Day-ahead Forecasting Electricity Spot Prices in Norway with ARIMA, XGBoost, and LSTM Models

1st Markus Wiik Jensen
School of Economics, Innovation and Technology
Kristiania University College
Oslo, Norway
email:maje012@student.kristiania.no

2nd Huamin Ren
School of Economics, Innovation and Technology
Kristiania University College
Oslo, Norway
email:huamin.ren@kristiania.no

3rd Andrii Shalaginov
School of Economics, Innovation and Technology
Kristiania University College
Oslo, Norway
email:andrii.shalaginov@kristiania.no

***Abstract*—This paper comprehensively explores univariate and multivariate forecasting models for the Norwegian Elspot markets. As a leading renewable energy supplier with a high reliance on hydropower, Norway offers valuable insights into balancing renewable sources. The volatility of its electricity market, influenced by broader European trends, underscores the need for accurate forecasting. Day-ahead electricity price forecasts from the Elspot market are crucial for electricity producers and market operators, informing supply bids and dispatch schedules. This research includes experiments with advanced forecasting methods, combining machine learning and time series analysis to improve accuracy. We compare three models—ARIMA, XGBoost, and LSTM—across Norway's six Elspot markets. LSTM outperforms the other models in three specific zones, demonstrating its superior predictive performance. Future research will focus on enhancing model generalization.**

***Index Terms*— Green Energy; Electricity Price Forecasting; Elspot prices; XGBoost; LSTM.**

I. Introduction

Electricity and energy are integral to modern society, driving economic growth, technological advancement, and overall quality of life. Energy consumption, closely linked to factors such as wealth, health, and infrastructure, has been steadily increasing due to population growth, industrialization, and technological development. As the world transitions away from fossil fuels and seeks sustainable solutions, understanding and forecasting energy markets becomes crucial. Accurately forecasting market trends and price fluctuations is of paramount significance for a diverse range of stakeholders, including investors, businesses, and policymakers [10] [14] [16] [31]. The Norwegian electricity markets, characterized by deregulation and high renewable integration, present unique research opportunities. Recent market disruptions, marked by volatile prices and increased uncertainty, underscore the need for advanced forecasting techniques. This research aims to enhance understanding of Norway's electricity markets by investigating key price drivers and evaluating electricity price forecasting (EPF) methods. Such predictions are crucial for electricity producers, consumers, and market operators to effectively plan their production, consumption and trading activities [3].

The NordPool spot (Elspot) market is a day-ahead market, where the price of power is determined by supply and demand. Such spot prices are the actual prices for electricity for the next day, and will be set at NordPool Elspot. Our primary focus is on day-ahead price forecasting using known spot prices. This forecasting directly informs bidding strategies for the upcoming day [19]. Due to the distinct characteristics of electricity markets, each forecasting challenge is unique across different markets and necessitates bespoke model developments [24]. We propose a framework for evaluating forecasting methods for all six Elspot markets of Norway while comparing three different numerical approaches to the problem of extrapolating prices in both univariate and multivariate configurations, facilitating the identification of region-specific models and model configurations. By examining the factors influencing price dynamics and comparing various forecasting methodologies, this study seeks to improve predictive accuracy and interpretability. The findings will provide a framework for future research and support decision-making in modeling and market analysis. In Section II, we dive into electricity markets and existing literature on EPF. Section III presents the methodologies employed. Section IV discusses the conducted experiments, and in Section V we conclude with an analysis of the obtained results.

Figure 1. Hydropower reservoir.

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

II. Background and Literature Studies

In this section we review the various market mechanics characterizing electricity markets and existing literature concerning EPF.

A. Background

Electricity is produced only moments before consumption, so unlike other commodities, electricity must be balanced between production and consumption at all times [17]. In a deregulated market environment, determining the unconstrained Market Clearing Price (MCP), commonly referred to as the spot price of an electricity pool typically involves the following steps:

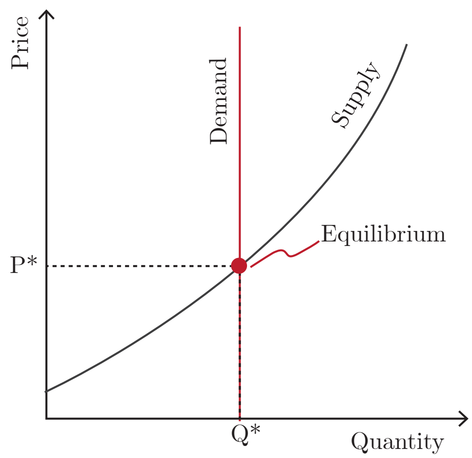


Figure 2. Equilibrium curve to determine the MCP of a bidding-pool.

- Generating companies bid prices for supplying energy, creating a supply curve.
- The demand curve may be set at a value derived from a forecast of the load due to short-term inelasticity for demand of electricity, resulting in a vertical line at the forecasted load value.
- Spot price is found where supply and demand curves intersect, signifying the market equilibrium.

The spot price is set at the equilibrium between supply and demand as seen in Figure 2 for each hour of the following day after accounting for the bids received within the deadline as illustrated in Figure 3 [14].

Like many goods and services, electricity demand exhibits daily, weekly, and seasonal fluctuations. Consumption typically peaks during late afternoon and early evening when people return home from work and school, activating their lighting and appliances. This period, known as "peak demand," necessitates increased electricity generation to meet the higher demand. Demand patterns also vary by location, influenced by local weather conditions and regional consumption behaviors [31]. For instance, in warmer climates during summer, electricity demand rises due to increased use of air conditioning. Conversely, in colder regions during winter, higher demand is driven by heating requirements. Understanding these demand factors is crucial for utilities and policymakers to ensure a reliable and sustainable electricity supply [14]. Electricity generation encompasses various methods, including coal, natural gas, nuclear power, and renewables such as wind, solar, and hydro power. Due to the non-storability of electricity and the need for load management, generators must continuously adjust their output to balance supply and demand. This task is particularly challenging for renewable sources like wind and solar power that are inherently intermittent and uncontrollable. Wind and solar energy production cannot be precisely controlled and is dependent on current weather conditions, resulting in variable output. In contrast, nuclear power provides a stable, continuous supply but lacks the flexibility to adjust output quickly in response to demand changes. Technologies capable of responding to rapid fluctuations, such as flexible hydro-power and liquid natural gas (LNG) generation, are essential for maintaining grid stability. Additionally, the value of storable production sources, such as fossil fuels and hydro, is influenced by their convenience yield—an extra benefit from holding the commodity beyond potential financial gains from its sale.

