Energy Strategy Reviews 54 (2024) 101436

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: www.elsevier.com/locate/esr

Electricity Price Forecasting in the Irish Balancing Market

Ciaran O'Connor [a], Joseph Collins [b], Steven Prestwich [c], Andrea Visentin [c,*]

[a] SFI CRT in Artificial Intelligence, School of Computer Science & IT, University College Cork, Ireland
[b] School of Mathematical Sciences, University College Cork, Ireland
[c] Insight Centre for Data Analytics, School of Computer Science & IT, University College Cork, Ireland

ARTICLE INFO

Dataset link: https://github.com/ciaranoc123/Balance-Market-Forecast

Keywords:
Day-ahead market
Balance market
Electricity Price Forecasting
Machine learning
Deep learning

ABSTRACT

Short-term electricity markets are becoming more relevant due to less-predictable renewable energy sources, attracting considerable attention from the industry. The balancing market is the closest to real-time and the most volatile among them. Its price forecasting literature is limited, inconsistent and outdated, with few deep learning attempts and no public dataset. This work applies to the Irish balancing market a variety of price prediction techniques proven successful in the widely studied day-ahead market. We compare statistical, machine learning, and deep learning models using a framework that investigates the impact of different training sizes. The framework defines hyperparameters and calibration settings; the dataset and models are made public to ensure reproducibility and to be used as benchmarks for future works. An extensive numerical study shows that well-performing models in the day-ahead market do not perform well in the balancing one, highlighting that these markets are fundamentally different constructs. The best model is LEAR, a statistical approach based on LASSO, achieving a mean absolute error of 32.82 €/MWh, surpassing more complex and computationally demanding approaches with errors ranging from 33.71 €/MWh to 44.55 €/MWh.

1. Introduction

Accurate price forecasts are challenging given the characteristics of short-term electricity market prices, i.e. high volatility, sharp price spikes, and seasonal demand. The continuing deployment of renewables and battery energy storage systems is likely to lead to increased price volatility [1,2].

The *Balancing Market* (BM) is the last stage for trading electric energy, exhibiting far higher volatility compared to both the *Day-Ahead Market* (DAM) and *Intra Day Market* (IDM). It plays an essential role (in particular in regions where storage of large quantities of electric energy is not economically convenient [3]) as production and consumption levels must match during the operation of electric power systems. The growing importance of accurate forecasts of BM prices to participants is outlined in Ortner and Totschnig [4], where forecast errors of variable renewable electricity will drive demand for BM participation.

Historically, the focus on the DAM is intuitive, given that it is a cornerstone of the European electricity market. In addition, the datasets required for forecasting the DAM are widely available. The lack of analysis of the BM is likely the result of a combination of factors including not all jurisdictions having a BM, the rules governing it can differ from region to region and the identification and acquisition of the relevant datasets can be complicated and expensive (with no open access dataset).

In recent years, given access to additional datasets and increasing GPU speeds, the application of *Deep Learning* (DL) models has become an attractive option. Short-term *Electricity Price Forecasting* (EPF), in particular, has seen an increasing number of publications pertaining to the DAM, with interest moving away from statistical approaches and towards *Machine Learning* (ML) approaches, [5–7]. While DL techniques have gained prominence in recent DAM EPF studies, their evaluation in the context of BM forecasting remains limited in the existing literature. Hence, our emphasis is on creating a benchmark using high-performing models from DAM EPF research to assess the performance of DL methods in the considerably more volatile BM setting.

1.1. Motivation and contributions

Accurate price forecasting is of particular interest to generators, buyers, and energy traders, including quick-response participants like battery energy storage systems that mitigate real-time supply–demand volatility. Non-physical financial traders, such as Net Imbalance Volume chasers [8], also capitalize on price spreads between ex-ante

* Corresponding author.
E-mail addresses: 119226305@umail.ucc.ie (C. O'Connor), 98718584@umail.ucc.ie (J. Collins), s.prestwich@cs.ucc.ie (S. Prestwich), andrea.visentin@ucc.ie (A. Visentin).

https://doi.org/10.1016/j.esr.2024.101436
Received 3 October 2023; Received in revised form 9 May 2024; Accepted 31 May 2024
Available online 12 June 2024

markets and the BM. Recent publications advocate best practices, promoting open-access DAM-related datasets and models [9,10]. Building on this, our paper extends the approach to the BM, presenting an analogous framework. Our contribution is to:

- Define a replicable framework to train and evaluate forecasting models on the BM. This includes clear instructions on the hyper-parameter optimization for each model, including providing the accompanying dataset. The data and code used for this work are publicly available.
- Benchmark the performance of prominent statistical, ML, and DL models, originally designed for the DAM EPF domain, when applied to the BM context.
- Compare model performance between BM and DAM price forecasting, offering insights into handling volatile price movements.
- We explore the unique difficulties that DL models encounter within the significantly more volatile BM. Our investigation reveals that DL models struggle with complex price fluctuations in the BM, resulting in elevated prediction errors.
- We conduct a comprehensive analysis to examine how changing the training data size impacts forecast accuracy across various predictive models.

The rest of this paper is structured as follows. In Section 2, we refer to some of the more recent & salient DAM and BM EPF publications. Section 3 provides an overview of the structure of a typical European electricity market and describes the peculiarities of the Irish BM. In Section 4, we present the specifics of the forecasting problem and the associated analytical framework. Section 5 describes the models utilized in the paper. The experimental results and their analysis are presented in Section 6. Finally, Section 7 concludes the paper with observations and future works.

2. Literature review

We highlight some of the more recent and salient short-term EPF publications, whether they relate to the DAM, IDM or the BM. In doing so, we broadly classify the publications as being either statistical, ML or DL approaches, with particular attention to DL ones, given their recent prevalence. In this work, we consider ML approaches all the AI techniques that do not involve DL. A broader picture of the existing EPF literature can be found in Weron [5], Nowotarski and Weron [11].

2.1. Day ahead market forecasting

Literature regarding the forecasting of the DAM has received considerable attention in the past. Lago et al. [9] covered multiple markets, presenting an extensive range of statistical, ML, and DL models, particularly hybrid DL models. Findings indicated that *deep neural networks (DNNs)* exhibit superior performance, closely trailed by hybrid, RNN, and ML models, with statistical models lagging behind. The authors suggested that linear statistical models or their variants are likely inadequate in DAM & BM EPF due to potential non-linearities in explanatory datasets (e.g., supply–demand curves in short-term markets with abrupt jumps & transitions). RNNs have received much attention in time series forecasting due to their ability to hold/incorporate relevant information from past inputs when generating forecasts, e.g. Anbazhagan and Kumarappan [12], Chen et al. [13]. LSTMs and GRUs, in particular, operate differently from standard RNNs in their ability to forget or persist certain information across time steps. For example, Ugurlu et al. [14] compared DNN and statistical methods for electricity price forecasting, highlighting GRU models' superiority, particularly in Turkish DAM predictions. Lagged price values and exogenous variables improved GRUs' accuracy consistently across months. The study emphasized data availability and uncertainty estimation. Notably, three-layer GRUs outperformed other models. ANNs and LSTMs displayed high accuracy compared as well. In Lago et al. [10], the *LASSO Estimated AutoRegressive (LEAR)* model stood out due to its low computational complexity and ease of implementation, alongside its accuracy. Results highlighted that Ensembles of LEAR models and ensembles of DNN models produced the most precise forecasts, outperforming any of the individual models. Their research underscored the considerably higher computational cost of the DNN model compared to LEAR, despite only marginal improvement in overall forecast accuracy. The paper not only outlined best practices and reference datasets for DAM EPF but also offers a LEAR implementation, transparent codebase, and a DNN application in the DAM context. Li and Becker [15] explored LSTM-based forecasting of Nordic DAM, emphasizing feature selection and cross-market effects on price prediction. The study examined hybrid LSTM architectures for EPF, highlighting the impact of different feature selection methods on model performance. The evaluation involved statistical measures and Diebold–Mariano tests, revealing the significance of selection methods, superiority of specific autoencoders, and the value of cross-border market features for precise predictions. The research underscored EPF's importance for power markets and policy implications for European cross-border trading enhancement. In a recent work, [16], study the influence of macroeconomic factors on DAM prices using ML techniques. They propose a forecasting method that adapts training intervals based on price volatility and combines multiple ML models to achieve improved accuracy, particularly during turbulent periods like the COVID-19 pandemic and geopolitical conflicts. Their approach enhances decision-making in electricity markets by effectively handling random events.

