CENTRALE NANTES

ENGINEERING PROJECT REPORT

CENTRALE NANTES — APPLIED MATHEMATICS DEPARTMENT

MODELLING, FORECASTING & OPTIMISATION OF AN INTRADAY TRADING STRATEGY ON THE ELECTRICITY MARKET

ROBIN GUICHON

30/11/2024 - 31/03/2025

Abstract : *This project develops a complete machine-learning pipeline for short-term electricity price forecasting and systematic trading. An enriched dataset combining Iberian market fundamentals and meteorological variables is constructed, and several predictive models—including XGBoost and LSTM-based architectures—are evaluated. After extensive optimisation, a feature-driven XGBoost model proves the most reliable and is integrated into six algorithmic trading strategies. A month-long backtest highlights the importance of filtering weak signals and applying disciplined position-sizing rules. The study demonstrates that, even in volatile markets, accurate forecasting combined with robust allocation logic can yield consistent and profitable trading performance.*

Keywords : *Electricity markets, Intraday trading, Price forecasting, Machine Learning, XGBoost, LSTM, Time series modelling, Feature engineering, Mispricing, Kelly criterion, Systematic strategies, Algorithmic trading, Backtesting, Risk analysis, Predictive analytics*

Abstract

This project presents the development of a complete end-to-end machine-learning framework designed to forecast intraday electricity prices in the Iberian market and to convert these forecasts into fully systematic trading strategies. The work integrates several disciplines—data engineering, time-series forecasting, optimisation, and algorithmic trading—into a coherent methodological pipeline capable of transforming heterogeneous information into actionable decisions. Electricity markets exhibit volatility, fast dynamics, and strong dependence on meteorological and fundamental conditions, making forecasting challenging but also offering opportunities for data-driven approaches capable of uncovering short-term structure in price formation.

The first component of the project focuses on data construction. A comprehensive dataset was assembled by merging hourly market fundamentals (generation by technology, system load, day-ahead prices) with high-resolution meteorological variables from several major Spanish cities. A dedicated preprocessing pipeline was implemented to cleanse and synchronise the data, treat missing entries, align timestamps, and create an enriched set of temporal features. Beyond simple lags, the dataset incorporates rolling averages, rolling standard deviations, and additional volatility-based indicators designed to encode short- and medium-term temporal dependencies. The resulting dataset provides a robust foundation for both feature-driven models like XGBoost and sequence-driven models like LSTMs.

The predictive modelling stage evaluates two major families of algorithms: gradient-boosted decision trees (XGBoost) and recurrent neural networks (LSTM, GRU and attention-based architectures). Although recurrent architectures have demonstrated excellent performance in large sequential datasets, the limited historical depth available in this project prevented them from effectively learning long-range temporal structure. As a result, LSTM-based models suffered from poor generalisation and unstable behaviour. In contrast, the feature-rich XGBoost models adapted extremely well to the available data. The model combining lagged features with rolling-window statistics achieved the best predictive performance, demonstrating that, under data scarcity, gradient boosting remains a highly competitive approach. These forecasts were subsequently embedded into six systematic trading strategies.

Three strategies operate without explicit capital allocation (directional, mispricing, threshold-based) and three incorporate dynamic position sizing (proportional signal, sigmoid confidence sizing, and the dynamic Kelly criterion). A 30-day backtest (720 hourly trades) enabled a detailed comparison of profitability, win-rate, drawdown behaviour, volatility, and Sharpe ratio. The no-capital strategies showed that filtering weak signals through a threshold rule significantly improves robustness by preventing trades in uncertain conditions. Among capital-based methods, sigmoid confidence sizing produced the most stable and reliable results: its smooth nonlinear allocation mechanism limited excessive exposure while still exploiting strong predictive signals. Proportional sizing, although profitable, amplified forecast noise, while the Kelly strategy proved too sensitive to estimation errors in win probabilities and payoff ratios. The findings highlight that trading performance depends not only on forecast accuracy but also on the alignment between the predictive model and the structure of the trading rule. A strong model can perform poorly when combined with an aggressive allocation scheme, while a more modest model can yield excellent results within a disciplined decision framework. The systematic comparison carried out in this work confirms that the combination of a robust forecasting model (XGBoost with lag and rolling features) and risk-aware position sizing (the sigmoid strategy) offers the best compromise between return, stability, and risk control.

Finally, the project identifies several avenues for future improvement. Expanding the dataset to several years would allow sequence-based architectures—particularly modern Transformer models—to capture long-range patterns and potentially surpass feature-driven approaches. Probabilistic forecasting, uncertainty quantification, and market-regime detection could further enhance trading decisions by incorporating risk and structural dynamics into allocation rules. Overall, the project demonstrates that machine-learning models, when integrated with appropriate trading logic, can provide powerful tools for short-term decision support in electricity markets, while also revealing the methodological challenges inherent to modelling highly volatile, real-world time series.

Table of Contents

List of Figures

List of Tables

Glossary

aFRR	Automatic Frequency Restoration Reserve
AI	Artificial Intelligence
API	Application Programming Interface
Backtesting	Historical simulation used to evaluate trading strategy performance
BM	Balancing Mechanism
BRP	Balance Responsible Party
DA	Day-Ahead Market
DA (ML)	Directional Accuracy
DL	Deep Learning
EDA	Exploratory Data Analysis
EUPHEMIA	European algorithm used to clear the day-ahead auction
FCR	Frequency Containment Reserve
Forward Market	Market for hedging electricity prices weeks to years ahead
GRU	Gated Recurrent Unit
HMM	Hidden Markov Model
ID	Intraday Market
Kelly Criterion	Rule determining optimal position size based on expected return and variance
KC	Kelly Criterion
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
mFRR	Manual Frequency Restoration Reserve
ML	Machine Learning
Mispricing	Difference between predicted and actual market price
NEMO	Nominated Electricity Market Operator
PnL	Profit and Loss
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
R^2	Coefficient of Determination (R-squared)
Sigmoid Allocation	Nonlinear exposure function mapping signal strength to position size
TF	Transformer (model)
TSO	Transmission System Operator
XGBoost	Extreme Gradient Boosting Algorithm

Framework and Market Foundations

I.1 Project Background and Strategic Objectives

I.1.1 Context and Motivation

Short-term electricity markets have become increasingly complex environments, shaped by the rapid rise of renewable energy sources, the decentralisation of power generation, and the growing interconnection of European systems. These structural shifts have amplified price volatility, particularly in Day-Ahead and Intraday markets where supply–demand imbalances emerge within very short time horizons. Forecasting these fluctuations has therefore become a central challenge for market participants, from producers and suppliers to quantitative traders and financial actors seeking to exploit short-term inefficiencies.

In this context, data-driven approaches provide a promising avenue for anticipating price movements by leveraging the large amount of meteorological, operational and market information now available. Machine learning models — especially gradient boosting and recurrent neural networks — offer the ability to capture complex nonlinear patterns and to integrate heterogeneous data sources in a unified predictive framework. The motivation of this project is rooted in this evolution: understanding whether these techniques can meaningfully improve short-term forecasts and, crucially, whether such predictions can support robust and systematic trading strategies capable of generating consistent returns in volatile electricity markets.

I.1.2 Scope of the Study

The scope of this study covers the full development of a modelling and trading pipeline applied to the Spanish electricity market, using hourly and sub-hourly historical datasets that include meteorological forecasts, renewable generation indicators, system fundamentals and market clearing prices. The project encompasses the entire modelling chain: data acquisition, preprocessing and feature engineering; construction and optimisation of predictive models based on both gradient boosting (XGBoost) and deep learning architectures (LSTM); generation of mispricing signals through the comparison of predicted and realised prices; and ultimately, the design of algorithmic trading strategies that translate these signals into decisions. The work focuses on the financial dimension of short-term markets rather than operational or regulatory considerations. It does not aim to reproduce the behaviour of physical assets or the full complexity of power system operations. Instead, the objective is to evaluate whether predictive analytics can be combined with disciplined trading rules to create a coherent and data-driven systematic strategy. This ensures both scientific rigour and practical relevance, while keeping the study centred on the core relationship between forecasting accuracy and trading performance.

I.1.3 Challenges of Short-Term Electricity Price Forecasting

Forecasting short-term electricity prices presents unique challenges that distinguish this task from conventional financial or time-series prediction problems. Electricity is a non-storable commodity, which forces continuous and instantaneous balancing of the system — a constraint that results in sharp, unpredictable price spikes whenever supply or demand deviates from expectations. Moreover, a significant share of electricity production now depends on weather-driven renewable sources, making meteorological conditions a dominant driver of price volatility. These factors contribute to non-linear dynamics, high-frequency fluctuations and distributional asymmetries that classical statistical models struggle to capture.

Another difficulty arises from the heterogeneous and high-dimensional nature of the data itself, which includes meteorological variables, lagged price patterns, rolling statistics, seasonal components and market fundamentals. Models must extract signals without overfitting, generalise across seasonality regimes and remain robust to structural breaks. Intraday markets introduce an additional layer of complexity, as new information becomes available continuously throughout the trading horizon, causing price formation mechanisms to evolve rapidly within a single day.

These challenges justify the use of machine learning techniques that can absorb large volumes of structured and

unstructured data, model nonlinear relationships and adapt to the fast-changing dynamics of electricity markets. Developing such models and integrating them into a coherent trading framework forms the central ambition of this project.

I.2 Structural Overview of European Power Markets

I.2.1 Chronological Architecture of Power Markets

European electricity markets are organised as a sequence of complementary mechanisms, each playing a specific role in the formation of prices and the balancing of the system. This chronological structure begins with long-term hedging instruments and progressively incorporates shorter-term adjustments as updated information becomes available. At the earliest stage, futures markets allow producers, suppliers and financial actors to hedge price risk weeks, months or years ahead of delivery. These contracts do not aim to reflect real-time system conditions, but rather provide price signals for the medium-term evolution of supply, demand and fuel costs.

The next layer is the Day-Ahead market (DA), which has become the cornerstone of the European market design. Operated through a daily auction, it aggregates supply and demand bids for each hour (and increasingly for each 15-minute interval) of the following day. The auction is cleared by the EUPHEMIA algorithm, designed to ensure a welfare-maximising dispatch while accounting for network constraints and cross-zonal exchanges. The Day-Ahead price serves as a reference for most actors, as it incorporates large volumes of information regarding generation availability, demand forecasts, and expected renewable production.

The Intraday market (ID) bridges the gap between Day-Ahead scheduling and real-time operations. Trading begins shortly after the DA auction and continues up to one hour before delivery. It is structured as a combination of discrete intraday auctions (IDA1, IDA2, IDA3) and a continuous order book where participants can react to updated forecasts, unplanned outages, and real-time system needs. Intraday markets are particularly relevant in systems with high renewable penetration, since wind and solar forecasts are refined throughout the day. These markets exhibit strong price sensitivity, fast-changing conditions and localised imbalances, making them attractive for short-term trading.

Finally, the balancing market ensures real-time stability by activating flexibility resources, such as Frequency Containment Reserve (FCR), automatic Frequency Restoration Reserve (aFRR), and manual Frequency Restoration Reserve (mFRR). These mechanisms guarantee that the grid remains at 50 Hz, and they operate under the responsibility of Transmission System Operators (TSOs). Balancing prices are not meant to be trading opportunities for financial players, but their behaviour influences expectations and provides an additional layer of complexity to short-term price formation.

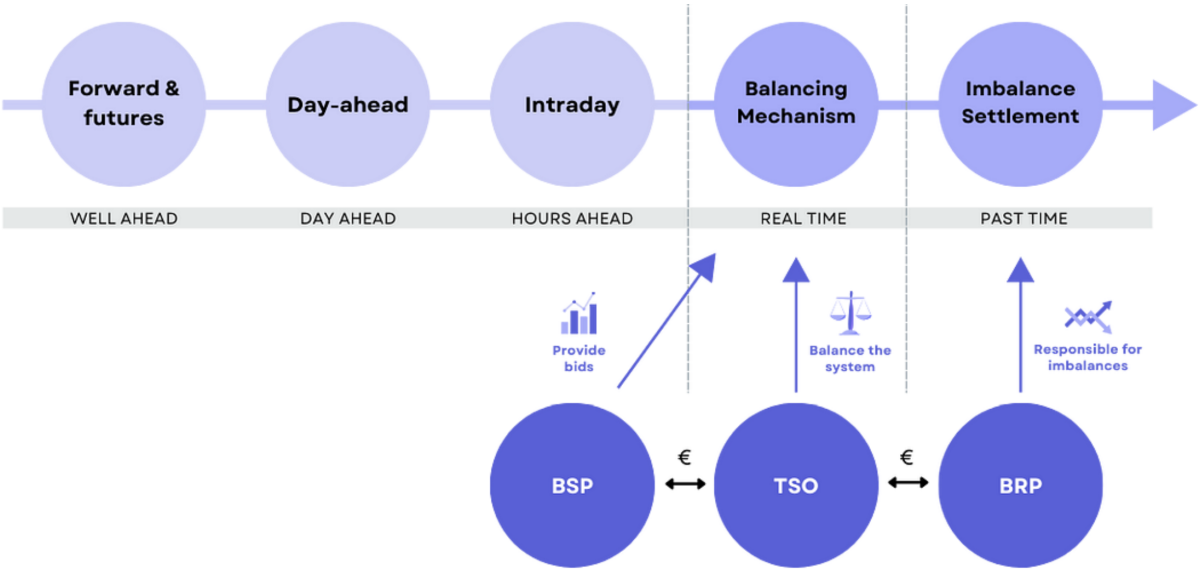

Figure I.2.1 – Structural Timeline of the European Electricity Markets

BSP (Balancing Service Provider) refers to the market participants capable of delivering balancing energy or capacity to the Transmission System Operator (TSO) in exchange for remuneration.
BRP (Balance Responsible Party) is the entity financially responsible for any imbalance between its contracted positions and its actual injections or withdrawals during real-time operations.

I.2.2 Key Market Participants and Operational Roles

The functioning of European electricity markets relies on a diverse set of actors whose interactions collectively determine price formation. At the core are **producers**, ranging from traditional thermal units to renewable generators, each submitting bids according to their marginal costs or opportunity costs. Thermal units typically bid based on fuel prices, carbon costs and operational constraints, while renewable producers bid close to zero due to their negligible marginal costs, creating strong downward price pressure during periods of high solar or wind output.

Suppliers and aggregators represent end-consumers or portfolios of distributed assets. They participate in the Day-Ahead and Intraday markets to secure energy volumes matching expected demand. Their forecasts directly shape market liquidity and price volatility. **Large industrial consumers** may also participate directly, either to optimise their procurement strategy or to monetise flexibility.

On the trading side, **financial actors** play an increasingly influential role. These include quantitative traders, utilities' proprietary trading desks and hedge funds seeking to exploit short-term inefficiencies. Their strategies range from statistical arbitrage to liquidity provision, and they significantly contribute to market depth and price discovery, especially in the Intraday continuous market.

The operational stability of the system is ensured by **Transmission System Operators (TSOs)**, such as RTE, REE, or ENTSO-E-coordinated entities at the European level. TSOs activate reserves and manage cross-border flows to maintain system reliability. **Market operators**, or NEMOs (Nominated Electricity Market Operators), such as EPEX SPOT, Nord Pool or OMIE, facilitate DA and ID trading, run the auctions, and ensure transparency and reliability of market processes.

These participants form a highly interdependent ecosystem where each category influences market dynamics at different time scales. Understanding their roles is essential for analysing price formation and designing predictive models capable of integrating the complexity of the system.

I.2.3 Market Dynamics and Volatility Drivers

Electricity price dynamics are shaped by a combination of structural factors, operational uncertainties and weather-dependent fluctuations. Unlike financial assets, electricity cannot be stored economically at scale, which forces markets to instantaneously reflect supply–demand imbalances. Even small deviations between expected and realised conditions can generate pronounced price spikes, particularly in intraday horizons where adjustments must be made rapidly.

Weather conditions are among the most powerful drivers of volatility. Solar and wind generation forecasts evolve throughout the day, and errors in these predictions directly impact residual demand. A sudden drop in wind or an unexpected reduction in solar irradiance can trigger rapid price increases, especially in systems with high renewable penetration. Similarly, temperature affects both demand (heating, cooling) and generation efficiency, creating nonlinear relationships that machine learning models must capture.

Operational uncertainties also contribute to volatility: unplanned outages, network constraints, maintenance events and cross-border flows all influence the tightness of the system and therefore short-term price formation. These factors are often difficult to model explicitly but can be indirectly inferred from historical data and properly constructed features.

Intraday markets amplify these dynamics by incorporating updated information almost continuously. As a result, price distributions tend to be leptokurtic, with heavy tails and sharp asymmetries. Forecasting in this context requires flexible models capable of learning complex temporal dependencies and adapting to frequent regime shifts. These characteristics explain both the difficulty and the strategic value of high-quality short-term forecasts.

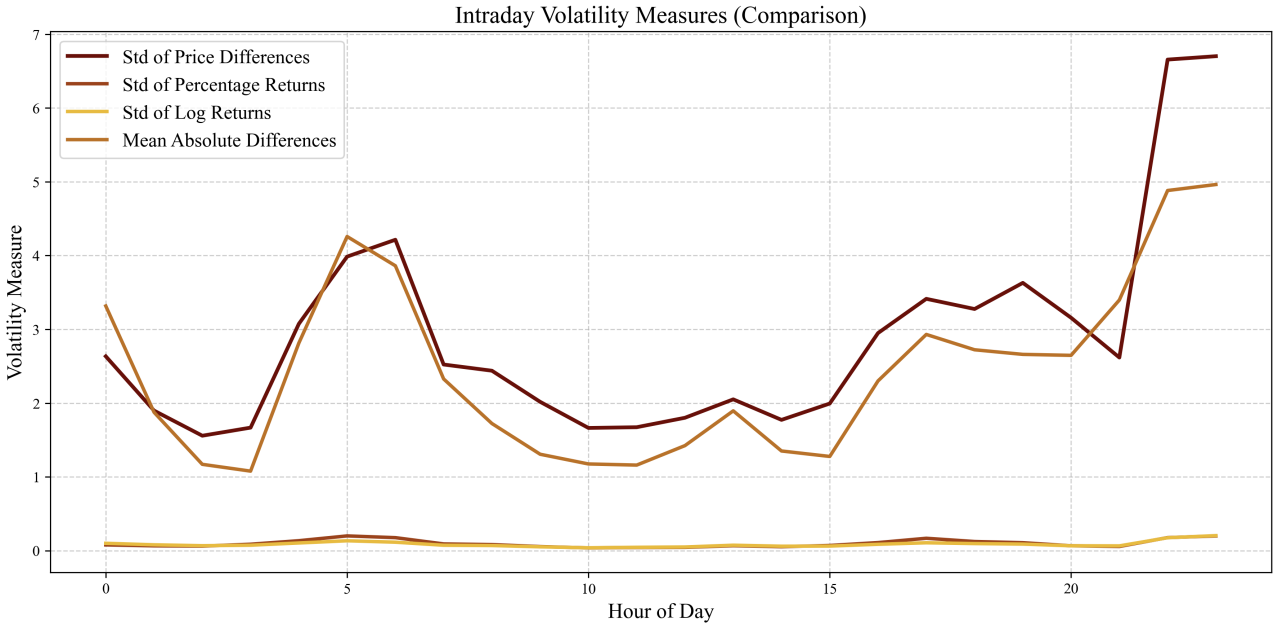

Figure I.2.2 – Volatility Dynamics of Intraday Electricity Markets

I.3 Organisation of the Report

The remainder of this report is structured to provide a coherent and progressive presentation of the methodology, modelling decisions, and analytical results that form the foundation of this study. Following the introductory elements of Section I, which outline the project's motivations, scope and the functioning of European electricity markets, Section II introduces the complete data and modelling framework developed for this work. It begins with a detailed description of the datasets used, their preparation and the feature engineering process that transforms raw information into model-ready inputs. The section then presents the predictive models implemented — from gradient boosting techniques to recurrent neural networks — and explains how their outputs are converted into quantitative trading signals. Finally, it describes the design principles behind the systematic trading strategies evaluated in the study.