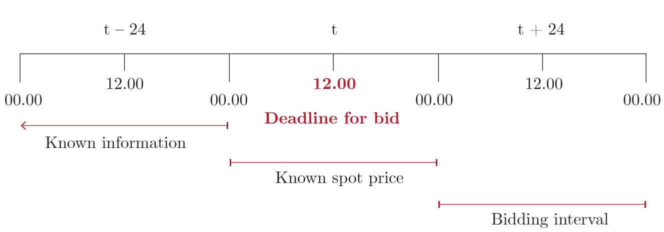


Figure 3. Deadline for bids in the Elspot markets.

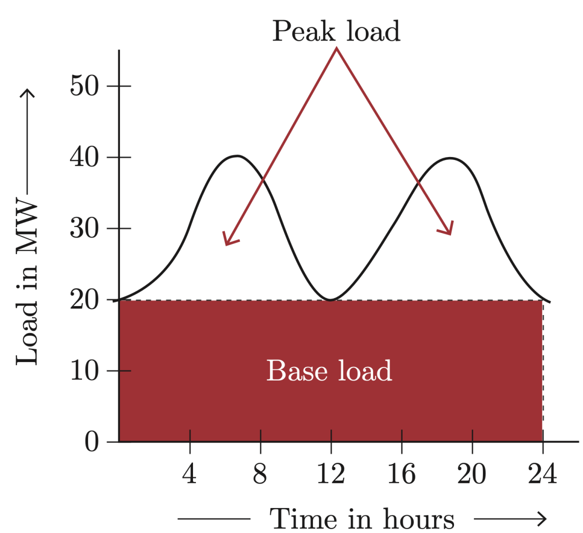


Figure 4. Base-load vs. Peak-load.

Mechanics such as electricity market and pricing, electricity production and consumption are crucial to understand the complexity of EPF. The electricity market is influenced by a multitude of factors, including supply and demand dynamics, changing industrial and household consumption,

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

multiple seasonality, weather conditions, regulatory policies, fuel prices, the integration of renewable energy sources, and the rapid diffusion of price-anomalies [3] [10] [14] [16] [31]. Understanding the key drivers of price movements aids in feature selection for predictive models. For instance, if weather patterns or economic variables significantly affect prices, incorporating these into a model may improve accuracy. The choice of methodology should also consider the nature of price drivers, as incorporating these considerations guides model selection. Furthermore, accurate price forecasts coupled with an understanding of their drivers provide valuable market insights.

A time series is defined as a series of data points indexed in time order [34]. Commonly expressed as:

$$X = X_{t}{}_{t=1}^{\infty} = (X_1, X_2, \dots) \tag{1}$$

where X_t denotes the observation at time t, and the sequence of observations is indexed by t ranging from 1 to infinity.

Accurately extrapolating the future electicity prices poses unique challenges due to several constraints imposed by time order. Some of these constraints include look-ahead bias, stationarity, auto-correlation, seasonality, trend and noise. Time-series data, characterized by sequential observations over time, requires specialized methodologies that can capture temporal dependencies and patterns. Time series forecasting (TSF) attempts to predict future outcomes based on historical context and has direct applications to many domains, including science, policy and business. Because TSF is based on historical data it can be useful for planning future actions based on previous actions, by measuring the statistical correlations between variables over time to predict the future it is possible to also explore meaningful patterns in the data that would otherwise remain inconceivable.

B. Literature Studies

In the domain of EPF, selecting appropriate input variables, historical data duration, and modelling techniques is crucial. Most efforts that focus on forecasting day-ahead prices typically experiment with an inference horizon of 1-4 weeks [4] [5] [12] [13] [15] [19] [20] [24] [25] [28] [32] [35]. Historical data spanning at least a year is commonly employed to capture yearly seasonality [4] [15] [20] [26] [35]. Input variables encompass a range of factors, including past prices [4] [5] [7] [8] [11] [12] [13] [15] [18] [19] [20] [23]- [29] [32] [35], system loads [15] [19] [23] [25]- [28] [32], weather variables [7] [15] [20] [26] [33], fuel costs [5] [7] [21] and sector indices [30]. Preprocessing and data transformations are essential to handle missing values and outliers that can affect model performance. Techniques like normalization [7] [8] [32], decomposition [8] [12] [20] [25] [27] [35], and differentiation [13] are used to improve data quality and model accuracy. Statistical models, such as econometric methods, like Linear Regression [15] [23] [25] [33] and Auto-Regressive models [5] [12] [13] [15] [18] [20] [32] [35], offer interpretability and insights into correlations. Algorithmic models like Deep Learning (DL) [8] [15] [18] [19] [21] [23]- [27] and Ensemble models [5] capture complex and nonlinear patterns.

As highlighted in numerous studies, the process of building a forecasting model involves decisions on input selection, forecasting horizons, preprocessing and feature engineering techniques, model choice, parameter estimation, and accuracy evaluation. However, guidelines for navigating these complexities are limited, with much variation in reported approaches. Given the specific nature of EPF, establishing baselines and ensuring rigorous reporting is critical for advancing research in this field.

The process of determining critical design decisions vary, Amjady and Hemmati [3] emphasize that most input-variable selections are based on forecaster heuristics rather than a systematic approach, while Aggarwal et al. [2] note that advancements in technology, market optimization, and data availability continue to alter the landscape of optimal variable selection. A universal set of price drivers is unlikely to emerge, given the diverse nature of electricity markets. Despite the growing number of studies on EPF, many lack transparency and statistical rigor, making it difficult to compare different approaches. Studies focusing on advanced statistical techniques often compare these only to basic machine learning (ML) methods, while ML-based studies typically contrast against simple statistical techniques, further complicating cross-study comparisons. Major review publications have pointed out that inconsistent datasets, implementations, error measures, and problem definitions exacerbate this issue, making it hard to assess the transferability of findings to other markets or future developments [2] [3] [24] [32]. The failure to control for issues like data contamination and look-ahead bias is common, as many studies do not specify details such as test-train splits, input variables, or data transformations. Lago et al. [24] stress the importance of ensuring that the test dataset is always the last segment of the full dataset, with no overlap with training data. Moreover, a significant number of studies neglect to benchmark new methods against simpler, well-established models such as naive heuristics, as they are crucial for evaluating the true generalization performance of complex models. The lack of such baselines can lead to spurious conclusions about model performance, even in otherwise well-conducted research. Future studies should rigorously incorporate these practices to enhance reproducibility, validity, and the significance of results.