2.2. Intraday market

Research on IDM is crucial for understanding trading dynamics, given its similarities to the BM. This area focuses on EPF, trading strategy optimization, and market efficiency, especially in relation to renewable energy integration. The emphasis is on accurate forecasting and effective bidding strategies to enhance market performance. Monteiro et al. [17] introduces novel intraday session models designed for hourly price forecasts in the Iberian electricity market. These models are analyzed using various combinations of input variables, aiming to identify the most effective models for forecasting IDM electricity prices. The study compares errors from different models to determine the key variables for accurate forecasting, providing insights into short-term price forecasting, which could benefit market agents and the electric energy industry. Shinde and Amelin [18] conduct a detailed literature review on intraday electricity markets, focusing on trading dynamics and price forecasting in the context of renewable energy integration. They emphasize the need for comprehensive research on bidding strategies, market structures, and price accuracy. The review identifies gaps and suggests future directions, including simulation modeling and global market comparisons, especially with the evolving European Cross-Border Intraday Market. Narajewski and Ziel [19] introduces a novel approach to forecasting hourly intraday electricity prices in the German Intraday Continuous Market. By simulating price trajectories using a mixture model of Dirac and Student's t-distributions, the study enhances intraday trading efficiency. The model integrates autoregressive effects, load, wind, and solar generation forecasts, capturing market complexities. Results demonstrate superior performance over benchmarks, providing insights into market behavior and the impact of market integration initiatives. The methodology extends beyond the German market, advancing intraday trading strategies and dispatch management. Birkeland and AlSkaif [20] conduct a systematic literature review on European IDM research, analyzing 132 primary studies. They categorize research by area, methodology, dataset context, and date, identifying six major groups: Bidding, Market Modeling, Price Forecasting, Market Design, Forecast Errors, and Market Abuse. Their review suggests using closer-to-live datasets, exploring market abuse, and expanding national market studies beyond Germany. This work aims to guide future research and benefit academics, industry, and regulators by organizing existing knowledge and identifying research gaps.

2.3. Balancing market

The BM within the European electricity market framework serves a critical role in ensuring grid stability and reliability by correcting supply–demand imbalances in real time. While considerable progress has been made in understanding and optimizing BM operations, challenges persist in accurately modeling and forecasting BM prices due to the unique characteristics and complexities of this market segment.

Balancing market: Market overview

Zachmann et al. [21] outlines the ongoing efforts in Europe to reform electricity market designs, emphasizing the need for efficient and resilient market mechanisms amidst increasing renewable energy integration. The recommendations underscore the importance of maintaining short-term price incentives for demand response and avoiding excessive intervention in flexibility incentives. Understanding these market design considerations is crucial for contextualizing the challenges faced in modeling BM prices. European electricity balancing markets have been a subject of extensive research, with studies such as [22] focusing on the design challenges and trade-offs involved in maintaining grid balance. The complexity of balancing market design variables and their interlinkages with broader market dynamics underscores the need for nuanced modeling approaches tailored to the BM context. Ortner and Totschnig [4] explores the evolving role of BMs in Europe amidst increasing renewable energy penetration. While BMs are expected to experience growing demand for balancing energy, the monetary volume of BMs remains relatively small compared to DAMs. This highlights the importance of understanding the drivers and dynamics unique to BMs to improve forecasting accuracy. Efforts to integrate European electricity balancing markets, as discussed in Roumkos et al. [23], aim to enhance market efficiency and competition across Member States. However, challenges remain in harmonizing national rules and integrating different types of balancing products. Understanding the implications of these integration efforts is crucial for developing robust forecasting models that account for cross-market dynamics and regulatory complexities. Rosales-Asensio et al. [24] investigates the challenges and opportunities associated with integrating distributed energy resources (DERs) within the BM framework. The study emphasizes the need for flexible and responsive market structures to accommodate the evolving energy landscape, highlighting the importance of considering technological advancements and regulatory frameworks in BM modeling. Research such as Eicke et al. [25], Van Der Veen et al. [26] delves into the behavioral aspects of market participants in response to BM incentives and pricing mechanisms. Understanding how market participants interact with BMs and the impact of regulatory design on market outcomes is essential for developing accurate forecasting models that capture the dynamics of supply–demand interactions in real time. Poplavskaya et al. [27] examines the impact of regulatory changes on the efficiency of European electricity balancing markets using agent-based models. The findings underscore the importance of market design in mitigating strategic behavior among market participants, highlighting the need for continued regulatory efforts to enhance market efficiency and competition. In summary, while significant progress has been made in understanding the complexities of BM operations and market dynamics, challenges persist in accurately modeling and forecasting BM prices. By integrating insights from the literature on BM design, market integration efforts, behavioral economics, and regulatory frameworks, researchers can develop more robust forecasting models that capture the nuances of BM price dynamics and contribute to the efficient operation of electricity markets.

Balancing market: Electricity price forecasting

The study by Klæboe et al. [28] compared models for predicting electricity market behavior, focusing on state-aware models like SARMA and ARM versus purely time series-based models like ARMA and ARX. It found that incorporating contextual information improved forecasting accuracy, with SARMA excelling in short-term predictions and ARM performing better in day-ahead forecasts. Its forecasts of the BM were limited to 1 h ahead. In a Belgium electricity market setting, Dumas et al. [29] introduced a new probabilistic method for predicting imbalance prices in the energy market. This approach used historical data to estimate net regulation volume transition probabilities, which are then used to forecast imbalance prices based on reserve activation and marginal prices. The method outperformed other techniques like Gaussian Processes and Multi-Layer Perceptron in terms of probabilistic error measures. Further improvements and integration into bidding strategies are suggested for future work. In Bunn et al. [30], they examine the predictability of BM prices in the British BM employing a regime-switching model to capture non-linear dynamics. By analyzing fundamental drivers such as wind and solar forecast errors, scarcity indicators, and lagged prices, the study demonstrates that BM prices exhibit predictable behavior contrary to efficiency conjectures. Results indicate that regime-switching models outperform linear benchmarks, suggesting market participants can benefit from forecasting and trading strategies informed by fundamental econometric relationships. This research contributes valuable insights into price formation dynamics in mature and liquid electricity markets. Lucas et al. [31] employed *Gradient Boosting*, *Random Forest (RF)* and *Extreme Gradient Boosting (XGB)* for BM price forecasting in Great Britain, achieving accurate peak predictions but with limitations in predicting price bottoms, with XGB coming out on top. Notably, this study lacked a comparison of deep learning approaches. Recently, Narajewski [32] focused on predicting extreme price spikes in a volatile electricity market using short-term forecasting of BM prices. Various models have been explored: Naive, LASSO, Gamlss, and Probabilistic Neural Networks. Evaluation metrics include CRPS, MAE, and empirical coverage. Findings showed Naive and Gamlss models as strong performers, with Naive having the lowest errors, albeit with poor empirical coverage. Combining Naive and Gamlss slightly enhanced coverage. They found that market volatility has a negative impact on LASSO and normal distribution-based models' effectiveness.

