



# From the East-European Regional Day-Ahead Markets to a Global Electricity Market

Adela Bâra<sup>1</sup> · Simona-Vasilica Oprea<sup>1</sup> · Bogdan George Tudorică<sup>2</sup>

Accepted: 23 June 2023 / Published online: 31 July 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

The so-called black swans, COVID-19 and the invasion in Ukraine, have led to an unprecedented increase in electricity prices. Since 2021, after lockdowns, the electricity price has started to increase due to economic recovery, rising prices of the tCO<sub>2</sub> and other primary sources that become unavailable or at much higher prices. In this context, we noticed that the variation of electricity prices in one country can be explained by the price fluctuations of the previous day in the neighboring countries. For instance, the prices for the current day ( $d$ ) in the Romanian Day-Ahead Market is strongly correlated with the prices of the previous day ( $d-1$ ) on DAMS of its neighboring countries. It is worth mentioning that the target can be switched by the rest of the variables. Not only the price in Romania can be estimated using the proposed Electricity Price Forecast (EPF) method, but also the prices in other neighboring countries can be a target for prediction because the regional prices on similar markets contain most of society's distress. Another interesting aspect is that the proposed forecasting methodology is robust, as proved by testing it on a varied and longer time interval (from January 2019 to August 2022). Furthermore, the proposed price forecasting methodology includes the adjustment of training interval according to the price standard deviation and weighing the results of the five individual Machine Learning (ML) algorithms to further improve the prediction performance. The set consists of data collected between 1st of January 2019—one year before COVID-19 pandemic outburst and 31st of August—several months after the war has started in the Black Sea region.

**Keywords** Black swan · Electricity price forecast · Regional day-ahead markets · ML algorithms

## List of Symbols

Variable	Description
$h$	Time of the historical records used for training and testing [hours]

Extended author information available on the last page of the article

$d$	Day
$T$	Time interval for the training and testing of the ML algorithms [hours]
$W^d$	Day of the week
$RMSE, R^2, MAPE, MAE$	Evaluation indicators for ML algorithms
$Q_{RO}^h$	Traded quantity of electricity on DAM in Romania [MWh]
$P_{RO}^h$	Hourly electricity price on DAM in Romania [Euro/MWh]
$Q_{HU}^h$	Traded quantity of electricity on DAM in Hungary [MWh]
$P_{HU}^h$	Hourly electricity price on DAM in Hungary [Euro/MWh]
$Q_{SR}^h$	Traded quantity of electricity on DAM in Serbia [MWh]
$P_{SR}^h$	Hourly electricity price on DAM in Serbia [Euro/MWh]
$Q_{BG}^h$	Traded quantity of electricity on DAM in Bulgaria [MWh]
$P_{BG}^h$	Hourly electricity price on DAM in Bulgaria [Euro/MWh]
$P_{SK}^h$	Hourly electricity price on DAM in Slovakia [Euro/MWh]
$P_{CZ}^h$	Hourly electricity price on DAM in Czech [Euro/MWh]
$Flow_{RO-HU}^h$	Electricity flow from Romania to Hungary [MWh]
$Flow_{HU-RO}^h$	Electricity flow from Hungary to Romania [MWh]
$\mu P^{d-1}$	Mean price of the previous day [Euro/MWh]
$P_{max}^{d-1}, P_{min}^{d-1}$	Maximum and minimum price of the previous day [Euro/MWh]
$\sigma P^{d-1}, \sigma^2 P^{d-1}$	Standard deviation and variance of the price of the previous day [Euro/MWh]
$\widetilde{P}^{d-1}$	Median of the price of the previous day [Euro/MWh]
$P_{range}^{d-1}$	Range price of the previous day [Euro/MWh]
$P_{ratio}^{d-1}$	Ratio price index of the previous day
$\mu P^{h_3}, \mu P^{h_7}$	The averages of the previous hourly prices for 3 and 7 consecutive days [MWh]
$\Delta P^{h_1}, \Delta P^{h_2}, \Delta P^{h_3}$	Hourly variation of the prices for the previous 3 consecutive days
$X^h$	Input variable of the ML algorithms [array]

$y^h$	Actual target variable of the ML algorithms [Euro/MWh]
$\hat{y}_i^h, i \in \{RF, LGBR, HGBR, XGB, VR\}$	Output of the ML models [Euro /MWh]
$\widehat{P}_{RO}^{hf}$	Forecast of the hourly electricity price in Romania [Euro/MWh]
$\omega_i$	Weights for adjusting the output of the ML algorithms

## 1 Introduction

Variation in the cost of energy, in recent years, in the global markets, can be seen as the result of several major factors. Among these factors, we mention, in relative order of decreasing impact: the conflict in Ukraine, the COVID-19 pandemic, the significant changes in direction of US energy policy, the massive fluctuations in OPEC oil production, and the significant seasonal variations in renewable energy-based generation. Two of these factors, the Ukrainian crisis and the COVID-19 pandemic are worth mentioning because they have had a significant influence, which can be easily detected by comparing the state of the energy markets before the occurrence of the respective factors and at the present time. The war in Ukraine can be considered perhaps the most impacting factor that has considerably influenced the evolution of prices, directly for electricity and indirectly for the resources normally used in the European economy for electricity production—natural gas, oil, and some of its derivatives, coal, uranium (Butler, 2022; Elliot, 2022; Kolaczowski, 2022; Menon, 2022; Thompson, 2022a; Tolefson, 2022).

The share of energy products in the total energy used, in 2020, in the European Union (EU) was: total petroleum products—34.5%, natural gas—23.7%, renewable energy—17.4%, nuclear energy—12.7%, solid fossil fuels (coal)—11.5%, other—0.2% (Eurostat, 2020b). In 2020, 57.5% of the total used energy was coming from imports, in various forms—35.8% of solid fossil fuels (coal), 83.6% of natural gas, 96.2% of crude oil, and 99.8% of uranium (Edmond, 2022; Eurostat, 2020a; Supply Agency of the European Atomic Energy Community, 2021). For the entire EU, in 2020, 22.9% of energy imports, 45% of gas imports, 29% of crude oil imports, and 20% of uranium imports were coming from the Russian Federation (Jakob Faveile Adolfsen et al., 2022). Taking into account the above-mentioned percentages, we can roughly estimate that the price for more than 50% of the total energy used in the EU was directly and indirectly controlled by the Russian Federation, and the price for another 30% or more of the total energy used in the European Union was dramatically influenced as the entire global energy market reacted to the Ukrainian crisis. Even energy-producing technologies which, at first glance, should not be affected (various types of renewable energy sources) would somehow be influenced, as some of them are using resources partially provided by the Russian Federation (e.g., lithium, rare-earth elements, etc.). The result is that the prices for energy and for all the resources used for generating energy raised abruptly—for example the European Energy benchmark price averaged 191 €/MWh in the second quarter of

2022, which is 181% higher than the second quarter of 2021 according to (European Commission, 2022).