III. Proposed Methodology

In this research, the approach begins with selecting a baseline method that is heuristic-based. Building upon this baseline, the study conducts an empirical-driven progression to develop previously proven forecasting models in both univariate and multivariate configurations. Three distinct approaches are explored: a econometric method, an algorithmic ensemble approach, and a deep learning (DL) approach. This methodology is designed to ensure objectivity and standardization in the evaluation process. Given the unique and inconsistent

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

nature of electricity markets, EPF challenges vary significantly across locations and time frames, rendering cross-study evaluations potentially misleading and universal benchmarks logically unsound for this domain. Therefore, the methodology involves systematic steps, including literature review of related work, data collection and preparation, model development and rigorous testing against real world outcomes. Models are trained, validated, and tested in both univariate and multivariate configurations, enabling comparisons of the added value of incorporating exogenous variables. The expectation is that multivariate models should outperform univariate models, that rely solely on price data, to justify the increased complexity and computational cost. By comparing forecast results from both configurations, the study aims to shed light on the role of exogenous variables as price drivers. All the data-handling, -visualization and model-implementation and -evaluation was done using Python software.

A. Heuristic Baseline

The persistence forecast is utilized as a baseline for this study. This approach involves using the last observed value of the time series as the forecast for the corresponding day-ahead time step. In the context of day-ahead EPF, this would mean using the most recent price value as its prediction for the same hour the next day. Assuming we have a time series of electricity prices $p_t, p_{t+1}, p_{t+2}, ..., p_{t+n}$ where t is the current time step, the persistence model predicts the current price 24 hours ahead for each time step. In the context of day-ahead EPF, the persistence model serves as a sensible baseline. While more complex modelling methods may exhibit reasonable accuracy, they must be able to generalize beyond the explicit information provided in the input data. As a baseline the heuristic provides a reference point against which, more advanced models can be evaluated, ensuring that they genuinely contribute to improved forecasting performance. We can express the persistence model in mathematical notation as follows:

$$\hat{P}_t = P_{t-24} \tag{2}$$

where $\hat{P}_t$ denotes the predicted electricity price at time t and P_{t-24} is the observed value of the electricity price 24-time steps earlier.

B. Econometric

The ARIMA (Autoregressive Integrated Moving Average) (p,d,q) method is a popular time series forecasting technique that models the time series data as a combination of autoregressive (AR) and moving average (MA) components, with an additional differencing step to account for non-stationarity. The parameters p, d, and q are integers that represent the order of the AR, differencing, and MA components, respectively. The ARIMA(p,d,q) method can be specified by the following mathematical notation:

$$Y_t = c + \sum_{i=1}^{p} \Phi_i Y_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{3}$$

where Y_t is the value of the time series at time t, c is a constant term, $\Phi_1, \Phi_2, ..., \Phi_p$ are the AR coefficients, ϵ_t is the error term at time t, and $\theta_1, \theta_2, ..., \theta_q$ are the MA coefficients.

The AR component models the current value of the time series as a linear combination of its past values, with the weights determined by the AR coefficients. The MA component models the current value of the time series as a linear combination of the past errors, with the weights determined by the MA coefficients. While ARIMA models offer benefits such as capturing auto-correlation and seasonality, providing interpretability, they have limitations in handling non-linear relationships and the assumption of stationarity. These factors should be carefully considered when applying ARIMA-type models to forecast electricity prices.

C. Algorithmic Ensemble

Extreme Gradient Boosting (XGBoost) is a popular gradient-boosting algorithm that is commonly used in machine-learning applications for both classification and regression tasks. It is an ensemble algorithm that combines multiple weak models (decision trees) to make a strong prediction. XGBoost learns from examples by building a series of decision trees. Each tree tries to correct the mistakes made by the previous trees reducing the risk of overfitting, and leading to a more accurate prediction. To further identify the most impactful variables and account for non-linear relationships between targets and inputs, XGBoost's feature gain scores are employed. This metric measures the relative contribution of each feature to the objective function, with higher scores indicating greater importance in generating accurate predictions [8]. The objective function for XGBoost can be written as:

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

where Θ represents the set of model parameters, n is the number of training examples, y_i is the true value of the i-th example, $\hat{y}_i$ is the predicted value, $l(y_i, \hat{y}_i)$ is the loss function, K is the number of weak models, f_k represents the k-th weak model, and $\Omega(f_k)$ is the regularization term.

The weak models used in XGBoost are decision trees, and can be expressed as:

$$f(x) = \sum_{t=1}^{T} w_t q_t(x), \quad w \in \mathbb{R}^T, \quad q : \mathbb{R}^d \to \{1, 2, \ldots, T\} \tag{5}$$

where x is the input features, w is the vector of weights associated with each leaf node of the tree, T is the number of leaf nodes, and $q(x)$ is the function that maps the input features to the index of the corresponding leaf node.

D. Deep Learning

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that is commonly used for time-series forecasting. Unlike traditional RNNs, LSTM networks are designed to overcome the problem of vanishing gradients,

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

making it difficult for the network to learn and remember long-term dependencies in the data. In simple terms, the LSTM network is like a specialized memory unit that can selectively remember important information from the past and use it to make predictions about the future. It achieves this by using a system of gates to control the flow of information within the network.

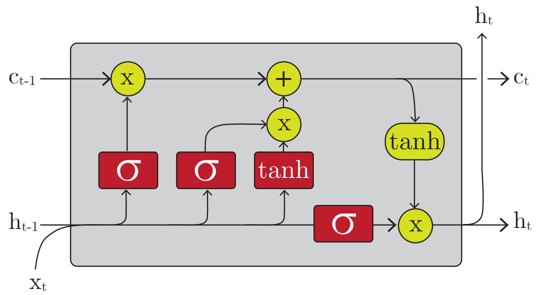


Figure 5. Long Short-Term Memory (LSTM) Network Diagram.

The LSTM network has three main types of gates as visualized in Figure 5: input gates, forget gates, and output gates. These gates allow the network to decide information that is important to keep, information to forget, and when to output its predictions [21].

IV. Experiments and Discussion

This section covers the datasets used, the experimental setup, and the ensuing presentation and discussion of results. The research aims to clarify the methods used to forecast Norway's contemporary electricity markets by examining both modelling approaches and relevant price drivers. Studying these aspects together is beneficial, as understanding key price drivers can guide the selection of inputs that enhance model performance. For example, incorporating influential factors like weather or economic conditions can improve forecast accuracy. Additionally, the choice of modelling method depends on how price drivers interact with electricity prices, particularly in cases of non-linearity or temporal dependencies. By aligning model selection with these dynamics, we aim to create more reliable and robust forecasts. While price drivers are not the primary focus, their consideration is essential for optimizing model inputs and improving the interpretability of results. This approach promises more accurate predictions and a clearer understanding of the factors shaping Norway's electricity market.

