

All participants with delivery or offtake obligations (generators, RES aggregators, consumers, LT-PTR holders) must nominate their quantities for Day D. These nominations are validated and then inserted into DAM as Priority Price-Taking Orders (sell for delivery, buy for offtake). Consequently, they are always accepted at minimum or maximum price levels. They ensure that physical obligations are honoured, but since they clear at extremes, they typically do not set the MCP unless the market is at its floor or ceiling.

Section 4.1.3.2 of the Rulebook clarifies that at the minimum price, PPT Supply (Sell) Orders are submitted by the HTSO, RES and GO Operator, Last Resort Aggregator, and HEnEx itself on behalf of participants. They cover Mandatory Hydro Injections, Commissioning units, Transmission losses, Feed-in RES (FiT, roof PVs, non-dispatchable RES), Physical delivery/offtake nominations, Imports/exports from LT-PTRs. Mandatory Hydro Injections are declared by hydro operators (for irrigation, overflow, ecological flows, etc.). Like RES FiT, they depress the MCP if demand is not enough, especially in low-load hours. RES and GO operator submits the forecasted production of FiT PVs, roof PVs, and other portfolios. They enter the DAM outside the bidding curve, at the minimum price, ensuring their priority dispatch.

On the other hand, the HTSO submits transmission losses of the Greek electricity grid, consumption of customers assigned to the Last Resort Supplier or the Universal Service Supplier, and auxiliary consumption of RES and CHP units included in its portfolios (where applicable) as PPT Demand (Buy) Orders. HEnEx, on behalf of Participants, submits Physical Offtake Nominations from forward/derivative contracts with physical settlement, Imports, or Exports from LT-PTRs (when nominated as offtake).

Paragraphs 10-12 of Section 4.1.3.2 state that if MCP is equal to the price floor with no competitive sell orders, the priority price is curtailed, while if MCP is equal to the price ceiling with no competitive buy orders, priority demand is curtailed. The curtailment procedure is dictated by the Regulatory Authority for Energy, Waste and Water (RAWEW).

Until 2021, Crete operated as an isolated system with its own balancing rules. Its production and consumption became part of the HEnEx DAM after the completion of the interconnection of Crete-

Peloponnese. Paragraphs 7 and 8 of Section 4.1.3.2 dictate that the RES and GO Operator submits non-dispatchable RES in Crete (mainly PVs and wind under FiT or FiP) as PPT Sell orders, while dispatchable generation in Crete (conventional units, dispatchable RES portfolios) participates normally in the DAM, through competitive orders.

Paragraph 9 of the same Section clarifies that HEnEx also sublits the forecasted consumption of Crete on behalf of suppliers. These buy-side priority orders are always cleared, unless the DAM MCP hits the price ceiling, in which case priority demand curtailment is applied. They do not set the MCP directly, but they increase effective demand and can push the clearing point upward.

2.3.2. HOURLY HYBRID ORDERS

Hourly Hybrid Orders are submitted by each participant for every MTU. According to Section 4.1.3.1, they are monotonic curves, increasing for Sell Orders and decreasing for Buy Orders, consisting of energy volume and their respective price points, and up to 50 segments per curve. In Figure 3, the curve's elements are illustrated:

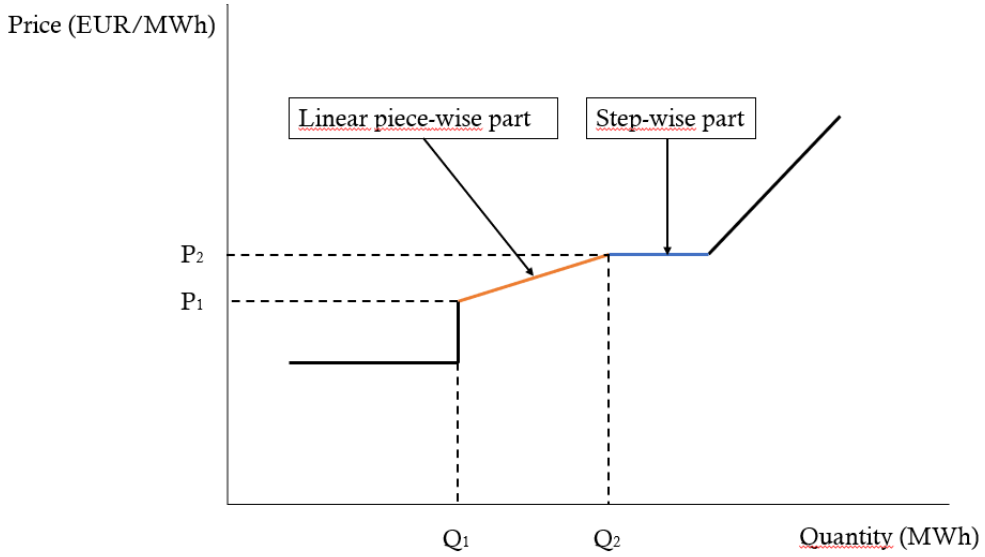


Figure 3. Hourly Hybrid Order curve.

According to Section 4.3.3, in order to be fully accepted, the Buy order curve's price segments must be higher than the MCP, while the Sell curves must be lower, while if a segment's price equals the MCP, it is partially accepted. Figure 4 illustrates the various acceptance cases for a segment of a Sell Curve Order.

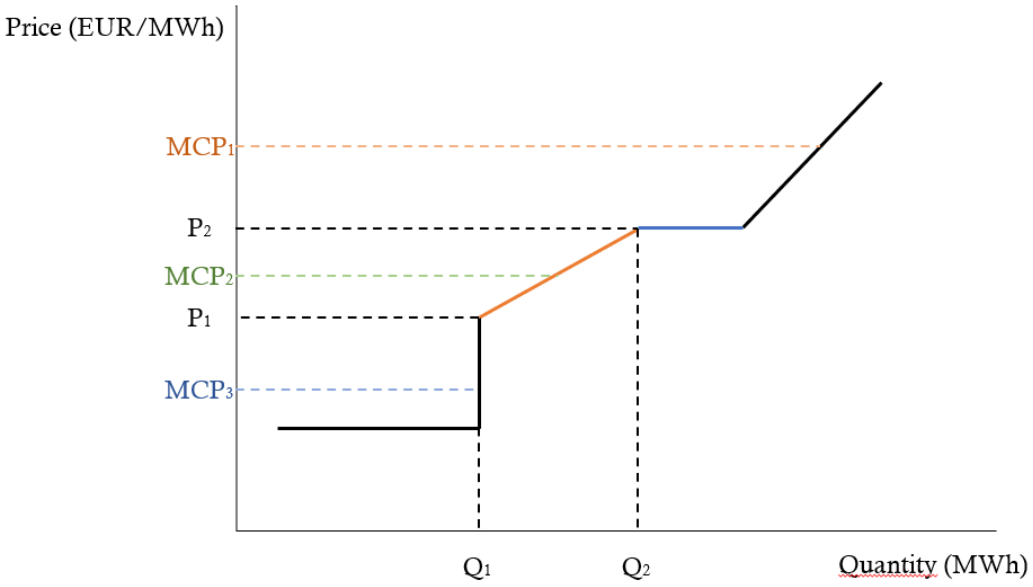


Figure 4. Acceptance Cases for a Hybrid Sell Order.

In this case, the linear segment is fully accepted for MCP_1, partially accepted for MCP_2, and rejected for MCP_3.

The aggregated accepted Buy/Sell curves from participants are used in the DAM clearing process. They combine all valid orders, partially or fully accepted for each MTU and include both price-sensitive competitive bids and PPT Orders. Their energy volumes and the corresponding cumulative price are combined, regardless of energy generation technology or energy demand type. Figure 5 below showcases the supply and demand hybrid curves for an indicative hour:

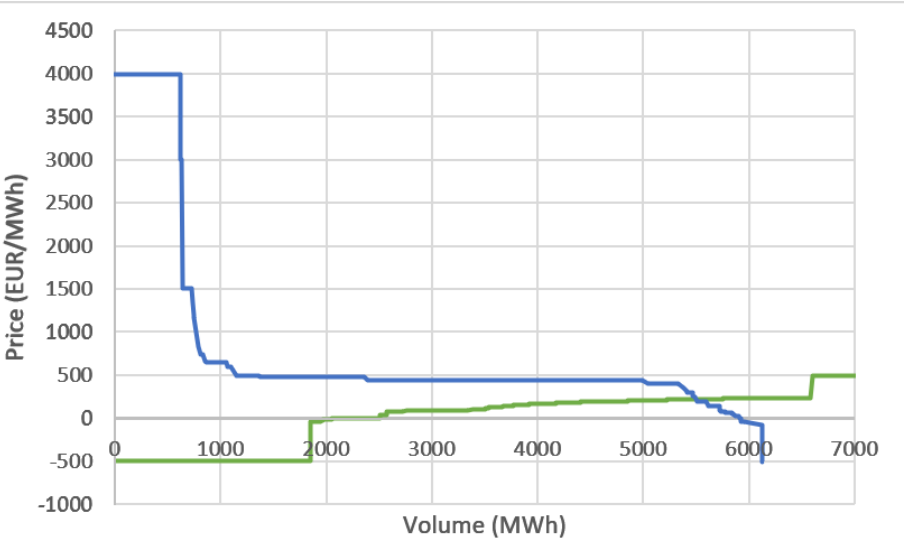


Figure 5. Supply and demand aggregated hourly hybrid curves.

The figure shows the supply and demand curves for September 1st, 2025, 00:00. The Figure shows that the PPT Sell orders at -500 EUR/MWh were a little less than 2000MWh, while the PPT Buy orders at 4000 EUR/MWh were less than 1000 MWh. The intersection is at 201 EUR/MWh, yet the MCP was lower, equal to 88.39 EUR/MWh, a paradox that will be explained in the next section.

Figure 6 shows the variability between curves of different hours.

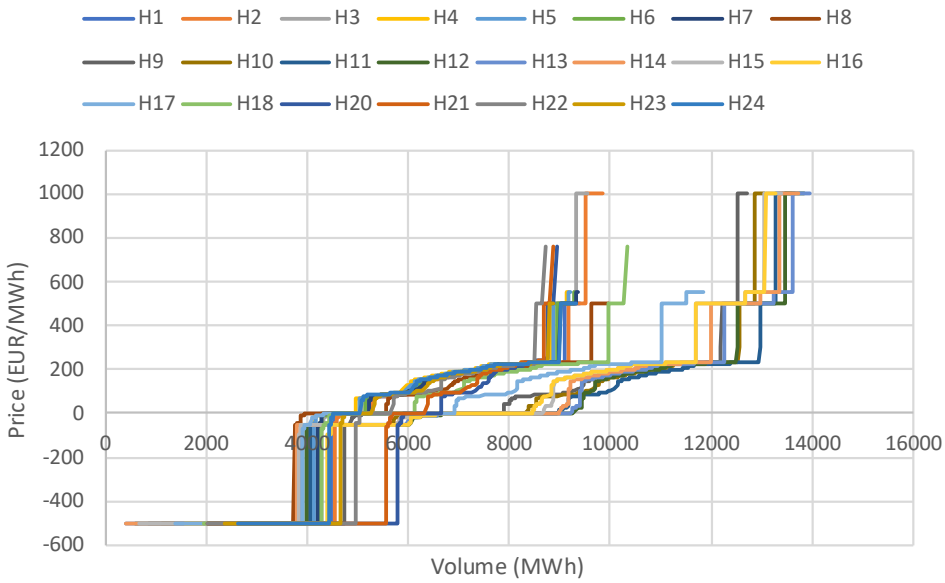


Figure 6. Daily aggregated supply hybrid curves.

For the specific day, the edges, corresponding to different PPT orders, which fluctuate within the day, yet there is a similarity in competitive orders so that each hour can be described by a displaced curve.

2.3.3. BLOCK ORDERS

Block Orders are, according to Section 4.1.3.1, bids that are characterised by the quantity of energy for sale/purchase, the quantity's price, the MTU (or multiple MTUs), and a Minimum Acceptance Ratio (MAR), where $0 \leq MAR \leq 1$. The energy quantity that can be accepted must be of an Acceptance Ratio (AR) greater than the MAR of the submitted quantity, while the MAR must be the same for all MTUs of the same Block Order. Block Orders introduce non-convexities in the market clearing problem, complicating the welfare maximisation process and potentially causing price paradoxes [12]. Figure 7 illustrates an example of a Sell Block Order spanning across four MTUs with the same price P and MAR.

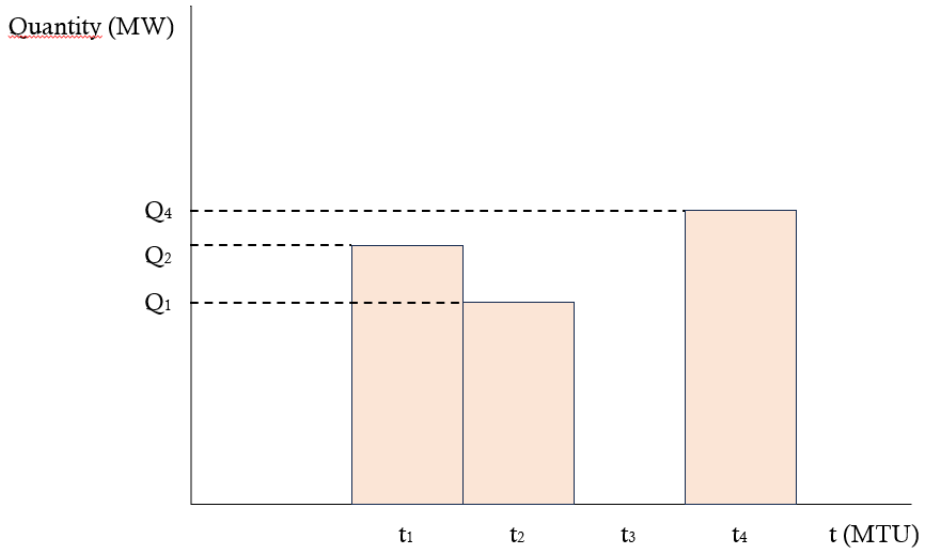


Figure 7. Example of a Sell Block Order.

For these MTUs, the EUPHEMIA algorithm determines the MCP and an AR for the Order, so the acceptance condition for the Sell Block Order is:

$$P \geq \frac{\sum_{t=1}^{24} MCP_t \cdot (Q_t \cdot AR)}{\sum_{t=1}^{24} Q_t} \quad (1)$$

The price P is compared with the weighted average MCP, and the Order is fully accepted if the price is less than the weighted average and partially accepted if it is equal.

Special types of Block Orders are Linked Block Orders and Exclusive Groups of Block Orders. A Linked Block Order consists of Block Orders that are linked via a "parent-child" situation. The "child" Block Order can only be accepted if the "parent" Block Order is accepted as well. In an Exclusive Group of Block Orders, the sum of the AR of the Block Orders can't exceed 1, meaning that if they have MAR=1, then only one at most can be accepted.

When a Block Order is accepted, it "displaces" the hybrid order Sell curve, since the Block's volume is added to the curve's energy volume to be cleared [13]. Practically, the Block Order acceptance or rejection explains why the official MCP can be lower than the initial intersection (if cheap blocks are accepted), or higher (if blocks at low prices are rejected, so the system "jumps" to a more expensive segment). Figure 8 depicts the shift in the hybrid curves caused by Block Orders for 28th September at 00:00.

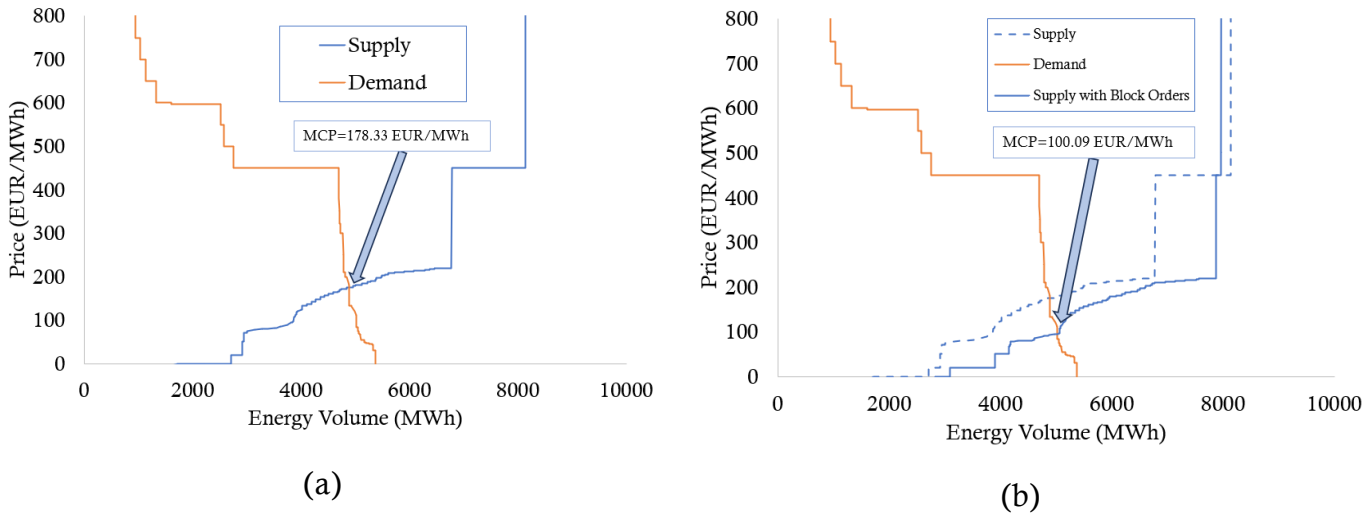


Figure 8. Impact of Block Orders on DAM Clearing on 28/09 00:00.