Some other energy markets are also highly dependent on the fossil fuel prices (Bernal et al., 2019). Li and Leung (2021) discussed how access to reliable energy services is increasingly seen as a prerequisite for well-being and human development. The article examined the relationship between fossil fuel dependence and energy insecurity and how this dependence can lead to energy insecurity and vulnerability. It also explored the potential of renewable energy sources to reduce energy insecurity and vulnerability. Mayer (2022) proposed a study that examines how natural resource dependence, measured as oil and coal production, impacts energy security at the country level. The study used entropy-balanced fixed effects models to analyze data from 137 countries over 15 years. The study found that fossil fuel-producing nations do not have better energy security outcomes than non-producing nations and may suffer from a natural resource curse or paradox of plenty. The study suggested that fossil fuel-producing nations should implement policies that would allow them to retain more wealth from fossil fuel production and invest in electric grid infrastructure. It also highlighted the need for energy justice and sustainability in the global energy system. Thompson (2022b) argued how energy researchers must grapple with the lessons of history to navigate the long road to net zero. The article explored the geopolitics of fossil fuels and renewables and how they reshape the world. It also examined the challenges and opportunities that lie ahead as the world transits to a low-carbon future.

The second significant influence on the evolution of energy markets was the COVID-19 pandemic. At the beginning of the pandemic, an immediate drop in oil prices was felt, which was particularly pronounced. For example, the price of Brent crude oil fell by 75% between February and April 2020, whereas the Dutch TTF gas price fell by 44%. After the initial timing, oil and gas prices recoiled, with gas prices reaching pre-pandemic levels in September 2020 and oil prices doing so around February 2021. Gas price growth was high in the second half of 2021 and intensified further in the first half of 2022—European gas prices increased by 145% compared to July 2021 levels, while oil prices increased by 46% over the same period. Both oil and gas prices have risen well above pre-pandemic levels, and European gas prices have reached a record high, in turn contributing to record-high wholesale electricity prices (Friderike Kuik et al., 2022). However, assuming the temporal overlap of the two factors (war in the Black Sea region and COVID-19 pandemic), it can be said that it is not possible to determine the exact contribution of each of the two factors to both the total change in energy prices and the change in the structure of energy consumption by component. Furthermore, massive fluctuations in energy prices, such as those described above, make it particularly difficult, if not impossible, to predict market developments with any degree of accuracy at such times.

Regardless the two major unpredictable influences, the electricity price forecast is a challenge for both researchers and operational planners as it traditionally relies on generation and consumption, meteorological parameters, power system contingencies, bids, and offers strategies of market players, renewables generation share (Lu et al., 2022), etc. However, it was proven that these features change their importance in time. More features have to be considered and analyzed, such as economic

growth, inflation, interest rate, prices of the certificates for CO<sub>2</sub> emissions, tie-lines flows, available capacity, and prices of previous day in neighboring markets. These last features have not been analyzed yet. Moreover, the recent intervals that encountered random events as pandemic and conflict in Ukraine have not been considered in the scientific research. The importance of the electricity price forecasting is indubitable in the context of the digitalization and energy transition (Tian et al., 2022). It facilitates numerous applications, including optimization, scheduling (Busse & Rieck, 2022), business strategies for market players (such as traders, suppliers, producers) to approach DAM and other electricity markets (Hubicka et al., 2019), investment strategies in generation technologies, operational planning, macroeconomic policies design (Lehna et al., 2022), etc.

The impact of the increasing share of variable Renewable Energy Sources (RES), especially wind power plants and electricity load on the short-term electricity markets (day-ahead, intra-day and balancing markets) was investigated in (Spodniak et al., 2021) emphasizing on the Nordic case (namely, Denmark, Sweden and Finland) for an interval between 2015 and 2017. Vector autoregressive (VAR) models were applied to study the relationship between prices and wind and load forecast accuracy, proving that markets closer to the real time are trading more energy due to the increasing share of RES. Another recent study that consider the spot prices in Nordic and Baltic countries and RES between 2013 and 2020 is (Solibakke, 2022). It further pointed out that RES forecasts considerably influence the spot prices. Such findings can improve the spot price forecast and reduce the confidence interval significantly. However, these studies analyzed time intervals before the COVID-19 pandemic and were unable to catch the effects of the restrictions or embargo imposed on Russia. Furthermore, the impact of RES or load forecast becomes less relevant in the current context.

A day-ahead electricity price prediction using a Long Short-Term Memory (LSTM) model, that approaches the nonlinear and complex issues in processing time series data, and feature selection algorithms taking into account market coupling was proposed in (Li & Becker, 2021). The influence of market coupling effects on electricity price forecast on Nordic DAM was investigated. The results demonstrated that feature selection and moreover features from other integrated markets have an essential impact on price forecast. For example, the German market prices are significant in the price forecast for the Nordic market. Out of numerous research studies and papers that approach the electricity price forecasting, (Tschora et al., 2022) is among the first papers that consider the history of the prices of the neighboring countries as input for the ML algorithms, such as random forest, support vector regressor, deep and convolution neural networks. They demonstrated the benefits of the contribution of each feature in the electricity price forecasting using Shapley values.<sup>1</sup> The authors also included production and consumption forecasts that may introduce additional error. However, they target Western European countries (France, Germany, Belgium, Netherlands, Spain, Switzerland). To support (Tschora et al., 2022), we strengthen on the fact that historical prices from neighboring

<sup>1</sup> [https://en.wikipedia.org/wiki/Shapley\\_value](https://en.wikipedia.org/wiki/Shapley_value)

countries are relevant. But in comparison, we provide a more reliable method that is based on adjusting training interval and weighting the results of the ML algorithms.