A. Dataset and Description

Following background theory and related work, a diverse range of independent variables that are identified as potential price-drivers for the Norwegian markets were selected, from fundamental variables such as operating data and weather variables that governs production, to macro variables such as oil-prices and international or regional trade. To collect and preprocess the data, a Python environment is utilized as it is capable of handling various sources, file types, and formats. The data, including unit measures, granularity and data sources are described in Table I. A total of six data-sets were created, each comprising time series data from one of the six bidding zones. Comparing electricity prices across different regions can provide insights into the factors driving price dynamics in a specific region and guide the development of region specific forecasting models. The data-sets consist of 14-16 variables each, with the amounts of variables varying depending on the number of exchange connections to neighbouring zones.

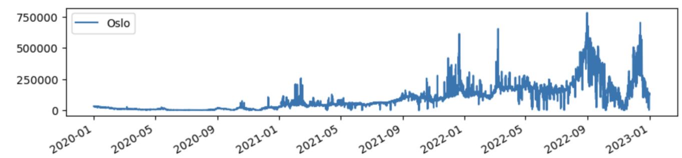


Figure 6. Historical Elspot prices for Oslo (NO1).

TABLE I. DESCRIPTION OF DATA (TARGET*).

Variable (units) [granularity]	Source
Elspot price (NOK/MWh) [h]	Nord Pool
Day-ahead Elspot price (NOK/MWh)[h]*	Nord Pool
Power production (MWh) [h]	Nord Pool
Power production prognosis (MWh) [h]	Nord Pool
Power exchange (MWh) [h]	Nord Pool
Power consumption (MWh) [h]	Nord Pool
Reservoir levels (GWh) [w]	Nord Pool
Reservoir capacity (GWh) [w]	Nord Pool
Gas price (NOK/mmbtu) [d]	Yahoo-finance
Oil price (NOK/barrel) [d]	Yahoo-finance
OSEBX price (NOK/OSEBX) [d]	Yahoo-finance
Air temperature (mean/degC) [d]	MET
Wind speed (mean/ms) [d]	MET
Percipitation (sum/mm) [d]	MET

Missing values occurred due to multiple reasons, such as changing time zones, observations at a lower frequency than the target values and stock exchanges being closed during weekends. Missing values due to these occurrences were appropriately imputed using interpolation, backward-fill or forward-fill. One example is the weather observation being recorded daily from hundreds of weather stations each day (see Figure 7), needing to be aggregated geographically to averages in each bidding-zone and filled for the 24-hours each day. Other preprocessing complexities include historical currency conversion of economic variables and handling large amounts of unstructured operational data.

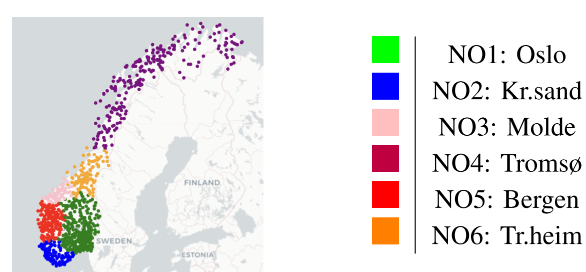


Figure 7. Locations of weather stations color-labeled by bidding zones.

The data is split into two sections, the first contains three years of data with 26 000+ price-observations and is allocated

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

for training and validation, the second is separated from the first and contains 4 months of recent and unseen data allocated for testing and evaluation. The date ranges are the following, 01.01.2020 00:00 - 29.12.2022 23:00 for train and validation, and 01.01.2023 00:00 - 30.03.2023 23:00 for the hold-out test set. Essentially, the train-test split contains the original time order and is not shuffled or re-ordered. Data is normalized using min-max scaling, this is done separately for the two sections in order to prevent introducing look-ahead-biases encoded in the scaling. In this research, all the data is scaled in order to help improve predictions and reduce risk of overfitting.

B. Data Analysis

To gain deeper insights into the relationships between independent and dependent variables, correlations and other descriptive statistics were computed. To further explore these relationships, Principal Component Analysis (PCA) was performed on the correlation matrices, projecting the variables onto a two-dimensional feature space, as illustrated in Figure 8. Singular Value Decomposition (SVD) provides a way to transform the data into a new coordinate system where the new axes (principal components) are linear combinations of the original variables, and the data can be represented in a lower-dimensional space with minimal loss of information. In all of the data-sets, the historical oil, gas and osebx prices are closely approximated in the feature-space. Weather variables, in particular precipitation is closely approximated to reservoir-levels in most bidding-zones and in some cases as in Oslo (Figure 8), precipitation and wind-speed is also relatively closely approximated to production.

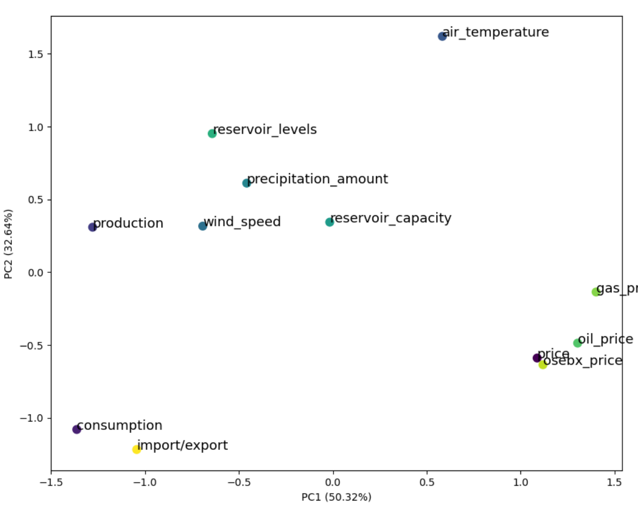


Figure 8. Approximation and projection of variables onto two-dimensional feature-space (PCA).

To better understand the price data we also conducted decomposition analysis and examined trends and seasonality, and potential memory effects within the time series by visualizing Auto-Correlation Functions (ACF) and Partial Auto-Correlation Functions (PACF). The stationarity of the data was confirmed using Augmented Dickey-Fuller (ADF) statistics, ensuring the data is suitable for further modeling and predictive analysis. It is important to recognize that in real-world markets, the assumption of a stationary time series can be misleading. Financial and economic data, such as electricity prices or stock prices, are influenced by a range of external factors like policy changes, market shocks, and seasonal effects, which introduce non-stationary behavior over time.