Section III focuses on the evaluation of these strategies through a comprehensive backtesting framework. This includes the simulation environment, the assumptions underlying order execution, and the portfolio metrics used to assess performance. The section then presents and compares the results obtained for the different strategies, before providing an advanced diagnostic analysis aimed at understanding the robustness, limitations, and behaviour of the models under varying market conditions. Particular attention is given to the link between forecasting accuracy, mispricing structure, and realised trading performance.

The report concludes with Section IV, which synthesises the main findings and discusses the implications of this work for future research or practical deployment. Potential avenues for improvement — including alternative modelling techniques, incorporation of order book data, or reinforcement learning approaches — are also outlined. A series of annexes provides additional material such as code excerpts, supplementary figures, technical derivations, and extended tables to support the reproducibility and transparency of the study.

Data Engineering, Predictive Analytics & Algorithmic Trading

II.1 Data Architecture and Feature Engineering

The objective of this section is to describe the complete data architecture underlying the predictive modelling framework developed in this study. Because electricity price forecasting is fundamentally a data-driven task, the construction of a coherent, reliable, and information-rich dataset constitutes a decisive stage of the project. The data engineering pipeline encompasses the identification and analysis of the raw sources, the rigorous cleaning and harmonisation of the variables, and finally the construction of several modelling datasets specifically adapted to the requirements of XGBoost and LSTM architectures. This section focuses first on the origin and nature of the raw data before presenting the merged dataset from which all subsequent modelling inputs were derived.

II.1.1 Data Sources and Metadata Description

The study relies on two independent datasets obtained from Kaggle, each capturing a distinct dimension of the Spanish electricity system. The first dataset describes the physical and economic state of the electricity system: hourly electricity generation for a wide range of production technologies, system load forecasts and measurements, and historical market prices. The second dataset contains meteorological observations for five major Spanish cities—Madrid, Barcelona, Bilbao, Seville and Valencia—providing hourly temperature, humidity, pressure, wind and precipitation variables. Although structurally different, the two datasets share a common hourly resolution and cover the same period between 2015 and 2019, enabling a rigorous temporal integration.

Electricity System Dataset

The first dataset used in this study is a Kaggle dataset describing the physical and economic state of the Spanish electricity system at an hourly resolution between 2015 and 2019. It contains more than 35,000 observations and provides detailed information on electricity generation, system load and wholesale market prices. Generation values are reported in MWh and disaggregated across a wide range of technologies, including biomass, fossil fuel categories (gas, hard coal, lignite, oil), nuclear, hydro units (run-of-river, pumped storage, reservoirs), wind (onshore and offshore), solar and other renewable technologies. This disaggregation is essential for modelling because each technology responds differently to weather conditions, fuel prices and operational constraints.

Alongside generation data, the dataset includes total load forecasts and actual load measurements, two variables that strongly influence short-term market behaviour. Crucially, it also provides both the day-ahead market price and the intraday price, the latter $P_t = priceactual(t)$ being the target variable of the forecasting problem. Although the dataset is extensive, several variables displayed missing or unusable information, such as the completely empty column *forecast wind offshore eday ahead*, which was later removed during the cleaning stage. Overall, this dataset forms the structural core of the modelling framework, as it contains both the target series and the fundamental system variables required for feature construction

Meteorological Dataset

The second dataset complements the electricity system data with detailed meteorological conditions recorded hourly for the five Spanish cities Madrid, Barcelona, Bilbao, Seville and Valencia over the same 2015–2019 horizon. With approximately 178,000 observations, it covers a broad set of atmospheric indicators, including temperature (actual, minimum and maximum), pressure, humidity, wind speed and direction, rainfall, snowfall and cloud coverage. These variables are critical because weather conditions directly affect electricity demand and renewable generation, making them indispensable predictors of short-term price movements.

To preserve geographical specificity, each weather variable was expanded into separate columns for each city—for example, *temp_Madrid(t)*, *wind_speed_Valencia(t)* and *clouds_all_Seville(t)*. This spatial detail allows the model to capture regional meteorological variations, which are especially relevant for wind and solar production patterns. As with the electricity dataset, several missing or irregular entries required systematic cleaning, which will be detailed in Section II.1.2. The meteorological dataset thus provides the essential exogenous component of the modelling framework, complementing system-level information with climate-driven drivers of price formation.

Construction of the Merged and Cleaned Dataset

The two Kaggle datasets described above constitute rich but heterogeneous sources of information. They differ in structure, dimensionality, and semantic scope: one characterises the behaviour of the Spanish power system, while the other captures meteorological conditions across five cities. Before any modelling could be performed, it was therefore necessary to consolidate these sources into a single, coherent, and analytically exploitable dataset. This consolidation required a rigorous selection of relevant variables, careful handling of missing or low-quality data, and the construction of a harmonised temporal index. The outcome of this process is the dataset *merged_cleaned*, which serves as the reference for all feature engineering and model training stages.

The first step consisted in identifying which variables from the two raw Kaggle datasets should be preserved. Although both datasets contained numerous variables, not all were relevant for short-term electricity price forecasting. Several columns in the generation dataset contained either no values or such a high proportion of missing entries that reliable imputation was impossible. For example, the variable *forecast wind offshore eday ahead* was entirely empty and removed. Certain meteorological indicators were also redundant across cities or exhibited limited variability, making them unsuitable for inclusion. A deliberate variable selection was therefore performed, guided by domain knowledge and preliminary data analysis.

The retained variables followed a clear methodological rationale. From the electricity dataset, all generation variables with consistent hourly coverage were kept, as they provide essential information on the supply side of the market. Load forecasts and actual load measurements were also preserved due to their strong influence on price formation. Both market prices—day-ahead and intraday—were included, the latter (P_t) being the target of the forecasting task. On the meteorological side, temperature, humidity, pressure, wind speed and direction, rainfall and cloud coverage were retained for each of the five cities, as these variables directly affect electricity demand or renewable generation forecasts, particularly wind and solar output.

Once the relevant variables had been identified, the cleaning stage involved transforming both datasets into compatible formats. This required harmonising timestamp formats, resolving delimiter inconsistencies in column names, and sorting each dataset chronologically so that the temporal index satisfied

$$t_1 < t_2 < \ldots < t_N$$

Following this alignment, each dataset underwent a structured imputation procedure. This consisted of linear interpolation for isolated missing entries,

$$x_t = x_{t-1} + \frac{1}{2}(x_{t+1} - x_{t-1})$$

Followed by forward-fill and backward-fill propagation to ensure that no temporal gaps remained. The choice of this hybrid approach ensured that imputed values remained smooth and coherent with temporal trends while avoiding artificial discontinuities.

After cleaning, the datasets were merged using an inner join on the timestamp variable. This operation produced a single dataset of 35,062 rows and 67 variables, representing the intersection of the time horizons of the two sources. The merged dataset contained, for each timestamp t, the complete set of selected meteorological and electricity-related variables. It is important to emphasise that this dataset does not merely constitute a convenient aggregation; it is the foundational information structure of the entire project. All models—whether gradient boosting or recurrent neural networks—ultimately receive their inputs from this merged dataset, either directly (LSTM) or after the application of engineered features (XGBoost). In that sense, *dataset_merged_cleaned* is the central artefact around which the entire methodological pipeline is organised.

To make the structure of this dataset explicit, a summary table listing all retained variables, grouped by category (generation technologies, load indicators, market prices, meteorological variables), will be inserted just below:

Category	Variables Included	Description
Generation (MWh)	generation_biomass / generation_fossil_brown_coal_lignite / generation_fossil_gas / generation_fossil_hard_coal / generation_hydro_pumped_storage_consumption / generation_hydro_run_of_river_and_poundage / generation_hydro_water_reservoir / generation_nuclear / generation_other / generation_other_renewable / generation_solar / generation_waste / generation_wind_onshore / generation_fossil_oil	Hourly electricity generation per technology, representing the national production mix.
System Fundamentals	total_load_forecast / total_load_actual	Forecasted and real system-wide electricity demand (MWh).
Market Prices	price_day_ahead / price_actual	Day-ahead market price and intraday market price P_t, used as prediction target.
Weather – Valencia	temp_Valencia / temp_min_Valencia / temp_max_Valencia / pressure_Valencia / humidity_Valencia / wind_speed_Valencia / wind_deg_Valencia / rain_1h_Valencia / clouds_all_Valencia	Meteorological indicators for Valencia, impacting solar generation and local demand variability.
Weather – Madrid	temp_Madrid / temp_min_Madrid / temp_max_Madrid / pressure_Madrid / humidity_Madrid / wind_speed_Madrid / wind_deg_Madrid / rain_1h_Madrid / clouds_all_Madrid	Central-region meteorology influencing heating/cooling demand and renewable fluctuations.
Weather – Bilbao	temp_Bilbao / temp_min_Bilbao / temp_max_Bilbao / pressure_Bilbao / humidity_Bilbao / wind_speed_Bilbao / wind_deg_Bilbao / rain_1h_Bilbao / clouds_all_Bilbao	Northern coastal climate, strongly correlated with wind generation variability.
Weather – Barcelona	temp_Barcelona / temp_min_Barcelona / temp_max_Barcelona / pressure_Barcelona / humidity_Barcelona / wind_speed_Barcelona / wind_deg_Barcelona / rain_1h_Barcelona / clouds_all_Barcelona	Mediterranean weather patterns affecting electricity demand and photovoltaic output.
Weather – Seville	temp_Seville / temp_min_Seville / temp_max_Seville / pressure_Seville / humidity_Seville / wind_speed_Seville / wind_deg_Seville / rain_1h_Seville / clouds_all_Seville	Southern-region meteorology with strong temperature-driven demand profiles.

Table II.1.1 – Variables retained in the merged dataset grouped by category, with their descriptions.

Finally, to illustrate the temporal behaviour of the target variable P_t, the next page presents the evolution of the intraday price from 2015 to 2019. This visualisation is particularly informative: it highlights the presence of pronounced seasonal cycles, sharp price spikes, long periods of stability punctuated by abrupt fluctuations, and episodes of high volatility. Such dynamics justify the subsequent use of temporal feature engineering—such as lagged variables, rolling statistics and volatility indicators—and explain why raw inputs alone are insufficient for tabular models such as XGBoost.

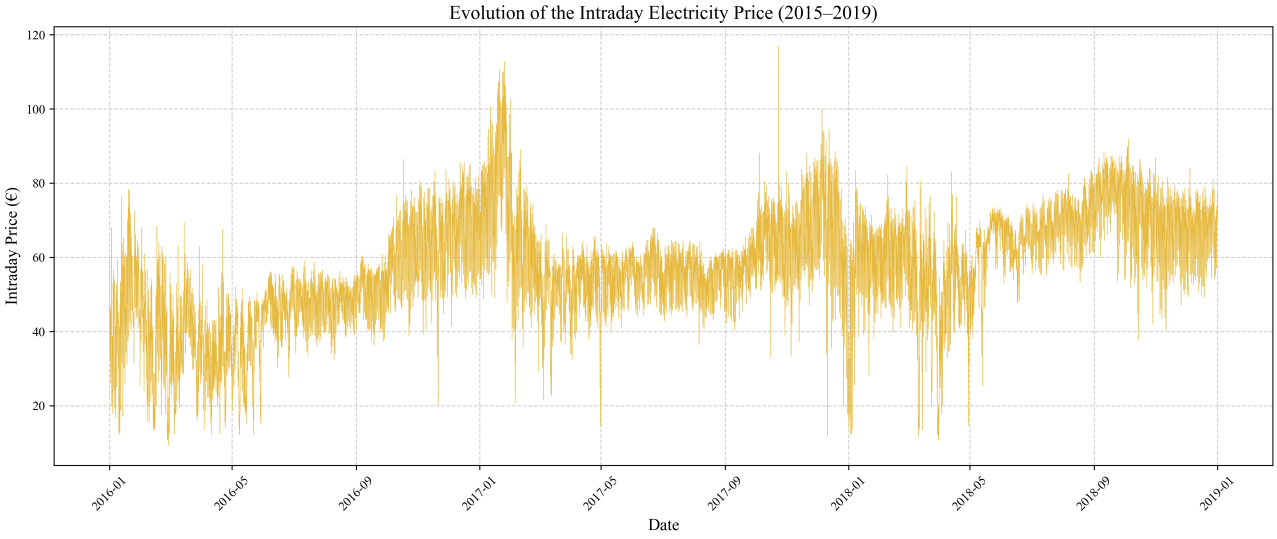

Figure II.1.1 – Historical evolution of the intraday electricity price between 2015 and 2019

II.1.2 Data Cleaning and Preprocessing Pipeline

Once the unified dataset had been obtained, a cleaning and preprocessing stage was applied to ensure that the data were structurally consistent, numerically stable and fully usable for feature engineering. At this stage, the objective was no longer to select variables or consolidate sources, but rather to transform the merged dataset into a complete, gap-free, and temporally coherent matrix suitable for both tabular and sequential learning models.

Handling of missing values

Although the dataset was already temporally aligned, several variables—particularly meteorological indicators—still contained isolated missing entries. Since short-term electricity price forecasting relies on continuous temporal dynamics, gaps in the data would compromise the construction of lagged or rolling features and introduce inconsistencies for sequential models such as LSTM. To address this issue, a three-stage imputation strategy was applied. First, linear interpolation was used whenever adjacent values were available,

$$x_t = x_{t-1} + \frac{x_{t+1} - x_{t-1}}{2}$$

ensuring that intermediate values followed the local trajectory of the signal.

Residual missing entries were then treated using backward-fill and forward-fill propagation:

$$x_t = x_{t+1} \quad \text{(bfill)} \qquad\qquad x_t = x_{t-1} \quad \text{(ffill)}$$

This hybrid approach produced smooth, leakage-free imputations while guaranteeing that all features remained fully defined over the entire 2015–2019 horizon.

Removal of duplicated timestamps

A structural audit of the merged dataset revealed occasional duplicate rows inherited from the raw data sources. These were systematically removed to enforce the strict temporal ordering condition

$$t_1 < t_2 < \ldots < t_N$$

which is essential for any model relying on time-dependent information, such as rolling statistics or recurrent neural networks.

Standardisation of column structure

Several column names inherited from the original Kaggle files contained inconsistent delimiters or formatting artefacts. These were corrected and normalised to ensure coherent naming across all variables. Additionally, variables that had been previously identified as irrelevant—such as the entirely empty column *forecast wind offshore eday ahead*—were removed during this step, along with a small number of other fields whose lack of usable information became evident only after merging.

Final cleaned dataset

After imputation, deduplication and structural harmonisation, the cleaned dataset consisted of approximately 35,000 hourly observations and 67 fully populated variables. Crucially, it contains no missing entries, no duplicated timestamps and no structurally inconsistent fields. This makes it suitable for the construction of lagged and rolling features used by XGBoost, as well as for the direct extraction of temporal sequences required by LSTM models. The cleaned dataset therefore constitutes the definitive input matrix from which all model-specific datasets were derived in the next step.

II.1.3 Feature Engineering and Construction of Model-Specific Datasets

After obtaining the cleaned and temporally consistent dataset described in Section II.1.2, the next step consisted in constructing the model-specific datasets used to train the forecasting architectures. Because the models explored in this study rely on fundamentally different learning paradigms, the feature engineering strategy had to be adapted accordingly. In particular, the XGBoost regressors require an explicitly engineered feature space encoding temporal dependencies, while the LSTM architecture operates on raw temporal sequences and therefore must not include handcrafted lagged or rolling features. This section presents the construction rationale and the mathematical definition of each set of engineered variables.

II.1.3.1 Lag-Based Dataset for XGBoost

To construct the first XGBoost dataset, the objective was to explicitly encode the temporal dependence structure of the intraday electricity price. Unlike recurrent neural networks, gradient boosting algorithms do not inherently capture time dynamics; instead, they treat each observation as an independent row in a tabular dataset. Temporal information must therefore be provided in the form of lagged variables, which represent the historical behaviour of the target price series.

Let P_t denote the intraday price at time t. For any integer lag $k > 0$, the corresponding lagged feature is defined as:

$$lag_k(t) = P_{t-k}$$

In practice, the following set of lags was selected:

$$k \in \{1, 2, 3, 6, 12, 24, 48, 168\}$$

These lags were chosen to capture multiple temporal scales relevant to electricity markets:

— 1–3 hours : ultra-short-term fluctuations
— 6–12 hours : intraday structure
— 24 hours : daily seasonality
— 48 hours : two-day persistence
— 168 hours : weekly cycles

To ensure that all lag values were properly defined, the first 168 rows of the dataset were removed. This operation avoids any undefined values in the engineered features and ensures that the design matrix contains only complete observations. After lag creation, the dataset was split into:

— a feature matrix X containing all variables except the target,
— a target vector $Y = P_t$

The resulting dataset (*X_final.csv*, *Y_final.csv*) constitutes the feature space for the first XGBoost model.

II.1.3.2 Lag + Rolling, Volatility and Trend Features for XGBoost

While lag features encode the direct temporal dependence of the intraday price, they do not fully capture more complex structures such as local smoothing patterns, volatility regimes, or multi-period drifts. To enhance the expressiveness of the XGBoost model, a second dataset was constructed by enriching the lagged feature space with rolling statistics, return-based indicators, volatility estimators and trend measures. Each of these engineered variables corresponds to a distinct temporal property of the price series and provides complementary information to the boosting algorithm. The subsections below present the rationale and the mathematical definition of each feature family.

Rolling mean and rolling standard deviation

Rolling statistics provide a local summary of the recent behaviour of the price series. They are particularly relevant for electricity markets, which exhibit rapid variations during peak periods and smoother patterns during off-peak hours. The rolling mean captures the central tendency over a window of length W, while the rolling standard deviation quantifies local variability.

$$\text{roll\_mean}_W(t) = \frac{1}{W} \sum_{i=1}^{W} P_{t-i}$$

$$\text{roll\_std}_W(t) = \sqrt{\frac{1}{W} \sum_{i=1}^{W} (P_{t-i} - \text{roll\_mean}_W(t))^2}$$

These features allow XGBoost to detect whether the price is entering a stable or unstable regime. Two window sizes were chosen:

— W=24 hours, which captures daily seasonality,
— W=168 hours, which captures weekly operational cycles.

Such windows are commonly used in energy modelling, where demand and renewable output follow strongly periodic patterns.

Return features

Return features quantify short-term changes, enabling the model to detect acceleration, reversal or jump behaviours that cannot be inferred solely from levels. They represent differences between lagged prices:

$$\text{return\_1h}(t) = P_{t-1} - P_{t-2}$$

$$\text{return\_24h}(t) = P_{t-24} - P_{t-48}$$

These returns encode rapid intra-hour fluctuations and day-to-day deviations. Because only past information is used, they are fully leakage-free. Return-based features are widely used in financial econometrics and energy forecasting because they often correlate more strongly with volatility than price levels.

Volatility indicators

Volatility, broadly defined as the dispersion of returns, is a key predictor of uncertainty in electricity markets. Periods of high volatility often correspond to renewable forecast errors, sudden changes in demand, or market stress. To estimate local volatility, we compute the rolling standard deviation of the 1-hour returns:

$$\text{vol}_W(t) = \text{std}(r_{t-1}, r_{t-2}, \ldots, r_{t-W}) \qquad \text{où } r_t = P_t - P_{t-1}$$

Two windows were again used (24h and 168h), enabling the model to distinguish between short-term market turbulence and slower structural variability. These indicators are analogous to realised volatility measures in financial markets and provide crucial information for forecasting abrupt price movements.

Trend indicators

Trend features measure directional movements of the price series by comparing the most recent price to its past values. They provide information on whether the market is trending upward, trending downward, or returning to previous levels.

$$\text{trend}_{24}(t) = P_{t-1} - P_{t-24} \qquad \text{trend}_{168}(t) = P_{t-1} - P_{t-168}$$

A positive trend suggests a sustained price increase (e.g., due to rising demand or decreasing renewable output), while a negative trend indicates the opposite. These variables also allow the model to distinguish periods with strong momentum from those exhibiting mean-reversion behaviour.