The contribution of this paper consists of the following:

- Targeting the East-European market region, tie-line flows and day-ahead markets in several neighboring countries that are closer to the conflict in the Black Sea: Romania, Bulgaria, Serbia, Hungary, Slovakia, and the Czech Republic;
- Identifying and extracting historical prices from national market operators and building up a new relevant data set that can be further used in an open research approach. Using these features, we aim to perform estimation in the context of random events or black swans such as COVID-19 pandemic and conflict in Ukraine, near the Black Sea region;
- Analyzing a longer and more recent interval, from January 2019 to August 2022, in which two major events took place. The interval is considerably generous, as it catches the context before and after the two major events;
- Proposing a forecasting methodology that includes the creation of new features by aggregation or derivation of the variables in order to capture the price fluctuations and spikes. The methodology also consists of adjusting the training interval for prediction, rotating the dependent variables, and weighting the results of the five individual ML algorithms to further improve the prediction performance;
- Adjusting the training interval based on empirical observation of the results in relation to the standard deviation of prices recorded in the previous month of the forecasting interval. By adjusting the training interval versus fixed training interval, a sensitivity analysis demonstrates the capacity of the proposed forecasting methodology;
- Rotating the dependent variable as the correlation among prices in the Eastern-European region is strong. The rest of the exogenous variables are shifted to the known values of day to  $d-1$ . This trait shows the capacity of the proposed methodology to obtain accurate results;
- Weighting the results of the five individual ML algorithms using two regressors to further improve the performance;
- Testing the methodology for stable and more turbulent intervals demonstrated its robustness though the results in disruptive years (2021, 2022) are slightly higher;
- Depicting interesting findings and insights related to the East-European market region in the current context and filling the obvious gap as most of the research is focusing on the Western-European market (Germany, Austria) (Busse & Rieck, 2022) and Australia, (Lu et al., 2022).

Our motivation to pursue with this study and propose a price forecasting methodology for DAM is related to the significant research gap. It consists in (a) the lack of studies using recent data sets that cover interval before and after major events (such as COVID-19 pandemic and war in the Black Sea region); (b) lack of studies that focus on the East-European region as most of the published papers concentrate on the Western or Nordic countries; (c) lack of the available recent data sets that allow price prediction—in this regard, we extracted and collected the variables from open

data sources and merged them to build up a relevant data set that can be further utilized to achieve more insights in an open research approach.

Considering the above-mentioned contribution, we notice that the existing approaches in terms of variables of the data sets focus on total consumption, RES share, economic features, and less on the adjacent markets that operate in a coupling mode. We also notice that the existing electricity price forecast methodologies do not consider the cumulative steps proposed in the current paper, namely correlation among prices in neighboring markets, adjusting the training interval based on the price standard deviation, rotating variables and weighting the results of the ML algorithms to enhance the output. Therefore, taking into account the research gap and the above exposed motivation, we propose an electricity price forecasting methodology that is relevant and robust for both stable and disruptive economic contexts.

The remaining of the paper is structured as follows: in the next section, a brief literature survey of the most recent and relevant research is presented; in Sect. 3, the input data is analyzed to understand the price evolution on DAM in the East-European region, the trend and its characteristics from January 2019 to the end of August 2020; the proposed methodology is depicted in Sect. 4, results in Sect. 5, and conclusion in Sect. 6.

## 2 Literature Review

The European market coupling started in 2006, when Belgium, France, and the Netherlands coupled their day-ahead markets for the first time to increase the trading potential of the market players and maximize the usage of the tie-lines capacity. It combines previously separated trading transactions and allocation of tie-lines capacity into an integrated electricity market. After four years, Germany and Luxembourg joined them and formed the Market Coupling Western Europe (MCWE) that involve both market operators and Transmission System Operations (TSOs). In parallel, in 2007, Spain and Portugal DAM started to operate as an integrated market. Also in 2007, Nordic countries were connected to Germany and since 2011 to Netherlands. Austria also joined the MCWE Group in 2013 and Switzerland became an observer.<sup>2</sup> The Price Coupling of Regions (PCR) is a system created in 2010 by the European countries that decided to couple their electricity markets and implement a single market algorithm (Van den Bergh et al., 2016).

In 2014, 15 European countries benefited from PCR, including the Baltic States, Great Britain, Nordic countries and Poland that joined the MCWE. One year later, Italy coupled its market to France, and in 2016 Austria and Slovenia also coupled their markets. These 19 countries represent 85% of the European electricity consumption. Furthermore, in 2019, Romania, Hungary, Czech Republic and Slovakia started to join this group and now their market operators are part of the Nominated Electricity Market Operator (NEMO) to use the same algorithms and secure the

---

<sup>2</sup> <https://www.next-kraftwerke.com/knowledge/market-coupling#milestones-on-the-way-to-the-european-single-market>

operation of the market coupling mechanisms in Europe together with the TSOs, market players and regulatory authorities (Lam et al., 2018).

Electricity price forecasting is important for maximizing economic benefits and mitigating power market risks. Electricity prices are highly volatile and depend on a range of variables, such as electricity demand and feed-in from renewable energy sources. Accurate electricity price forecasting can help to make informed decisions about energy consumption and production, and can help to reduce energy costs. Some of the benefits of using electricity price forecasting methods include: improved energy management, better decision making, reduced energy costs, improved energy efficiency, reduced carbon footprint.