C. Experiments

The experiments include a heuristic baseline and are compared against each other as opposed to previous experiments from related work. The persistence model is a naive approach that assumes the future price will be the same as the current price. In other words, it simply predicts the value of the target as the value of the target at time t-24. This model serves as a reference point for the performance of more complex models. The implementation of the persistence model is straightforward and can be easily achieved using any programming language or spreadsheet software. In this case, Python and the Pandas library was used to load and manipulate the data. The other models were trained and optimized in different ways due to their varying degree of complexity.

To fit an ARIMA model, optimal values for p, d, and q are typically determined by analyzing the data through various methods. This includes plotting the ACF and PACF, computing ADF statistics, and evaluating models using criteria like AIC or BIC. The optimal model is selected by fitting multiple models with different orders and choosing the one with the lowest AIC or BIC value.

$$AIC = -2\ln(\hat{L}) + 2k \tag{6}$$

$$BIC = -2\ln(\hat{L}) + k\ln(n) \tag{7}$$

Optimal parameters found:
p = 2, d = 1, q = 2

During training, XGBoost minimizes the loss function root mean squared error (RMSE), to improve prediction accuracy. Gradient boosting is employed to iteratively adjust the model's parameters and reduce the loss. One of the main advantages of XGBoost in this context is its ability to provide feature importance scores, which are derived from the model's decision trees. These scores offer interpretability in a TSF framework, allowing for an understanding of how different factors influence electricity prices over time. The feature importance scores were computed using the "gain" metric, it measures the contribution of each feature to reducing the loss function in the model. By analyzing these scores, it was possible to identify the most influential variables across different regions. In forecasting tasks, knowing the features have the most impact allows for targeted adjustments and improvements in the model. It also aids in validating and explaining model

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

predictions, enhancing transparency and trust in the model's outputs.

$$\mathcal{L}(\Theta) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\text{MSE}(y_t, \hat{y}_t)} \tag{8}$$

L2 Regularization:

$$\Omega(f) = \frac{\lambda}{2}\sum_{j=1}^{T} w_j^2 \tag{9}$$

where λ is the regularization strength.

Training neural networks involves feeding the model full cycles of training data, with each cycle called an epoch. After each epoch, the model updates its weights through a backward pass and optimization process to minimize the mean squared error (MSE) loss function. Configuring neural networks is complex due to the lack of a universal approach; instead, it requires systematic exploration of different configurations. This involves both dynamical exploration, assessing how the network behaves during training, and objective exploration, evaluating performance on validation or test sets. Finding the optimal number of epochs, number of hidden layers and number of neurons in each hidden layer must be explored. To address overfitting in neural networks, dropout regularization is applied to randomly drops weights and prevent excessive co-adaptation. Hyperparameter tuning, performed using the Optuna library, helps find the optimal configuration for aspects such as the number and shape of hidden layers, dropout rate, learning rate, batch size, and sequence length.

$$\mathcal{L}(\Theta) = \frac{1}{T}\sum_{t=1}^{T}\text{MSE}(y_t, \hat{y}_t) \tag{10}$$

Optimizer: Adam

First, the models are validated in the task of predicting the day-ahead hourly elspot prices on the validation set using a rolling forecast cross-validation (RFCV) scheme presented in Table II. These experiments provide information about the models' performance on a full year of daily-predictions with daily re-training. During validation, the error of the models is measured using RMSE. The errors are averaged by time of day; mornings (hours 6-12), mid-days (hours 12-15), evenings (hours 15-21) and nights (hours 21-6). An example of results from rolling forecasts origin validation with visualization from a sample period of 1 week including bar charts of aggregated time-of-day scores from the entire year are presented in Figure 9 (baseline results of aggregated RMSE are marked with red dashed lines for comparisons).

TABLE II. RFCV SCHEME (YYYY-MM-dd hh).

Fold	Train Start	Train End	Val Start	Val End
1	2020-01-01 00	2021-12-31 23	2022-01-01 00	2022-01-01 23
2	2020-01-01 00	2022-01-01 23	2022-01-02 00	2022-01-02 23
3	2020-01-01 00	2022-01-02 23	2022-01-03 00	2022-01-03 23
...	...	...	...	...
365	2020-01-01 00	2022-12-28 23	2022-12-29 00	2022-12-29 23

After validating the models on the last year of the train-set, they are then are evaluated in their ability to extrapolate 24 time-steps ahead from the known spot-price during a 4-month out-of-sample period on a recent hold-out test-set from all the bidding-zones, with their weights and hyperparameters determined from training and tuning on the previous 3 years of data. The results of these experiments are presented in Table III, allowing for comprehensive analysis and review of the different modelling approaches in relation to the bidding zones and the addition of exogenous variables. The evaluation scheme of model performance consists of four different error terms; Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Average Percentage Error (MAPE) and Residual Sum of Squares (RSS). To gain a comprehensive understanding of the models' capacity for generalization and their ability to navigate the bias-variance trade-off, we seek to offer diverse viewpoints on the models' performance.

D. Results

This section outlines the key findings from the research project, incorporating results from our conference paper [1], which summarizes the performance of various predictive models. The scope has been extended to address the econometric ARIMA model, highlighting the importance of different features to facilitate better interpretation of model outputs. Detailed discussion and interpretation of the results follow in subsequent sections.

The actual vs. predicted NO2 values for Kristiansand are presented for the time interval December 18th to December 24th, 2022. The top subplot (Figure 9a) shows the comparison between the actual and predicted values for a one-week period. The lower subplot (Figure 9b) aggregates the root mean square error (RMSE) by time of day for the entire year of 2022. Predictions from validation seem to be more accurate during mornings (6-12) and middays (12-15) as illustrated by the RMSE scores in Figure 9. However, none of the models consistently outperform the heuristic baseline across bidding zones and time-of-day during these experiments. The performance of different forecasting models (Heuristic, ARIMA, XGBoost, and LSTM) during out-of-sample evaluation is summarized in Table III, which reports the RMSE, MAE, MAPE, and RSS for each model with and without exogenous variables. The results cover the out-of-sample evaluation period from January 1st, 2023, to March 30th, 2023. The LSTM model in its multivariate configuration outperforms the other models for all aspects of error on the data-sets for bidding-zone NO2 and NO3. Surprisingly, the univariate LSTM outperforms the other models in all aspects of error for the bidding-zone NO4. The final model to outperform the baseline for all aspects of error is the multivariate XGBoost model for the bidding-zone NO6. For the remaining bidding-zones NO1 and NO5 there is no clear contender for best model performance. Among the models evaluated, LSTM and multivariate XGBoost models demonstrated superior performance, outperforming the baseline across all forecast criteria. These models successfully balanced the bias-variance trade-off, effectively capturing the

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE III. MODEL PERFORMANCE SUMMARY ON TEST SETS (01.01.2023 00:00 - 30.03.2023 23:00).