Leakage prevention

All engineered features were constructed using only past values of the target variable. Formally, each feature satisfies:

$$\text{feature}(t) = f(P_{t-1}, P_{t-2} \ldots)$$

ensuring strict causality and eliminating any risk of data leakage. Moreover, rolling and volatility windows require a certain number of past observations; thus, all rows for which any window was incomplete were removed. This ensures that the final dataset is fully populated and that each feature has a rigorous temporal interpretation.

Final dataset (Rolling version)

After enriching the dataset with the lag, rolling, return, volatility and trend features, all rows containing undefined values were dropped. The resulting dataset, stored as *X_clean_rolling_v2.csv*, contains several dozen engineered predictors explicitly encoding multiple temporal scales and structural regimes of the price series. The associated target vector is stored in *Y_clean_v2.csv*. This feature-rich representation is particularly well suited to XGBoost, whose gradient-boosted structure benefits from diverse and highly informative predictors.

II.1.3.3 LSTM Dataset: Raw Temporal Sequences Without Engineered Features

The construction of the LSTM dataset differs fundamentally from the approach used for XGBoost. Whereas gradient boosting models operate on a tabular representation and therefore rely on explicitly engineered temporal features, recurrent neural networks (RNNs)—and in particular Long Short-Term Memory (LSTM) networks—are designed to learn temporal dependencies directly from the sequential structure of the data. As a consequence, the LSTM architecture does not require lagged variables, rolling statistics, volatility indicators or trend features. Instead, it processes the raw exogenous and endogenous variables as ordered time sequences and internally infers multi-scale temporal patterns through its gating mechanisms.

Theoretical motivation: why LSTM does not require engineered temporal features

LSTM networks extend standard RNNs by introducing a memory cell c_t and three gating functions that regulate the flow of information through time: the input gate i_t, the forget gate f_t, and the output gate o_t. Given an input vector $x_t \in \mathbb{R}^F$ at time t, the LSTM cell updates its internal state according to:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f),$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \qquad \tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c),$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \qquad h_t = o_t \odot \tanh(c_t),$$

where h_t is the hidden state, $\sigma(\cdot)$ is the logistic activation, and $\odot$ denotes element-wise multiplication.
These equations show that *temporal information is recursively stored, transformed and propagated across time steps by the architecture itself.* Because the LSTM directly maintains a latent representation of past observations through (h_t, c_t), constructing explicit lag features such as P_{t-1}, P_{t-24} or rolling statistics would be redundant or even detrimental. Instead, the LSTM learns to extract short- and long-term temporal patterns automatically through its parametric gates.

Construction of the LSTM dataset

The LSTM dataset was constructed directly from the cleaned merged dataset, without adding any engineered temporal features. After removing the target variable, the remaining matrix $X \in \mathbb{R}^{N \times F}$ contains all meteorological, generation, load and price context variables, including the timestamp. The target vector is defined as:

$$Y_t = P_t = \text{price actual}(t).$$

In order to train the LSTM, the dataset was transformed into overlapping sequences of fixed length T. Each sample corresponds to a temporal window of T consecutive hours, represented as:

$$X^{(i)} = (x_{t-T+1}, x_{t-T+2}, \ldots, x_t),$$

with the corresponding forecasting target:

$$Y^{(i)} = P_{t+1}.$$

Values of T between 24 and 48 hours were used, enabling the network to learn daily and intra-day dependencies. This representation preserves strict causality and ensures that the model only receives information available before the prediction timestamp.

Why engineered features were explicitly excluded

The exclusion of lags, rolling averages or volatility features is justified by both theoretical and practical considerations:

1. **Redundancy:** Engineered features such as $\text{lag}_1(t) = P_{t-1}$ are already contained in the input sequence $(x_{t-T+1}, \ldots, x_t)$; adding them would duplicate information.
2. **Loss of temporal richness:** Rolling windows artificially smooth the data and may suppress high-frequency patterns that LSTMs are precisely designed to exploit.

3. **Risk of leakage through window behaviour:** Badly constructed rolling features may incorporate future information (e.g., centered windows). Using raw sequences avoids this risk entirely.
4. **Architectural suitability:** The recurrent update equations of LSTMs already implement a learnable form of lagged, smoothed and long-term memory. Thus, explicit hand-crafted features would constrain the model rather than enrich it.

Final sequential dataset

The final LSTM dataset was exported under the files *X_clean_LSTM.csv* and *Y_clean_LSTM.csv*. These files contain, respectively, the full matrix of explanatory variables and the target price series. During model training, these matrices are reshaped into three-dimensional tensors of dimension:

$$(N_{\text{samples}},\ T,\ F),$$

where N_{samples} is the number of usable windows, T the sequence length, and F the number of features.
This dataset is therefore fundamentally different in nature from the XGBoost datasets: instead of a tabular format with explicit temporal descriptors, it offers a raw sequential representation that allows the LSTM to infer temporal dependencies directly from the data.

II.1.4 Dataset Structuring and Train/Validation/Test Partitioning

Because electricity price forecasting is a strictly time-dependent problem, the dataset must be split chronologically to avoid information leakage. All models trained in this project—both XGBoost (lag-based and rolling-based versions) and the LSTM—use the **same temporal partition**, ensuring full consistency when comparing forecasting performance.

Temporal splitting strategy

Let the dataset be ordered as $\{(x_t, P_t)\}_{t=1}^{N}$. The split must satisfy:

$$t_{\text{train}} < t_{\text{val}} < t_{\text{test}},$$

ensuring that the model never observes future information during training or validation.
The dataset was divided into three contiguous blocks as follows:

— **Training set** (70%)**:** 2015 → end-2017
— **Validation set** (15%)**:** early-2018 → mid-2018
— **Test set** (15%)**:** mid-2018 → end-2018

configuration provides a sufficiently large training horizon while preserving meaningful validation and test periods reflecting recent market conditions.

Application to XGBoost datasets

For both XGBoost datasets (lags-only and lags+rolling), the split is applied **after feature engineering**, once rows affected by lag warm-up and rolling window initialisation have been removed. Each row in the final matrix corresponds to a single timestamp t, and the partition is simply:

$$\mathcal{D}_{\text{train}} = \{\, t < T_1 \,\}, \qquad \mathcal{D}_{\text{val}} = \{\, T_1 \le t < T_2 \,\}, \qquad \mathcal{D}_{\text{test}} = \{\, t \ge T_2 \,\}.$$

Application to the LSTM dataset

For the LSTM model, the temporal split is applied **before constructing the input sequences** of length T. A sequence is included in a subset only if:

$$(t - T + 1, \ldots, t) \subset \mathcal{D}_{\text{train/val/test}}.$$

This prevents any sequence window from crossing dataset boundaries and ensures strict causality in all samples.

II.2 Price Forecasting Models and Validation Protocols

Electricity price forecasting is a high-dimensional, nonlinear, and temporally structured learning problem. To address these challenges, this study combines gradient boosting methods and deep learning architectures to model the hourly intraday price P_t, each family offering complementary advantages.

***Boosted tree ensembles (XGBoost)** perform strongly on structured tabular datasets, handle heterogeneous predictors, and incorporate explicit regularisation. When provided with engineered lagged, rolling, and volatility features, they can effectively capture nonlinear interactions and medium-horizon temporal patterns.*

***Recurrent neural networks (LSTM, GRU)** and attention mechanisms learn temporal dependencies directly from raw sequences. Their gating structures and dynamic memory allow them to extract multi-scale temporal patterns without relying on handcrafted features.*

This section presents the forecasting models implemented in this study, outlines their theoretical foundations, and explains how they were adapted to the specific structure of the merged dataset. It then details the hyperparameter optimisation procedure, the evaluation metrics used to assess predictive performance, and concludes with a comparative analysis of the models.

II.2.1 Gradient Boosting Framework (XGBoost)

XGBoost is a high-performance implementation of gradient boosted decision trees designed to model complex nonlinear relationships in structured datasets. The algorithm belongs to the family of additive ensemble methods, where the prediction function is constructed as a sum of weak learners. In the context of electricity price forecasting, where interactions between meteorological, load-related and temporal signals exhibit strong nonlinearities, XGBoost provides an efficient balance between predictive flexibility, computational efficiency and robustness.

From a theoretical standpoint, gradient boosting builds a model iteratively by adding successive regression trees that minimise the residual errors of the previous ensemble. Let $\hat{y}_t^{(k)}$ denote the prediction at iteration k. The boosting update is defined as

$$\hat{y}_t^{(k)} = \hat{y}_t^{(k-1)} + f_k(x_t)$$

where f_k is a decision tree optimised to approximate the negative gradient of the loss function at iteration $k-1$. XGBoost distinguishes itself from classical implementations by relying on a second–order Taylor expansion of the objective function, using not only first–order gradients g_t but also second–order derivatives h_t. This yields a local approximation of the objective under the form

$$\mathcal{L}^{(k)} \approx \sum_{t=1}^{N} \left[g_t f_k(x_t) + \frac{1}{2} h_t f_k(x_t)^2 \right] + \Omega(f_k)$$

where Ω is a structural penalty acting on tree complexity. This formulation enables more stable optimisation steps, especially in noisy environments such as short–term electricity price series characterised by sharp variations and occasional outliers.

Another important feature of XGBoost lies in its explicit regularisation framework. The penalty term

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

where T is the number of leaves and w the vector of leaf weights, discourages excessively deep or irregular trees and promotes generalisation. Combined with subsampling of observations and features, this regularisation mitigates overfitting tendencies that can arise in high-dimensional settings. Furthermore, XGBoost incorporates a sparsity-aware split mechanism, which automatically learns the optimal direction for missing values in each node. This property is particularly relevant when modelling electricity markets, where meteorological variables occasionally contain gaps and strongly heterogeneous distributions.

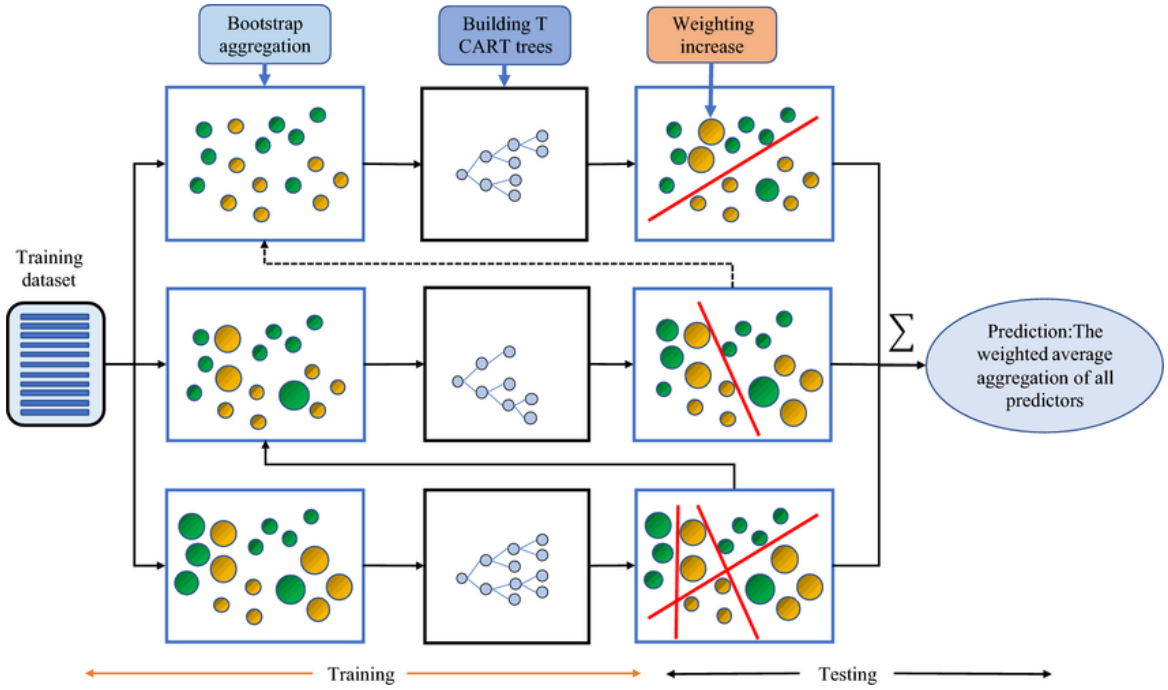

Figure II.2.1 – Schematic Representation of the XGBoost Boosting Process

Application to this study

In this project, XGBoost was used to model the intraday electricity price using engineered tabular representations derived from the cleaned dataset. Two variants of the model were trained: one relying exclusively on lagged values of the target, and another enriched with rolling statistics, volatility indicators and trend-based features. These feature sets were chosen to expose the algorithm to multiple temporal scales—short-term fluctuations, daily seasonality and multi-day structural dynamics—while preserving the tabular structure required by boosted trees.

The model was trained using the squared-error objective and a small learning rate, allowing the boosting process to incorporate information gradually. Regularisation parameters (α, λ, γ) and structural hyperparameters (tree depth, subsampling rates) were selected through Bayesian optimisation described later in Section II.2.3. Early stopping on a chronologically ordered validation set ensured that the final model retained the best generalisation properties without overfitting to recent fluctuations. The resulting predictors consist of additive ensembles of decision trees encoding the nonlinear and multi-scale relationships learned from the engineered features.

II.2.2 Deep Learning Architectures (LSTM, GRU, Attention)

Deep learning models provide a fundamentally different approach to time-series forecasting compared with tree-based methods. Instead of relying on handcrafted temporal descriptors such as lagged values or rolling statistics, recurrent neural networks (RNNs) learn temporal dependencies directly from raw sequential data. Their architecture is designed to retain information over extended horizons, making them particularly suitable for capturing the multi-scale temporal dynamics that characterise electricity prices.

Among recurrent architectures, Long Short-Term Memory (LSTM) networks have become a standard reference. LSTM units overcome the vanishing and exploding gradient problems that affect traditional RNNs by introducing a memory cell c_t regulated by gating mechanisms. At each time step, the input gate i_t, forget gate f_t and output gate o_t control how new information is integrated, how past information is preserved, and how the internal memory contributes to the output. Formally, the LSTM cell computes

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \qquad i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \qquad c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \qquad h_t = o_t \odot \tanh(c_t)$$

This structure enables the network to selectively retain or discard temporal information over long horizons, a desirable property for price series affected by daily seasonality, multi-day patterns and abrupt shocks.
A complementary recurrent architecture, the Gated Recurrent Unit (GRU), provides a more compact alternative. GRUs merge the forget and input mechanisms of the LSTM into a single update gate, resulting in fewer parameters and reduced computational cost, while still enabling effective modelling of medium-term dependencies. The combination of LSTM layers followed by GRU layers leverages the strengths of both architectures: the LSTM captures long-range patterns, while the GRU refines shorter and mid-range temporal structures.

To further enhance the extraction of informative patterns, an attention mechanism was incorporated in one of the models. Attention computes a set of trainable weights that quantify the relative importance of each time step in the input sequence, allowing the network to focus selectively on the most relevant periods. Given a sequence of hidden states $\{h_1, \ldots, h_t\}$, the attention module produces a context vector

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \qquad \text{context} = \sum_{t=1}^{T} \alpha_t h_t$$

where e_t is a learned relevance score. This mechanism is particularly suited to electricity prices, whose structure often depends on a limited subset of influential past hours (e.g., previous peaks, ramp-up periods, transitions between renewable and thermal dominance).

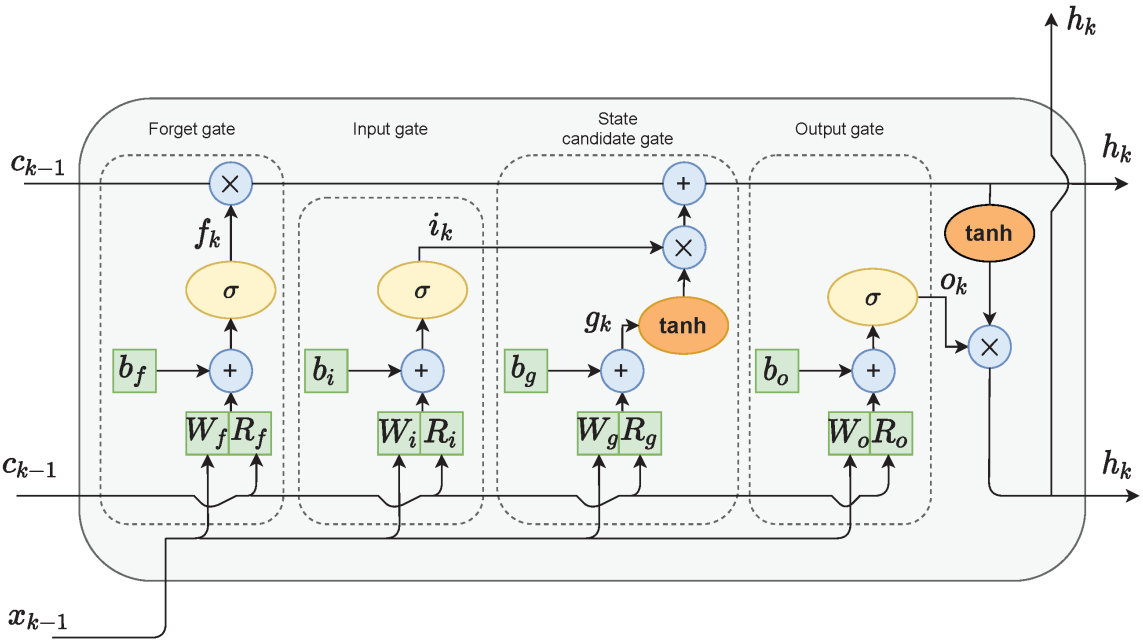

Figure II.2.2 – LSTM and GRU Neural Networks as Models of Dynamical Processes

Application to this study

In this project, the deep learning models operate directly on the cleaned multivariate dataset described earlier, without any manual temporal feature engineering. The input consists of sequences of 72 consecutive hours, each composed of all meteorological, load and generation variables. These sequences are standardised and fed to the network in the form of three-dimensional tensors.

Two architectures were explored. The first is a pure LSTM model composed of stacked recurrent layers with dropout regularisation, followed by a dense output unit. The second combines an LSTM layer, a GRU layer and an attention module, enabling the network to capture hierarchical dependencies across multiple time scales. Both models are trained with the mean-squared error loss and the Adam optimiser, using early stopping on a held-out validation set to preserve generalisation. The resulting networks learn temporal patterns directly from the raw sequences, without relying on explicit lag construction or rolling windows, offering a complementary modelling approach to the tree-based methods used earlier.

II.2.3 Hyperparameter Optimization and Regularization

Hyperparameter selection plays a central role in the performance and generalisation ability of machine learning models, particularly for algorithms such as XGBoost whose behaviour is highly sensitive to structural and regularisation parameters. In this study, hyperparameter optimisation was carried out primarily for the gradient boosting models, while the LSTM architectures were tuned through targeted empirical evaluation.

XGBoost

For XGBoost, model calibration was performed using Optuna, a modern optimisation framework based on Bayesian sampling strategies. Optuna iteratively explores the hyperparameter space by constructing a probabilistic model of performance and updating this model after each trial to favour regions associated with low validation error. This approach is particularly effective for algorithms with complex interactions between hyperparameters, such as learning rate, tree depth, regularisation coefficients, and subsampling ratios. The optimisation objective was defined as the validation RMSE computed on a strictly chronological train–validation split to prevent leakage.

Two separate optimisation campaigns were conducted. For the lag-only XGBoost model, 150 Optuna trials were executed over a wide search space encompassing learning rate, tree depth, minimum child weight, subsampling parameters and both L1/L2 regularisation terms. For the enriched model incorporating lag, rolling, volatility and trend features, a larger 300-trial optimisation was performed to account for the increased complexity and dimensionality of the feature space. In both cases, early stopping was used during each trial to avoid overfitting and to accelerate convergence. The resulting best parameter sets (to be inserted here as a table) constitute the configurations used in Section II.2.1 and serve as the final models evaluated throughout the study.

Hyperparameter	XGBoost (Lag Features Only) Optuna 150 Trials	XGBoost (Lag + Rolling Features) Optuna 300 Trials
Objective	Squared Error	Squared Error
Eval Metric	RMSE	RMSE
eta (learning rate)	0.011079248646397246	0.0052856998613360455
max_depth	10	9
min_child_weight	18	11
subsample	0.5007868939044449	0.5370858622948435
colsample_bytree	0.9985602201505985	0.9535031113477612
gamma	0.6291347114905599	2.2331918753326874
lambda	2.416950284412258	0.23912890666433612
alpha	0.5183556110228869	1.518891405996986

Table II.2.1 – Best hyperparameters obtained from Optuna optimisation for the two XGBoost models.