3. Electricity market structure

European electricity markets differ from country to country; however, recent years have seen a significant degree of convergence in the structure and operation of short-term electricity markets with several jurisdictions following a similar template [33], driven by the growing amount of variable renewable generation [1]. Therefore, we explain the configuration of a standard European market using the example of the *Irish-single electricity market* (I-SEM).

I-SEM serves as the wholesale electricity market for both Ireland and Northern Ireland. Forward energy markets span approximately 4 years to one month prior to delivery, while *spot* markets enable adjustments nearer to delivery. These spot markets include:

- *Day-Ahead Market* (DAM): The DAM consists of one pan-European auction at noon CET for 24 h of the next day.
- *Intra-Day Market* (IDM): After DAM clearance, additional trading opportunities include three scheduled auctions (IDA1/IDA2/IDA3) and a continuous market.
- *Balancing Market* (BM): The price is determined by averaging energy balancing instructions issued by the system operator for a 5-minute imbalance pricing period in I-SEM, applied over a 30-minute settlement period.

The I-SEM also includes *capacity* and *ancillary service* markets, similar to other jurisdictions (see Fig. 1). However, these markets are beyond the scope of this paper.

The primary purpose of the BM market is to enable the system operator to perform balancing actions (i.e. matching demand and supply close to or in real-time). Hence, the BM is an important component of short-term electricity markets in a number of jurisdictions. Some features of the balancing market include:

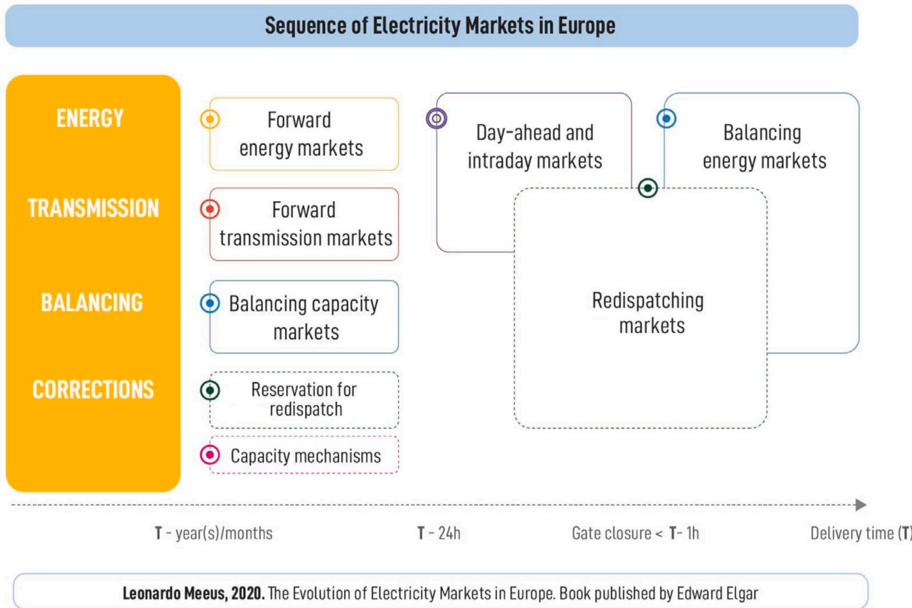


Fig. 1. Schematic overview of the typical sequence of existing electricity markets in the EU. Markets in dotted lines are optional [34].

- *Price Quantity Pairs*: These are the bids and offers that BM participants submit to the system operator ensuring that they are compensated for any instructed deviations to their physical notifications (e.g. their contracted positions in the ex-ante markets).
- *Flagging and Tagging*: The process by which the system operator excludes non-energy actions when calculating the BM price.
- *De Minimis Acceptance Threshold* (DMAT): Energy actions below the DMAT threshold are excluded to not unduly impact the BM price.

For a more in-depth discussion of the BM, we refer the reader to Bharatwaj and Downey [35].

4. Analytical framework

This section presents in detail the framework used. In particular, the following sub-section presents the data sources and preprocessing, Section 4.2 describes the dataset split and hyperparametrisation procedure, and finally, in Section 4.3, the metrics and the statistical test used are introduced.

4.1. Dataset

The data was sourced from the SEMO[1] and SEMOpx[2] websites, comprising historical and forward-looking data dating from 2019 to 2022. Our goal in this study is to predict the BM price for the next 16 open settlement periods. We choose this forecast horizon explicitly to align with the availability of the input explanatory variables.

Regarding the explanatory variables, they can be categorized into two types: **Historical data** and **Forward/future-looking data**. Historical data, like BM prices, utilizes the 48 most recent BM prices to predict the next 16. The dataset includes a rolling window of historical data, incorporating features like BM prices, BM volume, wind forecast deviations, interconnector flows (reflecting electricity transmission between regions or countries and affecting market prices based on supply–demand dynamics), and physical notifications (indicating changes in planned electricity generation/consumption, impacting market operations and pricing by adjusting supply–demand balance) from up to 48 time steps prior to the prediction target. In contrast, future-looking data includes forecast wind and demand (the forecast values relate to the BM settlement periods for which we are interested in predicting the price). It is possible for some explanatory variables, such as DAM prices, to appear in both a historical and future-looking context.

In our problem setting, the target is the BM price (BMP) for the proceeding, **open**, 16 settlement periods.

$$Y_t = [BMP_{t+2}, \dots, BMP_{t+17}]. \tag{1}$$

The forecast horizon begins at $t+2$ instead of t, as at time t when a forecast is being generated, balancing market periods t and $t+1$ are closed. If a participant wants to adjust its BM commercial order data, it can only do so for BM periods $t+2$ onwards, hence our focus on the $[t+2, \dots, t+17]$ horizon.

The historical data considered is:

- *Balancing Market Prices* (BMP) in €/MWh: This holds BM prices from the most recent & available 24 h, i.e. $[BMP_{t-51}, \dots, BMP_{t-3}]$.
- *Balancing Market Volume* (BMV) in MWh: Again, the most recent 48 observations of BM volume. It is given by $[BMV_{t-51}, \dots, BMV_{t-3}]$.
- *Forecast Wind - Actual Wind* (WDiff) in MWh: This is the difference between forecast & actual wind data for the most recent 48 available settlement periods. $[WDiff_{t-51}, \dots, WDiff_{t-3}]$.
- *Interconnector Values* (I) in MW: Interconnector flows from the previous 24 h. $[I_{t-50}, \dots, I_{t-2}]$.
- *DAM prices* (DAM) in €/MWh: DAM prices from the previous 24 h. DAM prices are published at an hourly granularity; as a result, we use the hourly price that is available for each half-hour settlement period. $[DAM_{t-48}, \dots, DAM_t]$.

The future-looking data provided by the system is as follows:

- *Physical Notifications Volume* (PHPN) in MWh: Sum of physical notifications for forecast horizon. That is $[PHPN_{t+2}, \dots, PHPN_{t+17}]$.
- *Net Interconnector Schedule* (PHI) in MWh: Interconnector schedule for the forecast horizon $[PHI_{t+2}, \dots, PHI_{t+17}]$.