		RMSE		**MAE**		**MAPE**		**RSS**	
	Model	endog	w/ exog	endog	w/ exog	endog	w/ exog	endog	w/ exog
NO1	Heuristic	29469	/	19946	/	25.19%	/	$18.7e^{11}$	/
	ARIMA	29469	44124	19946	36332	**25.18%**	36.84%	$18.7e^{11}$	$41.1e^{11}$
	XGBoost	29156	27052	20268	**18838**	26.22%	26.92%	$17.9e^{11}$	$\mathbf{15.5e^{11}}$
	LSTM	29174	**21109**	21035	20134	29.21%	26.69%	$17.9e^{11}$	$17.9e^{11}$
NO2	Heuristic	29474	/	19943	/	25.19%	/	$18.7e^{11}$	/
	ARIMA	29474	37216	19943	27932	25.18%	30.86%	$18.7e^{11}$	$29.2e^{11}$
	XGBoost	29545	27259	20715	19266	26.77%	26.19%	$18.4e^{11}$	$15.7e^{11}$
	LSTM	28354	**26431**	20317	**18173**	29.09%	**24.81%**	$16.9e^{11}$	$\mathbf{14.9e^{11}}$
NO3	Heuristic	30448	/	21069	/	37.58%	/	$20.0e^{11}$	/
	ARIMA	30448	44907	21069	34139	37.57%	47.61%	$20.0e^{11}$	$42.6e^{11}$
	XGBoost	28469	29069	19687	19666	35.79%	31.79%	$17.1e^{11}$	$17.8e^{11}$
	LSTM	28438	**28381**	20462	**19228**	40.91%	**31.29%**	$17.1e^{11}$	$\mathbf{16.6e^{11}}$
NO4	Heuristic	21456	/	11705	/	25.28%	/	$99.4e^{10}$	/
	ARIMA	21456	30875	11705	22584	25.27%	45.46%	$99.4e^{10}$	$18.7e^{16}$
	XGBoost	20592	23149	11424	12507	25.43%	29.35%	$89.6e^{10}$	$11.3e^{11}$
	LSTM	**19448**	21675	**10519**	13155	**22.76%**	28.05%	$\mathbf{79.9e^{10}}$	$96.1e^{10}$
NO5	Heuristic	25240	/	**16953**	/	**15.75%**	/	$\mathbf{13.7e^{10}}$	/
	ARIMA	25240	27679	**16953**	19177	**15.75%**	17.28%	$\mathbf{13.7e^{10}}$	$16.1e^{11}$
	XGBoost	24950	**24156**	17137	17018	16.27%	15.94%	$13.1e^{11}$	$12.3e^{11}$
	LSTM	25427	24584	18391	18189	17.80%	17.49%	$13.6e^{11}$	$12.9e^{11}$
NO6	Heuristic	30448	/	21069	/	37.58%	/	$20.0e^{11}$	/
	ARIMA	30448	45333	21069	34465	37.57%	48.04%	$20.0e^{11}$	$43.4e^{11}$/
	XGBoost	28469	**28326**	19687	**19532**	35.79%	**31.70%**	$17.1e^{11}$	$\mathbf{16.9e^{11}}$
	LSTM	28438	30100	20462	22870	40.91%	48.58%	$17.1e^{11}$	$19.1e^{11}$

complex data dynamics of EPF. Conversely, ARIMA models, while strong in interpretability and simplicity, faced limitations in out-of-sample extrapolation. The ARIMA models we cofigured struggled with longer inference horizons, exhibiting underfitting with endogenous variables alone and overfitting when incorporating exogenous variables. This performance discrepancy underscores the challenges of using ARIMA models for time series with complex and long-term dependencies, as their reliance on lagged values may fail to adequately capture and project complex price patterns. Table IV displays the feature importance scores for the XGBoost model across different regions. Key features such as energy prices, weather conditions, and exchange variables are ranked based on their contribution to the model's predictions. The importance of these features varies across the regions, with price and oil price generally being the most significant predictors.

E. Discussion

The observed regional differences between southern (NO1, NO2, NO5) and northern (NO3, NO4, NO6) bidding zones in Norway reveal important insights into the dynamics of the electricity markets. The southern zones' strong correlation with economic factors, particularly macroeconomic variables such as oil prices and global energy markets, indicates a heightened sensitivity to external shocks. This could explain the increased volatility in electricity prices in these regions, as they are more exposed to fluctuations in global supply and demand for energy commodities. The XGBoost feature importance analysis in Table IV supports this by highlighting the prominence of oil prices and other market-driven factors in the price forecasts for southern zones like Oslo (NO1). These results suggest that economic policies and global market developments could have a disproportionate impact on electricity prices in the southern regions. In contrast, the northern zones show a more stable price formation process, driven by operational factors such as hydropower availability and local consumption patterns. This suggests that, despite the geographic proximity of the regions, the drivers of electricity prices differ significantly. The northern zones, less exposed to macroeconomic volatility, may experience more predictable and stable price trends, driven by supply-side considerations like hydropower generation and reservoir levels. This aligns with the relative stability observed in the northern zones, where local operational factors play a more substantial role. These regional distinctions underscore the importance of adopting tailored forecasting approaches for different bidding zones. For instance, models forecasting prices in southern regions could benefit from incorporating global economic indicators and commodity market trends, while models for northern regions should focus more on hydrological conditions and localized operational factors. Furthermore, the differing sensitivity of regions to price drivers has implications for policy-making, as energy regulation or economic policies that affect market dynamics may need to be region-specific to ensure stability and predictability in electricity prices across Norway. This analysis also raises questions about the resilience of the Norwegian electricity market to global economic shifts.

Lack of improvements over the baseline during validation seen in Figure 6 could be attributed to the disruptive prices in 2022, making it difficult for the models to fit the data comprehensively. Results from the out-of-sample evaluation

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

TABLE IV. FEATURE IMPORTANCE (GAIN) SCORES FOR XGBoost IN DIFFERENT REGIONS.