LSTM

In contrast, the LSTM architectures did not undergo full hyperparameter optimisation. Preliminary experiments were conducted to determine an appropriate sequence length by comparing windows of 24 h, 48 h, 72 h and 96 h. The 72-hour configuration consistently provided the most stable behaviour and the lowest validation error, and was therefore adopted for the remainder of the analysis. Other architectural choices—number of units, dropout rates, optimiser and learning rate—were selected according to common deep-learning practice for sequential models, with early stopping used to mitigate overfitting. Given the comparatively modest performance of the neural models in subsequent sections, a more exhaustive search was not pursued.

Overall, this optimisation procedure ensures that each XGBoost variant operates under an appropriately regularised and data-driven configuration, while the LSTM models rely on empirically validated design choices consistent with their application to multivariate electricity time series.

II.2.4 Performance Metrics and Forecast Evaluation

Evaluating the accuracy of a predictive model requires quantitative metrics capable of measuring how close the model's outputs are to the true values observed in the test set. In the context of electricity price forecasting, this comparison is essential: a model is considered effective if its predictions replicate the behaviour, amplitude and direction of the real intraday price as faithfully as possible. Because the target variable exhibits strong volatility, non-linearities and occasional abrupt spikes, multiple complementary metrics are required to obtain a complete and robust assessment of forecasting performance.

The evaluation framework used in this study relies on a set of classical error metrics—RMSE, MAE, MAPE and R^2—augmented with two additional indicators particularly relevant for time-series applications: MASE, which normalises forecast accuracy relative to a naïve benchmark, and Directional Accuracy (DA), which assesses the model's ability to predict the sign of price variations. All metrics are computed on the held-out test set to ensure a fair and leakage-free comparison between models.

Error-based metrics

The Root Mean Squared Error (RMSE) is one of the most common measures of accuracy in regression tasks. Defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}$$

it penalises large deviations more strongly than small ones, due to the quadratic term. This property makes RMSE particularly informative for electricity markets, where occasional price spikes can induce significant forecasting errors. The Mean Absolute Error (MAE), defined as

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N} | \, y_t - \hat{y}_t \, |$$

provides a complementary view by measuring average deviation without disproportionately emphasising extreme events. RMSE and MAE together offer a balanced understanding of amplitude errors.

To evaluate the model in relative rather than absolute terms, the Mean Absolute Percentage Error (MAPE) expresses the forecast deviation as a percentage of the true value:

$$\text{MAPE} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right|.$$

Although sensitive to values close to zero, MAPE is useful for assessing error proportionality,which is relevant when interpreting model performance across time periods with varying price levels.

Beyond absolute errors, the coefficient of determination R^2 measures the proportion of variance in the target that is explained by the model:

$$R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}.$$

A high R^2 indicates strong explanatory power and demonstrates that the model captures the underlying dynamics of the series rather than merely reproducing its average behaviour.

Time-series specific metrics

While these classical regression metrics quantify pointwise errors, they do not assess the model relative to simple heuristics or reflect its ability to anticipate directional movements, which are both crucial aspects in time-series forecasting. To address this limitation, the Mean Absolute Scaled Error (MASE) compares the model's MAE to the MAE of a naïve one-step-ahead forecast:

$$\text{MASE} = \frac{\frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t|}{\frac{1}{N-1}\sum_{t=2}^{N}|y_t - y_{t-1}|}.$$

A value below 1 indicates that the model improves upon the simple "last value" predictor. This benchmark is particularly relevant in electricity markets, where short-term persistence often represents a surprisingly strong baseline. Directional Accuracy (DA) evaluates the model's capacity to predict the sign of hourly price changes:

$$\mathrm{DA} = \frac{1}{N-1}\sum_{t=2}^{N} \mathbf{1}\big[\,\mathrm{sign}(y_t - y_{t-1}) = \mathrm{sign}(\hat{y}_t - \hat{y}_{t-1})\,\big] \times 100.$$

This metric is crucial when models are used for trading decisions, since capturing the correct direction of price movement can be more important than minimising pointwise errors. DA complements RMSE and MAE by providing insight into the qualitative behaviour of the forecasts.

Synthesis

Together, these metrics offer a comprehensive framework for evaluating model performance. RMSE, MAE and MAPE quantify the magnitude of prediction errors, R^2 captures explanatory power, MASE benchmarks the model against a naïve reference, and DA assesses its ability to anticipate directional trends. By applying this complete set of indicators to the predictions of the XGBoost and LSTM models on the test set, we obtain a rigorous and multi-dimensional evaluation of forecasting accuracy. The comparative analysis of these results is presented in the following section.

II.2.5 Comparative Assessment of Predictive Models

The aim of this section is to compare the predictive performance of the four forecasting models developed in this study—two XGBoost regressors and two recurrent neural network architectures—and to identify the model best suited for integration into the trading framework introduced in Section II.3. As emphasised earlier, the most suitable predictor is not necessarily the one achieving the lowest pointwise error but rather the one that most reliably anticipates the shape, direction, and temporal structure of price movements. Forecasting electricity prices for trading requires correctly capturing the short-term dynamics and daily cyclicality of the market, even if the predicted values do not perfectly match the exact price level.

To formalise this evaluation, all models were assessed using the performance metrics introduced in Section II.2.4, computed on the same test horizon. The following table summarises the performance of all four models across the six evaluation metrics:

Metric	XGBoost (Lag)	XGBoost (Lag + Rolling)	LSTM	LSTM + GRU + Attention
RMSE	2.4180	2.1538	10.0622	12.2190
MAE	1.8726	1.6158	7.9433	10.4553
MAPE	2.80%	2.43%	11.13%	14.78%
R^2	0.9069	0.9260	-0.6180	-1.3860
DA (%)	78.54%	76.65%	70.77%	74.89%
MASE	0.9489	0.8190	4.0245	5.2972

Table II.2.2 – Comparative performance metrics for the four predictive models on the test set.

This consolidated view highlights the contrast between the boosted-tree models and the neural architectures in terms of absolute accuracy, relative performance, and ability to capture directional dynamics.

Analysis of the Four Models

The comparative results reveal a clear separation between the performance of the boosted-tree models and the recurrent neural network architectures. Both XGBoost variants exhibit strong numerical accuracy, with low RMSE and MAE values and positive, high R^2 scores. The enriched model incorporating lag, rolling and volatility features performs best overall across the majority of metrics. Its MASE score below unity further confirms that it consistently outperforms the naïve persistence benchmark. Interestingly, the lag-only XGBoost

configuration achieves slightly better MAPE and Directional Accuracy, reflecting its sharpness in capturing the relative magnitude and direction of short-term fluctuations. However, these advantages remain marginal when compared with the substantial gains in variance explanation, absolute precision and robustness exhibited by the rolling-enhanced model. Given the objective of producing a stable and generalisable forecasting engine, these broader performance dimensions are more decisive, justifying the selection of the enriched configuration.

In contrast, the two LSTM-based architectures display significantly weaker performance, with large pointwise errors and negative R^2 values. These results may seem surprising in light of the extensive literature where LSTM models often outperform tree-based methods on large-scale, smooth or highly structured time-series datasets. Their underperformance in the present context is explained primarily by the characteristics of the available data. The dataset used in this study, although multivariate, remains relatively limited in size for training deep sequential models, which typically require tens or hundreds of thousands of sequences to learn stable temporal representations. Moreover, electricity prices exhibit abrupt fluctuations, regime changes and strong noise components, which degrade the ability of recurrent networks to generalise when the amount of training data is insufficient. XGBoost, on the other hand, is particularly well-suited to small- and medium-scale tabular datasets and tends to perform robustly even when the signal-to-noise ratio is low. As a result, the LSTM models fail to learn reliable long-range dependencies, which explains their high RMSE and MAE values, despite achieving reasonable Directional Accuracy.

Taken together, these observations highlight the superiority of the XGBoost framework for this forecasting task, with the rolling-enhanced variant providing the best overall compromise between accurate magnitude prediction, strong variance explanation and robust directional behaviour.

To illustrate these behaviours, the two following figures display the predictions produced by the XGBoost model with lag and rolling features over the entire test horizon and over the final week. These plots highlight the model's ability to reconstruct the daily cyclical pattern and to anticipate short-term variations. The prediction curves of the remaining three models are provided in Appendix A for completeness.

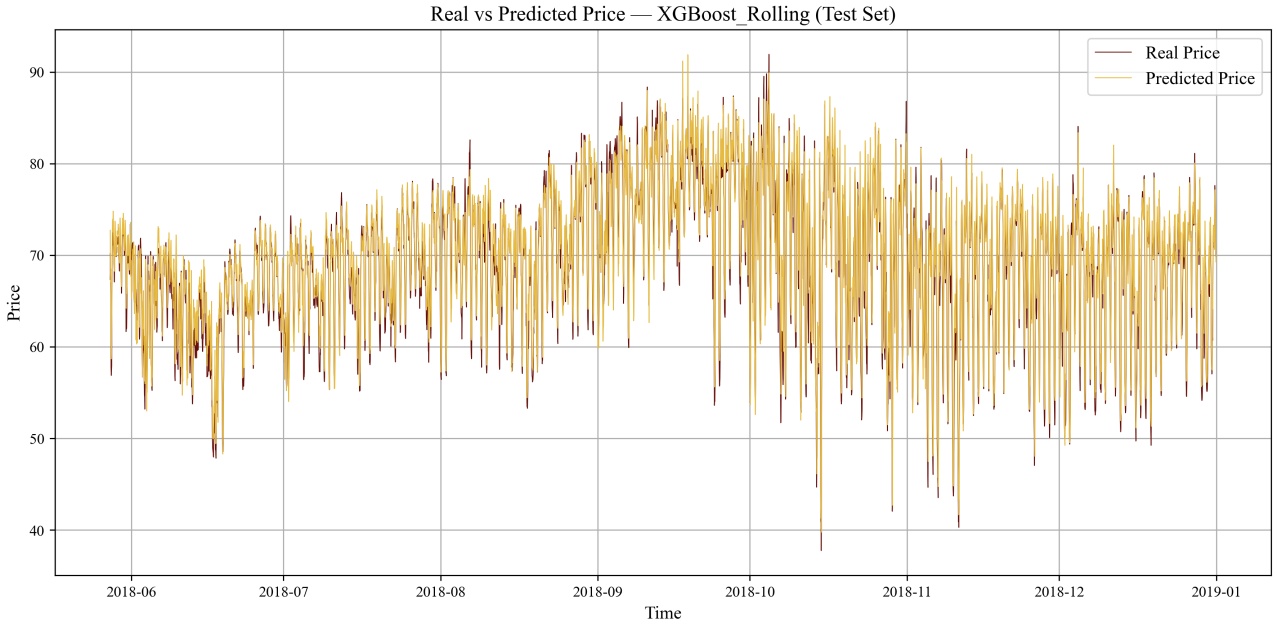

Figure II.2.3 – Real vs Predicted Electricity Price using XGBoost (Lag + Rolling), Test Set.

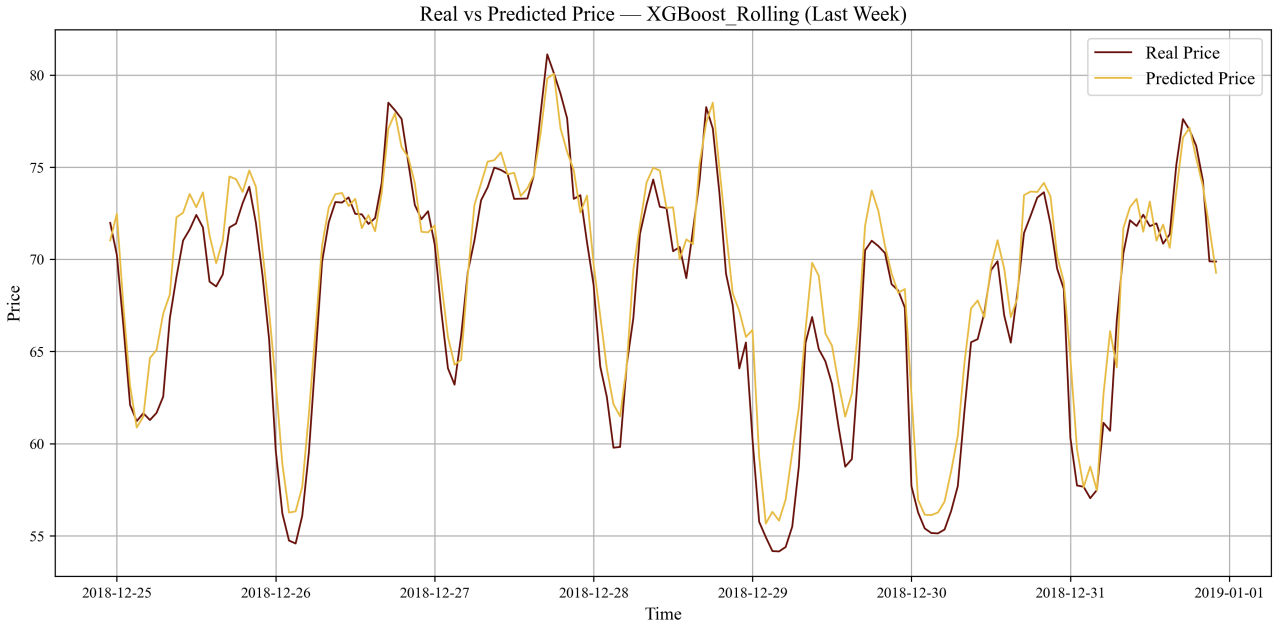

Figure II.2.4 – Real vs Predicted Electricity Price using XGBoost (Lag + Rolling), Final Week of the Test Set

Selection of the Best Model for Trading

Considering all numerical indicators and qualitative behaviours jointly, the XGBoost model employing lag, rolling and volatility features emerges as the most reliable and well-balanced forecasting tool. It achieves:

- high accuracy in absolute terms (low RMSE, low MAE),
- strong variance explanation (R^2 above 0.92),
- significant improvement over naïve persistence (MASE < 1),
- competitive directional performance (DA above 76%).

More importantly, it demonstrates a robust ability to reproduce the internal rhythm of the price series, including the daily cycles and short-term fluctuations that are critical for timing trading decisions. Its predictions faithfully follow the structure of real price movements, successfully anticipating upward and downward shifts even when the magnitude is imperfectly matched.

For these reasons, the **XGBoost (Lag + Rolling)** model is selected as the primary forecasting engine for the systematic mispricing-based trading strategies developed in Section II.3.

Trading Strategy, Backtesting Framework and Performance Assessment

III.1 From Forecast to Trading Signal

Electricity price forecasting provides numerical predictions $\widehat{P}_{t+1}$ for the next intraday hour, but these forecasts must be transformed into operational trading signals before any backtest or investment decision can be performed. This section formalises the transformation pipeline used throughout the study, which converts the output of the predictive model into a sequence of long/short positions interpretable by any trading strategy.

The XGBoost model selected in Section II.2 delivers, at each time step t, a one-hour-ahead point forecast $\widehat{P}_{t+1}$. The fundamental trading intuition is straightforward: if the predicted price increases in the next hour, a long position should be taken; if it decreases, a short position is appropriate. The raw forecasting signal is therefore defined as the predicted price change

$$s_t = \widehat{P}_{t+1} - P_t.$$

This scalar quantity encapsulates both the **direction** and **magnitude** of the expected market movement. Its sign determines the natural trading action:

$$\text{position}_t = \begin{cases} +1 & \text{if } s_t > 0, \\ -1 & \text{if } s_t < 0. \end{cases}$$

For more advanced strategies, the value of $|s_t|$ is also exploited as a proxy for **forecast strength**, enabling refined position-sizing rules (Sections 3.3.1–3.3.3). However, for the baseline direction-only strategies (Section 3.2), the model's output is converted into a simple binary long–short signal.

In addition to the forecast change, a second quantity proves essential for the mispricing-based mechanisms explored later:

$$m_t = \widehat{P}_{t+1} - P_{t+1},$$

representing the model's instantaneous prediction error. While unavailable in real time, this value is accessible during backtesting and allows the design of reversion-based strategies that exploit the discrepancy between predicted and realised prices.

Throughout the entire study, all trading strategies are evaluated on a one-month horizon representing **720 hourly observations**, ensuring consistent and comparable assessment across methods. The transformation described above acts as the common entry point for all subsequent trading rules; the differences between strategies arise solely from how they use s_t, m_t, and derived quantities to define positions and investment size.

III.2 Direction-Only Trading Strategies (No Capital Allocation)

Before introducing capital-based position sizing techniques, it is essential to evaluate the intrinsic quality of the predictive signals generated by the forecasting model. To do so, we consider a first family of trading rules—referred to as **direction-only strategies**—in which positions take values in $\{-1, +1\}$ and **no capital is allocated**. These strategies do not simulate a real financial portfolio; instead, they serve as diagnostic tools that quantify the model's ability to correctly anticipate the **direction** of short-term price movements.

In this setting, each trade corresponds to a change in price between two consecutive hours, and the profit-and-loss associated with a position is simply:

$$\text{PnL}_t = \text{position}_t \cdot (P_{t+1} - P_t)\,.$$

Because the position is always of unit magnitude, the PnL reflects **only the correctness of the directional prediction**, not the absolute economic gain of the strategy. This design removes all portfolio effects, leverage

choices or compounding, enabling a pure and unbiased assessment of the predictive model.

All strategies in this section are evaluated over the same one-month horizon, corresponding to **720 hourly observations**.

III.2.1 Pure Directional Strategy

The pure directional strategy is the most elementary transformation of the forecasting output into a trading rule. It relies solely on the predicted price change:

$$s_t = \widehat{P}_{t+1} - P_t.$$

The trading action is defined as:

$$\text{position}_t = \begin{cases} +1 & \text{if } s_t > 0, \\ -1 & \text{if } s_t < 0. \end{cases}$$

A long position is taken whenever the model predicts the next-hour price to increase, and a short position otherwise. This strategy effectively measures the **directional accuracy** of the model:

— if the model correctly identifies upward movements, the PnL is positive,
— if it mistakes the sign of the movement, the PnL is negative.

The pure directional strategy therefore acts as a **benchmark indicator** of the forecasting model's skill.

III.2.2 Mispricing Reversion Strategy

The mispricing strategy exploits the discrepancy between the predicted next-hour price and its realised counterpart:

$$m_t = \widehat{P}_{t+1} - P_{t+1}.$$

While this quantity is not accessible in real time, it can be used in backtesting to probe whether price deviations predicted by the model display systematic reversion patterns. The trading rule is constructed as:

$$\text{position}_t = \begin{cases} +1 & \text{if } m_t < 0, \\ -1 & \text{if } m_t > 0. \end{cases}$$

Interpretation:

— If the model believes the next price should be higher than it actually is ($m_t > 0$), then the realised price is "too low", suggesting upward mean reversion → long.
— Conversely, if the realised price is "too high" compared to the forecast, the strategy takes a short position.

This approach captures a different dimension of model behaviour: not its ability to predict direction, but its ability to forecast a fair value around which prices oscillate. Although not deployable in practice, it provides meaningful insight into whether the model embeds implicit equilibrium structure.

III.2.3 Threshold-Based Strategy

The threshold strategy introduces a selectivity mechanism that filters out weak or noisy signals. Instead of acting on every hour, the trader intervenes only when the forecasted movement exceeds an optimised threshold $\tau > 0$. The signal used is the same as in the directional case:

$$s_t = \widehat{P}_{t+1} - P_t,$$

but the trading action becomes:

$$\text{position}_t = \begin{cases} +1 & \text{if } s_t > \tau, \\ -1 & \text{if } s_t < -\tau, \\ 0 & \text{otherwise.} \end{cases}$$

Only sufficiently strong confidence signals lead to trades.

Advantages of this rule:

- reduces exposure to small, noisy fluctuations,
- increases robustness by avoiding ambiguous scenarios,
- typically improves the winrate at the cost of fewer trades.