The following scientific papers comment on the benefits of EPF and discuss various approaches:

- Alkawaz et al. (2022) proposed a new hybrid machine learning method for a day-ahead Electricity Price Forecasting (EPF) which involves linear regression Automatic Relevance Determination (ARD) and ensemble bagging Extra Tree Regression (ETR) models;
- Naumzik and Feuerriegel (2021) offered a machine learning-based approach to forecast electricity prices. The authors used a time series of electricity prices and weather data to train a machine learning model to predict electricity prices. The authors also performed a sensitivity analysis to identify the most important predictors of the model. The authors conclude that their machine learning-based approach outperforms traditional time series models in terms of forecast accuracy;
- Castelli et al. (2020) proposed a novel genetic programming approach to improve electricity price forecasting accuracy. The authors used a time series of electricity prices and weather data to train a machine learning model to predict electricity prices. The authors concluded that their machine learning-based approach outperforms traditional time series models in terms of forecast accuracy;
- Schnürch and Wagner (2020) employed machine learning algorithms to forecast German electricity spot market prices. The authors used bid and ask order book data from the spot market as well as fundamental market data like renewable feed and expected total demand to train a neural network model to predict electricity prices. The authors concluded that their machine learning-based approach outperforms traditional time series models in terms of forecast accuracy.
- Different time-horizons were considered for analyzing the electricity prices for DAM from ultra-short-term (4-h forecast horizon), short-term (day-ahead) (Zhang et al., 2022), and mid-term from one to four weeks (Busse & Rieck, 2022) to long-term approach (Gabrielli et al., 2022). Most of the research in electricity price forecasting—more than 50 were published in 2022—focuses on the traditional fundamental variables such as consumption, primary resource costs, transmission and distribution costs other fees (Dragasevic et al., 2021), generation, its breakdown, the level of integration of renewables in the power system (Maciejowska et al., 2021), buying and selling strategies, and various forecasting methods reviewed by (Lago et al., 2021), etc.



- The electricity price forecasting is approached from the spikes point of view using classification algorithms to detect normal and spikes. (Fragkioudaki et al., 2015) proposed a method to predict regular electricity prices on DAM and spikes in developing countries by means of classification and regression trees, using the historical prices of the European electricity markets and transmission capacities.
- -Keles et al. (2016) proposed a methodology based on Artificial Neuronal Networks (ANN) to forecast the electricity prices. They reinforced the idea that the selection and preparation of fundamental variables (such as load, wind, and solar generation, available capacities) have a significant impact on the price forecast. The selection and preparation of input variables were partially performed by means of several cluster algorithms. However, the proposed forecasting model was tested for a shorter interval with almost no random event from January to September 2013.
- A model that integrates four components: the improved empirical mode decomposition, exponential generalized autoregressive conditional heteroscedasticity, autoregressive moving average with exogenous terms, and adaptive network-based fuzzy inference system is proposed (Zhang et al., 2019) using the electricity price and demand from Spain and Australia for 2013–2016. Furthermore, (Ziel & Steinert, 2016) considered the auctioning data, using dimension reduction and lasso based estimation methods. They combined traditional features, such as renewables generation and the bidding structure for the day-ahead electricity prices in Germany and Austria. The main drawback is that the individual bids and offers are not open data. Additionally, the investigated interval was limited to November 2014—April 2015. From the performance point of view, we outperformed the analyzed similar works, especially we compared the results with the most recent and relevant ones (Ziel & Steinert, 2016), (Zhang et al., 2019), (Maciejowska et al., 2021) and (Tschora et al., 2022).
- -Gabielli et al. (2022) proposed a market data-driven model for the long-term prediction of electricity prices using Fourier analysis. It decomposed the price into two components: base evolution depicted by its amplitudes of the main frequencies of the Fourier series, and spikes or high price volatility that are predicted with a regression model using the electricity generation, prices, and demands. The data sets were collected from Eurostat, British Petroleum, Department for Business, Energy & Industrial Strategy BEIS, European Commission and National Grid, and the analyzed interval spanned from 2015 to 2019. Electricity demand, generation, imports, generation mix, wind, solar, biomass, hydro, geothermal, nuclear, natural gas, oil, coal generation, fossil fuel prices, natural gas, oil, coal, carbon price are the fundamental features, and the forecast is performed for the United Kingdom, Germany, Sweden, and Denmark.
- A comparison between time series and neural network models with external regressors in order to estimate the day-ahead electricity prices is provided in (Lehna et al., 2022). The German electricity price is predicted using the Seasonal Integrated Auto-Regressive Moving Average (SARIMA) model and Long-Short Term Memory (LSTM) neural network models, Convolutional Neural Network-LSTM and Vector Auto-Regressive model (VAR), including external variables, such as: consumption, fuel and CO<sub>2</sub> emission prices, average solar radiation and

wind speed. The data set spanned from October 2017 to September 2018 and the forecast horizons are variate from one day and seven days to one month. The LSTM model was the best on average, but VAR follows closely, being better for shorter forecasting horizons. The authors found that a combination of both types of forecasting methods outperforms individual models.

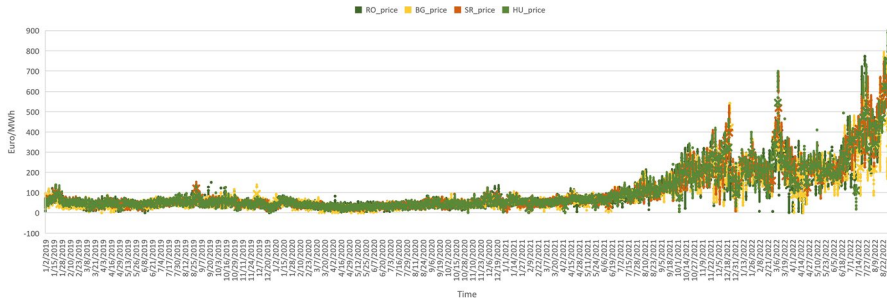
- Lu et al. (2022) proposed conditioning time series generative adversarial networks on external factors to estimate the electricity prices on the Australian DAM for a large interval 2000–2020. By changing the dimensionality of random noise input, the model was transformed into a probabilistic price forecasting model. The proposed model is compared with other models, such as ARIMA, LSTM, LASSO, etc. The Australian National Electricity Market provided as data input the historical prices, reserve capacity, consumption, renewables proportion, and tie-line flow from January to December 2019. A Multivariate Logistic Regression was proposed (Liu et al., 2022), which was compared with the Multi-Layer Perceptron (MLP) and Radical Basis Function neural network. An efficient forecasting solution consists of combining the extreme learning machine and building a hybrid model (Zhang et al., 2022). It was tested to predict the electricity prices of Australia from December 2019 and March 2020 and Spanish DAMs from August 2019 and March 2020.

Compared to the above papers, this paper proposes a machine learning-based approach to EPF, as can be seen in other similar papers, some of them highlighting the advantages of using ML over conventional prediction processes. But the proposed methodology has several contributions highlighted in the previous section.