The chart demonstrates the procedure where the "displaced" final hybrid Sell curve intersects with the hybrid Buy curve and yields the final MCP. In general, import orders would contribute to the final Sell curve, but for the specific hour, they were zero, so the Block Orders and the aggregated Sell hybrid order consist of the entirety of energy generation.

2.3.4. IMPORT AND EXPORT ORDERS

In the DAM, cross-border electricity trading occurs through implicit and explicit import and export orders. These mechanisms enable market participants to transact across bidding zones, ensuring that electricity flows efficiently according to both energy prices and available transmission capacities. Implicit orders integrate energy and capacity allocation within the market coupling algorithm, while explicit orders rely

on pre-allocated PTRs nominated by Participants before market clearing. In Figure 9, the operational and planned interconnections are depicted:

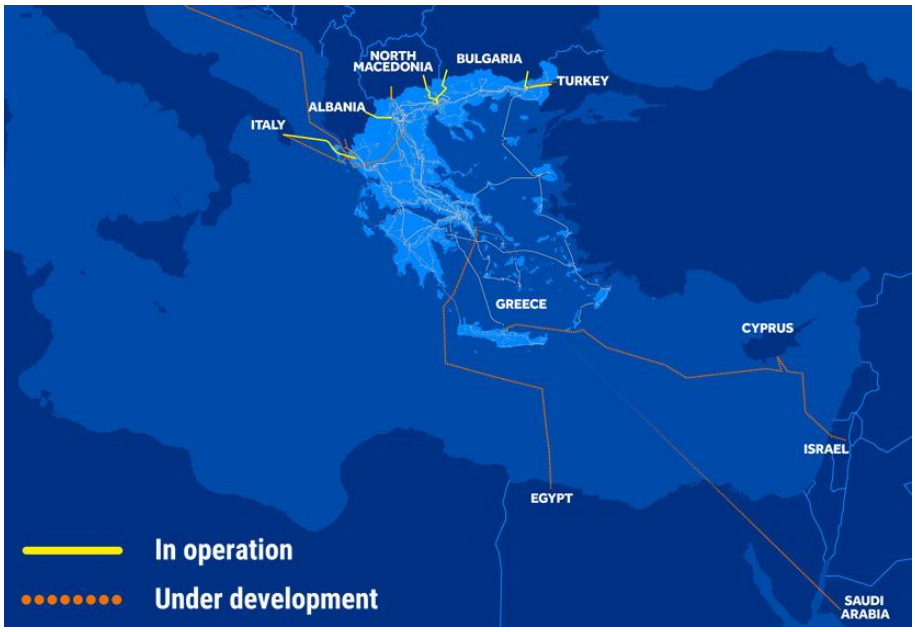


Figure 9. Map of Greece's Interconnections [14].

Sections 4.2.21 and 5.10.44 of the Rulebook describe implicit capacity allocation as the simultaneous matching of energy and cross-border capacity within the market coupling process. Under this model, participants submit regular energy buy and sell orders, and the Price Coupling Algorithm determines both the MCP and cross-zonal energy flows automatically. Implicit allocation is performed in cooperation with Coordinated Capacity Calculators and the Capacity Management Module within the Single Day-Ahead Coupling and Single Intra-Day Coupling frameworks. The Shared Order Book manages all cross-zonal bids and ensures that matching respects available capacity limits. When an order results in a cross-zonal trade, the required capacity is implicitly allocated through the Capacity Management Module.

This system ensures that trades across bidding zones (e.g. Greece-Italy, Greece-Bulgaria) occur automatically and optimally within the DAM or Intra-Day Market, without participants needing to explicitly nominate capacity.

Until September 2025, the interconnections between Greece and the countries that submit implicit orders are described in Table 2:

Table 2. Interconnections between Countries that submit Import Orders.

Country	Total Import Capacity	Characteristics
Bulgaria	1350 MW	1st Interconnection: 400kV overhead line, 1400MVA, Thessaloniki – Blagoevgrad 2nd Interconnection: 400kV overhead line, 2000MVA, Nea Santa – Maritsa

Italy	500 MW	1st Interconnection (GR.IT.1): HVDC, submarine cable, 500 MW 2nd Interconnection (GR.IT.2): under construction

Explicit import and export orders rely on the prior acquisition of LT-PTRs, which grant participants the right to schedule cross-border flows independently of the coupling algorithm. HEnEx submits PPT Sell Orders on behalf of Traders, Suppliers, and Self-Supplying Consumers for imports using confirmed nominated LT-PTRs not already committed through Physical Delivery Nominations. PPT Buy Orders are submitted for exports under similar conditions.

Per Rulebook's Section 4.1.4, these orders are priced at the minimum or maximum admissible DAM limits, ensuring that energy linked to long-term contracts is cleared regardless of market price variations.

3. TIME-SERIES FORECASTING

This section describes the forecasting methodology for key order types that influence the MCP and, therefore, the MCP itself, using publicly available data from January to September 2025. The Market's mechanisms and the Market participants' behaviour influence the forecasting methodology selected for implementation, aiming to achieve a successful month-ahead time-series forecast. Data from October 2025 and on were not taken into account, since the aggregation of their 15-minute time step to hourly, in order to match the previous time step, could skew forecast results. Data from 2024 were neglected, as they capture Market behaviour that is not present in 2025, such as zero imports and exports from Italy due to maintenance, etc.

As an overall technique, the data set covers the first case, January-August 2025 and the second, January-September 2025, with a hold-out scheme, where the training set includes the time interval between the first month and the second to last month targeted for forecasting, while the month previous the forecasted one is kept as a validation window, leading to the testing phase using the actual recorded data for the target-month.

3.1. METHODOLOGY FOR BLOCK ORDERS

3.1.1. PROBLEM FORMULATION

In order to forecast the hourly Block Order volumes y_t (MW) that would be accepted and scheduled for physical transmission in the Greek DAM for August and September 2025, the total time series is initially studied. The Figure depicts the time series that contains the historical data that will serve as input.

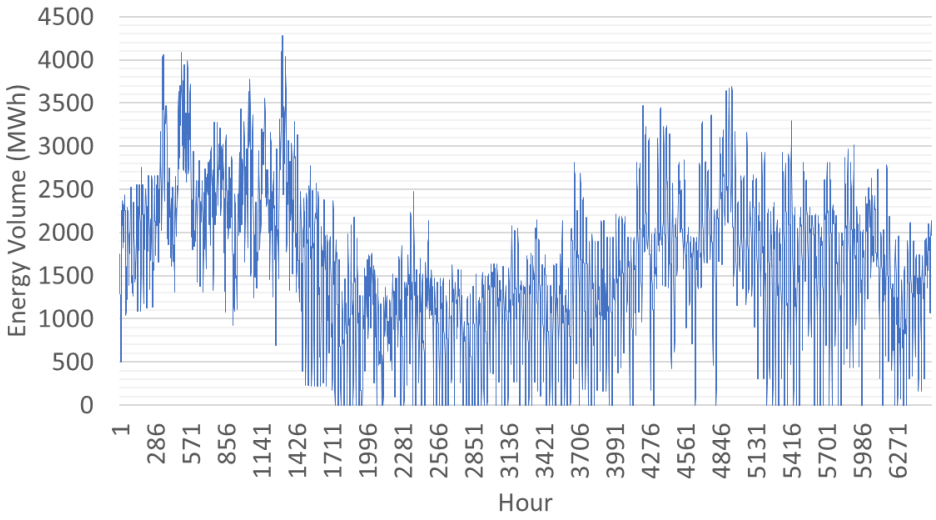


Figure 10. Accepted Block Orders from January to September 2025.

The time-series exhibits strong daily periodicity, possibly containing higher frequencies as well. For many hours of the year, the values reach zero, while the monthly peak deviates, without a visible trend component. The methodology must take these factors into account, as well as exogenous parameters that could influence the series progression, such as weather and electricity demand. Regression models are useful for this task; however, the considerable number of chosen variables requires a slightly more complex model than ordinary linear regression (OLS). In OLS, the model is expressed as:

$$y_t=\beta_0+\beta_1\cdot x_{1,t}+\beta_2\cdot x_{2,t}+\ldots+\beta_p\cdot x_{p,t}+\varepsilon_t \quad (2)$$

In this equation, y_t is the target, $x_{j,t}$ are the predictors, ε_t is noise, and the goal is to determine the coefficients β_j that best fit the data. The above equation can be alternatively written in matrix form:

$$y=\Phi\cdot\theta+\varepsilon \quad (3)$$

Where Φ is the design matrix, and θ is the vector of coefficients. The OLS solution that minimises the sum of squared errors $\|y-\Phi\cdot\theta\|^2$, is equal to:

$$\hat{\theta}_{OLS}=(\Phi^T\Phi)^{-1}\Phi^T y \quad (4)$$

This approach could lead to overfitting, since the features that potentially influence the Block Orders are numerous, but finite. For example, the hour, day, and month of the year, participation shares from other types of market participants or the total demand could influence the accepted Block Orders as well. These predictors could also be correlated (for example, the total demand and renewable production). In that case, $\Phi^T\Phi$ becomes nearly singular, which makes $(\Phi^T\Phi)^{-1}$ potentially unstable, and, therefore, the coefficients are huge. In other words, the predictors become sensitive to small changes in data, leading to large deviations in each forecasting period.

3.1.2. RIDGE REGRESSION

Potential overfitting issues are addressed using ridge regression. Ridge regression adds a penalty on the size of the coefficients by adding an additional term to be minimised [15]:

$$\text{minimise}_{\theta}\left(\|y-\Phi\cdot\theta\|^2+\lambda\cdot\sum_{j=1}^{p}\theta_j^2\right) \quad (5)$$

The first term ensures a good fit, while the second discourages large coefficients. The parameter $\lambda\geq 0$ is the regularisation parameter controlling the balance. In matrix form, the solution to the new minimisation problem becomes:

$$\hat{\theta}_{ridge}=\left(\Phi^T\Phi+\lambda R\right)^{-1}\Phi^T y \quad (6)$$

The vector $\Phi=[B \quad X]$ is the design matrix, while $\theta=\begin{bmatrix}\beta\\ \gamma\end{bmatrix}$ is the coefficient matrix. What ridge regression achieves in this way is the restriction of coefficient values to limits that keep the forecast stable by using the matrix λR, even if it worsens the fit, making it better at generalising to unseen data attributed with shifting behaviour over months, while addressing at the same time overestimating phenomena in case of multicollinearity between predictors. The penalty matrix R was set equal to:

$$R=\begin{bmatrix}0 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 1\end{bmatrix} \quad (7)$$

Keeping the intercept unpenalised. Ridge-regressed OLS handles the deterministic forecast components. In a residual correction paradigm, the target series is expressed as:

$$y_t=f_{det}(t,X_t)+r_t \quad (8)$$

Where $f_{det}(t,X_t)$ is the deterministic component modelled via regression on a calendar, seasonal, and exogenous features, and r_t is the residual component, optionally captured by a stochastic time-series model.

The deterministic part implements the Fourier transform to capture all seasonal frequencies within the time-series, and it can be written as [16]:

$$f_{det}(t,X_t)=\beta_0+\sum_k\left[a_k\sin(2\pi f_k t)+b_k\cos(2\pi f_k t)\right]+\sum_h\beta_h^{HOD}\cdot 1+\sum_d\beta_d^{DOW}\cdot 1+\sum_m\beta_m^{MON}\cdot 1+\gamma^T X_T \quad (9)$$

The term β_0 is the intercept, a constant term that represents the baseline level of the Block Order volume in MW. It is the model's estimate of the average Block Order volume when all regressors are at their reference level, that is, when no Fourier terms are active, and $X_T=0$. The rest of the coefficients that originate from the matrix θ are the Fourier coefficients, HOD, DOW, monthly coefficients, while γ that multiplies the exogenous variables contains the estimated ridge regression coefficients linking these features to the expected Block Order volume, capturing the economic responsiveness of Block Orders to the exogenous variables. The variables under summation contain the Fourier terms that capture daily, weekly and seasonal patterns, the hour of day, day of the week, and month dummies (B), while the vector X_T contains the standardised exogenous regressors (load, RES, nonlinear terms, and dummy variables).

The vector X_T is basically the "snapshot" of market conditions at the hour *t* that explains why the accepted Block Orders look the way they do. The vector includes $Load_t$, which is the system demand forecast, RES_t, which is the renewable energy forecast, $Net_t=Load_t-RES_t$, which is the net demand. When $Load_t$ increases, Block Order bids tend to rise. When RES_t increases (due to abundant solar or wind generation), Block Order bids tend to fall. The engineered nonlinear transformations of those include $Load_t^2$, RES_t^2 that capture

curvature effects (nonlinear response to demand or RES penetration). They also include the interaction term $Load_t \cdot RES_t$ that captures how the combination of high RES and load affects bidding. The term $RES_{Share}=\frac{RES_t}{|Load_t|+|RES_t|}$ captures the proportion of renewables to demand satisfaction, while the term $Load_{t,pos}=max(0,Load_t-q_{80})$ represents the "kink" capturing behaviour under very high demand. The seasonal categorical variables within X_T represent behavioural shifts that depend on the time of hour, day, or month. Even though they are time-based, they are treated as regressors and include weekday dummies, monthly dummies, hour-of-day (HOD) dummies, holiday dummies (e.g. for August 15th±1 day), suitable for Market operation, including cases of 11:00 or 12:00, when there is a high PV injection to the grid, and the Fourier terms and HOD dummies adjust accordingly. Also, when Net_t is high, the deterministic part pushes the forecast forward.

The residual r_t accounts for short-term deviations, unexpected events, or behavioural noise not captured by these structured events.

3.1.3. DETERMINISTIC REGRESSION DESIGN

As aforementioned, the regression design matrix Φ includes seasonal terms such as Fourier seasonality, calendar dummies that capture behaviour outliers during significant days of the year, potential trend terms, and influence attributed to exogenous variables. Smooth intraday and weekly periodicity are captured using Fourier harmonics expressed as:

$$\phi_k^{cos}(t)=\cos\left(\frac{2\pi k}{24}t\right),\ \phi_k^{sin}(t)=\sin\left(\frac{2\pi k}{24}t\right),\ k=1,\ldots,K \tag{10}$$

Optional weekly harmonics at a period of 168 hours and bi-weekly at a period of 336 hours were accounted for. Candidate sets up to K=7 are tested, and the optimal number is preserved.

As calendar dummies, weekday effects were captured in the form of binary dummies for Monday-Saturday (Sunday was used as a baseline), and monthly effects were included as binary dummies for December to February, with January serving as a baseline. The hour-of-day effects were represented by dummies for hours 1-23, 00:00 serving as the baseline, and lastly for holidays, a dummy for August equal to 1 for 14th-16th of August (15th of August is a national holiday), and none for September, while a linear trend term $\phi_{trend}(t)=t$ captures slow drift across months.

Exogenous regressors L_t (load forecast) and R_t (RES generation) are standardised and used on the training set:

$$z_{L,t}=\frac{L_t-\mu_L}{\sigma_L},\ z_{R,t}=\frac{R_t-\mu_R}{\sigma_R},\ z_{N,t}=\frac{(L_t-R_t)-\mu_N}{\sigma_N} \tag{11}$$

The engineered features include quadratic terms $z_{L,t}^2$, $z_{R,t}^2$, interaction terms $z_{L,t}\cdot z_{R,t}$, share terms $\frac{z_{R,t}}{(|z_{L,t}|+|z_{R,t}|+\varepsilon)}$, allowing the capture of nonlinear and asymmetric sensitivity to demand and renewable conditions. For further stabilisation, the target on the training set is normalised using:

$$y_{norm}=\frac{y-\mu_Y}{\sigma_Y} \tag{12}$$

Where σ_Y the standard deviation of the sample and μ_Y the mean value. Ridge then estimates β by using:

$$\hat{\theta}_{ridge}=(\Phi^T\Phi+\lambda I)^{-1}\Phi^T y_{norm} \quad (13)$$

And predictions are mapped back:

$$\hat{y}=\mu_Y+\sigma_Y\cdot(\Phi\hat{\theta}_\lambda) \quad (14)$$

Given Φ and target y, the regression coefficients are estimated via ridge regression:

$$\hat{\theta}_\lambda=\arg\min_\theta\left\{\|y-\Phi\theta\|_2^2+\lambda\|\beta_{-0}\|_2^2\right\} \quad (15)$$

Where β_{-0} excludes the intercept. A grid of $\lambda\in[0, 10^{-2}]$ is tested, with $\lambda=0$ representing the standard case of OLS.

For bias calibration, the mean residual over the last 168 training hours is added to the testing month deterministic forecast after fitting to correct for short-term level drift:

$$\tilde{y}_{det,\,test}=\hat{y}_{det,\,test}+\bar{r}_{last,\,V} \quad (16)$$

3.1.4. RESIDUAL MODELLING

Residuals $r_t=y_t-\hat{y}_t$ are optionally modelled with ARIMA [17]. Optionally means that ARIMA attempts to capture additional elements that may not have been detected by the deterministic model, such as intra-day autocorrelation or residual stochastic structures, and filter out white noise. The model expresses the residual time-series as:

$$\phi(B)(1-B)^d r_t=\theta(B)\Theta(B^{24})\varepsilon_t \quad (17)$$

The Backshift operator acts on the series as expressed in Equation X:

$$(1-B)^d r_t=\Delta^d r_t \quad (18)$$

Where the setting d=0 assumes that the residuals are already stationary. The autoregressive polynomial $\phi(B)$:

$$\phi(B)=1-\phi_1 B-\ldots-\phi_p B^p \quad (19)$$

It aims to reflect how slowly Block Order participation adjusts through time. Combined with the backshift operator, the left-hand term of the equality filters the residuals in the hope of transforming them to a filtered white noise, thus removing any remaining structure from the time-series. The moving average polynomial $\theta(B)$ captures short-memory noise correlation and simulates the gradual vanishing of shock effects.

$$\theta(B)=1+\theta_1 B+\ldots+\theta_q B^q \quad (20)$$

The seasonal moving average polynomial of order 1 at period 24 $\Theta(B^{24})$ expressed as:

$$\Theta(B^{24})=1+\Theta_1 B^{24} \quad (21)$$

This polynomial is added to capture periodic bidding strategies that repeat every 24 hours, and combined with the moving average polynomial, the total right-hand term of the equation represents the rapid shock term of the previous hour, the shock term found the day before at the same hour, as well as the hour before that, acting on white noise ε_t.