Feature	NO1	NO2	NO3	NO4	NO5	NO6
Price	33.05	41.01	8.04	4.54	33.97	8.6
Oil Price	12.2	4.61	3.73	0.51	3.76	3.26
Osebx Price	3.71	3.52	1.28	0.47	2.91	1.37
Gas Price	2.61	2.4	1.51	0.65	2.01	1.63
Reservoir Levels	1.94	2.77	1.35	0.68	1.21	1.05
Wind Speed	0.83	0.69	0.83	1.7	1.09	0.5
Air Temperature	0.45	1.24	0.77	1.23	0.59	2.1
Production	0.33	0.85	0.43	0.35	0.7	0.55
Production Forecast	0.51	1.24	0.91	1.03	0.66	0.62
Precipitation Amount	0.78	0.61	0.77	0.62	0.52	0.77
Consumption	0.36	0.23	0.92	0.46	0.32	0.43
Exchange NO1-NO2	0.24	0.26	-	-	-	-
Exchange NO1-NO3	0.11	-	0.28	-	-	-
Exchange NO1-NO5	0.29	-	-	-	0.24	-
Exchange NO1-NO6	-	-	-	-	-	0.25
Exchange NO1-SE3	0.23	-	-	-	-	-
Exchange NO2-NL	-	0.61	-	-	-	-
Exchange NO2-NO5	-	0.36	-	-	0.29	-
Exchange NO3-NO4	-	-	0.63	0.86	-	-
Exchange NO3-NO5	-	-	0.13	-	0.35	-
Exchange NO3-SE2	-	-	0.59	-	-	-
Exchange NO4-SE2	-	-	-	0.75	-	-
Exchange NO4-SE1	-	-	-	0.42	-	-
Exchange NO6-NO4	-	-	-	-	-	0.54
Exchange NO6-NO5	-	-	-	-	-	0.63
Exchange NO6-SE2	-	-	-	-	-	0.25

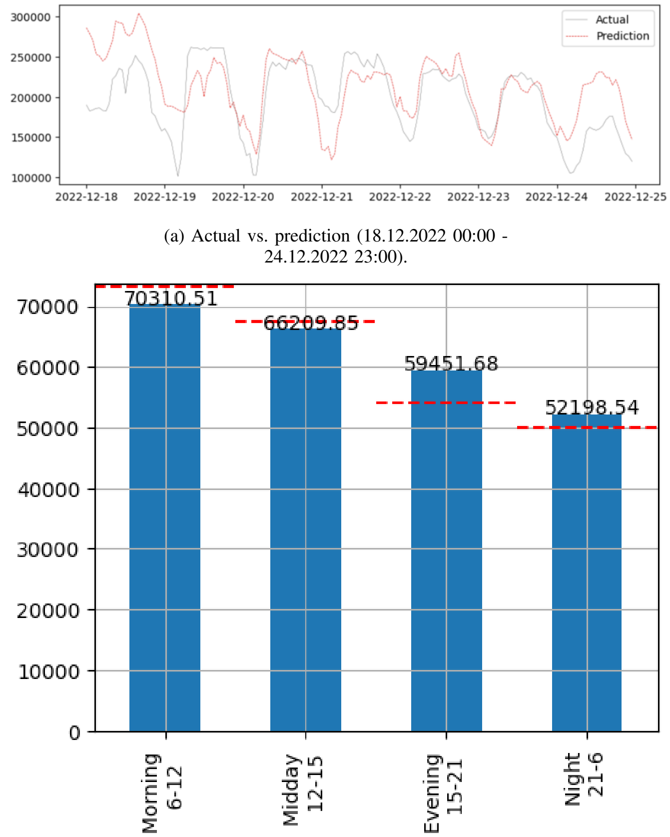
exhibit more promising improvements over the baseline. As seen in Table III, the LSTM and XGBoost models outperform the baseline across all evaluation criteria for most of the bidding-zones, meaning that they are able to balance between capturing price nuances while maintaining robustness to outliers. These results ultimately emphasize the potential of DL and ensemble ML techniques for capturing the complexities of EPF. The model performance across different bidding zones shows a mixed picture. The naive baseline model often performs well in terms of MAPE and RSS, suggesting it may be a strong benchmark for some zones. XGBoost generally excels in RMSE, indicating robust prediction accuracy, while also showing improvements with exogenous variables. The LSTM models, though slower to train and complex, tend to offer competitive performance, particularly in terms of RMSE and generalization, especially when configured with multiple variables. ARIMA shows variable results, sometimes overfitting or underfitting depending on the configuration. Regarding simplicity the ARIMA models stand out as contenders, often surpassing the baseline in validation. They offer a straightforward approach to TSF and ease of interpretation. However, challenges arise when extending these models to further out-of-sample inference. The interpretability of tree-based models like XGBoost provides significant advantages, particularly in understanding complex, non-linear relationships within the data. Unlike many black-box models, XGBoost offers clear insights into the features that are most influential in driving predictions. This interpretability is crucial for stakeholders who need to make informed decisions based on the model's findings and ensures that the model's behavior aligns with domain knowledge and expectations. The variability in results across the different data-sets highlights the presence of unique characteristics for the distinct bidding-zones, with varying predictability, model-performances and optimal model-configurations. Highlighting the need for region-specific price-modelling and improving model generalization.

V. Conclusion and Future Work

Forecasting day-ahead electricity prices plays a pivotal role in strategizing and balancing the supply and demand for the subsequent day, making it an essential area to delve into. In this paper, we introduce a framework to assess forecasting techniques across all Elspot markets in Norway, intimidating heuristic methods with more complex ARIMA models, advanced XGBoost and LSTM deep learning networks. Various models, including XGBoost and LSTM, show varying effectiveness across different bidding zones. XGBoost's interpretability aids in understanding non-linear relationships, while LSTM models demonstrate strong predictive capabilities they offer little insight into the patterns it learns from the data. Overall, the research underscores the importance of combining detailed analysis of price drivers with sophisticated modeling techniques to enhance the understanding and prediction of electricity markets.

The study's focus on the Norwegian market limits the generalizability of the findings. Future work should explore the applicability of the developed models to other electricity markets to assess their robustness across different contexts. Computational constraints and the omission of extensive

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org



(a) Actual vs. prediction (18.12.2022 00:00 - 24.12.2022 23:00).

(b) Aggregated RMSE (01.01.2022 00:00 - 29.12.2022 23:00).

Figure 9. Rolling Forecast Origin Cross-validation of multivariate LSTM for Kristiansand (NO2).

feature engineering also highlight areas for improvement. Incorporating additional data sources and exploring hybrid models that combine various forecasting approaches could further refine prediction accuracy. The variability in results across the different data-sets highlights the presence of unique characteristics for the distinct bidding-zones, therefore, model generalization will be the focal point of our future research endeavors. In conclusion, this research contributes valuable insights into the Norwegian electricity markets and forecasting methodologies. Addressing the identified limitations and exploring future research directions will enhance the development of more accurate and reliable forecasting models, benefiting both researchers and practitioners in the field.