Another element of novelty is that, to the best of our knowledge, the East-European market region was not analyzed for such longer interval from January 2019 to August 2022, that faced the pandemic, lockdowns, the economy slowdowns and acceleration after restrictions, and the conflict in Ukraine. Most of the studies analyzed data sets that spanned up to March 2020, but exactly the following months led to higher fluctuations and electricity price soaring. In addition, interesting insights and conclusions for both researchers, planners and decision makers are underlined in the current paper. It provides a robust method for electricity price forecast that is reliable for both more stable years (2019, 2020) in terms of random events and turbulent ones (2021, 2022).

### 3 Input Data

At the beginning of 2021, when most of the lockdowns were finished, the economic activities sharply increased in intensity and people started to travel and companies make business in an attempt to recover the wasted time and losses during the pandemic. This was also doubled by the increasing drought in European countries and the strategies to replace polluting technologies (thermal power plants based on coal) with sustainable ones. The prices for CO<sub>2</sub> certificates increased imposing a supplementary burden on the electricity price. Furthermore, the transition to new green technologies implies costs that give an impetus to the increase in prices. In



**Fig. 1** Daily prices from 2019 to 2022 in Romania and its neighboring countries

Figs. 1 and 2, the price evolution from January 2019 to August 2022 is depicted for Romania and its neighboring countries. These countries are very close to the conflict and were similarly affected in terms of the electricity prices in DAM. The data was extracted from the market operators of Romania (OPCOM),<sup>3</sup> Hungary (HUPX),<sup>4</sup> Bulgaria (IBEX)<sup>5</sup> and Serbia (SEEPX).<sup>6</sup> To extract data sets for Romania and Serbia, web scraping (using *BeautifulSoup* and *Selenium* Python libraries) was necessary as a download or APIs were not an option. In Fig. 3, the input variables are presented in a heatmap. The prices (in Romania, Serbia, Hungary, Bulgaria, Slovakia and Czech Republic); traded quantities (in Romania, Bulgaria and Serbia); flows RO-HU and HU-RO are depicted from the Pearson correlation point of view. Similarities can be noticed among the electricity price evolution on DAM in the four countries.

The much higher demand for goods and services led to inflation. This context was favorable for an accelerated increase in electricity prices in Romania from an average of 39.06 in 2020 to 112.15 Euro/MWh in 2021. This trend was also noticed in electricity DAMs of the neighboring countries, the stress of the pandemic and the conflict in Ukraine affected the entire region regardless the affiliation of the country to the integrated market coupling system. The prices further increase to an average of 266.62 Euro/MWh up to August 2022 (as in Fig. 4). Furthermore, an impressive deviation of the Romanian electricity prices took place in 2021 and 2022. It increased from 17.4 in 2020 to 80.14 in 2021 and 144.96 in 2022 (as in Fig. 5).

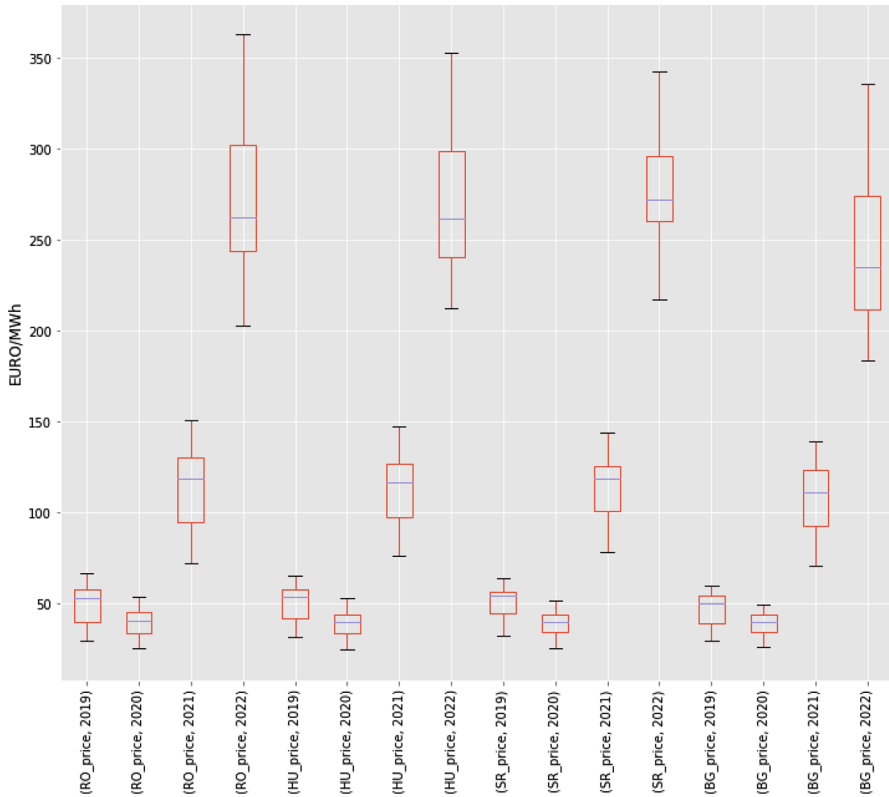
The electricity prices on DAMs are strongly correlated. Additionally, the electricity price in the Romanian DAM for day  $d$  is strongly correlated with prices of neighboring countries recorded in the previous day. With these coefficients (as in Table 1), we could rotate the target and predict the electricity price for each DAM.

<sup>3</sup> [https://www.opcom.ro/pp/grafice\\_ip/raportPIPsiVolumTranzactionat.php?lang=ro](https://www.opcom.ro/pp/grafice_ip/raportPIPsiVolumTranzactionat.php?lang=ro)

<sup>4</sup> <https://hupx.hu/en/market-data/dam/historical-data>

<sup>5</sup> <https://ibex.bg/markets/dam/day-ahead-prices-and-volumes-v2-0-2/>

<sup>6</sup> <http://seepex-spot.rs/en/market-data/day-ahead-auction>



**Fig. 2** Distribution of electricity prices in DAM (2019–2022) in Romania and neighboring countries

However, the results indicate that the random events are impacting the prices and reflect the distress of the society in terms of rising prices and scarcity of primary resources. As the tests indicated, no additional variable such as price index, interest rate, or power system parameters (such as total consumption, generation, or its breakdown) is necessary for the prediction process.