The ARIMA (p, d, q), including the seasonal MA (24) iteration grid, was set as:

$$p\in\{0,1,2,3\},\ d\in\{0,1\},\ q\in\{0,1,2,3\},\ SMA(24)\in\{0,1\} \tag{22}$$

Each residual model is estimated via maximum likelihood and forecast into the validation horizon. The validation MAE selects whether ARIMA improves over the no-ARIMA baseline. After selecting the best spec, refit on the training window (using the Fourier set and λ) is applied. Bias calibration of the deterministic forecast is executed using the last K residuals in training (for this model, K≤168):

$$bias=\frac{1}{K}\sum_{i=1}^{K} r_{LTx-i+1},\ \tilde{y}^{det}=\hat{y}^{det}+bias \tag{23}$$

And the ARIMA forecast:

$$\hat{y}_t=\tilde{y}^{det}+\hat{r}_t \tag{24}$$

3.1.5. FINAL FORECAST AND EVALUATION

Each candidate configuration (Fourier set, λ, ARIMA spec) is evaluated on the testing month with specific metrics [18]. The first is the Mean Absolute Error (MAE), which expresses how "off" the forecasts are, on average. The Root Mean Squared Error (RMSE) has a similar function to MAE, but it penalises larger errors more, making it more reliable as a metric. The Mean Absolute Percentage Error (MAPE) provides, as a percentage, the deviation of the forecasts from the actual values, yet the existence of zero values makes its use problematic. For that reason, its non-zero version ($MAPE_{nz}$) is preferred, as well as the symmetric MAPE (sMAPE). The Weighted Average Percentage Error (WAPE) is complementary to these, and Mean Absolute Scaled Error (MASE) compares the model's error to that of a 24-hour seasonal naïve benchmark that assumes today's energy volume is going to be the same as yesterday's was at the same hour:

$$MAE=\frac{1}{H}\sum_{t}|y_t-\hat{y}_t| \tag{25}$$

$$RMSE=\sqrt{\frac{1}{H}\cdot\sum(y_t-\hat{y}_t)^2} \tag{26}$$

$$WAPE=\frac{\Sigma|y_t-\hat{y}_t|}{\Sigma|y_t|} \tag{27}$$

$$sMAPE=200\cdot\frac{1}{H}\sum\frac{|y_t-\hat{y}_t|}{|y_t|+|\hat{y}_t|} \tag{28}$$

$$MAPE_{nz}=\text{MAPE over non-zero } y_t \tag{29}$$

$$MASE=\frac{MAE}{\text{MAE of naïve 24-h seasonal}} \tag{30}$$

The use of all these metrics could seem as an "overmeasure", but depending on the case their interpretation can be misleading and therefore their complimentary nature provides a more accurate evaluation of the algorithms' performance.

The best model is re-fit on the training window and used to forecast the desired month. The final prediction is:

$$\hat{y}_{t+h}=\hat{y}_{det,\,t+h}+\hat{r}_{t+h},\ h=1,\ldots,H \tag{31}$$

With $\hat{r}_{t+h}$, h=0 if no ARIMA is selected. Finally, a 90% interval band is constructed as:

$$\hat{y}_{t+h} \pm 1.645 \cdot \hat{\sigma}_{r,h} \quad (32)$$

The interval band serves as the forecast's error band.

3.2. METHODOLOGY FOR HYBRID ORDERS

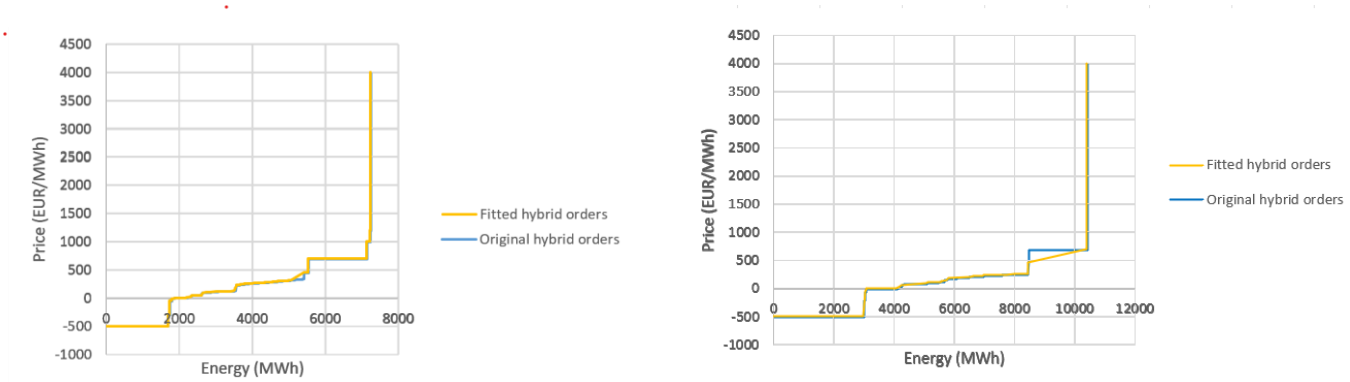
3.2.1. DATA PREPARATION

Before implementing the algorithm, the hybrid order curves had to be grouped by date and sorted into a single table. Since the hourly hybrid orders exhibit different price steps both in total number and in value, despite having, due to regulatory design, the same global price extrema (GP_{min}, GP_{max}), a matching process was performed to assign each hybrid order's volume $Q_{hour,price}$ to the closest price axis that was set for all orders. The final table that served as input had the form shown in Table 3:

Table 3. Table Format of Fitted Hybrid Orders.

	GP_{min}	GP_1	…	GP_{max}
01/01/2025 00:00	$Q_{1,min}$	$Q_{1,1}$	…	$Q_{1,max}$
01/01/2025 01:00	$Q_{2,min}$	$Q_{2,1}$	…	$Q_{2,max}$
…	…	…	…	…
30/09/2025 23:00	$Q_{6552,min}$	$Q_{6552,1}$	…	$Q_{6552,max}$

Since all hourly energy volumes are attributed to a single price grid, slight deviations from the original hybrid curves were used by the algorithm for the forecasting process. Figure 11 compares indicatively for a) January 1st, 06:00 and b) June 10th, 00:00 2025, the difference between the original and the fitted hybrid curve:

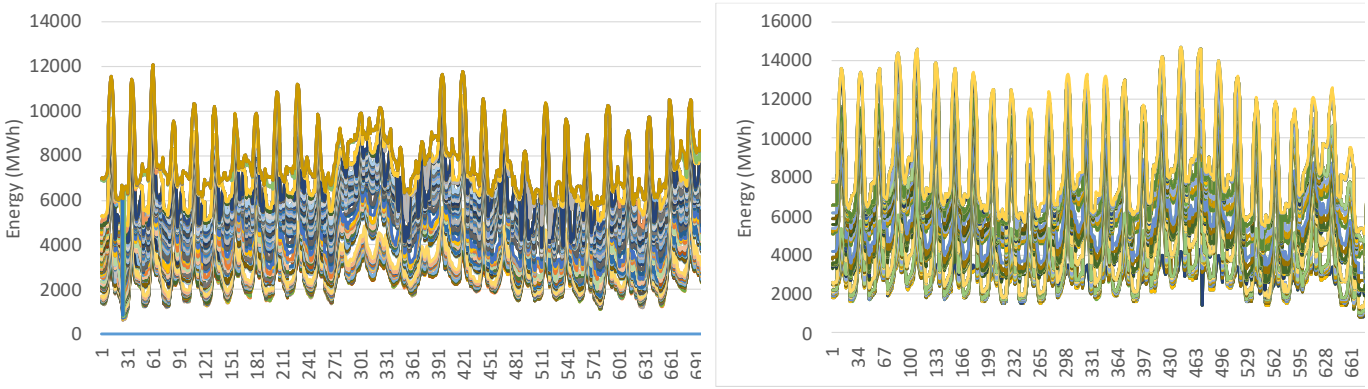


(a) Comparison for January 1st, 06:00 2025

(b) Comparison for June 10th, 00:00 2025

Figure 11. Original vs Fitted aggregated hybrid order curve.

By plotting each column across all hours under study, a cluster of time series can be studied, as shown in Figure 12, indicative of January and September of 2025:



(a) (b)

Figure 12. Fitted Price Grid time-series for (a) January and (b) September 2025.

There is a distinct periodic pattern across the price grid, yet there are differences in amplitude and frequency during different seasons, a crucial property for forecasting methodology selection.

3.2.2. HYBRID MODEL FORMULATION

Mathematically, each hybrid supply curve can be defined as a non-decreasing function $f_t(p)$ for each hour t, mapping price $p \in [p_{min}, p_{max}]$ to the cumulative energy volume offered up to that price:

$$f_t(p) = \sum_{i:P_i \leq p} Q_i(t) \tag{33}$$

Where $Q_i(t)$ denotes the quantity (in power units) offered at the price P_i.

The hybrid curve, thus, reflects the system's instantaneous marginal cost structure, combining zero-marginal renewable bids with thermal generation bids, imports, and other supply offers [19]. For modelling and forecasting purposes, each hourly curve $f_t(p)$ is treated as a functional observation sampled on a fixed price grid $\{p_1, p_2, \ldots, p_M\}$, forming a vector:

$$y_t = [f_t(p_1), f_t(p_2), \ldots, f_t(p_M)]^T \in \mathbb{R}^M \tag{34}$$

Stacking all hourly curves yields a matrix $Y \in \mathbb{R}^{T \times M}$, where each row corresponds to a specific hour and each column to a particular price level. The forecasting task consists of predicting the future trajectories of y_t for unseen hours using historical data, seasonal structure and exogenous variables.

3.2.3. FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS (FPCA)

Recent studies have modelled day-ahead supply curves directly rather than scalar prices, capturing richer structural information [20]. Similarly, FPCA provides a dimensionality-reduction technique that captures the dominant modes of variability in a set of curves [21], [22]. Functional approaches have been successfully applied to electricity price and load modelling, capturing complex temporal structures beyond scalar time-series representations [23].

Before applying FPCA, the deviations from the mean curve are analysed by computing the mean curve across time:

$$\mu=\frac{1}{T}\sum_{t=1}^{T} y_t \tag{35}$$

Next, each curve is centred by subtracting the mean shape:

$$y_t^{(c)}=y_t-\mu \tag{36}$$

Stacking these centred curves gives the centred data matrix:

$$Y_c=\begin{bmatrix}(y_1-\mu)^T\\ \vdots \\ (y_T-\mu)^T\end{bmatrix}=Y-1\mu^T \tag{37}$$

Once the curves have been centred, FPCA applies a singular value decomposition (SVD) [24]:

$$Y_c=U\Sigma V^T \tag{38}$$

Where $U \in \mathbb{R}^{T\times K}$ contains the temporal scores (time-dependent coefficients), $\Sigma \in \mathbb{R}^{K\times K}$ is a diagonal matrix with singular values, $V \in \mathbb{R}^{M\times K}$ is the matrix that holds the principal components (eigen-curves in the price domain).

The equation can be rewritten as:

$$Y_c^T Y_c=V\Sigma^2 V^T \tag{39}$$

Where V can be calculated by diagonalising the matrix $Y_c^T Y_c$ whose eigenvectors are the columns of V and eigenvalues $\sigma_1^2, \sigma_2^2,\ldots,\sigma_r^2$ the diagonal elements of Σ^2. Then, Equation 38 can be solved for U. The rank is calculated as r=min(T, M).

The matrix V is found such that the columns form an orthonormal basis of directions in the price-space that maximise variance in Y_c. Its first column, $V_{.,1}$, is the direction of maximum variation across price. The second, $V_{.,2}$ is the orthogonal direction with the next-largest variance, while the rest exhibit minor levels of variance.

The matrix Σ scales how much variance each of these directions explains, while U gives the coordinates of each hourly curve in these directions, meaning how much of each pattern is present at each time.

Thus:

$$Y_c=\sum_{k=1}^{r}\sigma_k U_{.,k} V_{.,k} \tag{40}$$

Each term $\sigma_k U_{.,k} V_{.,k}$ is a rank-1 matrix that captures one independent pattern of variation, reconstructing each hourly curve from the mean into its final form:

$$y_t=\mu+\sum_{k=1}^{K} s_{t,k} v_k \tag{41}$$

Where $s_{t,k}$ is the score for the component k, and v_k is the corresponding eigen-curve. This equation means that FPCA expresses every hourly supply curve as a mean shape plus a weighted sum of a few orthogonal deformation patterns, where the mean curve represents the average marginal cost structure.

In the HEnEx DAM model, producers submit orders, of which only a part is "relevant" at a given hour. The aggregated curves are standardised and capped based on admissible market supply. So, the operational supply stack depends on which units are online, how much power each unit offers, the technical constraints that are influenced by the net load, grid topology, and the availability of hydro and RES units that "push" others out. In this context, v_1 captures horizontal shifts of the curve to which total system loading is attributed. The reason is that when the electricity load is low, only base, cheap units are operating, so the curve stays low and flat. When the load is moderate, mid-merit units join, and the curve shifts up and steepens. At last, when the load is high, expensive units are needed, so the curve is much higher and steeper.

The second most important direction, v_2, expresses the slope changes reflecting marginal cost differences. It often represents curve steepening or flattening, most probably attributed to the balance between RES and thermal generation. The higher v_k components represent more subtle deformations, such as curvature, tilting, shoulder effects, etc, yet mathematically, K components in total are used during the reconstruction.

The total variance of all the curves (after centring) is proportional to the sum of the squared singular values σ_k. To determine how many components K are enough, the cumulative share of variance explained by the first K components is calculated, a ratio that describes what percentage of the total variability in all the hourly curves is captured by these components:

$$\frac{\sum_{k=1}^{K} \sigma_k^2}{\sum_{k=1}^{M} \sigma_k^2} \geq \eta \tag{42}$$

In this study, η=99% was selected as an adequate percentage of curve variability.

Once decomposed, the temporal evolution of each score series $s_{t,k}$ is modelled separately. In this framework, each score is assumed to consist of:

$$s_{t,k} = g_k(x_t) + r_{t,k} \tag{43}$$

Where $g_k(x_t)$ is a deterministic function of explanatory variables x_t (Fourier terms, calendar effects, exogenous signals), and $r_{t,k}$ captures short-term autocorrelated noise.

To capture seasonal cycles for their inclusion in x_t, the deterministic regressors for each score are constructed using Fourier basis functions:

$$\phi_h^{(P)}(t) = \sin\left(\frac{2\pi ht}{P}\right), \psi_h^{(P)}(t) = \cos\left(\frac{2\pi ht}{P}\right) \tag{44}$$

Where P denotes the period (24h, 48h, etc.) and h the harmonic order. A deterministic matrix $X_{Fourier}$ that compactly represents all seasonal effects is formed as:

$$X_{Fourier} = [\phi_1^{(24)}, \psi_1^{(24)}, \phi_2^{(24)}, \psi_2^{(24)}, \phi_1^{(168)}, \psi_1^{(168)}, \ldots] \tag{45}$$

Additionally, binary dummies $d_{weekend}$ and $d_{holiday}$ account for systematic weekend and public-holiday shifts, while a month-end indicator captures end-of-month redispatch effects.

To enforce market dynamics into the curve evolution, system load (L) and renewable generation (R) time-series are incorporated as exogenous regressors. Each exogenous series is standardised, smoothed by a 3-point moving average, and augmented with several temporal lags $L_t^{(1)},L_t^{(2)},L_t^{(24)},L_t^{(168)}$, where $L_t^{(1)}=L_{t-1}$ and so on, similarly for RES $R_t^{(1)},R_t^{(2)},R_t^{(24)},R_t^{(168)}$.

These lags represent immediate, short-term, daily and weekly dependencies, respectively, ensuring that the persistent effects of system balance are captured. The inclusion of lags also allows current curve shifts to depend on the most recent load and RES dynamics rather than contemporaneous values, otherwise cause and effect would happen simultaneously and the regression would be biased.

The resulting design matrix for score regression thus combines Fourier, calendar, dummies and exogenous terms in the form:

$$X_t=\left[1,\ \text{Fourier terms, calendar dummies},L_t^{(l)},R_t^{(l)}\right] \tag{46}$$

Where $l\in\{1,2,24,168\}$.

Each FPCA score $s_{t,k}$ is regressed on the design matrix X_t using ridge regression to prevent overfitting:

$$\hat{\beta}_k=\arg\min_{\beta}\|s_k-X\cdot\beta\|^2+\lambda\|\beta\|^2 \tag{47}$$

Where λ is the ridge penalty parameter.