In this work, the threshold τ is selected through a grid search optimising cumulative PnL over the test window.

III.3 Capital-Based Position Sizing Strategies

While the direction-only strategies of Section 3.2 test the intrinsic ability of the model to predict market movements, they provide no insight into the behaviour of an actual trading portfolio.
To simulate a realistic trading framework, this section introduces **capital-based strategies**, where each trade deploys a fraction of a 1 000 000 € portfolio.

These strategies evaluate how the model's forecast can be exploited under **risk constraints**, **position caps**, and **compounded capital dynamics**.

All approaches share the recursive update rule:

$$\text{capital}_{t+1} = \text{capital}_t + \text{PnL}_t,$$

and the PnL expression:

$$\text{PnL}_t = \text{position}_t \cdot \text{exposure}_t \cdot \left(\frac{P_{t+1} - P_t}{P_t} \right).$$

The differences between the strategies lie in *how the exposure is determined* and *how confidence is interpreted*.

III.3.1 Proportional Signal Allocation

The proportional strategy is the most intuitive of the three. It assumes that the **stronger the model's conviction**, the more capital should be deployed. Conviction is quantified through a composite intensity score:

$$I_t = |s_t| \cdot |m_t|,$$

which combines:

- the expected magnitude of the next-hour change,
- the deviation between forecast and realised price (interpreted as a "valuation error").

A raw allocation is formed:

$$a_t = \frac{I_t}{\kappa},$$

and capped:

$$a_t^{\text{final}} = \min(a_t, a_{\max}).$$

Exposure is further restricted by:

$$\text{exposure}_t = \min \left(a_t^{\text{final}} \cdot \text{capital}_t,\ E_{\max} \right).$$

Direction remains deterministic:

$$\text{position}_t = \text{sign}(s_t).$$

This strategy behaves as a **linear amplifier** of the signal: doubling the perceived strength doubles the capital at risk (up to the cap). It is simple, transparent, and reacts proportionally to the model's confidence, but its purely linear nature makes it more sensitive to outliers in the signal. It serves as a baseline for more sophisticated position sizing rules.

III.3.2 Sigmoid Confidence Allocation

The sigmoid strategy introduces **non-linearity** and **soft saturation**, addressing the main limitation of the proportional rule.
Instead of scaling exposure linearly, it transforms the composite score:

$$S_t = |s_t| \cdot |m_t|,$$

into a confidence level bounded between 0 and 1:

$$c_t = \frac{1}{1 + e^{-aS_t}}.$$

The resulting allocation is:

$$a_t = c_t \cdot a_{\max},$$

and the exposure:

$$\text{exposure}_t = \min(a_t \cdot \text{capital}_t,\ E_{\max}).$$

Direction is again:

$$\text{position}_t = \text{sign}(s_t).$$

he sigmoid curve introduces a **smooth, risk-controlled scaling**:

- — weak signals lead to very small allocations,
- — medium signals increase allocation progressively,
- — very strong signals saturate near the maximum allocation, preventing explosive behaviour.

This strategy therefore offers **robustness and stability**, reducing sensitivity to noise while still exploiting strong market opportunities. It is particularly well suited in environments where the model provides informative but imperfect forecasts, as is usually the case in electricity markets.

III.3.3 Kelly Criterion Allocation

The Kelly strategy is the most theoretically grounded and sophisticated. It stems from information theory and aims to **maximise long-run capital growth** under uncertainty.
Unlike the first two strategies, Kelly explicitly balances **expected reward** and **risk**.
A local estimated win probability is derived:

$$p_t = \frac{1}{1 + e^{-aS_t}}.$$

A payoff ratio is computed as:

$$R_t = \frac{|s_t| + \varepsilon}{|m_t| + \varepsilon}.$$

The Kelly fraction is then:

$$f_t = p_t - \frac{1 - p_t}{R_t}.$$

After clamping to $[0,\ 1]$:

$$f_t^{\text{final}} = \min(\max(f_t, 0), 1),$$

the allocation becomes:

$$a_t = f_t^{\text{final}} \cdot a_{\max}.$$

Final exposure:

$$\text{exposure}_t = \min(a_t \cdot \text{capital}_t,\ E_{\max}),$$

with directional sign:

$$\text{position}_t = \text{sign}(s_t).$$

Kelly is unique because it estimates how "good" the model appears locally, not only how strong the movement is. It invests more aggressively when both the probability of being right and the expected payoff ratio are favourable. This makes Kelly potentially the most efficient strategy, but also the most sensitive to estimation errors. The capping mechanisms introduced here ensure that its behaviour remains realistic and financially stable.

III.4 Backtesting Framework and Comparative Evaluation

The previous sections have established the full set of trading rules derived from the forecasting model, ranging from direction-only strategies to capital-based position sizing mechanisms. We now turn to their empirical evaluation through a systematic backtest conducted on the same one-month test horizon, corresponding to 720 hourly observations.

The purpose of this section is twofold: define the performance metrics used to evaluate the behaviour of each strategy in realistic trading conditions and compare these strategies across two distinct groups: direction-only strategies without capital allocation and capital-based strategies simulating a real portfolio.

Since the strategies differ in structure, risk-taking, and exposure dynamics, their performance cannot be assessed from profits alone. Instead, we rely on a combination of risk-adjusted metrics, stability indicators, and behavioural diagnostics that collectively capture profitability, robustness, and sensitivity to market conditions.

III.4.1 Evaluation Metrics for Trading Strategies

Backtesting a trading strategy requires a rigorous and multidimensional set of evaluation criteria. In this study, all strategies are assessed on the same 720-hour horizon using the metrics detailed below. These indicators quantify profitability, stability, and risk exposure, and allow for fair comparison between strategies with and without capital allocation.

Cumulative Profit-and-Loss (PnL)

For direction-only strategies (unit positions), the PnL is expressed as:

$$\mathrm{PnL}_t = \mathrm{position}_t \cdot \left(P_{t+1} - P_t\right).$$

For capital-based strategies:

$$\mathrm{PnL}_t = \mathrm{position}_t \cdot \mathrm{exposure}_t \cdot \left(\frac{P_{t+1} - P_t}{P_t}\right).$$

Cumulative PnL provides a first indication of raw profitability, but does not account for volatility or risk.

Winrate

The winrate measures the proportion of trades yielding a positive PnL:

$$\mathrm{Winrate} = \frac{\sum_t \mathbf{1}(\mathrm{PnL}_t > 0)}{N_{\mathrm{trades}}}.$$

It reflects the **consistency** of a strategy rather than its magnitude of profits. A high winrate typically indicates that a strategy captures directional movements reliably.

Sharpe Ratio (risk-adjusted return)

The Sharpe ratio is the primary metric for evaluating risk-adjusted performance.

Using hourly returns:

$$\mathrm{Sharpe} = \frac{\mathbb{E}[r_t]}{\sigma(r_t)}\sqrt{24 \times 365},$$

where

$$r_t = \frac{\mathrm{PnL}_t}{\mathrm{capital}_t}.$$

A high Sharpe ratio indicates that the strategy produces stable profits relative to its volatility, making it a critical metric in the comparison of strategies with capital.

Maximum Drawdown

Maximum drawdown measures the largest sustained loss observed during the backtest:

$$\text{Drawdown}_t = \text{capital}_t - \max_{k \le t} \text{capital}_k,$$

$$\text{Max DD} = \min_t \text{Drawdown}_t.$$

It quantifies **worst-case risk**, capturing how severely the portfolio can fall from its peak. Lower drawdowns indicate greater robustness and less exposure to adverse price movements.

Number of Trades

The number of trades executed over the 720-hour window provides insight into:

- signal frequency,
- strategy selectivity (especially for the threshold strategy),
- transaction efficiency.

It is especially critical when comparing a filtered strategy (e.g., Threshold) with always-active strategies (Directional, Kelly, Sigmoid).

Behavioural Diagnostics (non-metric indicators)

Additional qualitative diagnostics—such as:

- hourly PnL distributions,
- heatmaps of hourly winrate,
- price vs prediction alignment,

are used to understand *how* a strategy behaves, beyond its cumulative performance. These tools help interpret whether profits arise from systematic structure (e.g., daily seasonality) or chance.

III.4.2 Performance of Direction-Only Strategies

The first group of trading approaches comprises the Directional, Mispricing and Threshold strategies, all implemented under a simplified assumption of constant exposure: each hourly position corresponds to a fixed notional of 1 MWh, so that cumulative profits directly represent the sum of hourly gains and losses. This modelling choice allows us to isolate the intrinsic quality of the trading signal produced by the forecasting model, independently of any capital-allocation rule. All strategies are evaluated over the same 30-day horizon (approximately 720 hourly observations), ensuring strict comparability.

The numerical results obtained for the three strategies are summarised in Table III.4.1, which reports the total PnL, win-rate, Sharpe ratio and maximum drawdown.

Strategies	Total Profit (€)	Winrate	Sharpe	Max Drawdown
Directional	+1177.68	76.39%	62.05	-14.53
Threshold	+1184.00	74.03%	63.08	-14.53
Mispricing	+693.36	63.61%	32.20	-23.49

Table III.4.1 – Performance of No-Capital Strategies (Directional, Mispricing, Threshold)

Comparative Interpretation

A detailed examination of the results reveals clear differences in behaviour and robustness across the three methods.

The **Directional strategy**, which simply takes a long position whenever the model predicts an upward movement and a short position otherwise, displays strong intrinsic predictive accuracy. Its win-rate reaches 76%, and the cumulative PnL grows steadily throughout the month. However, because this strategy trades **every single hour**, it also captures periods where the model's predictive signal is weak or noisy. This produces a larger number of very small profits and occasional losses, revealing that unconditional trading exposes the strategy to

unnecessary micro-volatility. Despite this, the directional method remains a solid baseline, with a high Sharpe ratio indicating that the underlying XGBoost predictor successfully captures intraday directional structure.

The **Mispricing strategy**, which trades on the discrepancy between the predicted price and the realised price, performs substantially worse. Its underlying assumption—namely that large prediction errors contain exploitable information about future corrections—does not align well with the statistical properties of our forecasting model. The resulting behaviour is more erratic: the win-rate drops to 63%, the Sharpe ratio is halved compared to the directional approach, and the maximum drawdown becomes significantly more pronounced. This degradation suggests that mispricing signals are dominated by model noise rather than genuine arbitrage-like deviations, making the strategy unreliable in this context.

In contrast, the **Threshold strategy** offers a markedly superior balance between selectivity and performance. By imposing the condition

$$|\widehat{P}_{t+1} - P_t| > \theta \qquad \theta = 0.1$$

The strategy filters out low-confidence predictions and trades only when the forecasting model exhibits a sufficiently strong signal. This selective mechanism reduces the number of trades from 720 to 687, but leads to **higher-quality entries**, almost eliminating weak or noisy trades. The result is a combination of the best total PnL (1184 €), one of the highest win-rates (74%), and the strongest Sharpe ratio (63.1). Moreover, the maximum drawdown remains minimal, illustrating the stability of the filtered strategy.

Taken together, these observations indicate that the threshold mechanism acts as an effective **noise-reduction layer** on top of the raw predictive signal. The strategy exploits only the most confident model outputs, thereby maximising profitability while limiting downside risk. Among all no-capital methods, the threshold approach emerges unequivocally as the most efficient and the most robust.

Trade-Level Inspection

To further illustrate the internal mechanics of the threshold rule, a sample of the executed trades is provided in Table III.4.2. This excerpt highlights how trades are triggered only when the predicted movement exceeds the threshold and shows the resulting stability of the cumulative PnL trajectory. The alignment of predicted and realised movements is visible row by row, demonstrating how selective entry enhances both precision and profitability.

time	price_t	price_tplus	pred_tplus	position	pnl	pnl_cum
2018-12-01 23:00:00	64.5	60.27	58.9387	-1	4.2299	4.2299
2018-12-02 00:00:00	60.27	53.37	56.0892	-1	6.9000	11.13
2018-12-02 01:00:00	53.37	51.32	51.1058	-1	2.0499	13.18
2018-12-02 02:00:00	51.32	50.03	49.9786	-1	1.2899	14.4699
2018-12-02 03:00:00	50.03	50.25	49.2545	-1	-0.2199	14.25
2018-12-02 04:00:00	50.25	50.98	51.3179	1	0.7299	14.9799
2018-12-02 05:00:00	50.98	51.73	54.8740	1	0.75	15.7299
2018-12-02 06:00:00	51.73	51.42	57.0430	1	-0.3099	15.4200
2018-12-02 07:00:00	51.42	52.01	56.4001	1	0.5899	16.0099
2018-12-02 08:00:00	52.01	55.41	58.2717	1	3.3999	19.4099
2018-12-02 09:00:00	55.41	60.8	62.7655	1	5.3900	24.7999
2018-12-02 10:00:00	60.8	63.12	64.9460	1	2.3200	27.1199
2018-12-02 11:00:00	63.12	63.85	66.5483	1	0.7300	27.85
2018-12-02 12:00:00	63.85	67.94	64.9096	1	4.0899	31.9399
2018-12-02 13:00:00	67.94	66.48	67.7393	-1	1.4599	33.3999
2018-12-02 14:00:00	66.48	62.07	64.3112	-1	4.4100	37.8099
2018-12-02 15:00:00	62.07	61.77	61.7375	-1	0.2999	38.1099
2018-12-02 16:00:00	61.77	66.78	64.8574	1	5.0099	43.1199

Table III.4.2 – Extract of Executed Trades for the Threshold Strategy

Visual Examination of the Threshold Strategy

The quantitative superiority of the threshold method is further corroborated by the visual diagnostics presented in Figure III.4.1. The cumulative PnL curve displays an almost perfectly monotonic ascent, with no prolonged flat phases or large negative excursions. The hourly PnL plot reveals a characteristic pattern dominated by frequent small gains and rare moderate losses, consistent with a filtering mechanism that captures clear directional signals. The PnL distribution exhibits a pronounced positive skew, while the heatmap of hourly win-rates indicates consistent performance across both time-of-day and calendar-day dimensions.

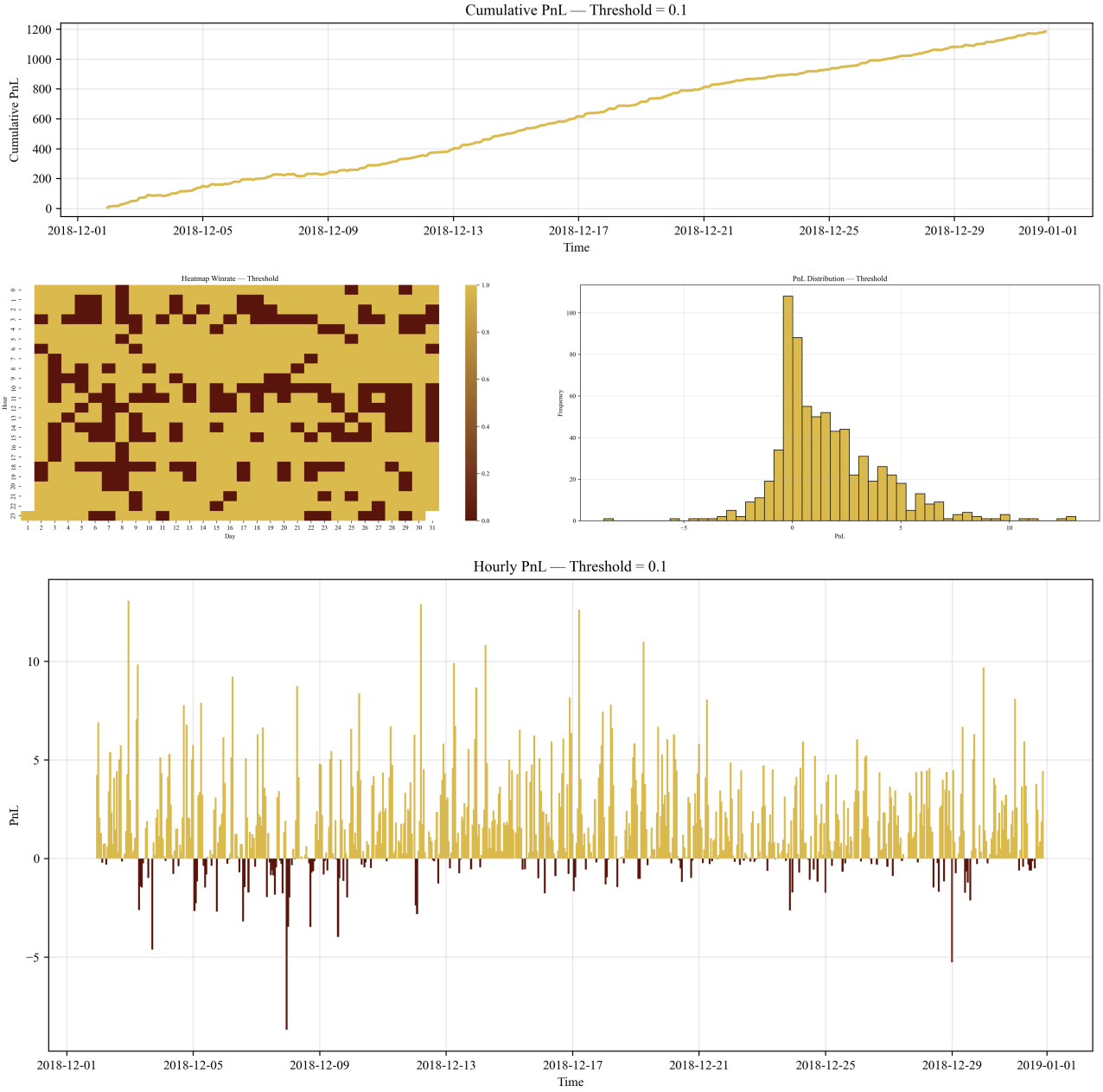

Figure III.4.1 – Comprehensive Performance Analysis of the Threshold Trading Strategy (= 0.1)

Across all no-capital strategies, the **Threshold strategy** provides the most compelling trade-off between profitability, consistency and risk control. By restricting trading activity to strong signals only, it significantly enhances both stability and efficiency, outperforming the directional baseline while clearly surpassing the mispricing approach. The Threshold strategy is therefore selected as the **benchmark no-capital method** for the remainder of the study. For completeness, all diagnostic figures corresponding to the Directional and Mispricing strategiesnare provided in Appendix B.

III.4.3 Performance of Capital-Based Strategies

The second family of trading strategies introduces dynamic position sizing, allowing the system to adjust monetary exposure based on the strength or reliability of the predictive signal. Unlike the no-capital methods analysed previously, these strategies operate on an initial capital of 1,000,000 €, and each trade invests a specific amount derived from model outputs. This framework enables an assessment not only of signal accuracy but also of allocation efficiency and risk management. All three strategies are evaluated over the same 30-day horizon, corresponding to 720 hourly observations.

Table III.4.3 summarises the performance of the *Proportional Signal*, *Sigmoid Confidence Sizing*, and *Dynamic Kelly Criterion* strategies. Although each generates positive returns, the differences in volatility, drawdown behaviour, and stability of the capital curve are significant.

Strategy	Total Profit (€)	Winrate	Sharpe	Drawdown
PropSignal	177,863 €	76.39%	58.28	–2,323 €
Sigmoid	181,661 €	76.39%	59.37	–2,323 €
Kelly	180,733 €	66.53%	**62.12**	–1,963 €

Table III.4.3 – Comparison of Position-Sizing Strategies (PropSignal, Sigmoid, Kelly)

Comparative Interpretation

The **Proportional Signal** strategy scales exposure directly with the magnitude of the predicted price increment. While conceptually intuitive, this mechanism does not suppress the amplification of spurious fluctuations in the forecast. Consequently, although the strategy yields a profit of approximately 178 k€, the capital trajectory exhibits sharper oscillations, and the drawdown is comparatively deeper. Its Sharpe ratio (around 58) remains respectable, but the method lacks any damping mechanism capable of reducing the impact of overconfident or noisy predictions.