[1] https://www.sem-o.com/
[2] https://www.semopx.com/market-data/

- *Renewable forecast* (PHFW) in MWh: Transmission system operator renewables forecast (for non-dispatchable renewables) for the forecast horizon. $[PHFW_{t+2}, \dots, PHFW_{t+17}]$.
- *Demand forecast* (PHFD) in MWh: Transmission system operator demand forecast for the forecast horizon. $[PHFD_{t+2}, \dots, PHFD_{t+17}]$.
- *DAM prices* (DAM) in €/MWh: DAM prices for the next 8 h. $[DAM_{t+1}, \dots, DAM_{t+16}]$.

For the historical data, the different time indices/intervals are solely due to data availability (i.e. the most recent and available 48 observations depend on the data source). We remove daylight savings days, and we impute missing variables with the mean of the previous 30 days or with the most recent data point, depending on the context.

4.2. Experimental configuration

In our analysis, the approach involves three main steps:

1. Hyperparameter tuning
2. Model training
3. Model forecast (on unseen data)

For each encountered model, we fit it using training datasets spanning 30, 60, 90, and 365 days. In model fitting and forecasting, we employ a walk-forward validation approach, which is a frequently used technique in short-term price forecasting [36]. In our configuration, the model is retrained every 8 h to incorporate the most recent data available. The hyperparameters for ML or DL models are determined using the performance of the validation sets. We used 25% of the training set as validation, capped at a maximum of 30 days for the 365 days dataset.

To accommodate diverse hyperparameter configurations for validation and training sets, we repeat step 1 (hyperparameter tuning) every quarter for both DL and ML. The process is as follows:

1. Given a model of interest and a starting time point t, obtain a validation set and determine hyperparameter settings.
2. Using hyperparameter settings from step 1, fit the model with training data.
3. Utilize the fitted model from step 2 to forecast using the test set.
4. Return to step 1, set $t = t + 90$ days, and repeat the process.

When performing the hyperparameter optimization step, we use various libraries/toolboxes. For ML models such as RF, XGB, and SVR we utilized the *Scikit-learn* library [37]; for XGB type models we utilized the *XGBoost* library [38]. For the DL models (SH DNN and MH DNN) we used *Tensorflow* [39] in conjunction with Talos [40]. For statistical model LEAR, we used the *epftoolbox* [10]. For ML and DNN hyperparameter searches, a broad range of values was tested to find optimal settings for each time period and training size.

4.3. Metrics

Commonly encountered metrics in the EPF literature used to evaluate forecast accuracy include Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics [41], as well as the symmetric mean absolute percentage error (sMAPE) metric [42]. sMAPE is preferred over MAPE due to its effectiveness in handling the frequent occurrence of prices near zero in the BM. It is defined as:

$$sMAPE = \frac{100}{n} \sum_{k=1}^{n} \frac{|y_k - \hat{y}_k|}{(|y_k| + |\hat{y}_k|)}. \quad (2)$$

When comparing performance metrics of models, the Diebold–Mariano (DM) test is commonly used in EPF literature to quantify statistical distinctions [43]. The DM test assesses null hypotheses, rejecting H_0, which indicates that a model has a statistically significant accuracy improvement over another.

5. Models

This section provides an overview of the models utilized for BM forecasting. Due to the limited exploration of BM forecasting, we decided to benchmark a wide variety of models selected from statistical, machine learning and deep learning approaches. Models were chosen based on their performance in recent EPF literature, encompassing both DAM and BM domains. Building upon studies such as [9,15,31] for DAM EPF and [28,31,32] for BM EPF, we identified consistent performers like Lasso/LEAR, XGB, and RF across both markets. Moreover, DNNs showed promising results specifically in the DAM setting. Leveraging these insights, our selection aimed to encompass models proven effective in diverse market conditions, ensuring robustness and relevance in our comparative analysis. For a more interested reader, the peculiarities of the models' implementation and hyperparameters range can be found in the paper's GitHub repository.[3] This contains all models and data used in this study. A high-level categorization of the models is the following:

- *Naive*: Utilizes the preceding 8 h of BM prices as a forecast for the subsequent 8 h, capturing immediate pricing dynamics crucial in the real-time BM environment. This approach outperformed alternatives like considering the same hour of the previous day or the same weekday, which may overlook such dynamics. It shares similarities with methods in Beigaite et al. [44], which employed the DAM values from the same days a year prior as a seasonal naive DAM forecast. This represents our baseline model.
- *Autoregressive Integrated Moving Average* (ARIMA): ARIMA is a time series forecasting method that models the relationship between a series of observations and the time lagged versions of itself. It incorporates three key components: Autoregression (AR), Differencing (I), and Moving Average (MA). The ARIMA model is specified by three parameters: p, d, and q, representing the number of autoregressive terms, the degree of differencing, and the number of moving average terms, respectively. We employed ARIMA to capture linear trends present in the data.
- *LASSO Estimated AR* (LEAR): It is a modified autoregressive time series approach that integrates LASSO regularization for enhanced performance and feature selection. In linear regression, the response variable Y_t at time point t is formed through a linear combination of n predictors. Variations of this model involve introducing a regularization term, such as the *Least Absolute Shrinkage and Selection Operator* (LASSO) [45], or its extension, the elastic net [46]. We adapted the implementation designed in Lago et al. [10] to the BM structure of the input/output data. Similar to [10], we performed daily hyperparameter tuning for LEAR, estimating λ with the LARS method and in-sample AIC. The optimal λ from LARS was then used to re-calibrate LEAR through traditional coordinate descent. LEAR is particularly beneficial for BM EPF, as it addresses the challenge of selecting relevant predictors in a high-dimensional and volatile market environment.
- *K-Nearest Neighbors* (KNN): KNN stands out as a non-parametric instance-based learning method renowned for its simplicity and effectiveness, particularly in scenarios where capturing intricate patterns within data is paramount. By leveraging the proximity of similar instances in the training set, KNN predicts an output by averaging the values of its "K" nearest neighbors. Its versatility in handling various distance metrics, such as Euclidean or Manhattan distance, enabling it to discern complex relationships within the data.

[3] https://github.com/ciaranoc123/Balance-Market-Forecast

- *Support Vector Regression* (SVR): It performs a nonlinear mapping of the data to a higher-dimensional space where linear functions are used to perform regression. SVR performs well in high-dimensionality space [47]. In our case, the radial basis function kernel was chosen every time.
- *Random Forest* (RF): It is an ensemble model that combines several regression trees to generate predictions. It is based on the principle of bagging — short for bootstrap aggregating, splitting the data into several subsets and choosing subsets randomly with replacement, thus combining individual models with low bias and high variance, which, when aggregated, reduce the variance and retain low bias.
- *Extreme Gradient Boosting* (XGB): Similarly to the RF, XGB combines several regression trees to generate forecasts, but it is based on the principle of boosting — where learners are learned sequentially, combining models with high bias and low variance to reduce the bias while keeping a low variance. This makes it suitable for capturing complex relationships in BM price data.
- *Single-Headed DNN Model* (SH-DNN): A simple extension of the traditional MLP, the first DL model for predicting DAM prices is a DNN used in Lago et al. [9]. Its adaptation to the BM market is described in the following section. The chosen format for the single-headed DNN model is sequential, allowing for further hyperparameter optimization with a range for the number of layers to be added for each block (where a block is comprised of between 1 to 3 connected layers) and activation function. The first layer flattens the data so the input data, followed by several dense layers, including a single dropout layer after block 3. Their output is fed to a series of DNN blocks. The SH-DNN model is chosen for its ability to effectively capture intricate non-linear patterns present in the data.
- *Multi-Headed RNN DNN Model* (MH-RNN): A non-sequential network that incorporates historical and future-looking data using dense and recurrent layers, emphasizing its ability to capture both temporal dependencies and high-dimensional relationships in the BM data. This hybrid approach aims to incorporate recurrent layers (LSTM) for capturing sequential patterns in time series data, along with fully-connected layers (DNN) focused on the future-looking data, as in Lago et al. [9]. These branches have their output concatenated into one vector, and this vector is fed into a regular output dense layer containing 16 neurons for the prediction of 8 h (or 16-time stamps). For clarity on the outline of the network, see Fig. 2. The obtained optimal hyperparameters for the Multi-Headed RNN DNN are 3 layers for both the LSTM and DNN networks, making for an 8-layer deep network with the concatenated and final layer. A more detailed structure of the network configuration can be found in the shared repository.