Ridge regression provides numerical stability since the design matrix includes correlated harmonics and exogenous terms, and it reduces coefficient variance without introducing bias in the intercept term. The fitted values yield the deterministic component of each score:

$$\hat{s}_{t,k}^{(det)}=X_t\hat{\beta}_k \tag{48}$$

After calculating the deterministic component, the residual scores are calculated using:

$$r_{t,k}=s_{t,k}-\hat{s}_{t,k}^{(det)} \tag{49}$$

To capture any potential short-term autocorrelation, each residual is optionally modelled with an ARIMA(p, 0, q) process. In this scenario, the backshift operator was neglected (d=0) since trends should be captured by the deterministic component. Therefore, residual expression is expressed as:

$$r_{t,k}=\phi_1 r_{t-1,k}+\ldots+\phi_p r_{t-p,k}+\varepsilon_t+\theta_1\varepsilon_{t-1}+\ldots+\theta_q\varepsilon_{t-q} \tag{50}$$

Where $\varepsilon_t\sim N(0,\sigma^2)$, per the standard assumption in linear time-series regression and ARIMA models.

Model orders $p, q\in\{0,1,2\}$ are selected via the Akaike Information Criterion (AIC) for each component k. The AIC is a metric that evaluates a statistical model based on a) the model fit, meaning how well the model matches the data, and b) the model complexity, meaning how many parameters the model uses [25]. Its purpose is to prevent overfitting by penalising overly complicated models that add little predictive value. For a model with the number of estimated parameters equal to k and log-likelihood l, the AIC is defined as:

$$AIC=2\cdot k-2\cdot l \tag{51}$$

The first term of the equation represents the penalty for model complexity, while the second is the measure of model error.

The log-likelihood l measures a statistical model's ability to explain observed data, and is equal to the logarithm of the probability that the model assigns to the observed data. If we suppose that for y_t observations, the model has a parameter vector θ, the assumed probability density of each observation is $f(y_t \mid \theta)$, then the likelihood is:

$$L(\theta)=\prod_{t=1}^{n} f(y_t \mid \theta) \tag{52}$$

With log-likelihood:

$$l(\theta)=\log L(\theta)=\sum_{t=1}^{n} \log f(y_t \mid \theta) \tag{53}$$

The logarithmic metric is particularly useful in such cases, since it transforms the products into sums and allows simpler derivatives, making it much easier to both compute and optimise.

By fitting the FPCA score model $s_{t,k}=X_t\hat{\beta}_k+r_{t,k}$, with the standard linear model assumption of $\varepsilon_t \sim N(0,\sigma^2)$, FPCA removes the mean function, dominant structure, and seasonal components through regression. Therefore, the residuals are expected to be of Gaussian behaviour. Thus, the probability density function is considered:

$$f(s_{t,k} \mid X_t, \hat{\beta}_k, \sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s_{t,k}-X_t\hat{\beta}_k)^2}{2\sigma^2}} \tag{54}$$

And the log-likelihood becomes:

$$l=-\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(s_{t,k}-X_t\hat{\beta}_k)^2 \tag{55}$$

The term k is the total number of free parameters estimated in the model. In the case of FPCA, for each score $s_{t,k}$, the regression model includes the Fourier seasonal regressors, where each harmonic contributes two parameters $a_h \cos(\omega_h t)+b_h \sin(\omega_h t)$, while the exogenous regressors contribute one regression coefficient, a parameter from the intercept, and a parameter due to noise variance. In this way, AIC forces the model to "justify" each parameter in use.

The residuals, then, pass a Ljung-Box test for whiteness to test the Null Hypothesis, which dictates that the residuals are non-correlated up to lag H. To do so, the accumulated sample autocorrelation between the residuals and their lag at k is calculated as:

$$\hat{\rho}_k=\frac{\sum_{t=k+1}^{n}(r_t-\bar{r})\cdot(r_{t-k}-\bar{r})}{\sum_{t=1}^{n}(r_t-\bar{r})^2} \tag{56}$$

Where $\bar{r}$ the mean value of the residual time-series. In this equation, the nominator is equal to the autocovariance at lag k, while the denominator is the total variance.

With the calculation of $\hat{\rho}_k$ the Ljung-Box test statistic Q is calculated [26]:

$$Q=n\cdot(n+2)\sum_{k=1}^{H}\frac{\hat{\rho}_k^2}{n-k} \tag{57}$$

Where n is the sample size, and H is the number of lags being tested. For the implementation of this methodology, the number of lags to be tested was selected as H=24, assuming that in the hourly time-series the short-term leftover structure would be repeated daily.

The metric for the validity of the Null Hypothesis is the p-value, which is calculated using the Probability Density Function of the X_{df}^2 distribution

$$f(x;k)=\frac{1}{2^{k/2}\cdot\Gamma(k/2)}x^{k/2-1}\cdot e^{-x/2} \tag{58}$$

Where k is the degrees of freedom and Γ is the Gamma function. Since the Ljung-Box test was deployed before ARIMA (p=q=0), the degrees of freedom df=H-p-q were set equal to the number of lags H, 24. The p-value is equal to:

$$p=\int_{Q}^{\infty} f(x;k)\,dx \tag{59}$$

After its calculation, the check for the Null Hypothesis is selected as (p>0.1), which, if true, signifies that the residuals do not contain significant statistical structure.

The ARIMA layer can be omitted in case the residuals pass the Ljung-Box test, thereby reducing complexity and avoiding overfitting. For the forecast horizon h=1, …, H, each component's deterministic and ARIMA-predicted scores are summed to obtain future FPCA scores:

$$\tilde{s}_{t+h,k}=\hat{s}_{t+h,k}^{(det)}+\hat{r}_{t+h,k} \tag{60}$$

The predicted hybrid supply curve is then reconstructed as:

$$\hat{y}_{t+h}=\mu+\sum_{k=1}^{K}\tilde{s}_{t+h,k}v_k \tag{61}$$

The reconstructed curves must adopt the characteristics of the original curves, that is, monotonicity, lack of absurd spikes, and non-negative volumes. To ensure this systematic alignment, a bias correction term is added, computed as the mean residual over the last B days of training, calibrated equal to the number of days in a month (B=30):

$$\hat{y}_{t+h}^{(final)}=\hat{y}_{t+h}+a\bar{\varepsilon}_B \tag{62}$$

Where $a\in[0, 0.25, 0.5, 0.75, 1]$ is the shrinkage factor tuned on validation data by iterating through its set of values and calculating for which value $\hat{y}_{t+h}^{(final)}$ yields the optimal RMSE.

Finally, the curves are constrained to physical feasibility:

$$\hat{y}_{t+h,i}=\max\left(0,\min\left(\hat{y}_{t+h,i},q_{0.98}\left(y_{.,i}\right)\right)\right) \tag{63}$$

That way, positive volume quantities are ensured max(0,.). The term $\min\left(\hat{y}_{t+h,i}, q_{0.98}\left(y_{.,i}\right)\right)$ is responsible for the prevention of sudden peaks within the curve. Its purpose is to iterate through each price column, looking at all the historical energy volumes submitted at that price and compare their 98th percentile with the forecast. The minimum value is stored, so the forecast for a given price doesn't exceed 98% of the recorded volumes, providing a realistic curve ascent. The comparison between the forecast and the maximum value recorded for the same price column is avoided in case of outliers that could have been recorded at some point within the year.

The monotonicity of each curve is preserved by using:

$$\hat{y}_{t+h,i}=\max\left(\hat{y}_{t+h,i},\hat{y}_{t+h,i-1}\right) \tag{64}$$

Thus, ensuring non-decreasing volume offers across prices.

The reconstructed curves are evaluated using multiple complementary error metrics that capture both overall accuracy and shape fidelity. They jointly quantify the model's ability to reproduce both the shape and the scale of the hybrid supply curves. The metrics in use are the coefficient of determination (R^2):

$$R^2=1-\frac{\Sigma_{t,i}\left(y_{t,i}-\hat{y}_{t,i}\right)^2}{\Sigma_{t,i}\left(y_{t,i}-\bar{y}\right)^2} \tag{65}$$

As well as MAE, RMSE, and sMAPE.

3.3. METHODOLOGY FOR IMPLICIT ORDERS

3.3.1. OVERVIEW OF IMPLICIT IMPORT ORDERS

The implicit import orders time series represents the hourly cleared volume of cross-border energy imports (in MW) in the Greek DAM. In Figure 13, the recorded time-series from January to September 2025 are illustrated.

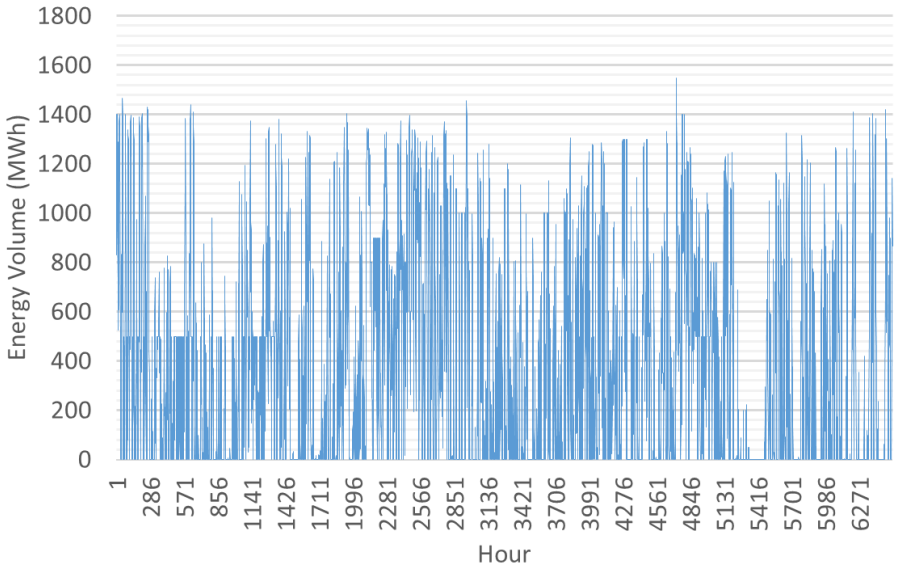


Figure 13. Implicit Import Orders from January to September 2025.

Unlike Block Orders, the time-series does not exhibit distinct seasonality, trend, or calendar effects. The data appear to be characterised by high intermittency, non-Gaussian variability, reflecting sudden activations of interconnection flows driven by price spreads and transmission capacity. Classical linear models (e.g., ARIMA, exponential smoothing) fail to adequately capture these properties because they assume continuous, homoscedastic, and autoregressive dynamics.

To address these limitations, a probabilistic regime-switching model was developed, combining logistic activation, Fourier-based deterministic structure, and ridge-regularised regression for magnitude estimation.

3.3.2. MARKOV MODEL

Let y_t denote the implicit import volume at the hour t, and $s_t \in \{0,1\}$ represent the latent regime variable indicating inactivity ("OFF") or import activation ("ON"). This formulation separates the problem into two parts: a) When imports happen ("ON"/"OFF" state system), and b) how large they are when they happen (meaning the magnitude when they happen). This two-layer structure is inspired by regime-switching or Markov-switching models, which are widely used in econometrics [27], [28], [29]. The formulation, then, becomes:

$$\hat{y}_t = P(ON) \cdot E[y \mid ON] \tag{66}$$

The reason that P(ON) is needed, and not just the forecasted magnitude E[y|ON] is that the observed time-series exhibits strong "switching-off" behaviour for many hours of the year, which a continuous model would not capture. The probabilistic model is based on the conversion of the historical time-series y_t into a binary state:

$$s_t = \begin{cases} 1, & y_t \geq \tau \\ 0, & y_t < \tau \end{cases} \tag{67}$$

The threshold τ is a parameter that determines which import volumes are included in regime training. This step transforms the series into a simple pattern of 0 and 1, which makes it possible to detect persistence, meaning whether imports tend to stay "ON" for several hours once they begin. Trades that are too small in volume might interfere as noise, and are not taken into account as a "switching ON" state. Since the threshold τ is dynamic in the physical world (associated with fluctuating capacity levels of interconnections), it is tuned by a set of values of the 60^{th}, 70^{th}, 80^{th} percentiles of the data, and the maximum between the 85^{th} and the fixed level of 150MW, below which the import is considered insignificant as a price setter.

The next step is to estimate the likelihood of switching between "ON" and "OFF". This is modelled with the use of a Markov chain, which assumes that the next state only depends on the current one, formulated as:

$$P(s_t \mid s_{t-1}, s_{t-2}, \ldots) = P(s_t \mid s_{t-1}) \tag{68}$$

In matrix form, the transition matrix is designed as:

$$T = \begin{bmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{bmatrix} \tag{69}$$

Where $p_{01} = P(s_t = 1 \mid s_{t-1} = 0)$ the probability to turn "ON" after being "OFF", and $p_{10} = P(s_t = 0 \mid s_{t-1} = 1)$ the probability of turning « OFF » after being "ON".

The stationary (long-run) "ON" probability is calculated using:

$$\pi_1^* = \frac{p_{01}}{p_{01}+p_{10}} \tag{70}$$

This showcases how often, on average, the system is "ON". The Markov chain provides a simple probabilistic model for import timing, but it lacks responsiveness to time-of-day or other cyclical patterns.

3.3.3. LOGIT MODEL

In this step, the probability of being "ON" at each hour, based on predictive features X_t is calculated:

$$p_t^{emit} = P(s_t=1 \mid X_t) = \sigma\left(\beta_0 + X_t^T \beta\right) \tag{71}$$

Where σ symbolises the sigmoid function:

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{72}$$

This function is suitable for turning any real number into a value between 0 and 1. In other words, a linear combination of features (X_t) is computed and turned into probabilities by using the sigmoid function. The vector X_t includes several features that describe typical patterns in the data, such as daily cycles ($\sin(2\pi t/24)$, $\cos(2\pi t/24)$), lagged proxy states, meaning persistence ($s_{t-1}, s_{t-2}, s_{t-24}$), or recent magnitudes capped to reduce overconfidence caused by outliers.

The capping is performed by winsorising the historical data, which means replacing extremely high values with the 99.5th percentile $q_{0.995}$ of the series, resulting in the winsorised time-series:

$$y_t^{(w)} = \min(y_t, q_{0.995}) \tag{73}$$

The lagged values $y_{t-1}^{(w)}$, $y_{t-24}^{(w)}$ are used as features for the Logit regression model.

In a binary classification problem like this, the model could favour predicting "ON" or "OFF" states when one of these dominates for most hours, to increase its overall prediction accuracy, therefore sacrificing realistic predicted behaviour. To prevent this, a weighted LASSO-logit regression is used so that large coefficients are penalised, and rarer events get a higher weight, so the model doesn't ignore them [30].

A few assumptions have to be made before introducing the final objective function of the LASSO-logit optimisation problem. Having calculated the binary series by using Equation 67, a Bernoulli distribution is applied to calculate the probability of the "ON" and "OFF" states [31]. For a random variable S has the value S=1 with probability p, and S=0 with probability 1-p, with a probability mass function:

$$P(S=s) = p^s \cdot (1-p)^{1-s} \text{ for } s \in \{0, 1\} \tag{74}$$

Based on the given features X_t, the predicted probability is $p_t = P(s_t=1|X_t)$ and the likelihood of observing s_t is:

$$P(s_t \mid X_t) = p_t^{s_t} \cdot \left(1-p_t\right)^{1-s_t} \tag{75}$$

By assuming independence across time, given the features, the likelihood function is equal to:

$$L = \prod_{t=1}^{n} (s_t \mid X_t) = p_t^{s_t} \cdot \left(1-p_t\right)^{1-s_t} \tag{76}$$

As an optimisation process, it would be easier to use in the objective function the log of the likelihood function:

$$\log L = \sum_t [s_t \cdot \log p_t + (1-s_t) \cdot \log(1-p_t)] \tag{77}$$

The LASSO term dictates which β coefficients are needed and how much, adjusting to the complexity of the problem. To integrate this in the objective function, an a priori Laplace distribution is applied to the β coefficients:

$$p(\beta_j) = \frac{\lambda}{2} e^{-\lambda|\beta_j|} \tag{78}$$

This distribution encourages the coefficients to shrunk values, while still allowing higher magnitudes of λ through the most influential features, when supported by the historical data.

The final objective function that matches the s_t with the "ON" probability p_t is a LASSO-logit weight assignment that is minimised, designed to capture the "ON" states in case they represent the minority of the two states:

$$\min_{\beta,\beta_0} \left[-\sum_t w_t [s_t \log p_t + (1-s_t) \log(1-p_t)] + \lambda \sum_j |\beta_j| \right] \tag{79}$$

The first term of the equation is the negative log-likelihood of the Bernoulli problem. It rewards high model probabilities p_t when the observed states $s_t = 1$, multiplied by the fixed weight w_t. The weight is assigned different values, depending on the value of s_t, to "reward" the "ON" states with a high probability of occurrence:

$$w_t = \begin{cases} w_{pos}, & s_t = 1 \\ 1, & s_t = 0 \end{cases} \tag{80}$$

The second term is the minus log of Equation 77, the LASSO penalty that shrinks unnecessary coefficients to zero. Each coefficient describes how the feature affects the odds of the system being "ON". If $\beta_j > 0$, when the feature increases, the probability of being "ON" increases. If $\beta_j < 0$, then, as the feature increases, the probability decreases. The larger $|\beta_j|$, the stronger the influence.