REFERENCES

[1] M. W. Jensen, H. Ren, and A. Shalaginov, "Day-ahead Electricity Price Forecasting of Elspot Markets in Norway", ENERGY 2024, The Fourteenth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies, vol. 14, pp. 1–6, 2024.

[2] S.K Aggarwal, L. M Saini and Ashwani Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation", International Journal of Electrical Power & Energy Systems, vol. 31, pp. 13-22, 2009.

[3] N. Amjady and M. Hemmati, "Energy price forecasting - problems and proposals for such predictions", IEEE Power and Energy Magazine, vol. 4, pp. 20–29, 2006.

[4] R. Beigaite, T. Krilavicius and K. L. Man, "Electricity Price Forecasting for Nord Pool data", International Conference on Platform Technology and Service (PlatCon), pp. 1–6, 2018.

[5] K. Bitirgen and Basaran Filik, "Electricity price forecasting based on XGBoost and ARIMA algorithms", BSEU Journal of engineering research and technology, vol. 1, pp. 7–13, December 2020.

[6] L. Breiman, Classification and regression trees, Routledge, 2017.

[7] M. Castelli, A. Groznik, and A. Popovic, "Forecasting electricity prices: A machine learning approach", Algorithms, vol. 13, pp. 119, May 2020.

[8] Z. Chang, Y. Zhang and W. Chen, "Electricity Price Prediction based on Hybrid Model of Adam optimized LSTM neural network and Wavelet Transform", Energy, pp. 187, 2019.

[9] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system", In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, August 2016.

[10] Y.-k. Chen, A. Hexeberg, K. E. Rosendahl and T. F. Bolkesjø, "Long-term trends of nordic power market: A review", WIREs Energy and Environment, vol. 10, pp. 413, 2021.

[11] A. Ciarreta, M. P. Espinosa and C. Pizarro-Irizar, "Is green energy expensive? empirical evidence from the Spanish electricity market", Energy Policy, vol. 69, pp. 205–215, 2014.

[12] A. Conejo, M. Plazas, R. Espinola and A. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models", IEEE Transactions on Power Systems, vol. 20, pp. 1035–1042, 2005.

[13] J. Contreras, R. Espinola, F. Nogales and A. Conejo, "Arima models to predict next-day electricity prices", IEEE Transactions on Power Systems, vol. 18, pp. 1014–1020, September 2003.

[14] A. Creti and F. Fontini, Economics of electricity, Cambridge University Press, 2019.

[15] A. Cruz, A. Munoz, J. L. Zamora and R. Espınola, "The effect of wind generation and weekday on spanish electricity spot price forecasting", Electric Power Systems Research, vol. 81, pp. 1924–1935, 2011.

[16] C. Defeuilley, "Retail competition in electricity markets", Energy Policy, vol. 37, pp. 377–386, 2009.

[17] G. Erdmann, "Economics of electricity", EPJ Web of Conferences, vol. 98:06001, January 2015.

[18] G. Gao, K. Lo and F. Fan, "Comparison of ARIMA and ann models used in electricity price forecasting for power market", Energy and Power Engineering, vol. 09, pp. 120–126, January 2017.

[19] P. S. Georgilakis, "Market clearing price forecasting in deregulated electricity markets using adaptively trained neural networks", SETN: Hellenic Conference on Artificial Intelligence, vol. 3955, pp. 56–66, 2006.

[20] L. Grossi and F. Nan, "Robust forecasting of electricity prices: Simulations, models and the impact of renewable sources", Technological Forecasting and Social Change, vol. 141, pp. 305–318, 2019.

[21] J.-J. Guo and P. Luh, "Selecting input factors for clusters of Gaussian radial basis function networks to improve market clearing price prediction", IEEE Transactions on Power Systems, vol. 18, pp. 665–672, June 2003.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural computation, vol .9, pp. 1735–1780, 1997.

[23] Z. Hu, L. Yang, Z. Wang, D. Gan, W. Sun and K. Wang, "A game-theoretic model for electricity markets with tight capacity constraints", International Journal of Electrical Power & Energy Systems, vol. 30, pp. 207–215, 2008.

[24] J. Lago, G. Marcjasz, B. De Schutter, and R. Weron, "Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark", Applied Energy, vol. 293, pp. 116983, 2021.

[25] C. Li and S. Wang, "Next-day power market clearing price forecasting using artificial fish-swarm based neural network", ISNN: Advances in Neural Networks, pp. 1290–1295, 2006.

[26] P. Mandal, T. Senjyu and T. Funabashi, "Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market", Energy Conversion and Management, vol. 47, pp. 2128–2142, 2006.

[27] M. S. Nazar, A. E. Fard, A. Heidari, M. Shafie-khah and J. P. Catalao, "Hybrid model using three-stage algorithm for simultaneous load and price forecasting", Electric Power Systems Research, vol. 165, pp. 214–228, 2018.

[28] E. Raviv, K. E. Bouwman and D. Van Dijk, "Forecasting day-ahead electricity prices: Utilizing hourly prices", Energy Economics, vol. 50, pp. 227–239, 2015.

[29] K. Skytte, ”The regulating power market on the nordic power exchange nord pool: an econometric analysis”, Energy Economics, vol. 21, pp. 295–308, 1999.

[30] B. A. Souhir, B. Heni and B. Lotfi, ”Price risk and hedging strategies in nord pool electricity market evidence with sector indexes”, Energy Economics, vol. 80, pp. 635–655, 2019.

[31] R. Weron, Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach, John Wiley Sons, 2006.

[32] R. Weron and A. Misiorek, ”Forecasting spot electricity prices with time series models”, Proceedings of the European Electricity Market EEM-05 Conference, pp. 133–141, May 2005.

[33] R. Weron and M. Zator, ”Revisiting the relationship between spot and futures prices in the nord pool electricity market”, Energy Economics, vol. 44, pp. 178–190, 2014.

[34] Wikipedia, ”Time series,” *Wikipedia, The Free Encyclopedia*, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Time_series&oldid=1193976963. [Accessed: 2024-02-09].

[35] Z. Yang, L. Ce, and L. Lian, ”Electricity price forecasting by a hybrid model, combining wavelet transform, arma and kernel-based extreme learning machine methods”, Applied Energy, vol. 190, pp. 291–305, 2017.

2024, © Copyright by authors, Published under agreement with IARIA - www.iaria.org