These models cover different approaches to time series forecasting: a simple naive prediction, popular EPF statistical models (ARIMA and LEAR), ML techniques both instance-based (KNN) and model-based (XGB, RF and SVR) and two DL approaches. Benchmarking such a broad spectrum of methodologies allows us to investigate the peculiarities of the Irish BM market better.

6. Results

In this section, we present an in-depth analysis of the performance of the models presented above. The first part focuses on the accuracy of the prediction and compares the impact of different training horizons. The results are examined in relation to the performances on the DAM. The models are then compared to see if their differences are statistically significant. Later, we analyze how the prices and their predictions change during the day to understand which parts of the BM are harder to forecast. Finally, we analyze the models in terms of the computational effort required.

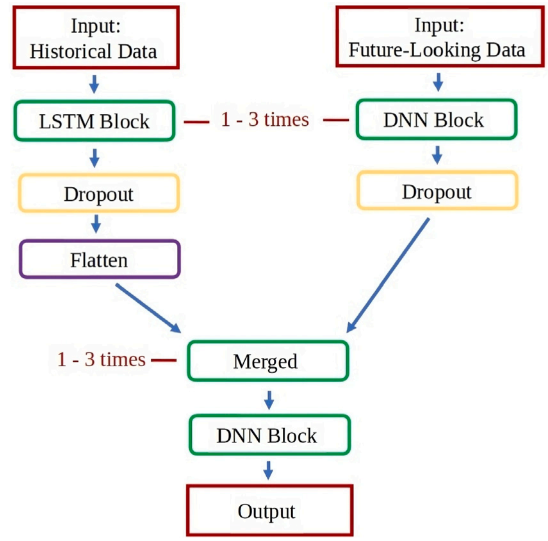


Fig. 2. MH RNN DNN Model.

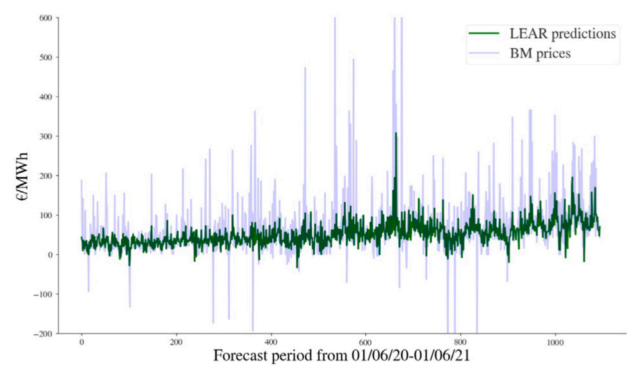


Fig. 3. LEAR Forecast for BM.

6.1. Accuracy

Table 1 offers a comprehensive overview of MAE, RMSE, and sMAPE errors for different models and training sizes. Comparisons are also made with DAM metrics for SH DNN and LEAR models over identical time frames, with noteworthy differences emerging between DAM and BM predictions.

LEAR emerges as the top-performing model for the BM, with limited exceptions for shorter training periods. Its subtle adjustments led to relatively smoother and lower variation in the predictions. Fig. 3 displays a comparison between the BM prices and LEAR's predictions. We can see that the model does not frequently attempt to predict significant spikes. This seems to be the reason for its good accuracy; the model returns forecasts that might not consistently predict the exact direction of BM prices but avoids producing highly inaccurate predictions. Notably, increasing the training set size had a significant impact on LEAR's performance. The MAE decreased by 3.31 when transitioning from a 30-day training period to 310 days.

ML models, including KNN, RF, XGB, and SVR displayed varied performance. XGB stood out by closely approaching the accuracy of LEAR and outperforming KNN, SVR, and RF. Both XGB and RF outperformed the DL models, while both KNN and SVR struggled to capture data patterns, with KNN failing to outperform ARIMA. These findings align with UK BM paper [31], confirming XGB's superiority over RF in a similar setting.

DL models exhibit inferior performance, consistently underperforming across all metrics, unsurprising considering their struggles

Table 1
Balancing Market prediction accuracy for the different training sizes. In bold, the best-performing model. MAE is measured in €/MWh.

Training size	MAE					RMSE					sMAPE				
	30 days	60 days	90 days	Max Data	DAM	30 days	60 days	90 days	Max Data	DAM	30 days	60 days	90 days	Max Data	DAM
Model															
Naive	51.96	51.96	51.96	51.96		84.27	84.27	84.27	84.27						
LEAR	35.95	**34.12**	**33.57**	**32.82**	10.15	**57.89**	**56.38**	**56.34**	56.53	18.84	72.14	**67.19**	**66.20**	**64.93**	22.68
ARIMA	43.15	42.72	42.44	41.51		68.54	67.61	67.20	66.00		76.26	75.82	75.50	74.24	
XGB	**35.35**	34.40	34.29	33.71		60.46	58.03	56.89	57.03		69.71	67.62	67.12	66.89	
RF	36.15	35.24	34.92	34.66		58.62	57.27	56.66	**56.22**		**68.67**	68.55	68.45	67.18	
SVR	41.43	38.01	37.70	37.39		69.81	66.12	65.58	64.92		78.35	69.85	68.97	68.09	
KNN	45.85	44.94	44.71	44.55		69.79	68.88	69.20	68.59		79.83	77.97	76.98	74.57	
SH DNN	40.14	38.67	38.31	37.41	**7.34**	64.13	68.91	62.25	61.95	**13.86**	77.85	72.10	72.05	70.81	**19.01**
MH RNN DNN	41.42	40.89	40.35	38.28		64.83	64.02	64.42	61.96		79.77	77.20	76.68	74.84	

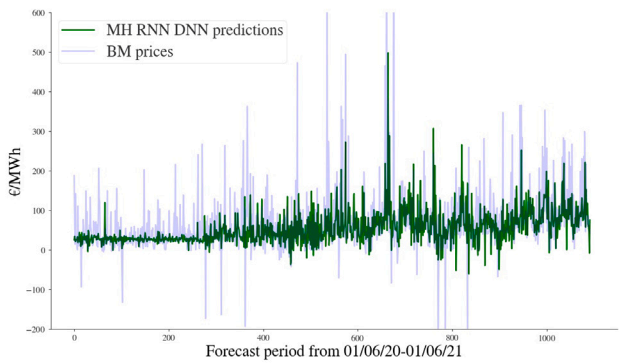


Fig. 4. MH RNN DNN Forecasts for BM.

Table 2
DB test statistical significance.