As part of the methodology, the parameter λ is calibrated by using a five-fold cross-validation where the training window is divided into five folds. Each fold k is withheld in turn for validation while the other four minimise Equation 78 for a fixed λ that yields the LASSO coefficients. For the held-out fold, the probabilities are calculated using the resulting LASSO coefficients through Equation 70, and then its weighted deviance:

$$Dev(\lambda) = -2 \cdot \sum_{t \in k} w_t [s_t \log p(\lambda)_t + (1-s_t) \log(1-p(\lambda)_t)] \tag{81}$$

The final λ parameter is the one that yields the minimum five-fold-average deviance value. The final probability is determined through their linear combination:

$$p_t = a \cdot p_t^{Markov} + (1-a) \cdot p_t^{emit} \tag{82}$$

Where p_t^{Markov} is the Markov model, which is based on state persistence that has already been calculated:

$$p_t^{Markov}=p_{01}(1-p_{t-1})+(1-p_{10})p_{t-1} \tag{83}$$

While p_t^{emit} is the Logit-LASSO model probability. The parameter $a \in [0,1]$ has an optimal value that is chosen by testing different values and picking the one that minimises forecast error on recent data.

3.3.4. MAGNITUDE MODEL

After the estimation of whether imports are likely "ON", the magnitude of imported energy needs to be estimated. For all hours where $s_t=1$, a ridge regression model is used:

$$E[y_t \mid s_t=1]=Z_t^T\gamma \tag{84}$$

Where Z_t contains smooth seasonal and calendar effects, avoiding the use of multiple, autocorrelated terms. The vector contains only the Fourier daily harmonics and monthly dummies. The ridge regression minimises:

$$\min_{\gamma}\|y_{ON}-Z_{ON}\cdot\gamma\|_2^2+\lambda\|\gamma\|_2^2 \tag{85}$$

Where only "ON" observations are used, and λ penalises large coefficients to avoid overfitting. Before fitting, winsorisation is deployed to prevent single spikes from dominating the ridge coefficients and preserve a realistic profile.

After the magnitude model, the final forecast is calculated:

$$\hat{y}_t=\hat{p}_t\cdot(Z^T\hat{\gamma}) \tag{86}$$

3.3.5. HOLDOUT METHOD

The forecast of the desired month is achieved by firstly initialising the state probability p_0 by using the last filtered training probability of the previous month's same hour. Since the future states aren't known yet, proxy lags are created by repeating the last 24-hour pattern of observed states to fill s_{t-1}, s_{t-2}, s_{t-24} for the start of the forecast month.

The next step is the calculation of logit probabilities p_t^{emit} for all hours of the month. The Markov-logit fusion is investigated iteratively to determine the optimal a and τ. For each τ, the s_t states are calculated, then the T matrix and the weighted LASSO-logit fit. The magnitude model is also trained on the same window. During the last V hours of the validation window, these variables are evaluated for each parameter a, until the optimal MAE is found. The month-to-be-forecasted period is held out for genuine out-of-sample testing. Performance is evaluated using the metrics MAE, RMSE, WAPE, sMAPE, $MAPE_{nz}$, and MASE.

4. RESULTS

4.1. BLOCK ORDERS FORECASTS

4.1.1. FORECASTING SETUP FOR AUGUST

The model that was trained from January to June, and tested in August, provided the model parameters shown in Table 4 as the best-performing configuration combined:

Table 4. Forecast Parameters for August Block Orders.

Term	Outcome	Meaning
Daily Fourier terms	K=5, frequencies= (1/24, 2/24, 3/24, 4/24) h^{-1}	Three daily harmonics that represent the diurnal load and bidding rhythm
Weekly/Biweekly Fourier terms	K=2, frequencies= (1/(24·7), 2/(24·7)) h^{-1}	Two weekly frequencies that add to the overall monthly bidding strategy
ARIMA residuals	(0,0,0)	No autocorrelation left to model
Ridge regularisation	λ=2050	Strong smoothing against overfitting

This setup balances flexibility in seasonality with stability from ridge regression, while validation confirmed that ARIMA residual modelling did not improve accuracy beyond deterministic and exogenous drivers.

In Figure 14, the first plot's blue line shows the training residuals, leftover from removing the deterministic part of the training window. The red segment represents the residual forecast, whose flat shape is attributed to the absence of an ARIMA correction for August. This confirms that the residuals exhibit no strong autocorrelation, so the deterministic structure captures most of the variance. The middle panel divides the time series into three parts: a) the grey part, which is the actual Block Order volume in the training set, b) the green part, which is the fitted deterministic component on the training data, and c) the magenta part, which is the deterministic forecast for August. The third panel shows the actual time series used for training in grey, while the black part shows the final August forecast, composed of the combination of the residual and deterministic components, accompanied by a shaded band representing the 90% prediction interval.

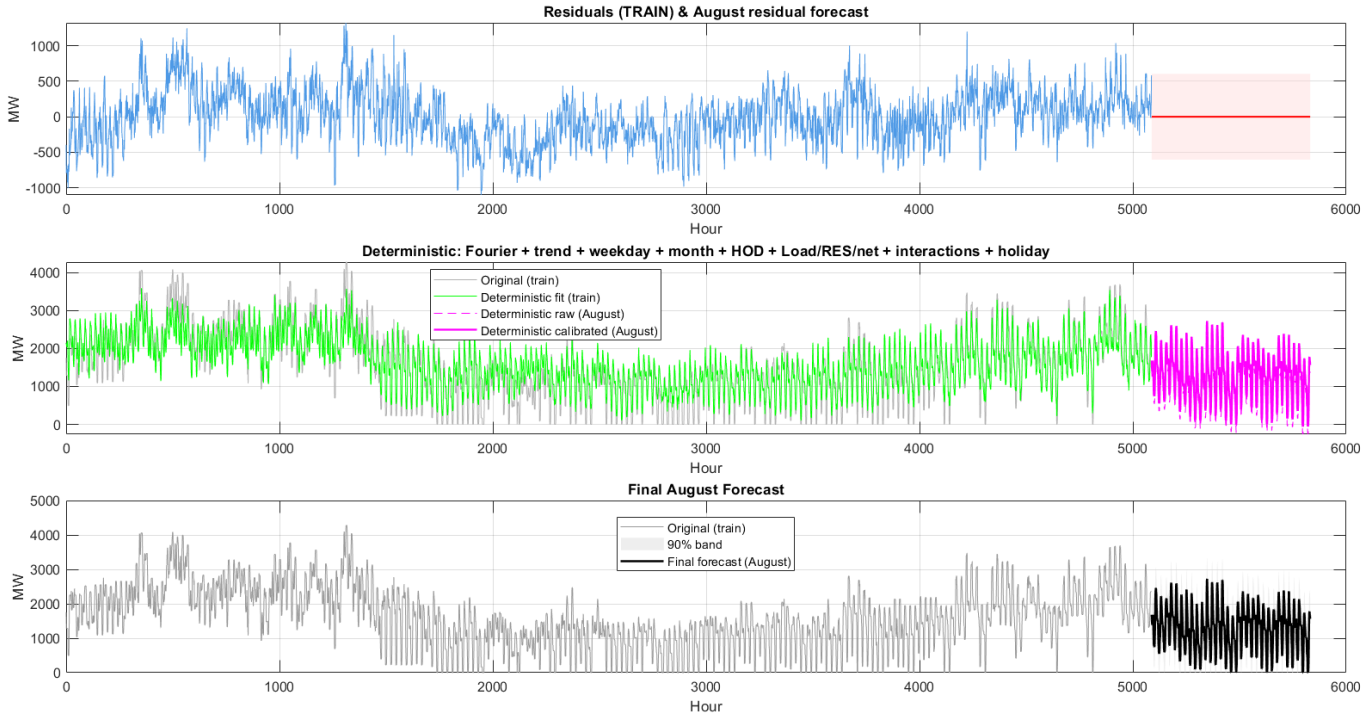


Figure 14. Accepted Block Orders Forecasting for August.

The August forecast extends seamlessly from July, preserving the temporal structure of peaks and valleys. The peaks reach about 3GW, while valleys approach zero, which is consistent with past observed bidding behaviour.

In Figure 15, the red forecast versus the actual black shows the alignment across August. The forecast preserves the shape of the actual data, yet it appears to be a little off-phase with it, while the amplitudes are slightly underpredicted, a typical outcome when $\lambda>0$ in ridge regression, since large magnitudes are penalised.

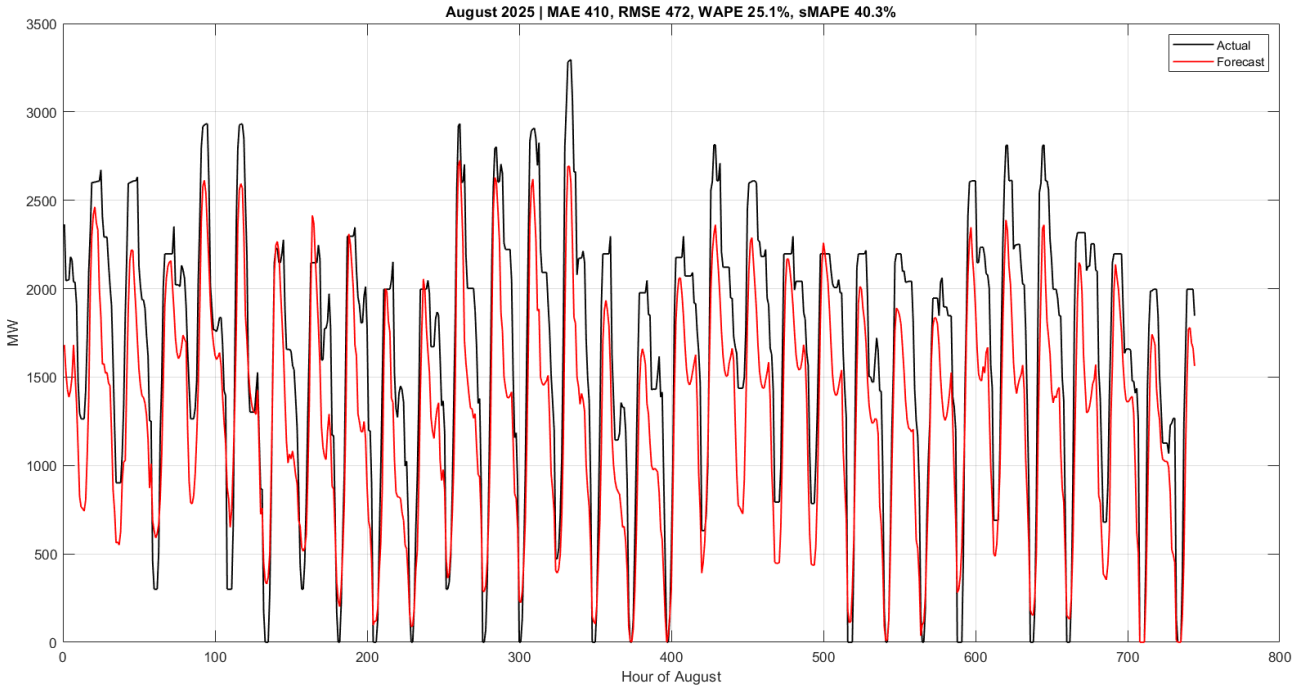


Figure 15. Accepted Block Orders Metrics for August.

A noticeable trait of the forecasts is the undervaluation of the minima. For many hours, the amplitude falls below the recorded values, occurring during 11:00-14:00 on these days when the conventional power systems contribute less due to RES dominance. The most frequent matching cases are when they both reach zero volumes, attributed possibly to positive value restrictions on the forecasts.

The metrics that were used to evaluate the forecast's fitness to the testing window are shown in Table 5:

Table 5. Forecasting Metrics for August Block Orders.

Metric	Value
MAE	410.4 MW
RMSE	471.9 MW
WAPE	25.1%
sMAPE	40.3%
$MAPE_{nz}$	28.3%
MASE	1.33

The metrics indicate that, despite shape consistency, the forecast performs overall poorly on the testing window. The clearest indicator is the MASE value that is greater than 1, meaning that the model performs worse than a naïve model.

4.1.2. FORECASTING SETUP FOR SEPTEMBER

The same process was conducted for September, with the actual Block Orders recorded in August serving as the validating window. The parameters of September's forecast are compiled in Table 6:

Table 6. Forecast Parameters for September Block Orders.

Term	Outcome	Meaning
Fourier terms	K=7, frequencies=(1/24, 2/24, 3/24, 4/24, 5/24, 6/24, 7/24) h^{-1}	Capture of finer intra-day oscillations
Weekly/Biweekly terms	K=2, frequencies= (1/(24·7), 2/(24·7)) h^{-1}	Two weekly frequencies that add to the overall monthly bidding strategy
ARIMA residuals	(0,0,0)	No autocorrelation left to model

Ridge regularisation	λ=2200	Strong smoothing against overfitting

For this month, the deterministic term is more complex, having captured 9 frequencies in total, while the residuals maintain their white noise behaviour, and the ridge regression coefficient maintains the same level of overfitting prevention. Figure 16's first panel indicates that residuals are centred near zero and show no strong autocorrelation, implying that the deterministic model is sufficient. The second panel illustrates continuity from summer to early autumn, while capturing the gradual reduction in block magnitudes as demand softens after August. The last panel demonstrates that the deterministic signal carries over smoothly, which closely tracks the observed September load pattern, reproducing day-to-day volatility well.

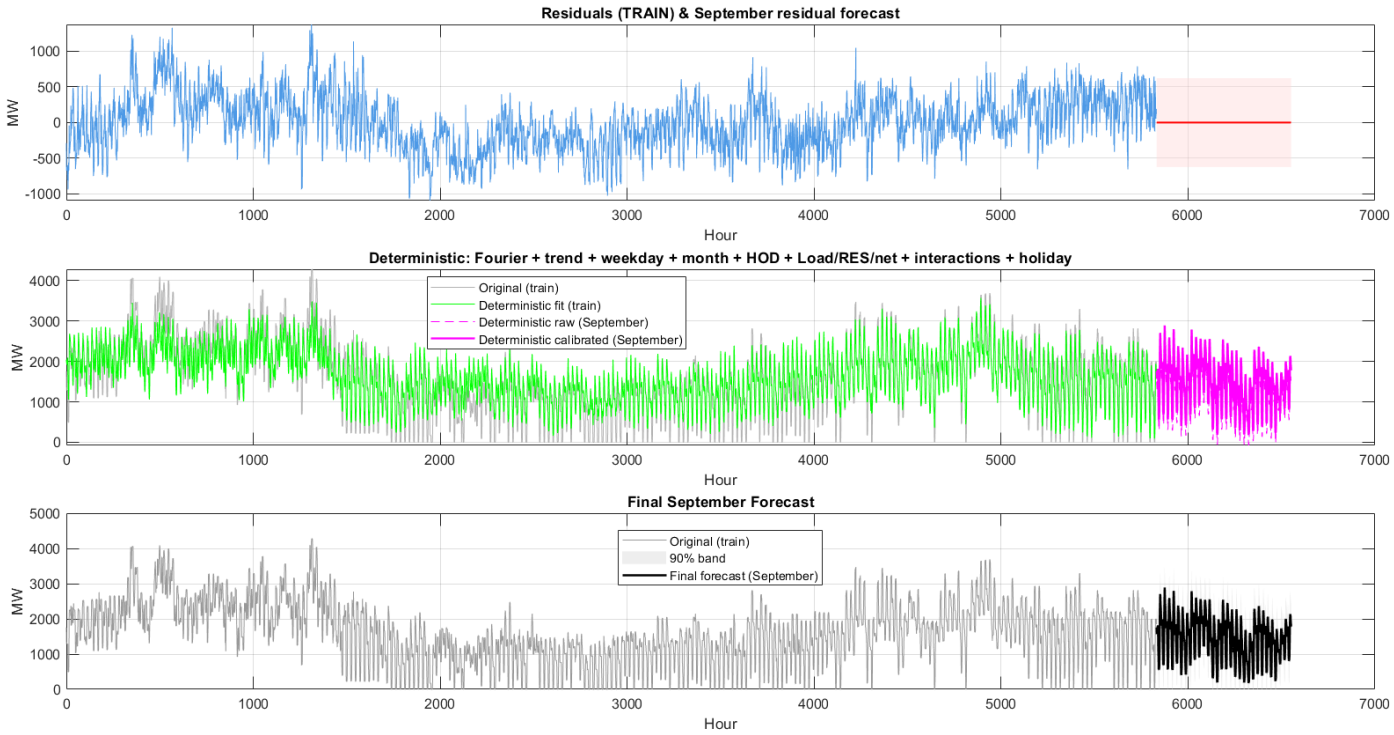


Figure 16. Accepted Block Orders Forecasting for September.

The residual patterns flattened, indicating that most of the temporal and exogenous structure, daily, weekly, and load/RES interactions, were successfully captured by the deterministic component, leaving minimal unexplained variance for ARIMA correction. The deterministic component uses dozens of predictors, yet the ridge regression stabilises them and distributes "credit" smoothly between overlapping variables.

As seen in Figure 17, the forecast follows the actual values with good phase alignment. Peaks and valleys are correctly timed, and amplitude underestimation is smaller than in August, showing that adding extra Fourier harmonics improved precision, with the ridge regression coefficient contributing successfully.