## 4 EPF Methodology

The proposed EPF method relies on a methodology that consists in several steps that first extract, prepare and configure the input data of the ML models and then train them, compute the individual and average forecast, evaluate the accuracy of the models, and then compute the weighted forecast using linear regression and decision tree regressor to adjust the individual forecast and weight it to further improve the results. For training, the interval is adjusted considering an empirical notice that the higher the deviation in the previous month, the shorter the interval. For all the five ML techniques described in a following section, we followed the usual approach for machine learning: Design

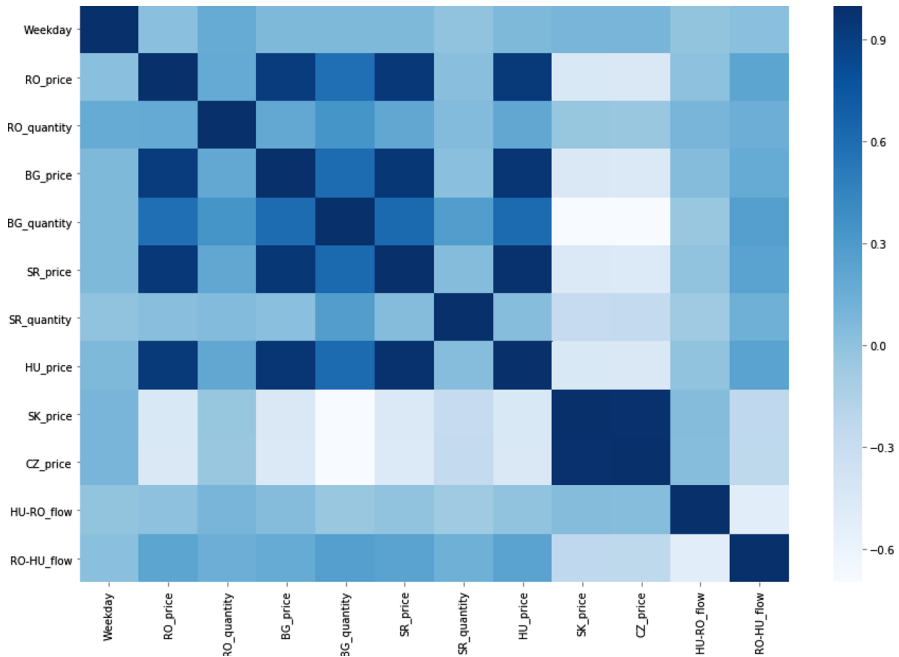


Fig. 3 Heatmap of the input variables

a specific question based on the available data and the required analysis; Convert the data into the required format; Search for noticeable anomalies and missing data points; Create the machine learning model; Set the desired baseline model; Train the data machine learning model; Provide an insight into the model with test data; Compare the performance metrics of both the test data and the predicted data from the model; Model improvement, if required; Data interpretation and report. ML techniques were previously successfully used in previous works in related fields (Irfan et al., 2021; Mas-saoudi et al., 2021; Punmiya & Choe, 2019; Wang et al., 2019).

### 4.1 ML Algorithms Configuration and New Features

The input of the ML algorithms is initially composed of the 14 raw features that represent the recorded hourly values of the previous day ( $h - 24 \in d - 1$ ). The weekday ( $W^d$ ) and the hour ( $h$ ) are also extracted from the date and added to the input variable to capture the time variations.

$$X^h = [Q_{RO}^{h-24}, Q_{HU}^{h-24}, P_{HU}^{h-24}, Q_{SR}^{h-24}, P_{SR}^{h-24}, Q_{BG}^{h-24}, P_{BG}^{h-24}, P_{SK}^{h-24}, P_{CZ}^{h-24}, Flow_{RO-HU}^{h-24}, Flow_{HU-RO}^{h-24}, W^d, h] \tag{1}$$

The input is filled with 13 more engineered features obtained by aggregating the previous day’s hourly prices and by determining the range and variations of the current prices versus the previous day’s prices.

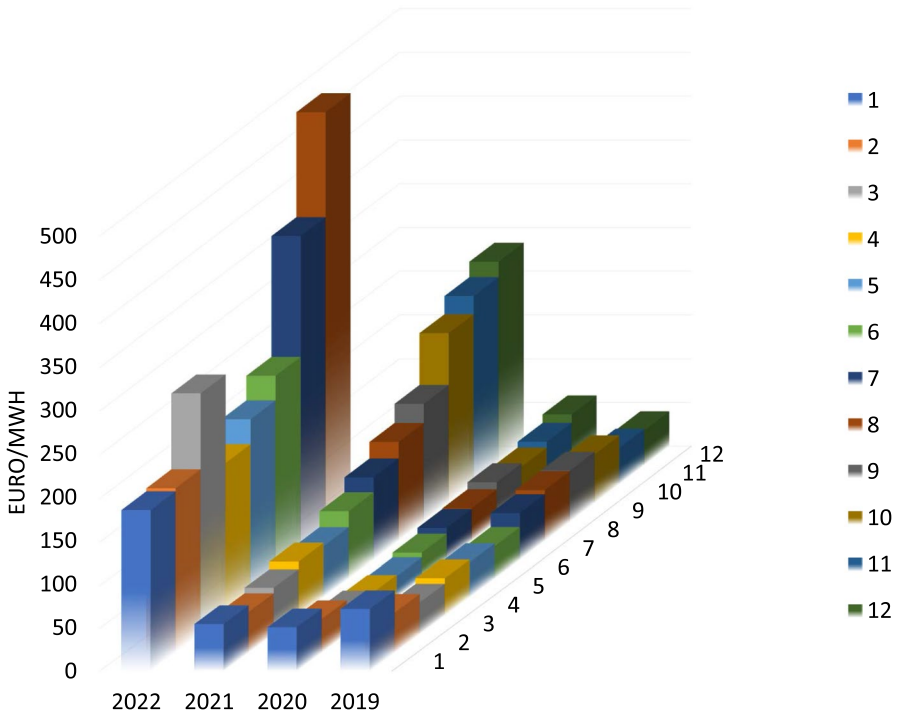


Fig. 4 Monthly average prices in Romania from 2019 to August 2022

The following six variables are calculated for the previous day ( $d-1$ ): minimum and maximum price, mean price, variance, and standard deviation, and median of the previous day's prices.