The **Sigmoid Confidence Sizing** strategy improves substantially on this behaviour by mapping the joint signal–mispricing magnitude through a smooth bounded transformation:

$$\text{allocation}_t = 0.3 \times \frac{1}{1 + \exp\left(-a\,|s_t m_t|\right)}.$$

Thus avoiding extreme exposure while still rewarding strong predictive agreement. This controlled non-linearity results in the most regular and stable capital curve among the three. With a profit of **181.7 k€**, the strategy narrowly outperforms the proportional method but does so with markedly lower drawdown and a **higher Sharpe ratio** (≈ 59). The combination of soft filtering and capped exposure produces a smoother distribution of hourly PnL and improved resilience to model uncertainty.

The **Dynamic Kelly Criterion** incorporates both estimated win probability and payoff ratio, theoretically maximizing long-run capital growth. In practice, however, the estimation of these quantities from model outputs introduces substantial uncertainty. While the Kelly strategy generates a profit similar to the sigmoid method (**180.7 k€**) and achieves the **highest Sharpe ratio** of the three (≈ 62), its win-rate is noticeably lower, and the behaviour of individual trades is more irregular. This indicates that Kelly-based allocations are more sensitive to model misestimation and may overreact to transient fluctuations in prediction confidence.

Overall, the comparative analysis highlights a critical trade-off: Kelly sizing offers theoretically optimal growth under perfect model calibration, whereas proportional sizing tends to over-amplify noise. The sigmoid approach occupies an advantageous middle ground, providing a stable transformation of the predictive signal that offers **both high performance and strong risk control**.

Visual Examination of the Sigmoid Strategy

Given its balanced performance across all risk–return dimensions, the **Sigmoid Confidence Sizing strategy** is selected as the preferred capital-based method. Its behaviour is illustrated in Figure IV.3.1, which reveal a consistently rising capital curve with very few periods of stagnation or negative drift. The hourly PnL distribution is characterised by numerous moderate gains and a limited number of controlled losses, while the heatmap of win-rates displays robust performance across most hours and days of the month. These visual diagnostics confirm the superior stability of the sigmoid approach relative to both proportional and Kelly sizing.

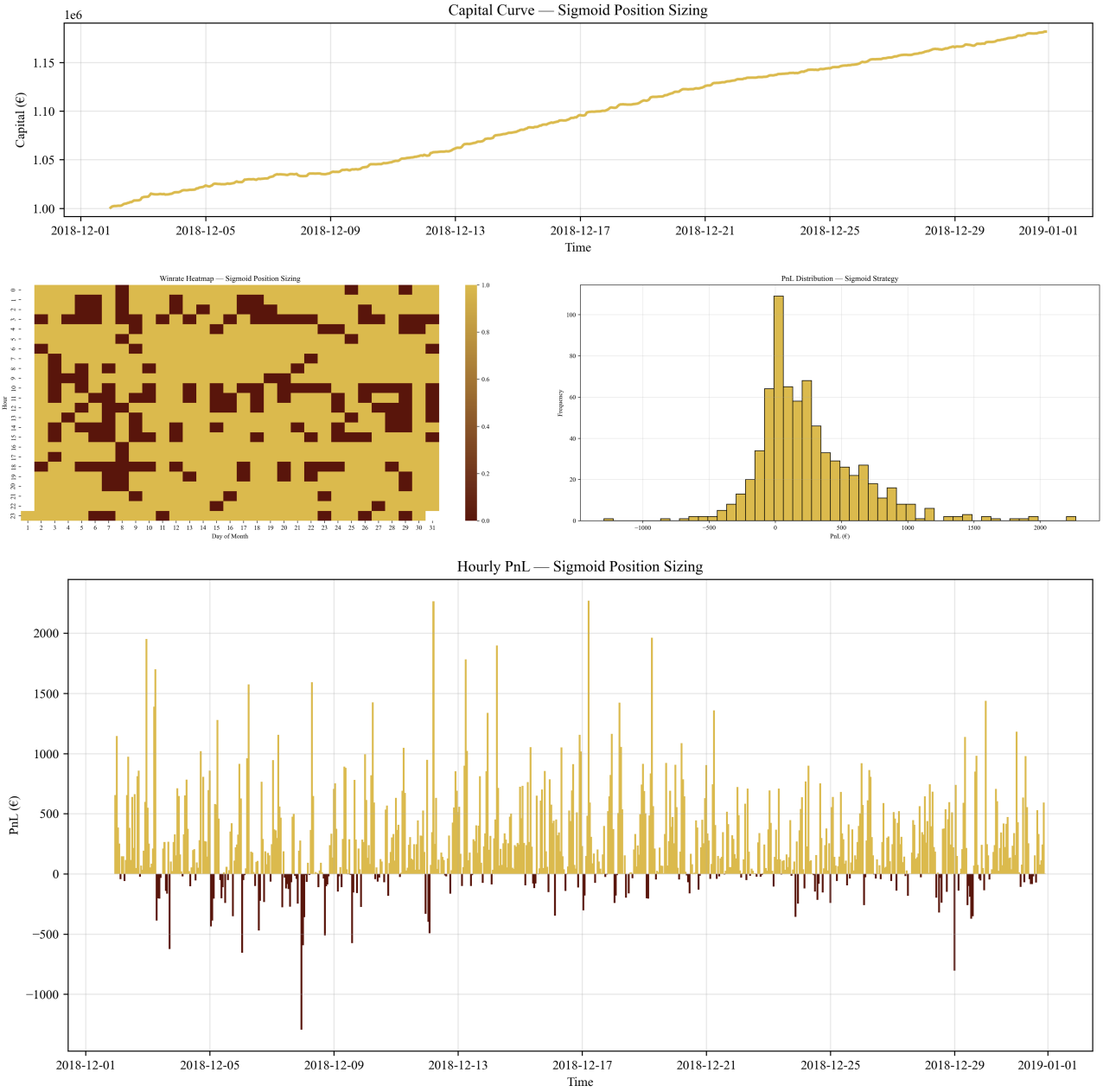

Figure III.4.2 – Diagnostic Visualisation of the Sigmoid Position Sizing Strategy

Trade-Level Inspection

To further illustrate how dynamic position sizing operates in practice, a sample of the executed trades for the sigmoid strategy is presented in Table III.4.4. The table shows the interaction between predicted movement, mispricing signal, allocation, and resulting PnL. It demonstrates how the sigmoid function scales exposure smoothly, preventing excessively large trades and thereby contributing to the strategy's stable capital trajectory.

Time	Price_t	Price_t+1	Pred_t+1	Signal	Mispricin	Score	Conf.	Alloc.	Pos.	PnL	Capital
2018-12-01 23:00	64.50	60.27	58.9387	-5.5613	-1.3313	7.4037	0.9759	0.2928	-1	655.81	1000655
2018-12-02 00:00	60.27	53.37	56.0892	-4.1808	2.7192	11.3684	0.9966	0.2990	-1	1144.84	1001800
2018-12-02 01:00	53.37	51.32	51.1059	-2.2641	-0.2141	0.4848	0.5603	0.1681	-1	384.11	1002184
2018-12-02 02:00	51.32	50.03	49.9786	-1.3432	-0.0514	0.0689	0.5086	0.1526	-1	251.36	1002436
2018-12-02 03:00	50.03	50.25	49.2545	-0.7755	0.3370	0.7719	0.5953	0.1786	-1	-43.97	1002392
2018-12-02 04:00	50.25	50.98	51.3179	1.0680	1.9656	0.3609	0.5449	0.1635	1	145.27	1002537
2018-12-02 05:00	50.98	51.73	54.8740	3.8940	3.1440	12.2430	0.9978	0.2993	1	147.11	1002684
2018-12-02 06:00	51.73	51.42	57.0431	5.3131	5.6231	29.8759	0.9999	0.3000	-1	-59.92	1002624
2018-12-02 07:00	51.42	52.01	56.4002	4.9801	4.3902	21.8639	0.9999	0.2999	-1	114.74	1002739
2018-12-02 08:00	52.01	55.41	58.2717	6.2617	2.8617	17.9194	0.9999	0.2999	1	653.72	1003393

Table III.4.4 – Extract of executed trades for the Sigmoid Strategy

Among the capital-based strategies, the Sigmoid Confidence Sizing approach clearly offers the most attractive balance between profitability, robustness, and drawdown control. While the Kelly Criterion achieves a slightly higher Sharpe ratio, its reduced stability and heavier reliance on precise model calibration make it less suitable for operational deployment. Proportional sizing, although profitable, amplifies noise excessively and exhibits a less favourable risk profile. For these reasons, the Sigmoid method is retained as the optimal capital-based strategy for the subsequent analysis. For completeness, all figures corresponding to the Proportional Signal and Kelly Criterion strategies are provided in Appendix C.

Conclusion

General Synthesis of the Work

The objective of this project was to design, implement, and evaluate a complete predictive and algorithmic trading framework dedicated to the short-term forecasting of intraday electricity prices. Starting from raw heterogeneous data sources and concluding with fully operational systematic trading strategies, the work brought together multiple methodological components: data engineering, time-series forecasting, model optimisation, and financial backtesting. This section summarises the central achievements and the scientific rationale that emerged throughout the study.

The project began with the construction of an enriched electricity–meteorology dataset, combining hourly Iberian market fundamentals with high-resolution weather observations from several Spanish regions. A rigorous preprocessing pipeline was designed to merge, clean, and structure these data, introducing lagged, rolling and volatility-based features to capture temporal dependencies. This step proved crucial, as electricity prices are driven simultaneously by meteorological variability, consumption patterns, and generation mix dynamics. The result was a robust explanatory matrix capable of feeding both tree-based machine learning models and sequence-based neural architectures.

The second part of the work focused on forecasting. Two modelling paradigms were explored in depth: gradient-boosted decision trees (XGBoost) and recurrent neural networks (LSTM, GRU and attention layers). Each family was analysed theoretically and evaluated empirically. Extensive hyperparameter optimisation, conducted with Optuna, revealed that tree-based models were particularly well adapted to the dataset at hand. Owing to their capacity to exploit engineered temporal features and to handle heterogeneous predictors, XGBoost architectures consistently achieved superior accuracy and stability compared with recurrent neural networks. In contrast, LSTM-based models, which typically perform well on large sequential datasets, struggled in this context due to the relatively limited training horizon and the moderate size of the available observations, which restricted their ability to learn complex temporal structures. The comparative assessment ultimately identified the XGBoost model enriched with lagged and rolling features as the best-performing forecasting tool.

Building upon this predictive foundation, the project then shifted to systematic trading applications. Six trading strategies were designed and backtested using the model's hourly price forecasts: three without explicit capital allocation (directional, mispricing, threshold-based) and three incorporating dynamic position sizing (proportional signal, sigmoid confidence, Kelly criterion). Each strategy was rigorously evaluated on a 30-day, 720-hour test window, enabling a fair comparison of their respective profitability, stability, and risk-adjusted performance. This systematic exploration demonstrated how different interpretations of the predictive signal—whether focusing on direction, deviation from fair value, or strength of confidence—translate into distinct trading behaviours and financial outcomes.

The study culminated in a comprehensive comparison of all strategies. Among the no-capital methods, the threshold strategy emerged as the most robust, achieving high profitability while filtering out low-quality signals. Among capital-based methods, the sigmoid confidence sizing strategy offered the best balance between return generation, drawdown control and forecast uncertainty mitigation, outperforming both proportional and Kelly-based allocations. These results confirm that the combination of a strong predictive model and a disciplined allocation mechanism can lead to consistent trading performance.

Overall, the project successfully delivered a complete and operational workflow, from data acquisition to trading execution, demonstrating both the feasibility and the relevance of machine-learning-augmented decision systems within the electricity market context. The integration of forecasting accuracy, risk management, and systematic evaluation provides a coherent foundation for future extensions and for real-world deployment scenarios.

Discussion and Critical Assessment

The purpose of this section is to critically examine the methodological choices, modelling assumptions, and empirical outcomes obtained throughout the project. While the study demonstrates that machine-learning-driven trading strategies can be effectively adapted to the intraday electricity market, a number of limitations became apparent. These limitations stem from the data structure, the intrinsic behaviour of electricity prices, and the properties of the algorithms employed. Understanding these aspects is essential both for contextualising the results and for identifying promising avenues for improvement.

IV.2.1 Modelling Constraints and Dataset Limitations

A central difficulty encountered in this project arises from the size and structure of the available dataset. Although the merged electricity–meteorology dataset is rich in explanatory variables, the effective training horizon remains relatively short compared with the needs of deep learning architectures. Recurrent models such as LSTMs and GRUs are known to require large quantities of sequential data to learn long-range temporal dependencies, calibrate internal memory states, and avoid overfitting. In large real-world energy datasets—spanning several years of hourly observations—LSTMs frequently outperform tree-based models by capturing recurrent seasonal cycles and complex dynamical patterns.
In contrast, in the present study, the model had access to only a limited number of months for training, which constrained the diversity of temporal regimes it could observe. As a consequence, the LSTM struggled to generalise and displayed unstable predictions, reflected in high RMSE and negative R^2 scores. Meanwhile, XGBoost—whose strength lies in exploiting richly engineered features rather than raw sequences—proved more robust under data scarcity. This confirms that, for relatively small time series with strong exogenous structure, feature-driven models outperform sequence-driven ones.

Another limitation concerns the fact that electricity prices exhibit discontinuities, spikes and abrupt regime changes that are difficult for neural networks to learn without massive data augmentation. Boosted trees, with their piecewise-constant structure, inherently adapt more easily to such irregularities. Yet they too have limitations: XGBoost predictions tended to under-react during extreme price swings, as seen in the final-week diagnostic figures. This behaviour reflects the regularisation pressure of the model: by construction, boosted trees tend to smooth large deviations unless sufficient historical examples exist, which was not the case here.

IV.2.2 Sensitivity of Trading Strategies to Forecast Noise

The evaluation of the six trading strategies reveals how different transformations of the predictive signal amplify or attenuate model imperfections. The directional and threshold strategies derived from price increments depend heavily on the correct forecasting of relative variations rather than absolute levels. This made them relatively resilient to slight bias in the prediction, but also vulnerable to small, noisy movements in the forecast. The threshold strategy addressed this issue by filtering weak signals, which is why it clearly outperformed the directional and mispricing strategies: by ignoring uncertain situations, it implicitly compensated for the model's limited precision during flat or noisy market phases.

The capital-based strategies, however, exposed more deeply the sensitivity of allocation mechanisms to forecast calibration. The Proportional Signal method magnified raw prediction errors, leading to some unnecessarily large exposures and consequently higher volatility. Although the strategy remained profitable thanks to the overall accuracy of the forecasting model, its behaviour illustrates the risk of linear scaling mechanisms when the underlying predictive score is itself noisy.

The Kelly Criterion strategy further emphasised this issue. In theory, the Kelly ratio is optimal when the probability of success and the payoff ratio are correctly estimated. In practice, these quantities were derived from the model's output, not from market structure. The estimated win probability was not a true calibrated forecast, and the payoff ratio fluctuated randomly with the magnitude of prediction errors. This led the Kelly strategy to overreact, producing unstable allocations and a lower win-rate despite achieving strong profitability. The method therefore demonstrated its classic weakness: Kelly sizing is extremely powerful under ideal calibration but excessively fragile when any part of the input is misestimated.

The Sigmoid Confidence strategy corrected many of these issues by transforming the predictive score through

a smooth, bounded function. This disciplined exposure rule limited overconfidence while still rewarding clear directional signals. As a result, it delivered the most stable capital trajectory and the best balance between risk and reward. The strategy effectively demonstrated that, in practical trading systems, the shape of the allocation function is as important as the accuracy of the underlying model.

IV.2.3 Implications for Model Robustness and Practical Deployment

The findings of this project highlight a number of general principles relevant to real-world electricity trading. First, predictive accuracy alone does not imply trading profitability: what matters is the alignment between the model's strengths and the structure of the trading rule. Some strategies were designed to extract value from directionality, others from magnitude or confidence; each interacted differently with the noise profile of the model. The best-performing strategies were those that compensated for the forecasting limitations through explicit filtering or controlled exposure rules.

Second, the project underscores the need for calibrating probabilistic forecasts when building capital-based strategies. Better calibration—through quantile regression, ensemble methods, or Bayesian uncertainty estimates—would significantly enhance the performance of dynamic allocation rules such as the Kelly Criterion. Similarly, incorporating uncertainty-aware models, such as Monte Carlo dropout networks or gradient-boosting ensembles with variance estimates, could improve risk management.

Third, although the backtest period of one month provides a fair comparison of strategies, it remains a limited window for assessing long-term robustness. Electricity markets are characterised by seasonal cycles, structural breaks, and exogenous shocks. For deployment in a real trading environment, walk-forward validation over multiple years, the inclusion of transaction costs, and stress-testing under extreme price regimes would be necessary.

IV.2.4 Overall Critical Perspective

While the project successfully demonstrated the feasibility of combining machine learning with systematic trading in the electricity market, several limitations must be acknowledged. The predictive accuracy is constrained by data availability, the trading rules rely on simplified assumptions (no transaction costs, perfect liquidity), and the strategies were evaluated over a single recent period. Nonetheless, the methodological pipeline developed here is modular, extensible, and well grounded in the principles of quantitative energy trading. Its strengths lie in its clarity, reproducibility, and the synergy between model design, optimisation, and risk-aware trading logic.

Perspectives and Future Work

The results obtained in this work demonstrate the feasibility of combining machine learning with systematic trading to extract value from short-term electricity price movements. Nevertheless, the methodological pipeline developed here can be enhanced in several directions. Future progress may arise not only from more expressive predictive models but also from more sophisticated trading rules capable of better capturing uncertainty, market regimes, and non-linear risk profiles.

IV.3.1 Advanced Predictive Modelling: From Sequential Networks to Transformers

While XGBoost proved highly effective on the present dataset, modern deep learning architectures—particularly sequence-to-sequence Transformers—offer promising opportunities for improving short-term electricity price forecasting. Transformers represent a significant conceptual shift from traditional recurrent networks such as LSTMs: instead of processing time steps sequentially, they rely on self-attention mechanisms that model global temporal dependencies in parallel. This architecture allows them to capture long-range relationships, sudden regime shifts, and cross-variable interactions more efficiently than recurrent methods.

Transformer-based models (e.g., Temporal Fusion Transformer, Informer, Autoformer, TranAD) have shown superior performance in numerous time-series domains, including financial forecasting, traffic prediction, and energy demand modelling. They are particularly well suited to datasets where temporal patterns are irregular, multi-scale, or influenced by numerous exogenous factors—precisely the characteristics of electricity markets.

Moreover, Transformers natively produce attention maps that offer a form of interpretability, helping to identify which past hours or which exogenous variables were most influential at a given prediction time.

However, the adoption of Transformer architectures requires substantially larger training datasets than those used in this project. To exploit their full potential, future work should consider expanding the historical window to multiple years of hourly data, incorporating additional markets (France, Portugal, Germany), and adding higher-frequency fundamental indicators (balancing market prices, renewable forecasts, congestion metrics). With such enriched datasets, a Transformer model could likely surpass XGBoost by modelling long-range seasonalities and exploiting richer cross-temporal patterns.