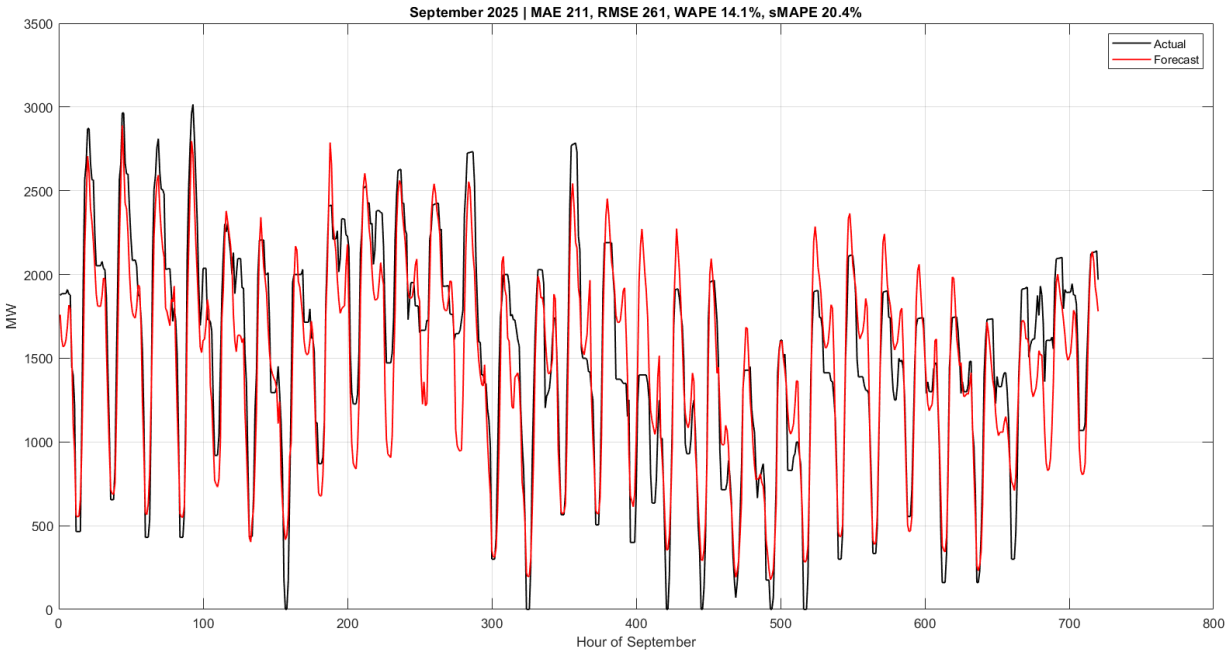


Figure 17. Accepted Block Orders Metrics for September.

The September Block Order forecast shows a marked improvement over August, both visually and in quantitative metrics, as seen in Table 7.

Table 7. Forecasting Metrics for September Block Orders.

Metric	Value
MAE	211.2 MW
RMSE	260.6 MW
WAPE	14.1%
sMAPE	20.4%
$MAPE_{nz}$	18.2%
MASE	0.69

From the Table, the greatest improvement seems to be the reduction of MASE to a value close to 0.5, meaning that the model outperforms the naïve forecast and validates its design. This improvement reflects the model's ability to generalise better once trained on a longer, more complete dataset and to capture the more stable, less volatile market dynamics of September compared to the summer peak month of August.

In order to both calibrate and test the model properly, the accumulation of further data is needed, especially during different seasons of the year, when the climate and general operational differences will serve as adequate "stress" tests for the algorithm.

4.2. HYBRID ORDERS FORECASTS

4.2.1. FORECASTED HYBRID ORDERS FOR AUGUST

The methodology's overall accuracy in forecasting future hybrid orders was tested by calculating the absolute value of error as $|\text{Error}|=|y_t-\hat{y}_t|$. The absolute error between the forecasted and actual hybrid curves for August is shown in Figure 18.

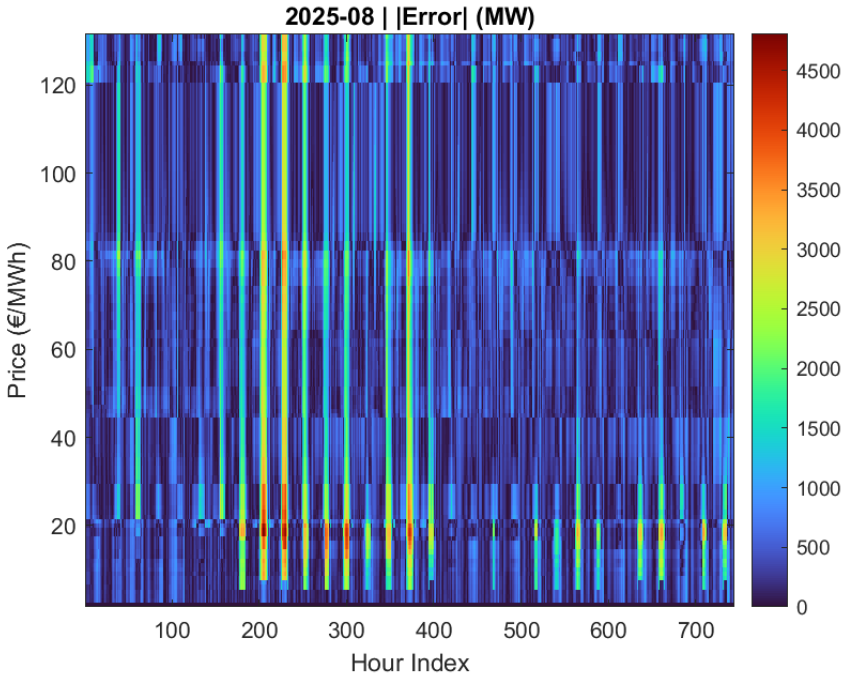


Figure 18. Hybrid Forecast Heatmap Error for August.

The vertical axis of the price range is limited to the area of the most significant errors. The dominant error values are between 0 and 500 MW, indicating the model's efficiency at capturing the general curve shape. For some hours, the higher errors occur simultaneously across many price levels, indicating a shift in the entire curve. Such a difference can be attributed to various factors, like atypical behaviour in the electricity load, sudden shifts in RES generation not captured by the regressors, abrupt changes in daily or weekly behaviour, or market events that cause a real shift in the hybrid curve. The highest errors are noticed at 20-30 EUR/MWh and 70-80 EUR/MWh, the transition points between RES and thermal generation or possible breakpoints between lignite and natural gas bids.

Figure 19 illustrates indicative hourly hybrid order forecasts across the month under test. Monotony and general shape are achieved across hours, but the difference between the forecasted and the actual hybrid could lead to a substantial difference between the recorded and forecasted MCP.

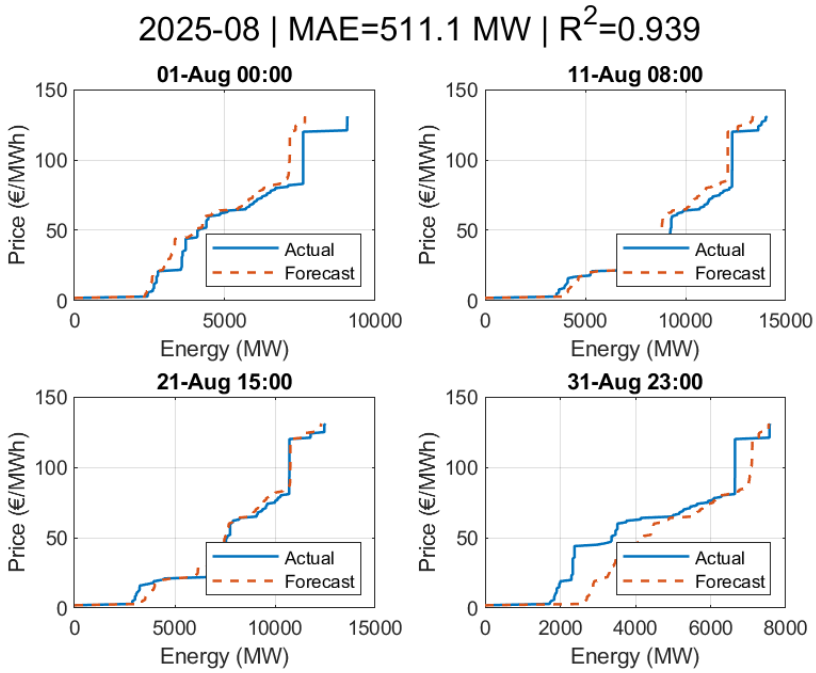


Figure 19. Indicative Hybrid Curve Forecasts in August.

The dashed forecast curve follows the original curves' slope and curvature, indicating that the FPCA eigenfunctions have accurately captured the structure of the curves, and the score regression has predicted the dominant modes of variability effectively, minimising major shape distortions. The algorithm struggles to predict the vertical jumps at 100-120 EUR/MWh, slightly underestimating the steepness or shifting it by a few hundred MW. For the AIC metric, the optimal number of parameters is recorded in Table 8:

Table 8. Optimal number of AIC parameters for August.

Parameters	Selected
Daily harmonics	2
Semi-daily	0
Weekly	1
48h	0
72h	0
Exogenous	Load, RES
Weekend dummy	Yes
Intercept	Yes

Noise variance	Yes
Total number	**11**

By using these parameters, the model achieves R^2=93.9% and metrics:

Table 9. Hybrid Order Metrics for August.

Metric	August
K	3
Explained var. (%)	99.62
MAE (MW)	511
RMSE (MW)	776
sMAPE (%)	7.5

The model achieves a satisfactory percentage of R^2, as well as sMAPE of less than 10%. The significantly higher value of RMSE compared to MAE validates the heatmap, where the significant errors are, while few in number, extreme in value, in the range of 4000-4500 MW.

4.2.2. FORECASTED HYBRID ORDERS FOR SEPTEMBER

The same methodology is repeated for September:

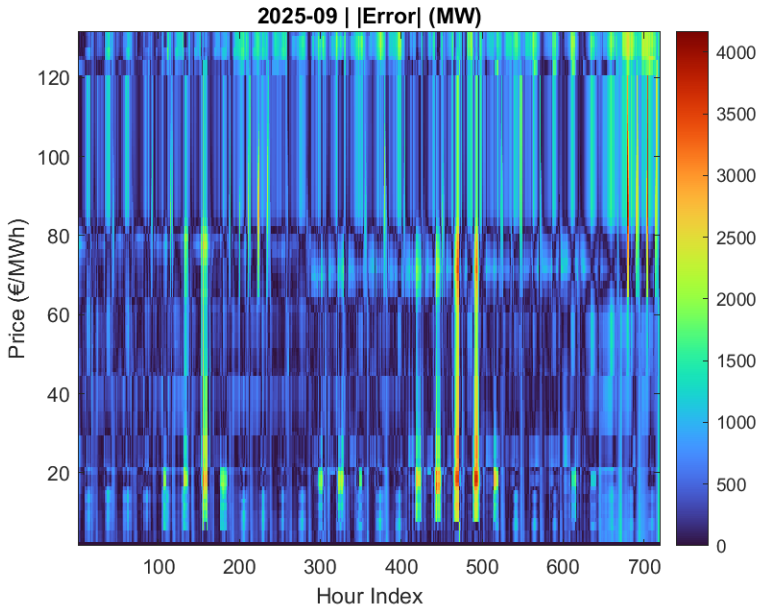


Figure 20. Hybrid Forecast Heatmap Error for September.

September also shows low errors across hours and price ranges, yet worse performance than August. The reasons could be increased price variability, more transitions between RES-dominated and thermal-dominated supply, and higher volatility in mid-price ranges. In these instances, the entire supply curve shifts up or down, attributed to load drops due to weather transitions, resulting in the inability of load regressors to track sudden events. Further investigation is required to determine if these sudden changes are caused by RES volatility, maintenance periods, hydro releases, etc. Around 60-80 EUR/MWh, there is a slightly increased error range, the price region corresponding to the mid-load thermal generation block. In this zone, the slope of the curve increases, and supply flexibility "shrinks", making it one of the hardest parts of the curve to model. Small RES changes or thermal redispatch cause notable shifts in that area, which the FPCA components capture reasonably well, despite the small deviations.

In Figure 21, indicative hybrid order curves are compared to their respective forecasts, where the monotony and increasing slope are once more achieved:

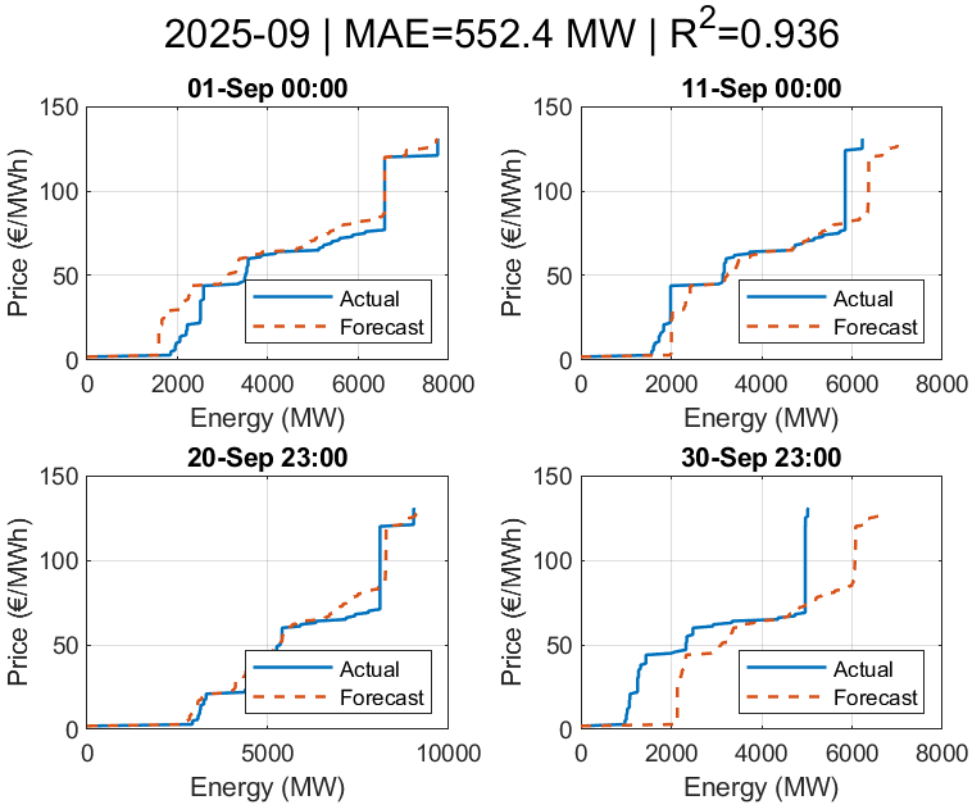


Figure 21. Indicative Hybrid Curve Forecasts in September.

The indicative hourly curves depicted demonstrate that the model tracks the shape of the curve, mirroring major steps and slopes. However, there is a consistent underprediction of energy, which means that FPCA is capturing structural modes correctly via the score regressors, but the forecast predicts more energy volume bid at the same price as the original curves. Various factors could lead to this, such as load regression that overestimates demand, or the second eigenfunction of V matrix may be over-amplified. Since September is a transitional month between summer and autumn, such behaviour is to be expected to some degree.

For the AIC metric, the optimal number of parameters is organised in Table 10:

Table 10. Optimal number of AIC parameters for September.

Parameters	Selected
Daily harmonics	3
Semi-daily	0
Weekly	1
48h	1
72h	0
Exogenous	Load, RES
Weekend dummy	Yes
Intercept	Yes
Noise variance	Yes
Total number	**15**

AIC required more parameters to achieve similar levels of R^2, explained by the increased volatility of bidding behaviour compared to August. Table 11 compiles the forecast metrics for September, with a total $R^2 = 93.6\%$:

Table 11. Hybrid Order Metrics for September.

Metric	September
K	3
Explained var. (%)	99.57
MAE (MW)	552
RMSE (MW)	739
sMAPE (%)	9.5

The FPCA methodology explains more than 99.5% of the curve variance, using only three principal components. Both monthly forecasts yield R^2 near 0.94, indicating that the model captures most of the

curve variability. Compared to August, sMAPE is increased, but still below 10%, while MAE and RMSE maintain, more or less, their levels.

4.2.3. OVERVIEW OF FPCA FORECAST PERFORMANCE

Both months demonstrated FPCA scores K=3 and consistent R^2 values, showing very little structural shift between late summer and early autumn. The combination of the deterministic Fourier and exogenous variables effectively replaced the need for strong ARIMA dependence, since residuals are near-white, thus ensuring temporal stability. The downside is that there is a slight underestimation during extreme scarcity, around 120 EUR/MWh and higher.

Overall, the final model forecasts hybrid curves with sufficient shape consistency and statistical accuracy. The exogenous time-series contributed substantially to this result, providing a mechanism that generalises across both months under study. Their incorporation proved to be decisive, since load acts as a positive driver of rightward curve shifts (higher energy volumes at each price level), while RES availability exerts the opposite effect by extending the zero-marginal segment and flattening the overall shape.

The hourly absolute error distributions for both months show a low overall magnitude and structured periodicity, rather than random dispersion. In August, the largest deviations occur intermittently at 20-40 EUR/MWh and around 80 EUR/MWh, corresponding to transitions between RES-dominated and conventional units. September displays a similar vertical error pattern, but with reduced amplitude, suggesting smoother system conditions and improved model generalisation. The vertical stripes in the heatmaps indicate weekly cycles of residual mismatch, not instability, representing systematic intra-week variations that could be potentially mitigated by extending the training horizon beyond 120 days or adding longer harmonics.