Model	DB Test - MAE								
	LEAR	XGB	RF	SVR	SH DNN	MH RNN DNN	ARIMA	KNN	Naive
LEAR		✗	✗	✓	✓	✓	✓	✓	✓
XGB			✗	✓	✓	✓	✓	✓	✓
RF				✓	✓	✓	✓	✓	✓
SVR					✗	✗	✓	✓	✓
SH DNN						✗	✓	✓	✓
MH RNN DNN							✓	✓	✓
ARIMA								✓	✓
KNN									✓
Naive									

in Dumas et al. [29], Narajewski [32]. While these models perform admirably in domains like the DAM, their effectiveness in predicting BM encounters challenges, including overfitting to recent price spikes and difficulties with complex BM dynamics. This divergence underscores the complexities inherent in electricity market forecasting. Fig. 4 shows the performances on part of the test set. Comparing them to LEAR's (Fig. 3), we can see that the DL model behaves over-confidently, predicting considerable spikes. If these predictions are incorrect, they are strongly penalized by the metrics compared to more cautious ones. The SH DNN model outperformed the MH RNN DNN model, showing improvements as the training set size increased (resulting in a 2.47 reduction in MAE). While the SH DNN model displayed better forecasting ability than XGB and LEAR during the least volatile period (first 3 months), its performance deteriorated as market volatility increased. Another possible reason behind the lower accuracy of the deep learning approaches is the limited amount of data. Their training would likely benefit from a bigger training set due to the high number of possible structural configurations and parameters. The MH RNN DNN model, which exhibited excellent results in the DAM [9], struggled to accurately forecast BM prices. This could be attributed to the current dataset's unsuitability for the split approach used, where data is fed into two separate branches. If past and future data heavily influence each other, both structures may fail to establish these relationships. In contrast, the SH DNN model avoids assumptions about input data, enabling more comprehensive relationship-building.

We investigate the impact of training data size on model performance. Having access to a higher amount of data strongly benefited all the models. Increasing the training data window to 365 days resulted in an average decrease of 2.39 MAE, 2.61 RMSE, and 5.12 sMAPE across models. Surprisingly, the DL models had a smaller benefit from the increased amount of data. The consistent enhancement of model performance highlights the importance of historical data.

6.1.1. Statistical test

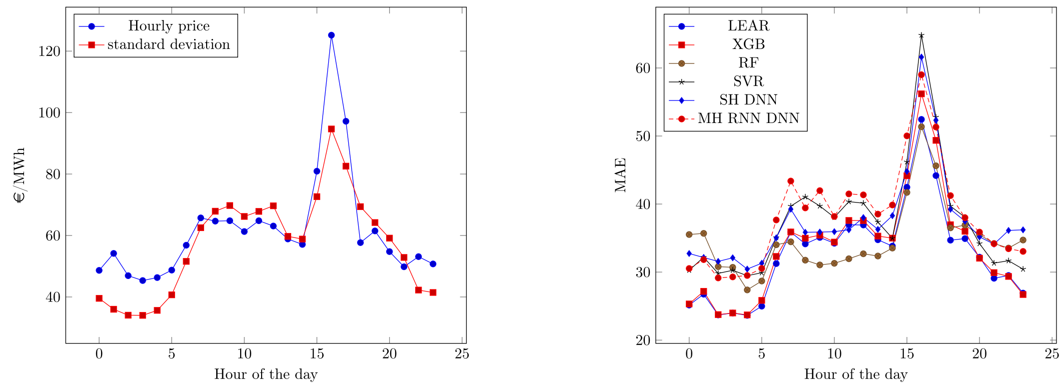
The differences in accuracy can be caused by the intrinsic stochasticity of the dataset. For this reason, a test comparing them in a statistically significant manner provides more robust insights. When comparing performance metrics of models M_1 and M_2, the Diebold–Mariano (DM) test is commonly used in EPF literature to quantify statistical distinctions [43]. Using forecasted BM price Y_t and errors $[\epsilon_1^{M_1}, \dots, \epsilon_N^{M_1}]$ and $[\epsilon_1^{M_2}, \dots, \epsilon_N^{M_2}]$, compute loss differentials. The DM test assesses null hypotheses:

$$\text{DM Test} = \begin{cases} H_0 : \mathbb{E}(d_k^{M_1,M_2}) \geq 0, \\ H_1 : \mathbb{E}(d_k^{M_1,M_2}) < 0. \end{cases} \tag{3}$$

Rejecting H_0 for model 1 implies a significant accuracy improvement in model 2. Table 2 presents the results of the DM test on the MAE, where column (model 2) and row (model 1) headers are organized by decreasing MAE errors. The ✓symbol signifies a statistical significance of 95% or more in accepting the alternative hypothesis from the DM test; otherwise, the comparison is marked with ✗. This implies that model 1's predictive accuracy is statistically significantly superior to that of model 2. The models' performances can be grouped into three groups. LEAR and the two ensemble ML models (XGB and RF) have similar results, significantly outperforming the others. The ARIMA, KNN, SVR and DL networks differences are not consistent enough to accept the alternative hypothesis with reasonable confidence. Finally, all the models considered are significantly better than the naive baseline.

6.1.2. Hourly analysis

Analyzing the average trends over the 24 h gives better insights into the reasons for the performance differences. Fig. 5(a) shows the average price and its standard deviation, while Fig. 5(b) shows the models' MAE both aggregated by the hour. We considered the models trained with 365 days of data. We can see that there are considerable variations in prices over the day; unsurprisingly, there is a strong peak in the 15:00–18:00 range. The same period is characterized by the highest variability as well. During the night, the prices and their standard deviation are considerably lower. The forecast error of the predictors follows a similar pattern, where higher prices and variabilities lead to less accurate predictions in absolute terms. These peculiarities also strongly affect the individual model performances. The DL models are more accurate on the night prices, with less uncertainty. However, their attempt to predict the peaks leads to higher errors during the



(a) BM Price mean and SD for each hour. (b) Models' MAE for each hour.

Fig. 5. Hourly breakdown of the BM prices and the models' forecast error.

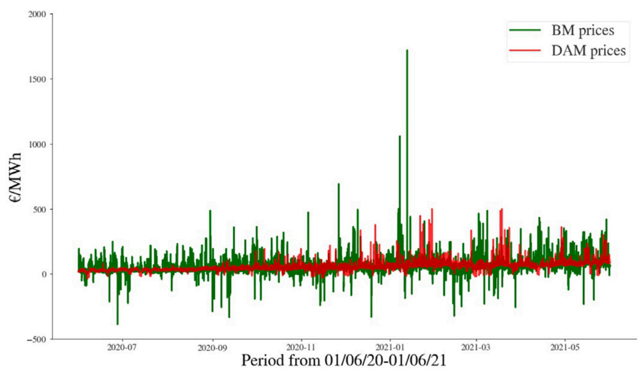


Fig. 6. DAM and BM prices.

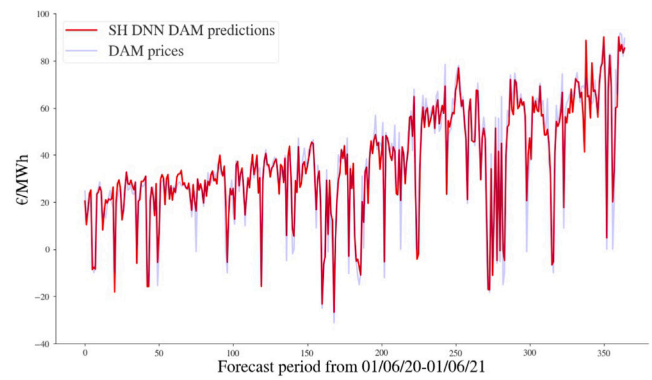


Fig. 7. SH DNN Forecast for DAM.

more volatile parts. On the contrary, LEAR and the best-performing ML models have a more stable error, being the most accurate over the volatile part. Techniques that more accurately predict lower and stable prices might be preferred over the best model on average for specific applications. For example, when having to choose which part of the night is better to charge a battery.