$$P_{min}^{d-1} = \min_{h|h \in d-1} P_{RO}^h; P_{max}^{d-1} = \max_{h|h \in d-1} P_{RO}^h \tag{2}$$

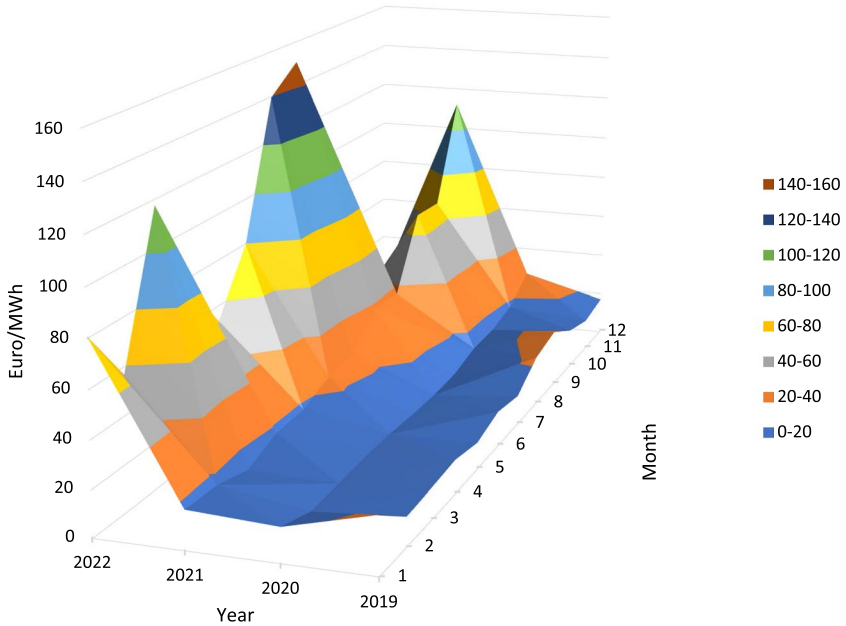
$$\mu^{P^{d-1}} = \frac{1}{24} \sum_{h|h \in d-1} P_{RO}^h \tag{3}$$

$$\sigma^2 P^{d-1} = \frac{1}{24} \sum_{h|h \in d-1} (P_{RO}^h - \mu^{P^{d-1}})^2; \sigma P^{d-1} = \sqrt{\sigma^2 P^{d-1}} \tag{4}$$

$$\widetilde{P}^{d-1} = \text{median}_{h|h \in d-1} P_{RO}^h \tag{5}$$

These variables are added to the initial input  $X^h$ :

$$X^h = [X^h, P_{min}^{d-1}, P_{max}^{d-1}, \mu^{P^{d-1}}, \sigma^2 P^{d-1}, \sigma P^{d-1}, \widetilde{P}^{d-1}] \tag{6}$$



**Fig. 5** Monthly standard deviation of the electricity prices on DAM in Romania from 2019 to August 2022

**Table 1** Correlation coefficients between the electricity price of Romania and its neighbors

	RO_price	SR_price	BG_price	HU_price
RO_price	1			
SR_price	0.941106	1		
BG_price	0.915517	0.952917	1	
HU_price	0.937248	0.984877	0.96055	1

Based on the aggregated values of the previous day, in this step, seven new variables are calculated and added to the model. Range price is obtained in Eq. (7) as the difference between the maximum and the minimum price of the previous day:

$$P_{range}^{d-1} = P_{max}^{d-1} - P_{min}^{d-1} \tag{7}$$

The ratio price index is obtained in Eq. (8) as the ratio between the maximum and the average price of the previous day:

$$P_{ratio}^{d-1} = \frac{P_{max}^{d-1}}{\mu P^{d-1}} \tag{8}$$

Previous hourly prices with a lag of 24–168 h are also calculated by shifting the prices:  $P_{RO}^{h-24}$ ,  $P_{RO}^{h-48}$ ,  $P_{RO}^{h-72}$ ,  $P_{RO}^{h-96}$ ,  $P_{RO}^{h-120}$ ,  $P_{RO}^{h-144}$ ,  $P_{RO}^{h-168}$ . The averages of the previous hourly prices for 3 and 7 consecutive days are calculated using Eq. (9):

$$\mu P^{h_3} = \frac{\sum_{i \in \{24,48,72\}} P_{RO}^{h-i}}{3} \tag{9}$$

$$\mu P^{h_7} = \frac{\sum_{i \in \{24,48,72,96,120,144,168\}} P_{RO}^{h-i}}{7} \tag{10}$$

The hourly variations of the previous prices for the last 3 days are determined using Eq. (10):

$$\Delta = P^{h_1} = 100 - \frac{P_{RO}^{h-24}}{P_{RO}^{h-48}} \times 100; \Delta P^{h_2} = 100 - \frac{P_{RO}^{h-48}}{P_{RO}^{h-72}} \times 100 \tag{11}$$

Finally, the input of the ML algorithms is completed with the above-calculated features:

$$X^h = [X^h, P_{RO}^{h-24}, P_{RO}^{h-48}, P_{RO}^{h-72}, P_{RO}^{h-96}, P_{RO}^{h-120}, P_{RO}^{h-144}, P_{RO}^{h-168}, P_{range}^{d-1}, P_{ratio}^{d-1}] \tag{12}$$

$$X^h = [X^h, \mu P^{h_3}, \mu P^{h_7}, \Delta P^{h_1}, \Delta P^{h_2}, \Delta P^{h_3}] \tag{13}$$

The actual electricity hourly prices constitute the target variable ( $y^h = P_{RO}^h$ ) of the ML algorithms.

### 4.2 Training and Evaluation of the ML Algorithms

The following five ML algorithms were used: Random Forest (RF), Light Gradient Boosting Regressor (LGBR), Histogram-Based Gradient Boosting Regressor (HGBR), eXtreme Gradient Boosting (XBR), and Voting Regressors (VR). We used the following ML packages for Python: SciKit Learn (sklearn),<sup>7</sup> XGBoost,<sup>8</sup> and LightGBM<sup>9</sup> (both XGBoost and LightGBM are wrapper interfaces for Scikit-Learn API).