4.3. IMPLICIT IMPORT ORDERS FORECASTS

4.3.1. OPTIMAL CONFIGURATION FOR AUGUST

The deployment of the methodology described in Section 3.3 for August yields the results compiled in Table 12:

Table 12. Markov-Logit Parameters for August

Model Parameters	
Parameter	Value
Validation window (h):	336
Threshold τ (MW):	800
Markov-Logit parameter α:	0.3
LASSO λ:	0.002

Selected Features	
Intercept β_0 coefficient:	-2.63
s_{t-24} coefficient:	0.95
s_{t-1} coefficient:	0.49
$\sin\left(2\pi t/24\right)$ coefficient:	0.02
$\cos\left(2\pi t/24\right)$ coefficient:	-0.15
$y_{t-1}^{(w)}$ coefficient:	0.005
$y_{t-24}^{(w)}$ coefficient:	0.0004
Metrics (Test Window)	
MAE (MW)	171.1
RMSE (MW)	227.2
sMAPE (%)	140.7
WAPE (%)	76.2
$MAPE_{nz}$ (%)	147.0
MASE	0.57

The threshold activation for "ON" states was calculated to be equal to 800MW, a significantly high value. The optimal Markov-Logit parameter α of 0.3 indicates that the model "leans" heavily on harmonic and lagged features for activation prediction, while Markov-like state persistence plays a minor role. The five-fold regularised LASSO coefficient is very low as well, indicating minimal shrinkage of the β coefficients. Six features are preserved in total, the most dominant by absolute value being the s_{t-24} coefficient. Both the value and its positive sign imply that if imports were "ON" at the same hour of the previous day, then they are more likely to be "ON" at the same hour of the next. The next significant feature is of s_{t-1}, and its coefficient value showcases the effect of short-term inertia within the time-series, as they tend to stay "ON" if they were so in the previous hour. The trigonometric features show a minor impact of daily cycles on import activation, while the winsorised lags are barely relevant.

The values of MAE and RMSE are decent, in contrast to sMAPE, WAPE, and $MAPE_{nz}$. This is attributed to the increased number of "OFF" states, where the forecast is zero, that increase the fraction within the

metrics. This means that in such cases, such metrics are obsolete, and must be substituted by different, more reliable ones.

As shown in Figure 22, the model's probability-weighted forecast closely matches both the frequency and amplitude of activation bursts. The top panel shows the blue staircase, which is the actual implicit import orders (in MW) across the training window (January-July). The red curve is the filtered probability of being "ON", meaning the model's estimate of how likely implicit imports are active at each hour. The middle panel shows the black trace, which is the actual trades from training. In the last panel, the black line shows the actual implicit orders during the forecasted month (holdout). The red line is the model's forecast (probability-weighted expected MW).

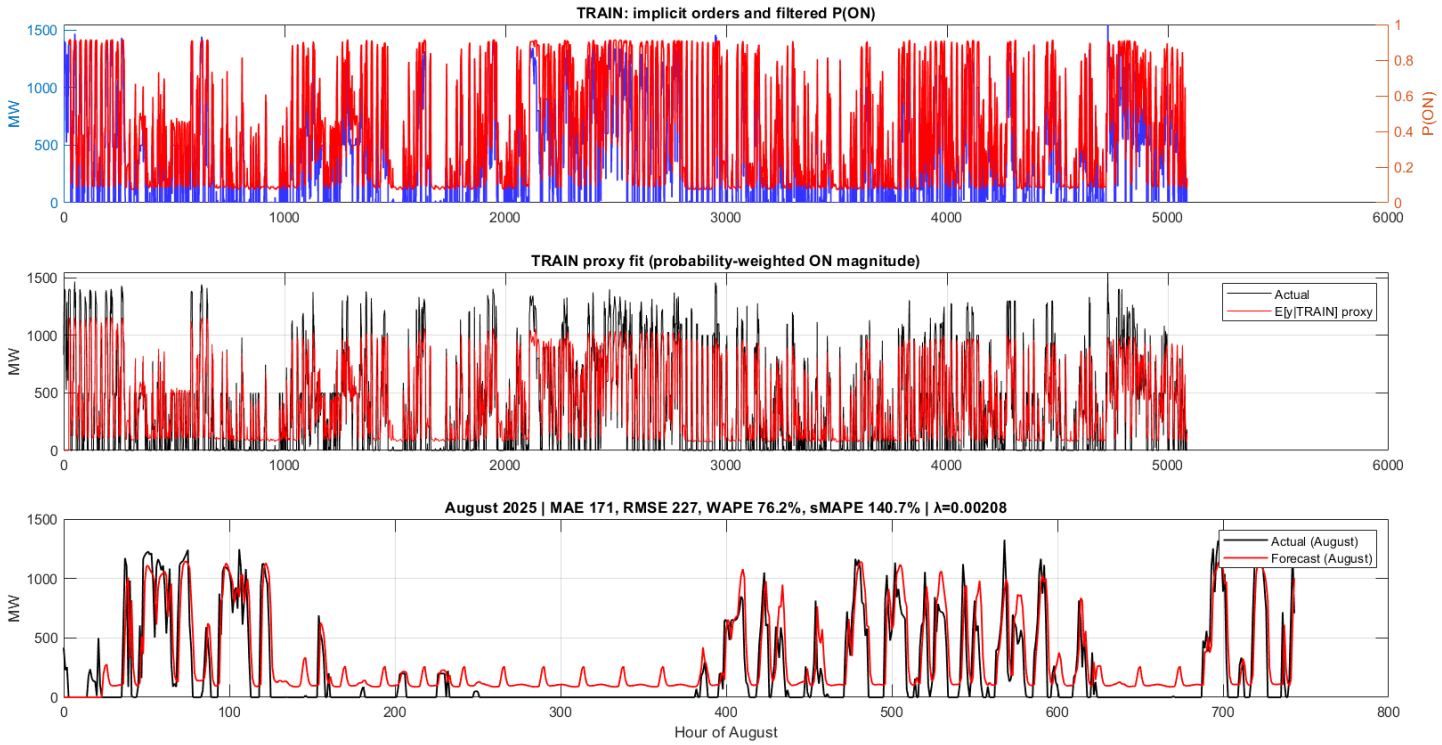


Figure 22. Implicit Import Orders Forecasting for August.

The seemingly random bursts are captured sufficiently, as well as their magnitude. However, in prolonged "OFF" states, the forecast's seasonal components persist, leading to an unwanted non-zero forecast profile, as supported by the trigonometric coefficient feature values from the previous Table.

4.3.2. OPTIMAL CONFIGURATION FOR SEPTEMBER

The repeated process for September leads to the parameter calculations as shown in Table 13:

Table 13. Markov-Logit Parameters for September

Model Parameters	
Parameter	Value
Validation window (h):	336

Threshold τ (MW):	911
Markov-Logit parameter α:	0.0
LASSO λ:	0.0017
Selected Features	
Intercept β_0 coefficient:	-3.24
s_{t-24} coefficient:	0.93
s_{t-1} coefficient:	0.94
$\sin\left(2\pi t/24\right)$ coefficient:	0.00
$\cos\left(2\pi t/24\right)$ coefficient:	-0.47
$y_{t-1}^{(w)}$ coefficient:	0.004
$y_{t-24}^{(w)}$ coefficient:	0.0006
Metrics (Test Window)	
MAE (MW)	196.7
RMSE (MW)	286.0
sMAPE (%)	140.6
WAPE (%)	73.2
$MAPE_{nz}$ (%)	161.6
MASE	0.68

In September, the threshold is higher, indicating higher polarisation between extremely high volumes and "OFF" states, implying potential interconnection operation at near-capacity levels. The Markov-Logit parameter α is zero, meaning that the model deems that Markov persistence does not contribute to the validation MAE, as Logit persistence features, s_{t-24}, s_{t-1}, describe persistence sufficiently, making Markov redundant. The LASSO coefficients of these two features are almost equal, leading to the expectation of daily, extended clusters of "ON" states. Regarding the daily Fourier coefficients, September was characterised by decreased LASSO values, suggesting that daily periodicity was not as present this month,

during which the participants' bidding behaviour changed daily. The metrics of the testing window show similar behaviour as August, with MASE remaining below 1, beating the naïve seasonal (lag 24) model both times.

Figure 23 illustrates the forecasted implicit import orders for September, where the above results are visualised in the constructed time-series:

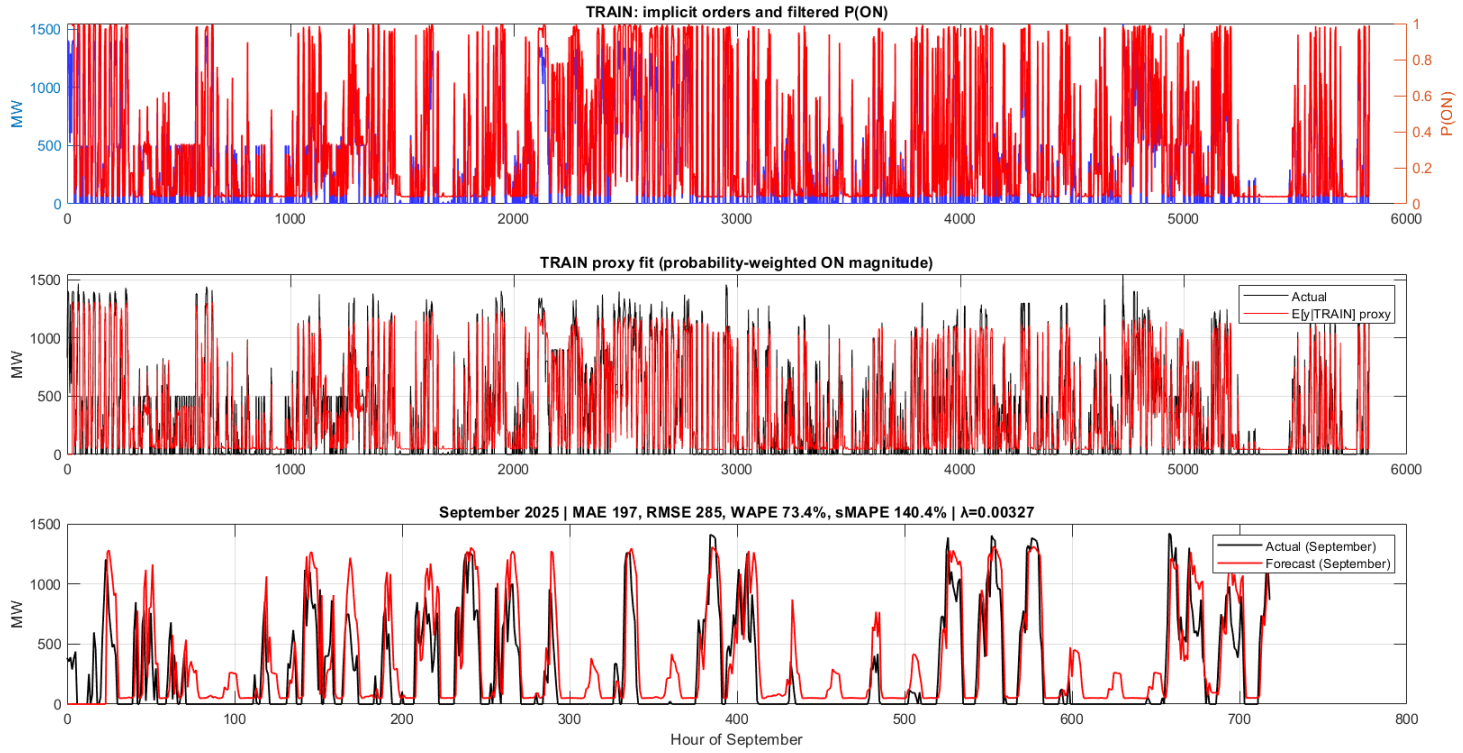


Figure 23. Implicit Import Orders Forecasted for September.

The "ON" persistence is sufficiently captured, a testament to the efficiency of the Logit model, as well as the series' magnitude. The Fourier contribution is smaller during this month, but still falsely assigned to consecutive "OFF" states, highlighting the need to restrict their use.

4.3.3. CONCLUSIONS AND FUTURE WORK

The implicit import forecasting process followed a Markov-Logit/LASSO model that aimed to predict the temporal presence and magnitude of energy orders. This framework was selected through observation of the historical time-series that demonstrated seemingly random activations, without a trend component. The methodology's implementation for two forecasted months, August and September of 2025, managed to capture the persistence in the historical data, whether through the Markov persistence matrix, or the Logit lagged coefficients, minor seasonal Fourier terms, while successfully avoiding overfitting. August exhibited a mixed profile, relying more heavily on the Logit model and a predicted daily pattern that was falsely predicted during prolonged "OFF" states. September exhibited more polarised behaviour of stronger ÖN" state persistence, explained exclusively by the Logit model, while the Markov component was deemed obsolete through the validation phase.

While effective, the current framework is mostly probabilistic, lacking direct exogenous market signals such as interzonal price differences or net-position indicators. In future work, these could be included as additional regressors in the logistic emission model. Alternatively, Bayesian hierarchical or hidden semi-

Markov formulations (Hamilton, 1994; Frühwirth-Schnatter, 2006) could capture non-stationary transition structures if persistence re-emerges at a higher temporal granularity.

4.3.4. COMPARISON WITH THE FOURIER-ARIMA HYBRID FRAMEWORK

The proposed model differs fundamentally from the earlier Fourier-ARIMA hybrid used for Block Order, and partially for hybrid order, forecasting. While all of them exploit deterministic periodicity through harmonic regressors, their underlying assumptions and objectives diverge sharply. The Fourier-ARIMA and FPCA approaches treat the target as a continuous, autocorrelated process, decomposing it into smooth deterministic trends and stochastic residuals modelled through autoregressive-moving-average components. They perform well for stable, quasi-stationary series, such as cleared energy blocks or market prices, where residual autocorrelation is informative and temporal inertia is present.

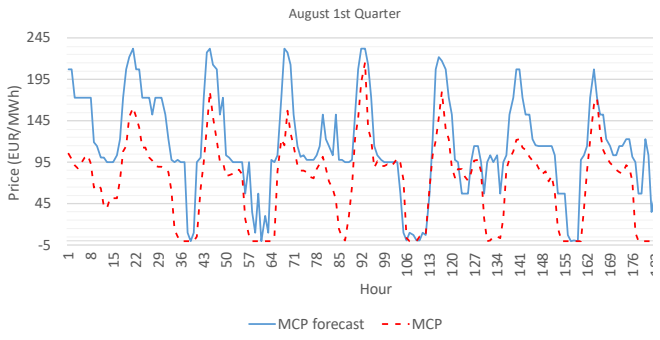
On the other hand, implicit import orders display abrupt, binary-like switching between inactive and active regimes, with fluctuating serial dependence, since activation persistence behaviour changes every month. For such an intermittent process, the ARIMA component adds little explanatory power and can overfit noise. Replacing the ARIMA residual model with a logistic activation mechanism reframes the problem probabilistically, so, instead of predicting continuous deviations around equilibrium, the model estimates activation likelihoods modulated by deterministic cycles.

The two methods are complementary, rather than competing. The Fourier-ARIMA hybrid remains suitable for smooth, autocorrelated market quantities, whereas the Markov/LASSO-logistic process is tailored to reactive, event-driven import flows. Together, they provide a unified forecasting architecture capable of handling both continuous and intermittent processes within the DAM context.

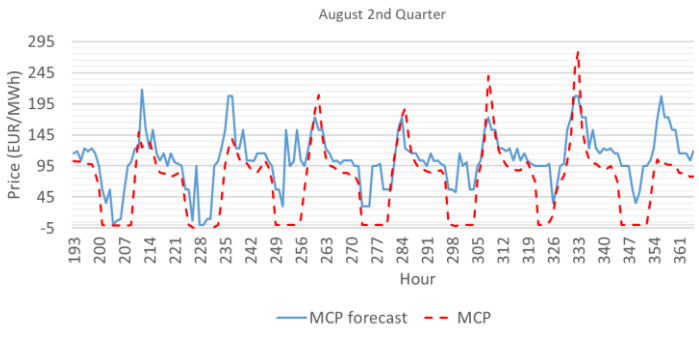
4.4. MCP FORECASTS

4.4.1. MCP FORECASTS OF AUGUST

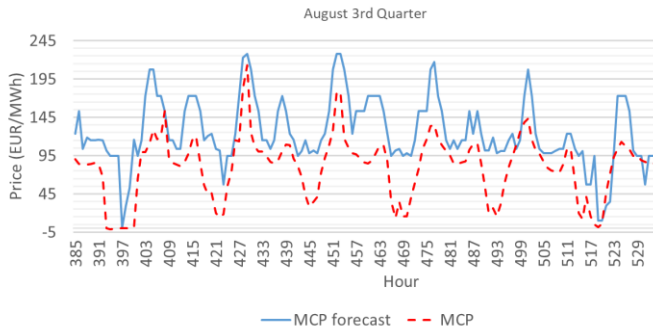
By shifting the forecasted hybrid curves via the forecasted accepted Block Orders and implicit imports, the final supply curves occur, and their intersection with the actual, recorded load for the two months under testing provides the forecasted MCP. In Figure 24, the forecasted MCP is depicted, along with the actual price, divided into four quarters for clarity.



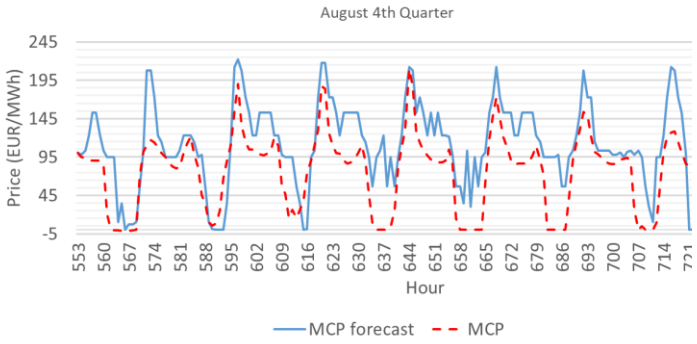
(a)



(b)



(c)



(d)

Figure 24. Forecasted and actual MCP in August.