6.2. Comparison with DAM

To appreciate the differences between the two markets, we deployed the forecasting models presented in Section 5 to the DAM. Fig. 6 gives an idea of the difference in price ranges and uncertainty of the two markets. Considering the substantial volatility observed in the BM, where contiguous trading periods often witness extreme price changes, both negative and positive, for example:

- For the 4 BM settlement periods on 27/06/2020 in the 18:30-20:30 time horizon, prices veered from €2.01/MWh to €-390.07/MWh to €-38.69/MWh before ending at €3.12/MWh.
- On the 07/01/2021 15:00-16:30 prices jumped from €359.89/MWh to €1222.89/MWh, then to €1059.07/MWh and back to €320.88/MWh. For context, within the related 12-hour time block, the BM price ranged from a low of -€98.42/MWh to the aforementioned high of €1222.89/MWh.

While the DL models encounter challenges in BM predictions, despite the same configuration as the DAM, they exhibit more accurate forecasts in the DAM domain, reflecting their ability to capture price spikes and market dynamics. DL models' superior performance in the DAM compared to the BM underscores the influence of market structure and data characteristics on model performance. The SH DNN outperforms LEAR by a significant €2.81/MWh, confirming the results present in the literature. In Fig. 7, we can see the predictions of the SH DNN in comparison to the DAM prices. Compared to Fig. 4, we can see that the predicted peaks are more accurate. As outlined above, the DNNs overfit to recent spikes in the BM and follow sharp increases in the BM with an increase in the next 8 h of predictions, which is frequently wrong. The more stable DAM market is better suited for the DL approaches. Moreover, the effectiveness of these models may have been influenced by relatively small training sample sizes. Models optimized on a year or less of training data experienced poorer outcomes, as emphasized in Lago et al. [9]. This comparison strengthens the importance of an investigation for better forecasting models tailored to the BM.

6.3. Computational efficiency

In practical terms, when generating forecasts in the live setting of the BM, participants should consider the time lag between available information (explanatory variables) and the computational time required for model output. This can vary based on factors such as the deployed model and available hardware. In our analysis, DL models were trained on a GPU (Nvidia GTX 1080 ti), while statistical and ML models were trained on the CPU (Intel(R) Core(TM) i7-6800K CPU @ 3.40 GHz).

Fig. 8 shows the training time required by all the models. DL models generally entail higher computational costs compared to ML and statistical counterparts. Notably, LEAR and XGB perform excellently in computational cost and forecast accuracy, making them well-suited for real-time applications. In contrast, ARIMA and KNNs prioritize speed over accuracy. As expected, DL model training requires a considerable amount of time and exhibits the highest variability. Many factors affect

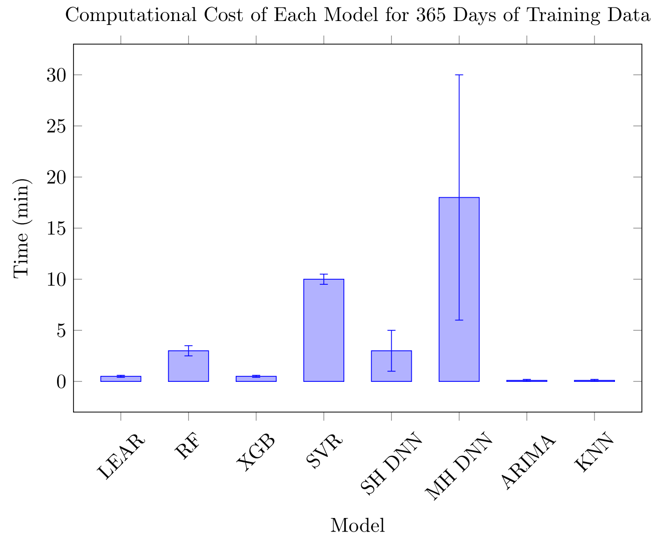


Fig. 8. Computational Time Required for Training.

their performance, such as the number of parameters, training data size, and number of epochs. The MH model's higher complexity and higher number of parameters considerably increase the time required for its training. The plot refers to training times; the models need to be retrained quite often to incorporate the most recent training data. The prediction time for a single point of the test set is negligible for all the models.

6.4. Limitations

Our models focus on providing forecasts that are accurate on average. This partially compromises their ability to identify extreme cases: really high or negative prices. This is due to the fact that an erroneous prediction of such extremes would be strongly penalized by the training metrics. In applications where the identification of these anomalies is crucial, even if this would lead to worse average accuracy, other approaches can be adapted from extreme events in time series prediction, e.g. Ding et al. [48]. Another way to deepen the scope of the prediction is the introduction of probabilistic forecasting. Alternatively, employing a different forecasting approach by utilizing more granular 5-minute datasets and integrating publicly available commercial order data into the model structure may enhance overall accuracy.

Furthermore, the findings are relative to the Irish market data from 2019 to 2022, and their generability decreases with the evolution of the Irish electricity generation and infrastructure. For example, wide adoption of battery storage systems could potentially reduce the volatility and likelihood of extreme price events, changing the ranking of our models. However, these changes gradually happen over long periods, and the framework presented herein can be easily adapted to new data. Moreover, a wider dataset could benefit some models, making them more competitive in that scenario.

7. Conclusion

This paper introduced a framework to benchmark predictive models for forecasting volatile electricity prices in the BM. In an extensive numerical analysis, we evaluated statistical, ML, and DL approaches that have been proven successful in the literature on similar tasks. The DL models that excelled in the more stable DAM struggled to forecast the uncertainty of the most volatile part of the BM, where simpler approaches shone. In particular, LEAR and tree ensemble techniques achieved the best accuracy due to their conservative predictions. The DL models encountered challenges in handling sudden price spikes, resulting in elevated errors. The simpler approaches, ARIMA and KNN, do not have the representational power to successfully model the dataset's patterns. The analysis also showed interesting insights on the Irish BM market. The original dataset, the models, and the framework are provided under an open-source license.

This BM price forecasting work suggests potential future directions. The hourly analysis suggests that an algorithm portfolio using different approaches for different parts of the day could leverage the differences between the models. Or, the different predictions can be combined in an ensemble to increase their robustness [16]. Exploring additional BM markets could generalize findings and highlight differences. Some BM operators offer more granular 5-minute data, allowing an alternative modeling approach. Moreover, higher data availability would allow the deployment of transfer learning approaches.

Further BM research avenues could encompass generating probabilistic forecasts with modern approaches (e.g. conformal prediction, [49]), treating price prediction as a classification challenge by employing DNNs' spike detection and leveraging forecasts for improved trading opportunities, potentially in conjunction with the DAM, to enhance financial outcomes for battery energy storage systems participants.

CRediT authorship contribution statement

Ciaran O'Connor: Conceptualization, Methodology, Software, Data curation, Visualization, Formal analysis, Investigation, Writing – original draft. **Joseph Collins:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Steven Prestwich:** Supervision, Resources, Writing – review & editing. **Andrea Visentin:** Supervision, Resources, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and the code have been made available in a public GitHub repository[4].

Acknowledgments

This work was conducted with the financial support of Science Foundation Ireland under Grant Nos. 18/CRT/6223, 16/RC/3918 and 12/RC/2289-P2 which are co-funded under the European Regional Development Fund. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

[1] C.B. Martinez-Anido, G. Brinkman, B.-M. Hodge, The impact of wind power on electricity prices, Renew. Energy 94 (2016) 474–487.
[2] Eurostat, Statistics explained, 2022, Statistics Explained. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Renewable_energy_statistics.
[3] N. Mazzi, P. Pinson, 10 - Wind power in electricity markets and the value of forecasting, in: G. Kariniotakis (Ed.), Renewable Energy Forecasting, in: Woodhead Publishing Series in Energy, Woodhead Publishing, 2017, pp. 259–278.
[4] A. Ortner, G. Totschnig, The future relevance of electricity balancing markets in Europe-A 2030 case study, Energy Strategy Rev. 24 (2019) 111–120.
[5] R. Weron, Electricity price forecasting: A review of the state-of-the-art with a look into the future, Int. J. Forecast. 30 (4) (2014) 1030–1081.