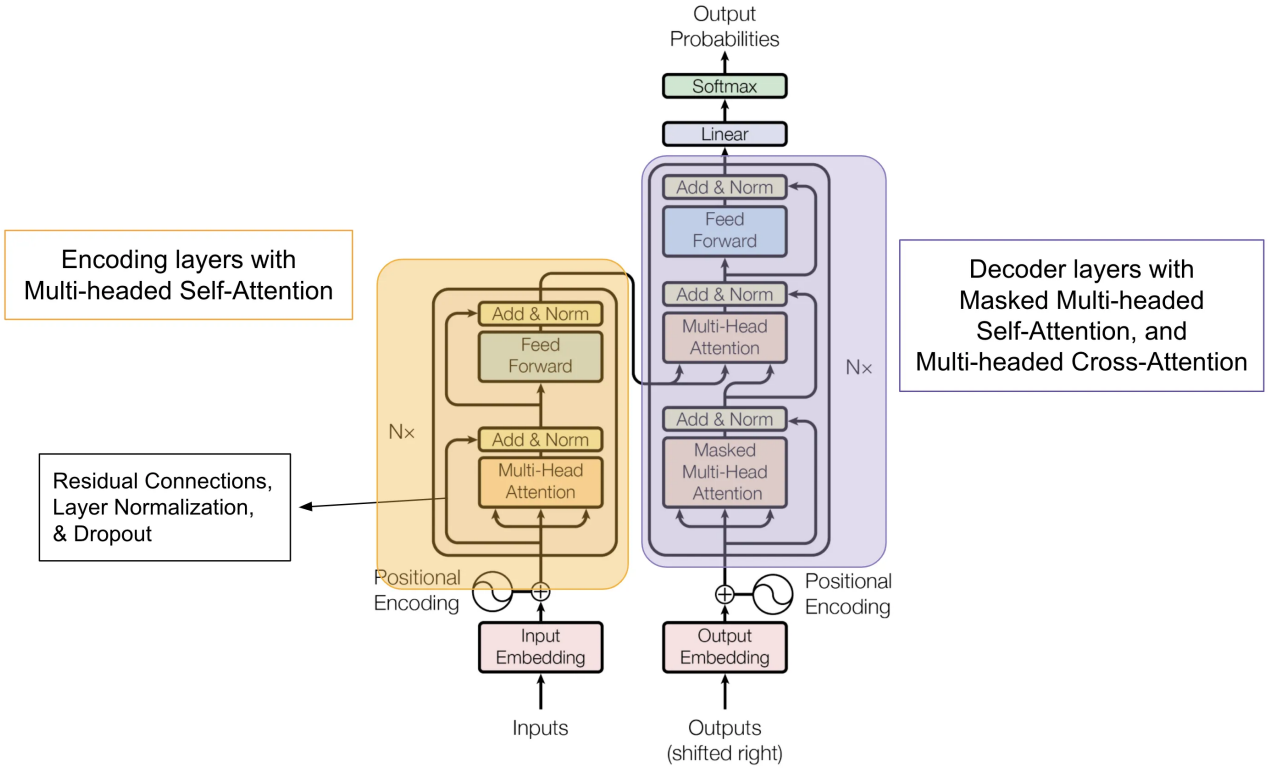

Figure IV.3.1 – Schematic Overview of a Transformer Encoder Architecture

IV.3.2 Towards More Robust and Adaptive Trading Strategies

The second avenue for extension concerns the trading strategies themselves. The methods explored in this report primarily rely on deterministic rules linking point forecasts to position decisions. While effective, these strategies can be made considerably more sophisticated in future work.

A first natural extension would be to incorporate probabilistic forecasting. Instead of relying solely on point predictions, the model could produce predictive distributions or quantile forecasts. This would allow strategies to explicitly account for forecast uncertainty, leading to improved risk management. For example, a trade could be executed only when the probability of a positive return exceeds a given threshold, or position size could scale with predictive confidence in a statistically calibrated manner. This would directly address the fragility observed in the Kelly strategy and greatly enhance the robustness of capital allocation.

Another extension would be to integrate regime detection into the trading logic. Electricity markets alternate between low-volatility and high-volatility regimes, influenced by renewable ramps, demand shocks, or market stress. A regime classifier—possibly built using clustering, hidden Markov models or Transformer encoders—could modulate trading intensity based on structural conditions. Signals would be trusted in stable regimes but filtered more aggressively during unstable periods.

In addition, advanced position-sizing frameworks such as risk-parity allocation, expected shortfall constraints, or convex optimisation of portfolios across multiple markets could be considered. Integrating multiple electricity markets (Spain, France, Germany) into a unified cross-market strategy would allow the system to exploit

arbitrage-like behaviours or structural correlations. Techniques from reinforcement learning could also be explored to discover allocation policies that optimise long-term capital growth under uncertainty.

Finally, a realistic deployment would require a more comprehensive backtesting environment that includes transaction costs, bid–ask spreads, liquidity constraints, and rolling-window re-training. These elements, while beyond the scope of the present work, would be essential for assessing the viability of the system in operational conditions.

Personal Contribution and Reflection

Beyond the technical contributions delivered throughout this project, the work represented a significant learning experience at both the methodological and intellectual levels. Developing a full predictive–trading pipeline from raw data to operational strategies required moving across several domains—energy economics, time-series modelling, machine learning, optimisation, and quantitative finance—and integrating them into a coherent system. This transversal structure shaped my understanding of the complexity inherent to real-world modelling and the subtle interplay between data, algorithms, and decision rules.
2mm A central takeaway from this project is the importance of empirical discipline and methodological humility. Working on electricity price forecasting highlighted how even sophisticated models can underperform when confronted with data scarcity, structural noise, or abrupt market changes. The difficulties encountered with sequence-based neural networks, which in theory dominate many large-scale forecasting benchmarks, taught me that model choice must always be grounded in the characteristics of the dataset rather than prevailing trends in the literature. This reinforced my appreciation for the practical strengths of feature-driven models like XGBoost, which ultimately proved more reliable in this context.

The design and evaluation of trading strategies also provided a valuable insight into the relationship between prediction quality and economic value. I realised that a machine-learning model does not need to be perfectly accurate to be financially exploitable, but it must align with the structure of the trading rule. Conversely, even a strong predictive signal can lead to poor financial outcomes when embedded into an inappropriate allocation mechanism, as illustrated by the fragility of the Kelly-based strategy. This clarified the importance of risk-aware decision rules and the necessity of balancing aggressiveness with robustness.

From a more personal standpoint, the project strengthened my ability to manage an end-to-end quantitative workflow: integrating heterogeneous data sources, engineering meaningful features, experimenting with algorithms, tuning hyperparameters, interpreting outputs, and constructing reproducible backtests. The process also familiarised me with the discipline of scientific writing—structuring arguments, articulating limitations, and maintaining methodological justification at every stage. These skills will be directly transferable to any future work in machine learning, data science, or quantitative finance.

Finally, the experience deepened my interest in the domain of energy markets, where uncertainty, volatility and system dynamics combine to form a rich field for data-driven modelling. The challenges encountered encouraged me to explore more advanced architectures such as Transformers, probabilistic forecasting tools, and reinforcement-learning-based decision systems, which I now see as natural extensions of this work. Overall, the project was not only technically enriching but also intellectually formative, shaping both my understanding of machine learning in complex environments and my motivation to pursue quantitative modelling at a higher level.

Appendices

Appendix A : Additional Prediction Curves

This appendix provides the full prediction curves for the three remaining forecasting models developed in Section II.2. These visualisations complement the main analysis by illustrating the qualitative behaviour of each model over the test horizon. They allow for a direct comparison between the real intraday electricity price and the corresponding predictions, thereby offering further insight into the strengths and limitations of each architecture beyond the numerical metrics presented in the main text.

V.1.1 XGBoost (Lag Features Only)

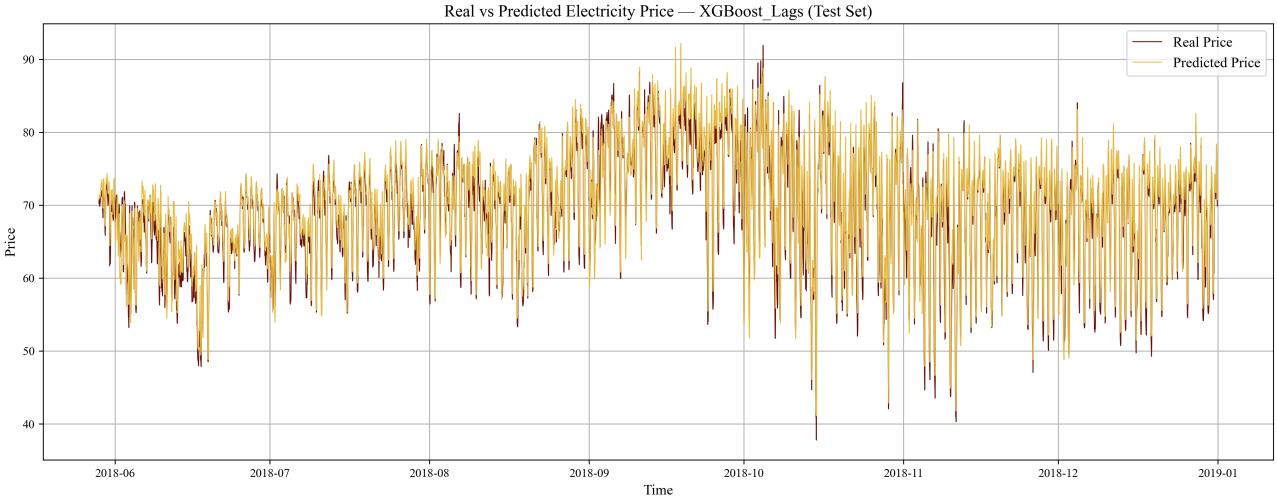

Figure V.1.1 – Real vs Predicted Electricity Price using XGBoost (Lag), Test Set.

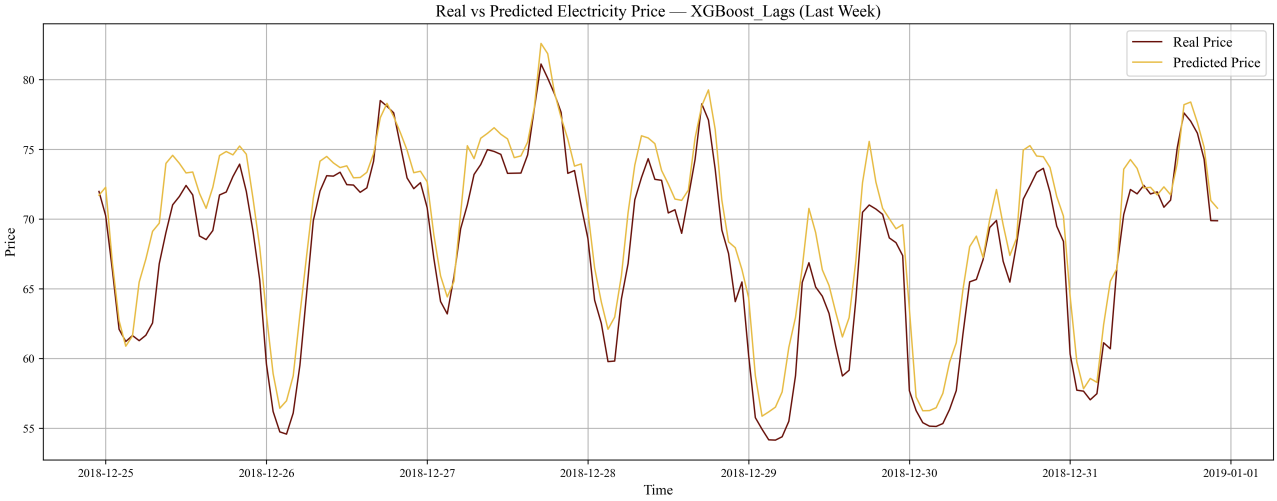

Figure V.1.2 – Real vs Predicted Electricity Price using XGBoost (Lag), Final Week of the Test Set

V.1.2 LSTM (72-hour Input Window)

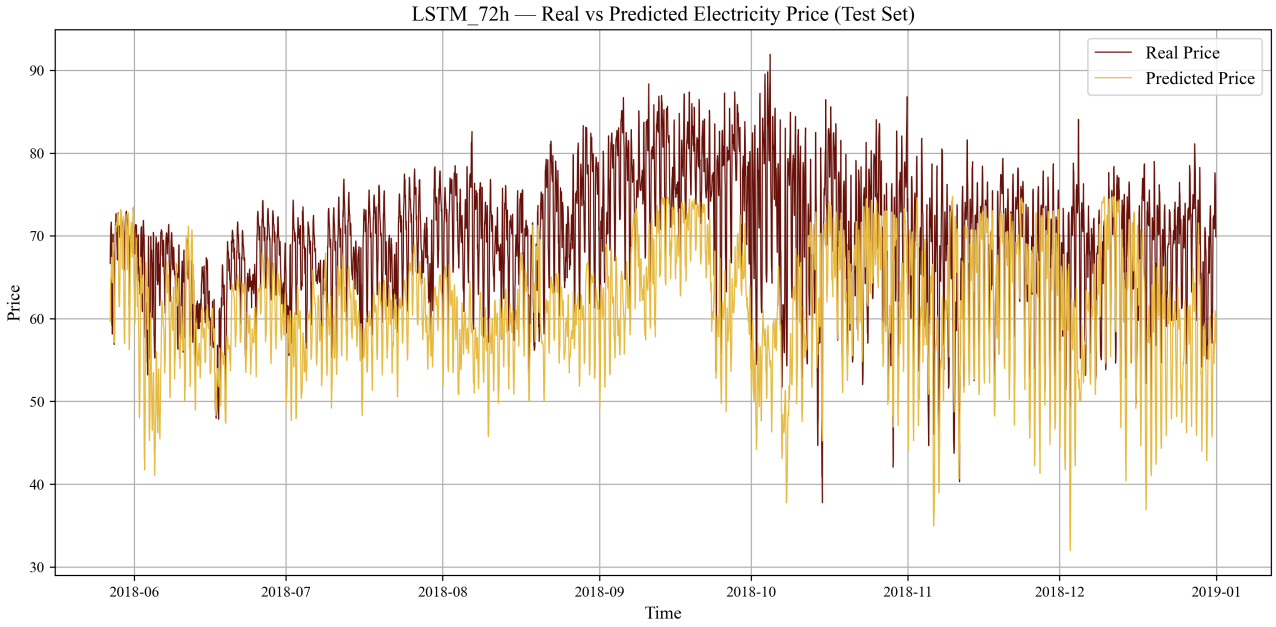

Figure V.1.3 – Real vs Predicted Electricity Price using LSTM, Test Set.

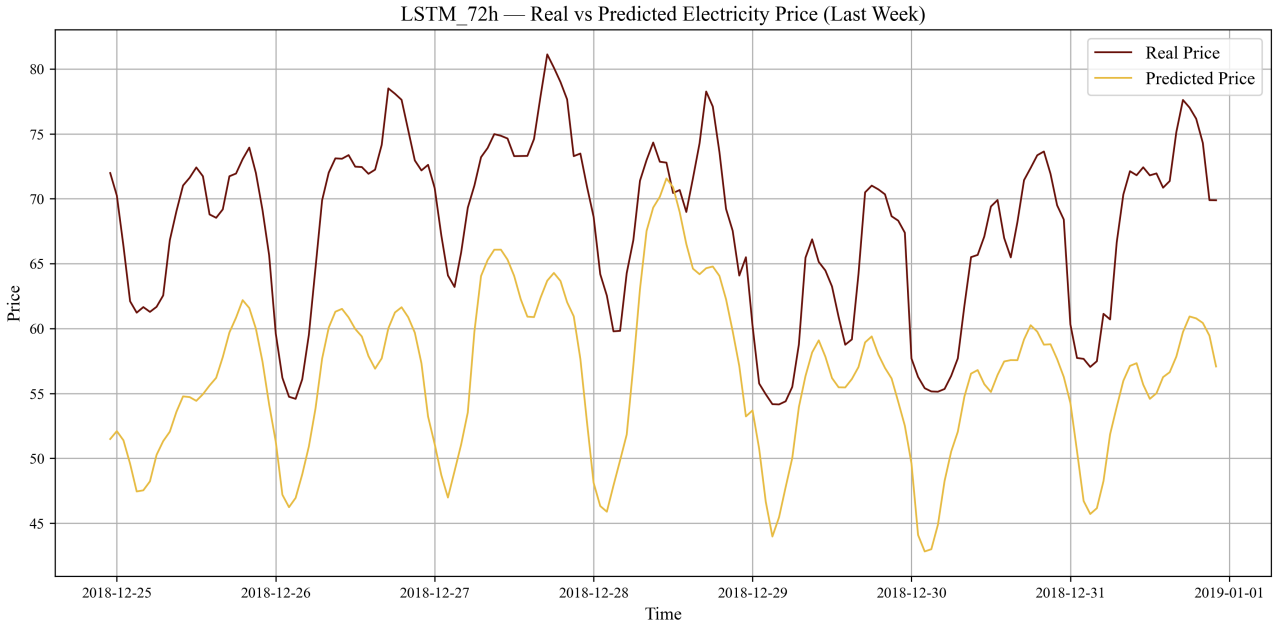

Figure V.1.4 – Real vs Predicted Electricity Price using XGBoost (Lag + Rolling), Final Week of the Test Set

V.1.3 LSTM + GRU + Attention

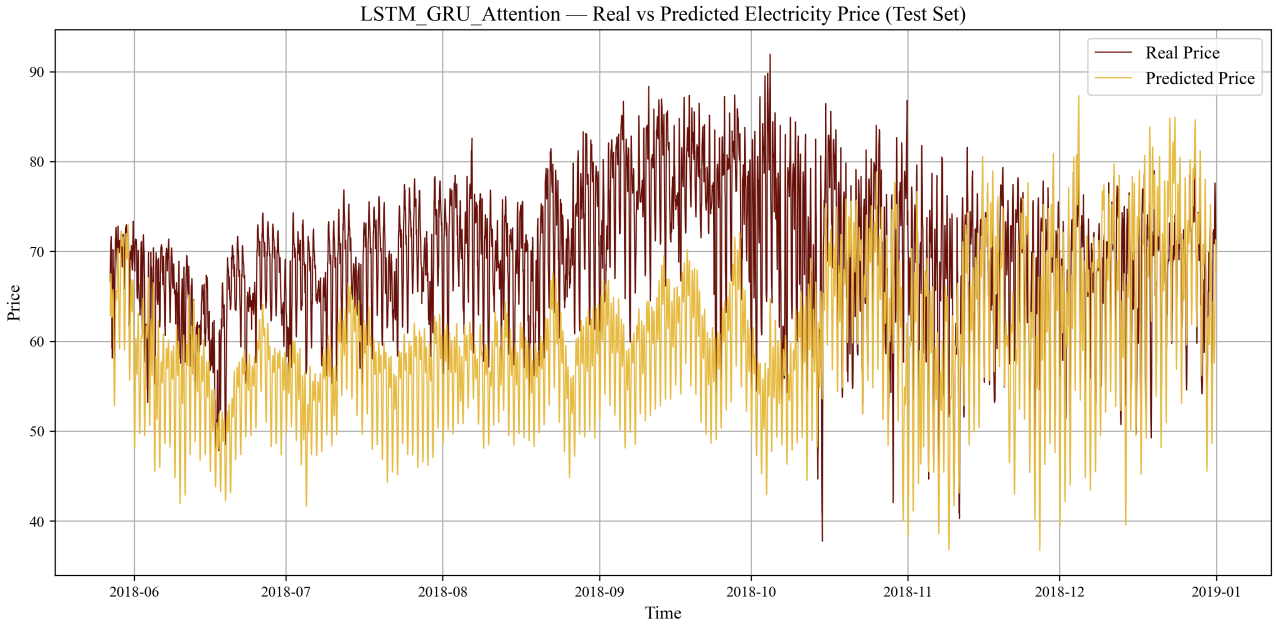

Figure V.1.5 – Real vs Predicted Electricity Price using LSTM + GRU + Attention, Test Set.