Overall, the intra-day pattern is captured very well across all four quarters, the peaks and troughs occur at the right hours, and the daily profile (evening peak, morning "shoulder", midday "valley") is consistently reproduced. The mid-range prices (60-130 EUR/MWh) are forecasted quite accurately, and in many hours, the two lines overlap. Yet, there are systematic issues in the extremes, where the very low or zero prices are almost always under-predicted, and some very high peaks are under-shot or over-shot, depending on the period. The model also predicts high volatility for a few hours per day, mostly during times of RES domination or transitional hours from late night to early morning.

More specifically, during the first quarter's first three days, the forecast is consistently higher than the actual MCP. Its peaks are higher, while its valleys are not low enough. As the week progresses, the timing aligns with the exact values, but the forecast remains "stuck" around 90-120 EUR/MWh, even when the actual MCP drops to nearly 0 EUR/MWh. This is a very strong expression of the aforementioned high bias, the final forecast supply curves are too expensive in low-price segments, especially early in the month, where patterns are slightly different from the training window of January to July. The FPCA smooths and "pulls" the extremely cheap days back towards the typical curve, and the intersection with the actual load lands in the 100-150 EUR/MWh range instead of the 0-20 EUR/MWh one.

In the second and third quarters, there is a much better overlap between forecasted and actual values. The peaks align in time and are closer in magnitude. Mid-range levels track each other well, yet the 0-price hours are not fully reproduced. At this stage, the supply and RES profiles are closer to what the model has been trained on in previous months, so the overall methodology works better. The remaining discrepancy is almost purely a failure to reproduce the RES surplus/curtailment regime [32].

In the fourth quarter, there are several days with very low MCP, often reaching 0 EUR/MWh for many consecutive hours. The forecast manages to get the transition in the right hour, but it never stays at 0 for long, and sometimes it reverts to the 80-100 EUR/MWh range, instead of the 0-20 EUR/MWh. Furthermore, on many days, the peaks are slightly under-forecast. Overall, the last quarter seems to be a combination of more intense RES-surplus periods that the model cannot capture in full, and there are very strong peak events that exceed the behaviour that the model has learned [33].

The influence of Block Order and implicit forecast is non-negligible. In all quarters, the night's low prices are not fully captured. The block forecast underestimates the amount to be validated during these times, failing to shift the hybrid curve enough to lower the final price at the intersection with the load. Furthermore, on days when the block forecast under-forecasts the volume peaks, the MCP is estimated to be lower than

the actual value. On the contrary, when the block volumes are over-predicted, the MCP forecast remains too high, failing to capture the decreasing effects of RES participation in the middle of the day. Despite this, the Block Order forecast captures the correct timing of the actual values, thereby transmitting this synchronicity between the final MCP estimation and the actual MCP values.

Ridge regression shrinks all coefficients except the intercept towards zero, which has a very specific effect on Block Order time-series. The large positive coefficients are shrunk, leading to under-predicted block peaks, while the negative coefficients are over-predicted, leading to a divergence between forecasted and actual block valleys. These differences in block forecast extrema are translated to the final MCP as differences in curve shift, changing the intersection point with the load, and misalignment with the actual MCP values. Despite these shortcomings, ridge regression is still necessary, since it manages to stabilise the fit in the presence of huge collinearity (Fourier, load, RES interactions), reducing the validation MAE over the bulk of hours instead of just a smaller set of hourly extremes, preserving the consistency of the final forecast. The optimiser suffices to sacrifice extreme peaks in order to produce stable values of lower average error.

The implicit imports, being in lower amplitude by comparison, don't seem to damage this much the MCP forecast. Examples like the 300-400 hour-range of August, the push that the imports cause, compared to the absence of the actual push from imports in those hours, is not suffered much, only in the hours when the Block Orders are ought to do it.

4.4.2. MCP FORECASTS OF SEPTEMBER

Similarly, for September, the forecasts are calculated by using the three forecasted time-series, leading to a final hourly bidding curve that intersects with the recorded demand to yield the forecasted hourly MCP:

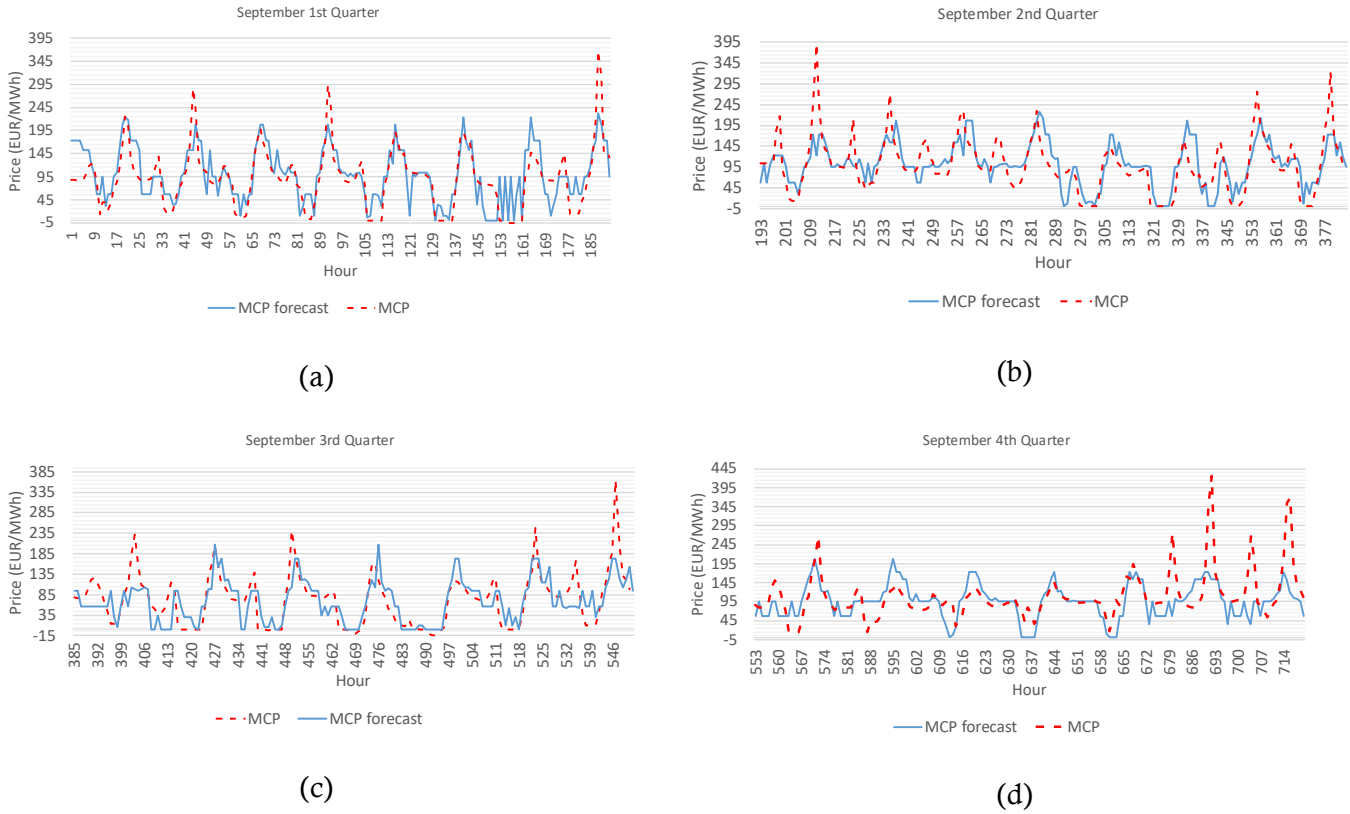


Figure 25. Forecasted and actual MCP in September.

Compared to August, the September forecast demonstrates a very different profile, where the model captures both the timing of daily cycles and the mid-price range very well. The peaks and troughs occur at the correct hours in all quarters, overlapping with the actual time-series in the price range of 70-140 EUR/MWh. The forecast captures low and zero prices more efficiently than it did in August, yet it systematically misses high MCP peaks. Volatile behaviour in the forecasts is still noted during RES domination, a behaviour that deviates from the recorded market behaviour.

In the first quarter, the daily cycles are well aligned with the actual time-series, matching the moderate prices quite well. The large peaks are severely underestimated, and the troughs are satisfactorily captured, though some erratic behaviour is observed during the hours 150-170. During this quarter, the FPCA basis vectors capture the typical shape but fail to reproduce the extreme right-tail behaviour of the supply curve, while the exogenous parameters of load and RES explain most of the moderate variation, but scarcity events that might be the cause of significant price volatility are not captured. Consistency is maintained during the second quarter as well, and the forecast matches the daily minima better, yet the peaks above 200 EUR/MWh are still underestimated, while some days are forecast too smoothly, the main contributor being the FPCA smoothing effect. Unlike August, the model does not overpredict valleys, with FPCA compressing rare patterns and Gaussian score regression smoothing out outliers. The third quarter represents the best quarter visually, with many days being forecast nearly perfect in both shape and amplitude. The persistence of underestimated peaks is evident here as well. During September's last quarter, the peaks were undervalued once more, while the forecast volatility appears lower than the actual from the real values.

While August exhibited a systematic overprediction of prices during RES-dominated hours, September's forecast undervalued the early evening peaks on almost all days. In the latest month, the model captured the daily profile, due to validation during August, improved accepted Block Order forecasts, and the decrease in RES contribution in September. It maintained its difficulty in capturing extreme behaviours of extremely high or negative prices.

5. CONCLUSION

The concept of this thesis was the study of specific forecasting techniques that would potentially forecast the future values of key Electricity Market Order types that form the final energy price (MCP), and therefore lead to the forecasting of the MCP itself. The decision of the forecasting algorithms was based on the nature of the recorded time-series of these order types, which are published by HEnEx, the Greek NEMO, and therefore their efficiency would not only provide valuable forecasts to be used by stakeholders, but emerge the underlying market mechanisms and Participant behaviour under various seasons and market events. The Order types under study were the accepted Block Order energy volumes, the Hourly Hybrid Orders, and the Implicit Import Orders. The training data spanned from January to August 2025, and the forecasted period was August and September 2025.

The selected methodology for Block Orders was ridge regression that aimed to capture the deterministic aspect of Participant behaviour, backed by ARIMA for the capture of leftover residual structure. For both months, the methodology performed well, in terms of shape and synchronicity preservation, but the amplitude of energy volume near the extrema was significant, even with the use of exogenous features of the recorded RES and Load values during the testing period, a setback that needs to be mended since Block Orders contribute significantly to the aggregated Hybrid Sell curve, and therefore to the price formation. For the Hourly Hybrid Orders, FPCA was deployed within an overall set of ARIMA and a deterministic

toolbox of exogenous features, Fourier transform and calendar events. The methodology succeeded in the realistic curve formation, managing to converge to the actual shape, monotonicity and scale of the recorded curve orders, avoiding overfitting at the same time and capturing the underlying patterns. Implicit Imports required a different approach than the other two, due to their stochastic nature, which led to the use of a Markov-LASSO/Logit algorithm that simulated the specific market behaviour as a probabilistic, activation-like participation with various degrees of persistence for each month. The algorithm managed to capture the majority of the imports' energy activation, yet it frequently underpredicted their magnitude and assigned activation times in cases when the imports were zero, a miscalculation of the algorithm's deterministic component.

Both the shortcomings and the successes of the forecasts were transferred to the final MCP forecast. The Block Order model consistently captured the timing of the actual Block commitments, which led to correctly timed MCP peaks across the testing phase. However, the differences between the forecasted and actual Block Order magnitudes translated directly into amplitude deviations in MCP. The forecasting process for the Hybrid Orders succeeded in capturing the transition from high to low MCP values and vice versa efficiently, adding to the successful capture of price variability on a daily scale. Although Implicit Imports are smaller in volume compared to Block Orders, their forecast's deviation from the actual values contributed to the distortion of the MCP forecasts, which, in combination with Block Order over- and underpredictions, introduced increased volatility in the MCP forecasts, a phenomenon that was missing from the actual values. Overall, the methodology of this thesis did manage to produce, to some degree, consistent and realistic MCP forecasts that seemed to improve in all aspects as the training window grew and the algorithms trained under more seasons and market events. On the other hand, the efficiency of the methodology was mostly improved with the use of the exogenous regressors of RES generation and Load demand for the months that were tested, which, in a realistic scenario, would not be available, only their forecasts. This means that the methodology could potentially decrease its accuracy further in a realistic forecasting process.

As future work, the entire process will be trained and assessed by using the time-series of the new 15-minute MTUs under which the Market operates since October 2025, testing its ability to capture the volatility caused by a shorter time step. Additionally, tuning of the ridge regression's lambda coefficient and ARIMA grids to fit the forecast shapes effectively to the actual values across all months under study is crucial for the perseverance of the methodology. Finally, the integration of additional, elementary exogenous regressors, like time-series of capacity allocation and weather events, could introduce market patterns that have been missing from the training process so far.

BIBLIOGRAPHY

[1] Kirschen, D., & Strbac, G. (2019). Fundamentals of power system economics (Second edition). Wiley.

[2] Bunn, D. W. (2000). Forecasting Loads and Prices in Competitive Power Markets. Proceedings of the IEEE, 88(2), 163–169. https://doi.org/10.1109/5.823996

[3] Weron, R. (2006). Modeling and Forecasting Electricity Loads and Prices: A statistical approach (1st Edition). Wiley. https://doi.org/10.1002/9781118673362

[4] Weron, R. (2014). Electricity Price Forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030–1081. https://doi.org/10.1016/j.ijforecast.2014.08.008

[5] Carmona, R., & Coulon, M. (2013). A survey of commodity markets and structural models for electricity prices. University of Sussex.

[6] SDAC PRESS RELEASE. (2025, July 10). Successful Implementation of 15-Minute Market Time Unit (MTU) in SDAC.

[7] Official Journal of the European Union. (2015, July 25). COMMISSION REGULATION (EU) 2015/1222.

[8] NEMO Committee. (2019, October 4). EUPHEMIA Public Description Single Price Coupling Algorithm.

[9] Clearing & risk management dam—Enexgroup. Accessed on 15 November 2025, at https://www.enexgroup.gr/web/guest/markets-commodities-clearing

[10] HELLENIC ENERGY EXCHANGE S.A. (2024). HEnEx SPOT TRADING RULEBOOK, Day-Ahead & Intra-Day Markets Trading Rulebook (Version 2.0).

[11] Διαχειριστής ΑΠΕ & Εγγυήσεων Προέλευσης Α.Ε. (ΔΑΠΕΕΠ). (2023). Κώδικας Διαχειριστή ΑΠΕ & Εγγυήσεων Προέλευσης [Code of the RES & GO Operator]. ΡΑΑΕΥ.

[12] Madani, M., & Van Vyve, M. (2015). Computationally Efficient MIP Formulation and Algorithms for European Day-Ahead Electricity Market Auctions. European Journal of Operational Research, 242(2), 580–593. https://doi.org/10.1016/j.ejor.2014.09.060

[13] Van Vyve, M. (2011). Linear prices for non-convex electricity markets: Models and algorithms. https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A91368

[14] Tsirekis, K. (2024, May 30). Interconnecting Greece: The role of IPTO. 3rd Tirana Energy Forum.

[15] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

[16] Harvey, A., Koopman, S. J., & Riani, M. (1997). The modeling and seasonal adjustment of weekly observations. Journal of Business & Economic Statistics, 15(3), 354–368. https://doi.org/10.1080/07350015.1997.10524713

[17] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). Time series analysis: Forecasting and control (Fifth edition). Wiley.

[18] Uhrig, M., Vignolo, L., & Müller, O. (2025). Modelado y predicción de la demanda eléctrica: Comparación de enfoques estadísticos y de aprendizaje automático. SADIO Electronic Journal of Informatics and Operations Research, 24(1), e067. https://doi.org/10.24215/15146774e067

[19] Aïd, R., Campi, L., Huu, A. N., & Touzi, N. (2009). A structural risk-neutral model of electricity prices. International Journal of Theoretical and Applied Finance, 12(07), 925–947. https://doi.org/10.1142/S021902490900552X

[20] Ziel, F., & Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-Model. Energy Economics, 59, 435–454. https://doi.org/10.1016/j.eneco.2016.08.008

[21] Ramsay, J. O., & Silverman, B. W. (2005). Functional data analysis. Springer New York. https://doi.org/10.1007/b98888

[22] Gao, Y., Shang, H. L., & Yang, Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. Journal of Multivariate Analysis, 170, 232–243. https://doi.org/10.1016/j.jmva.2018.10.003

[23] Liebl, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. The Annals of Applied Statistics, 7(3). https://doi.org/10.1214/13-AOAS652

[24] Strang, G. (2009). Introduction to linear algebra (4th Ed). Wellesley-Cambridge Press.

[25] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd edition). OTexts.

[26] Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

[27] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57(2), 357. https://doi.org/10.2307/1912559

[28] Krolzig, H.-M. (1997). Markov-switching vector autoregressions (vol. 454). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-51684-9

[29] UĞURLU, U. (2019). ESSAYS ON ELECTRICITY PRICE MODELING AND FORECASTING. ISTANBUL TECHNICAL UNIVERSITY.

[30] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[31] Bertsekas, D. P., & Tsitsiklis, J. N. (2008). Introduction to probability (2nd ed). Athena scientific.

[32] Pavlík, M., Kurimský, F., & Ševc, K. (2025). Renewable energy and price stability: An analysis of volatility and market shifts in the European electricity sector (2015–2025). Applied Sciences, 15(12), 6397. https://doi.org/10.3390/app15126397

[33] Pavlík, M. (2025). Ever more frequent negative electricity prices: A new reality and challenges for photovoltaics and wind power in a changing energy market—threat or opportunity, and where are the limits of sustainability? Energies, 18(10), 2498. https://doi.org/10.3390/en18102498