[4] https://github.com/ciaranoc123/Balance-Market-Forecast

[6] B. Neupane, W.L. Woon, Z. Aung, Ensemble prediction model with expert selection for electricity price forecasting, Energies 10 (1) (2017) 77.
[7] S. Luo, Y. Weng, A two-stage supervised learning approach for electricity price forecasting by leveraging different data sources, Appl. Energy 242 (2019) 1497–1512.
[8] S. Nyarko, Investigations into net imbalance volume chasing ('NIV chasing') in GB electricity market, 2019, Elexon BSC. URL https://www.elexon.co.uk/documents/groups/panel/2019-meetings/294-september/294-09-investigations-into-niv-chasing-in-gb-electricity-market-shadrack-nyarko/.
[9] J. Lago, F. De Ridder, B. De Schutter, Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms, Appl. Energy 221 (2018) 386–405.
[10] J. Lago, G. Marcjasz, B. De Schutter, R. Weron, Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark, Appl. Energy 293 (2021) 116983.
[11] J. Nowotarski, R. Weron, Recent advances in electricity price forecasting: A review of probabilistic forecasting, Renew. Sustain. Energy Rev. 81 (2018) 1548–1568.
[12] S. Anbazhagan, N. Kumarappan, Day-ahead deregulated electricity market price forecasting using recurrent neural network, IEEE Syst. J. 7 (4) (2012) 866–872.
[13] Y. Chen, Y. Wang, J. Ma, Q. Jin, BRIM: An accurate electricity spot price prediction scheme-based bidirectional recurrent neural network and integrated market, Energies 12 (12) (2019) 2241.
[14] U. Ugurlu, I. Oksuz, O. Tas, Electricity price forecasting using recurrent neural networks, Energies 11 (5) (2018) 1255.
[15] W. Li, D.M. Becker, Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling, Energy 237 (2021) 121543.
[16] A. Bâra, S.-V. Oprea, Predicting day-ahead electricity market prices through the integration of macroeconomic factors and machine learning techniques, Int. J. Comput. Intell. Syst. 17 (1) (2024) 10.
[17] C. Monteiro, I.J. Ramirez-Rosado, L.A. Fernandez-Jimenez, P. Conde, Short-term price forecasting models based on artificial neural networks for intraday sessions in the Iberian electricity market, Energies 9 (9) (2016) 721.
[18] P. Shinde, M. Amelin, A literature review of intraday electricity markets and prices, in: 2019 IEEE Milan PowerTech, IEEE, 2019, pp. 1–6.
[19] M. Narajewski, F. Ziel, Ensemble forecasting for intraday electricity prices: Simulating trajectories, Appl. Energy 279 (2020) 115801.
[20] D. Birkeland, T. AlSkaif, Research areas and methods of interest in European intraday electricity market research—A systematic literature review, Sustain. Energy Grids Netw. (2024) 101368.
[21] G. Zachmann, L. Hirth, C. Heussaff, I. Schlecht, J. Mühlenpfordt, A. Eicke, The Design of the European Electricity Market–Current Proposals and Ways Ahead, European Parliament, Luxembourg, 2023.
[22] R.A. Van der Veen, R.A. Hakvoort, The electricity balancing market: Exploring the design challenge, Utilities Policy 43 (2016) 186–194.
[23] C. Roumkos, P.N. Biskas, I.G. Marneris, Integration of European electricity balancing markets, Energies 15 (6) (2022) 2240.
[24] E. Rosales-Asensio, D.B. Diez, P. Sarmento, Electricity balancing challenges for markets with high variable renewable generation, Renew. Sustain. Energy Rev. 189 (2024) 113918.
[25] A. Eicke, O. Ruhnau, L. Hirth, Electricity Balancing as a Market Equilibrium, ZBW-Leibniz Information Centre for Economics, Kiel, Hamburg, 2021.
[26] R.A. Van Der Veen, A. Abbasy, R.A. Hakvoort, Agent-based analysis of the impact of the imbalance pricing mechanism on market behavior in electricity balancing markets, Energy Econ. 34 (4) (2012) 874–881.
[27] K. Poplavskaya, J. Lago, L. De Vries, Effect of market design on strategic bidding behavior: Model-based analysis of European electricity balancing markets, Appl. Energy 270 (2020) 115130.
[28] G. Klæboe, A.L. Eriksrud, S.-E. Fleten, Benchmarking time series based forecasting models for electricity balancing market prices, Energy Syst. 6 (2015) 43–61.
[29] J. Dumas, I. Boukas, M.M. de Villena, S. Mathieu, B. Cornélusse, Probabilistic forecasting of imbalance prices in the belgian context, in: 2019 16th International Conference on the European Energy Market, EEM, IEEE, 2019, pp. 1–7.
[30] D.W. Bunn, J.N. Inekwe, D. MacGeehan, Analysis of the fundamental predictability of prices in the British balancing market, IEEE Trans. Power Syst. 36 (2) (2020) 1309–1316.
[31] A. Lucas, K. Pegios, E. Kotsakis, D. Clarke, Price forecasting for the balancing energy market using machine-learning regression, Energies 13 (20) (2020) 5420.
[32] M. Narajewski, Probabilistic forecasting of German electricity imbalance prices, Energies 15 (14) (2022) 4976.
[33] D. Newbery, G. Strbac, I. Viehoff, The benefits of integrating European electricity markets, Energy Policy 94 (2016) 253–263.
[34] L. Meeus, The Evolution of Electricity Markets in Europe, Edward Elgar Publishing, 2020.
[35] V.N. Bharatwaj, A. Downey, Real-time imbalance pricing in I-SEM-Ireland's balancing market, in: 2018 20th National Power Systems Conference, NPSC, IEEE, 2018, pp. 1–6.
[36] M. Mörke, Marcos López de Prado: Advances in Financial Machine Learning: Wiley, 2018, 400 pp, USD 50.00, Approx. EUR 45, ISBN: 978-1-119-48208-6, Springer, 2019.
[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[38] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[39] F. Chollet, Deep Learning with Python, Simon and Schuster, 2021.
[40] L. Yang, A. Shami, On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing 415 (2020) 295–316.
[41] U. Ugurlu, O. Tas, U. Gunduz, Performance of electricity price forecasting models: Evidence from Turkey, Emerg. Mark. Finan. Trade 54 (8) (2018) 1720–1739.
[42] S. Makridakis, Accuracy measures: theoretical and practical concerns, Int. J. Forecast. 9 (4) (1993) 527–529.
[43] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econ. Statist. 20 (1) (2002) 134–144.
[44] R. Beigaite, T. Krilavičius, K.L. Man, Electricity price forecasting for nord pool data, in: 2018 International Conference on Platform Technology and Service, PlatCon, 2018, pp. 1–6.
[45] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[46] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2) (2005) 301–320.
[47] H. Drucker, C.J. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, Adv. Neural Inf. Process. Syst. 9 (1996).
[48] D. Ding, M. Zhang, X. Pan, M. Yang, X. He, Modeling extreme events in time series prediction, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 1114–1122.
[49] C. Kath, F. Ziel, Conformal prediction interval estimation and applications to day-ahead and intraday power markets, Int. J. Forecast. 37 (2) (2021) 777–799.