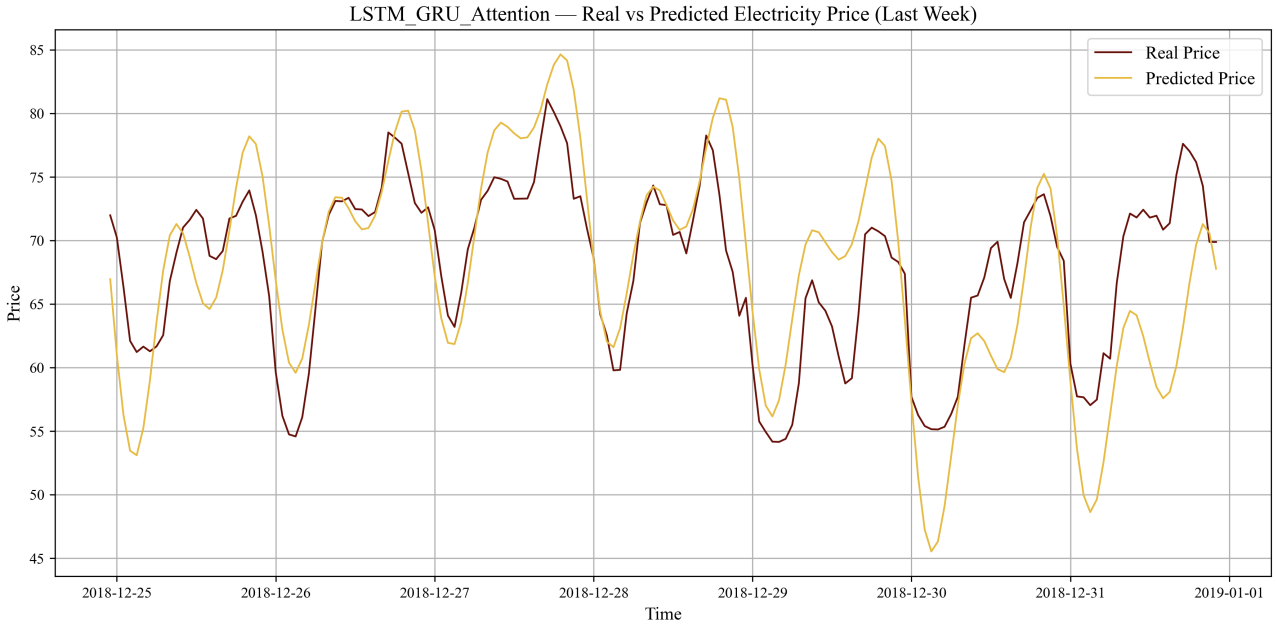

Figure V.1.6 – Real vs Predicted Electricity Price using LSTM + GRU + Attention, Final Week of the Test Set

Appendix B : Diagnostic Figures for No-Capital Trading Strategies

Appendix B provides the complete diagnostic visualisations for the no-capital trading strategies evaluated in Section III.4.2, namely the Directional, Mispricing, and Threshold strategies. These figures supplement the quantitative results presented in the main text by illustrating, in detail, the trade-by-trade dynamics underlying each method. The visualisations include hourly PnL, cumulative returns, distribution of profits and losses, and hourly–daily win-rate heatmaps. Together, they offer a qualitative understanding of how each strategy reacts to intraday price movements, the stability of its performance, and the nature of its risks. These graphical analyses provide additional insight into the behavioural differences between the strategies, helping to contextualise the comparative assessment discussed in Section III.4.2.

V.2.1 Directional Strategy — Diagnostic Visualisation

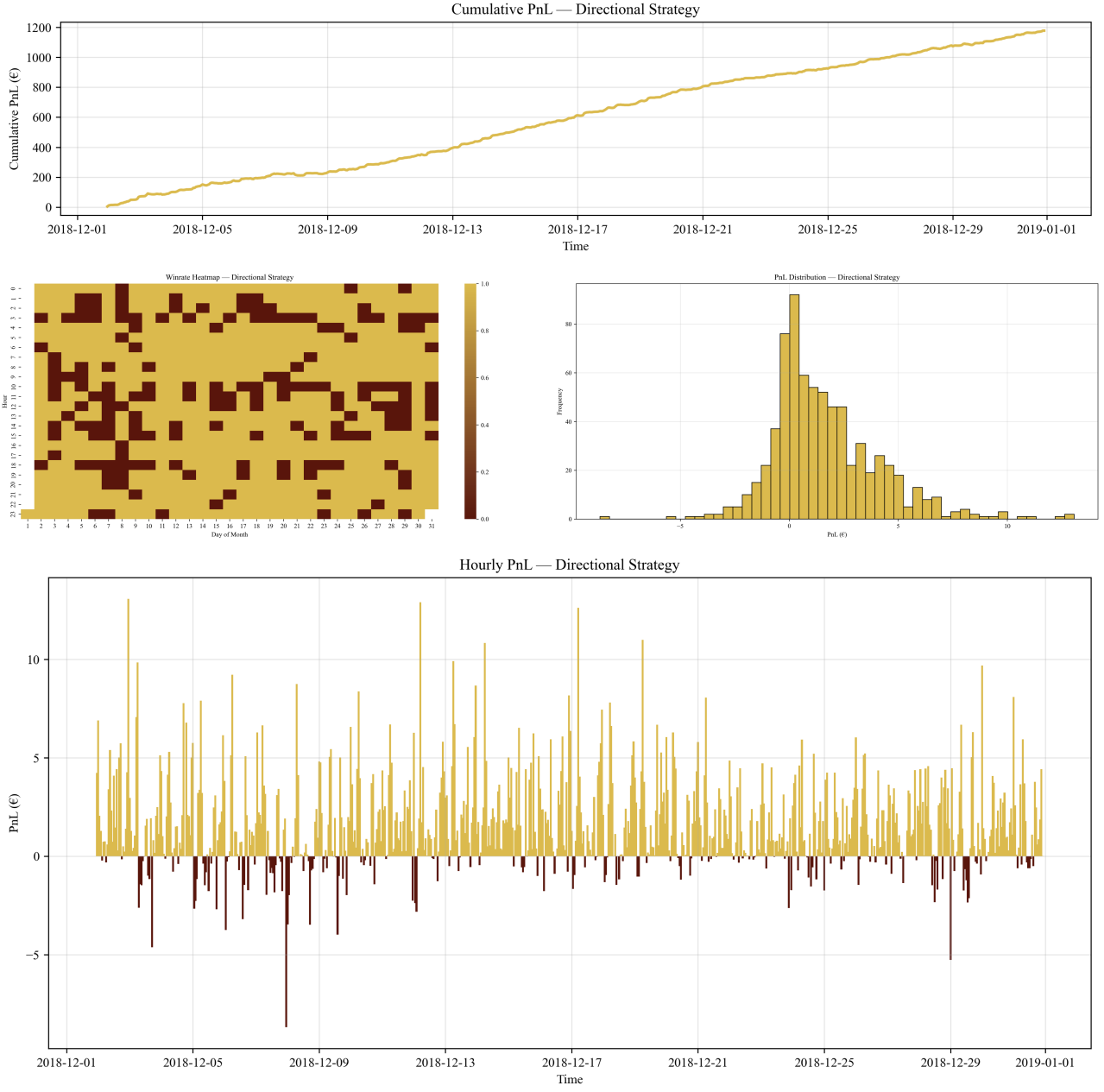

Figure V.2.1 – Diagnostic Visualisation of the Directional Strategy

V.2.2 Mispricing Strategy — Diagnostic Visualisation

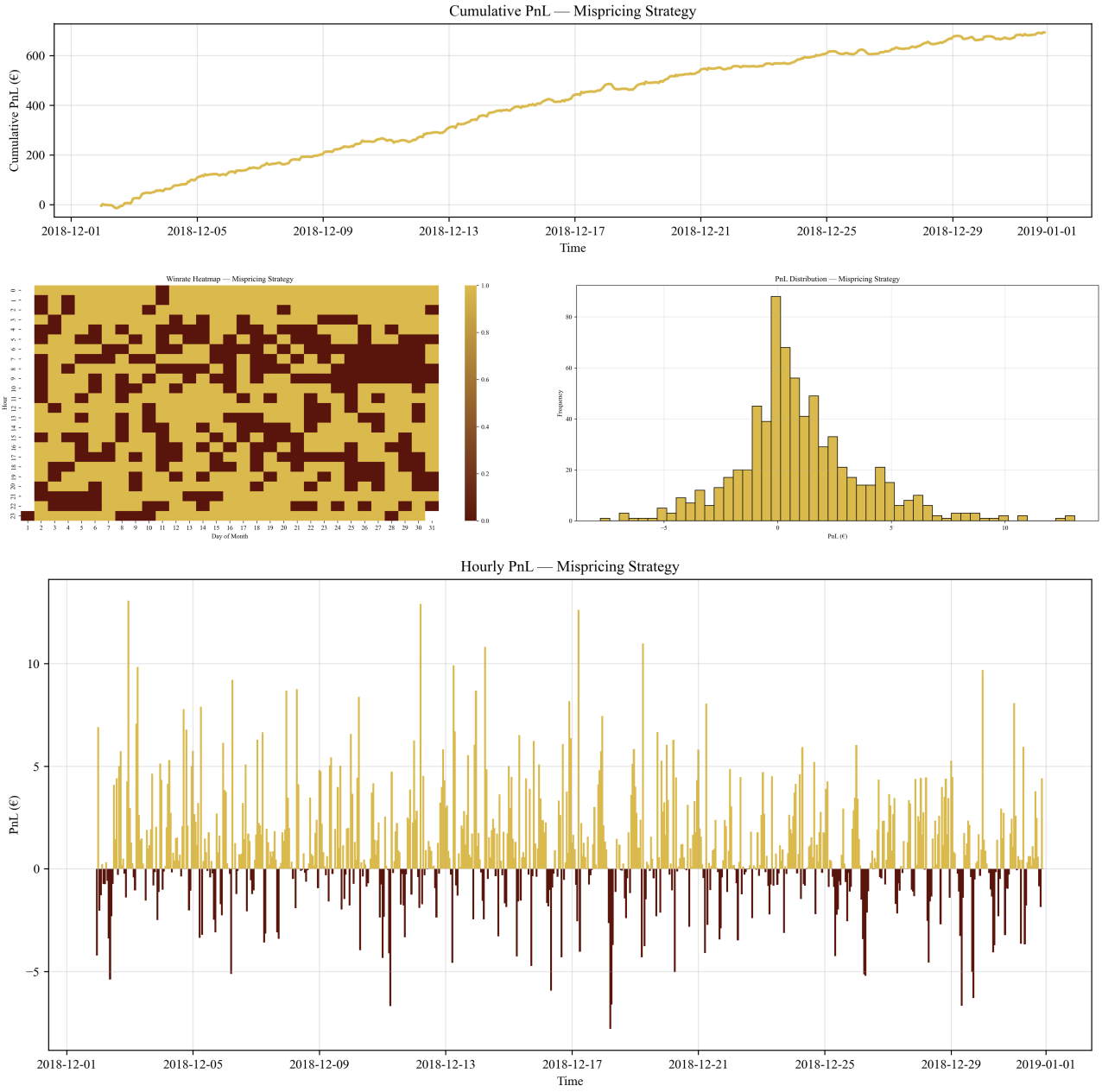

Figure V.2.2 – Diagnostic Visualisation of the Mispricing Strategy

Appendix C : Diagnostic Figures for Capital-Based Trading Strategies

Appendix C presents the full diagnostic visualisations for the capital-based trading strategies developed in Section III.4.3, including the Proportional Signal, Sigmoid Confidence, and Kelly Criterion approaches. These figures complement the numerical evaluation provided in the main text by displaying how capital evolves over time, how individual trades contribute to profit and loss, and how trading performance varies across hours and days. The combined visualisations (capital trajectory, per-trade PnL, PnL distribution, and win-rate heatmaps) allow for a deeper qualitative interpretation of each allocation mechanism's behaviour. They highlight the intrinsic trade-off between aggressiveness, robustness, and risk exposure, thereby enriching the comparative analysis carried out in Section III.4.3.

V.3.1 Proportional Signal Strategy — Diagnostic Visualisation

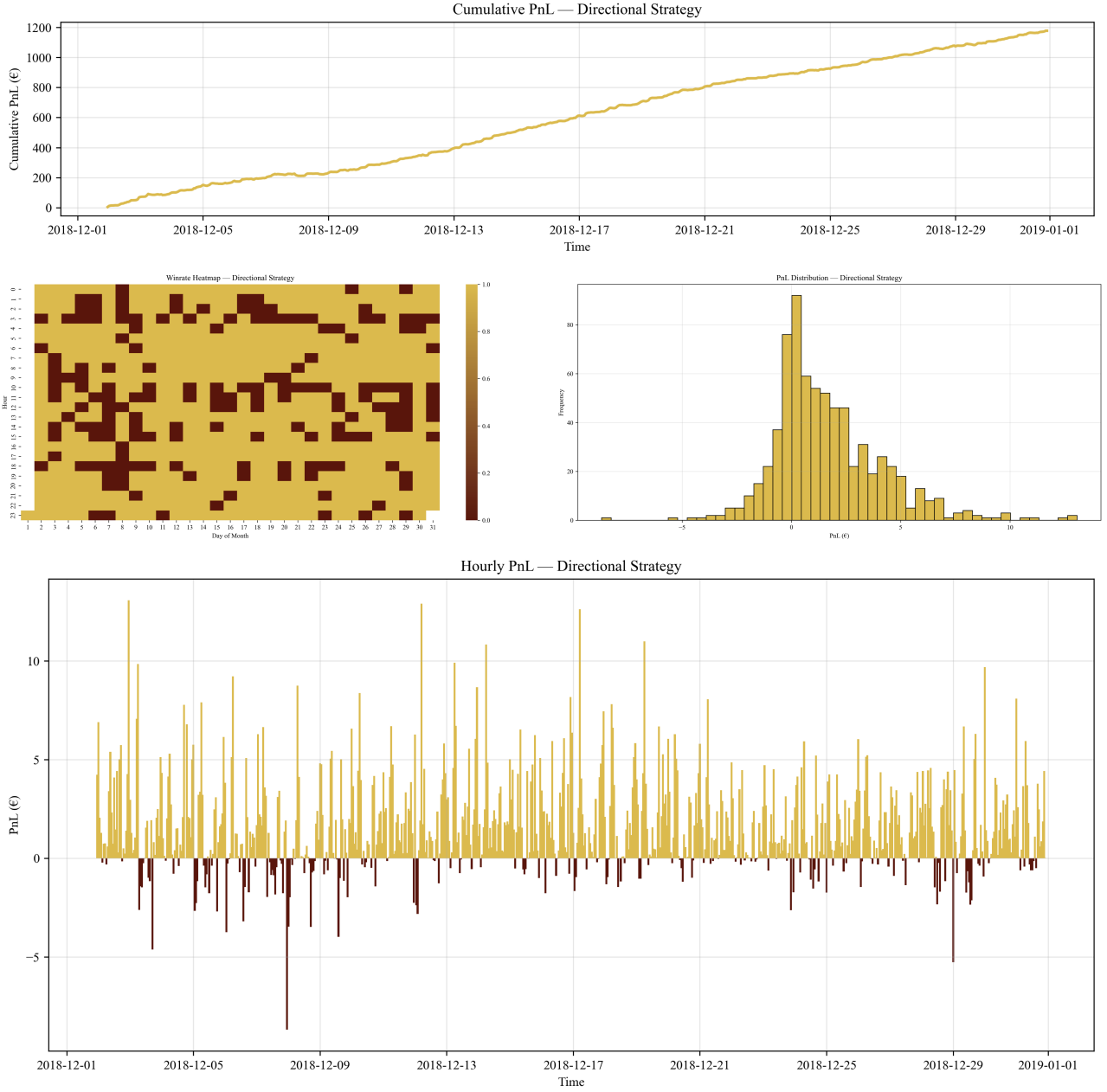

Figure V.3.1 – Diagnostic Visualisation of the Proportional Signal Strategy

V.3.2 Kelly Criterion Strategy — Diagnostic Visualisation

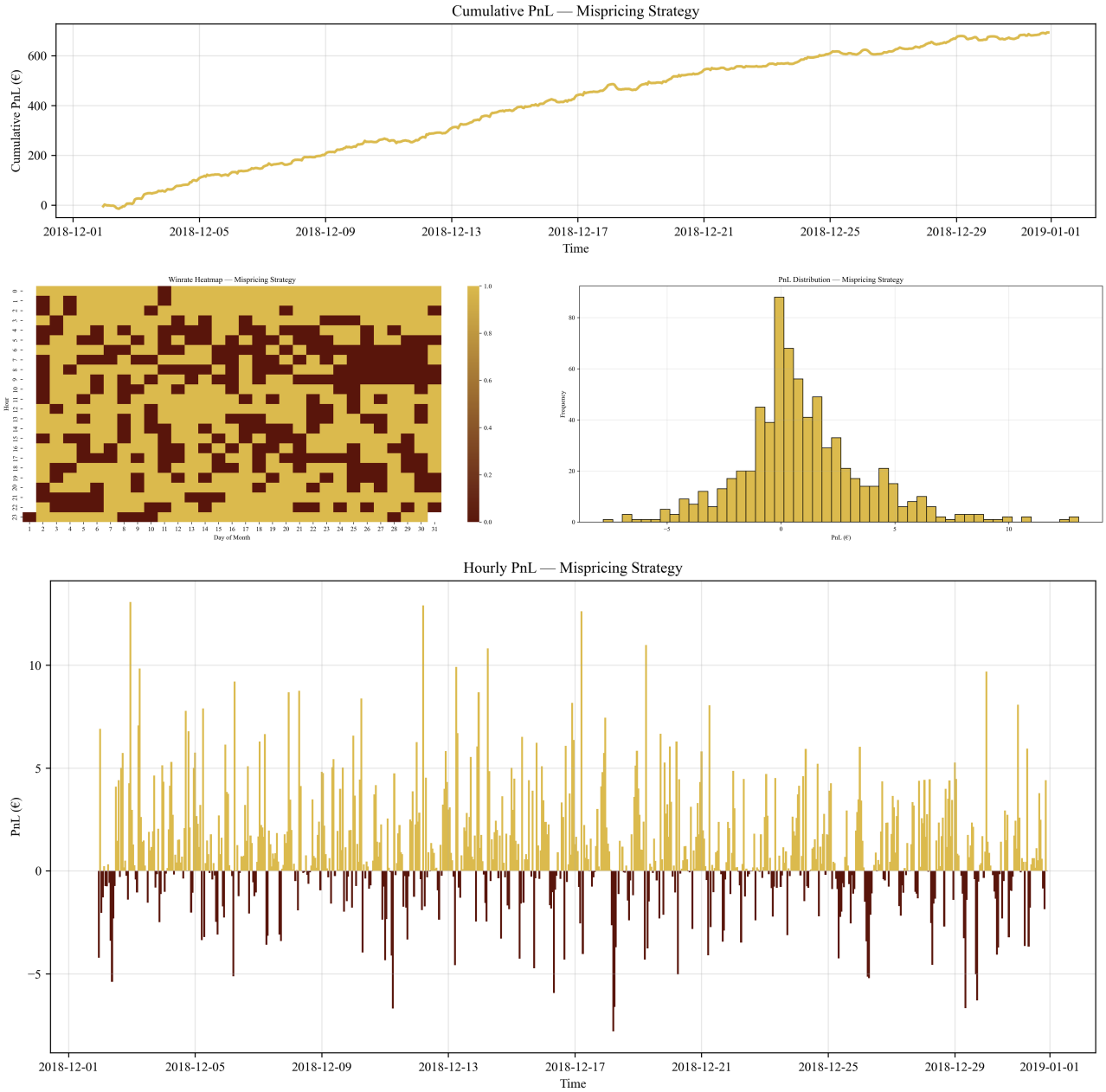

Figure V.3.2 – Diagnostic Visualisation of the Kelly Criterion Strategy

References

[1] J. Lago, F. De Ridder, B. De Schutter, *Forecasting Spot Electricity Prices: Deep Learning Approaches and Empirical Comparison*, Applied Energy, 2018.

[2] S. Hochreiter, J. Schmidhuber, *Long Short-Term Memory*, Neural Computation, 1997.

[3] T. Chen, C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, Proceedings of the 22nd ACM SIGKDD Conference, 2016.

[4] P. J. Weron, *Electricity Price Forecasting: A Review of the State-of-the-Art*, International Journal of Forecasting, 2014.

[5] B. Zhou et al., *Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting*, AAAI Conference on Artificial Intelligence, 2021.

[6] S. Makridakis, E. Spiliotis, V. Assimakopoulos, *The M4 Competition: 100,000 Time Series and 61 Forecasting Methods*, International Journal of Forecasting, 2020.

[7] A. Lago, N. De Somer, B. De Schutter, *Adaptive Load and Price Forecasting in Electricity Markets Using Hybrid Deep Learning Models*, Energy, 2019.

[8] G. Dudek, *Short-Term Load Forecasting Using Random Forests and Gradient Boosting*, Electric Power Systems Research, 2015.

[9] M. Cuturi, A. Doucet, *A Practical Introduction to Machine Learning for Trading*, Journal of Financial Data Science, 2020.

[10] B. Lim, S. Zohren, *Time Series Forecasting with Deep Learning: A Survey*, Philosophical Transactions of the Royal Society A, 2021.