Random Forest (RF) is an ensemble technique capable of performing both regression and classification tasks using multiple decision trees and a technique called Bootstrap and Aggregation, which is also commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final result, rather than relying on individual decision trees. RF has multiple decision trees as basic learning models. Random row sampling and feature sampling from the dataset forming sample datasets are performed for each model. This part is called Bootstrap

<sup>7</sup> <https://scikit-learn.org/>

<sup>8</sup> <https://xgboost.readthedocs.io/en/stable/index.html>

<sup>9</sup> <https://lightgbm.readthedocs.io/en/latest/index.html>



(Brownlee, 2020a; GeeksforGeeks, 2022; Great Learning, 2022). The purpose of these two randomizations is to decrease the variance of the forest estimator—the injected randomness in forests yields decision trees with somewhat decoupled prediction errors, which can be canceled out by taking an average of the predictions. Random forests achieve a reduced variance by combining diverse trees, sometimes at the cost of a slight increase in bias. In practice, the variance reduction is often significant, hence yielding an overall better model (SciKit, 2022). For our data analysis, we used the `RandomForestRegressor` class from `sklearn`.<sup>10</sup> In contrast to the original proposal, (Breiman, 2001), the `scikit-learn` implementation combines classifiers by averaging their probabilistic prediction, instead of letting each classifier vote for a single class (SciKit, 2022).

Gradient Boosting Decision trees (GBDT) are popular ML algorithms, although they are not very performant—each feature should scan all the various data instances to make an estimate of all the possible split points which is very time-consuming and tedious. To solve this problem, the Light Gradient Boosting Model (LGBM) is used. It uses two types of acceleration techniques. One of them directly replaces GBDT with the more efficient GOSS (Gradient-based One-Side Sampling), which will exclude the significant portion of the data part that has small gradients and only use the remaining data to estimate the overall information gain. The data instances that have large gradients actually play a greater role in the computation of information gain. GOSS can get accurate results with significant information gain despite using a smaller dataset than other models. The other one is Exclusive Feature Bundling (EFB), which puts the mutually exclusive features along with nothing but it will rarely take any non-zero value at the same time to reduce the number of features. This affects the overall result for an effective feature elimination without compromising the accuracy of the split point (Guolin Ke et al., 2017; Mondal, 2021). LGBM seems to be the first algorithm (chronologically) in a set of several competing techniques, all attempting to improve in various ways (mostly by binning) over the original GBDT algorithms. Two other algorithms working on the same principle, HGBM, and XGBM, are used for this study and presented further. For our data analysis, we used the `LGBMRegressor` class from `LightGBM`.<sup>11</sup>

`Scikit-learn` 0.21 introduced two new implementations of gradient boosting trees, which are named `HistGradientBoostingClassifier` and `HistGradientBoostingRegressor` (HGBM). These histogram-based estimators are estimated to be orders of magnitude faster than `GradientBoostingClassifier` and `GradientBoostingRegressor` when the number of samples is larger than tens of thousands of samples. These fast estimators first bin the input samples  $X$  into integer-valued bins (typically 256 bins), which greatly reduces the number of splitting points to consider and allows the algorithm to leverage integer-based data structures (histograms) instead of relying on sorted continuous values when building the trees. The API of these estimators is slightly different than that of `GradientBoostingClassifier` and `GradientBoostingRegressor`, and some of the old features are not yet supported, for instance, some loss functions (SciKit Learn, 2022a). For our data analysis, we

<sup>10</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

<sup>11</sup> <https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMRegressor.html>

used the `HistGradientBoostingRegressor` class from `sklearn`.<sup>12</sup> eXtreme Gradient Boosting (XBR) is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements ML algorithms under the Gradient Boosting framework. XBR provides a parallel tree boosting (also known as GBDT, GBM) that solves many data science problems in a fast and accurate way. The same code runs on some major distributed environments (Hadoop, SGE, MPI), and can solve problems with very large data sets (millions to billions of rows) (`dmlc XGBoost`, 2022). XBR gained significant favor in the last few years as a result of helping individuals and teams win virtually every Kaggle structured data competition (Vidia, 2022). While LGBR was the first to be introduced, and HGBR followed, XBR, the last to appear, is sought to be the most performant of the three algorithms (but in our study, due to the reasonable amount of data, there were no noticeable differences on this aspect). For our data analysis, we used the `XGBRegressor` class from `XGBoost`.<sup>13</sup>

A Voting Regressor (VR) is an ensemble meta-estimator that fits several base regressors, each on the whole dataset. Then it averages the individual predictions to form a final prediction. The idea behind VR is to combine conceptually different ML regressors and return the average predicted values. Such a regressor can be useful for a set of equally well-performing models in order to balance out their individual weaknesses (SciKit Learn, 2022b, c). A voting ensemble may be considered a meta-model, a model of models. As a meta-model, it could be used with any collection of existing trained ML models and the existing models do not need to be aware that they are being used in the ensemble. A voting ensemble is appropriate when two or more models, that perform well on a predictive modeling task, are used. The models used in the ensemble must mostly agree on their predictions (Brownlee, 2020b), which is true for our research, as proven by the results shown in Sect. 5. For our data analysis, we used the VR class from `sklearn`.<sup>14</sup>

#### 4.2.1 Adjusting the Training Interval and Weighing the Results of the Individual ML Algorithms

The over standard deviation evolution in time made us consider adjusting the trading interval  $T$  according to Eq. (14). If the deviation in the previous month was less than 25, the training interval was set to 75 days, while if the deviation is greater than 100, it was set to 20 days. Thus, the higher the deviation, the shorter the training interval.

$$T = \begin{cases} 75, & \text{if } \sigma P^{m-1} < 25 \\ 45, & \text{if } \sigma P^{m-1} \in [25, 50] \\ 30, & \text{if } \sigma P^{m-1} \in [51, 99] \\ 20, & \text{if } \sigma P^{m-1} \geq 100 \end{cases} \quad (14)$$

<sup>12</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingRegressor.html>

<sup>13</sup> [https://xgboost.readthedocs.io/en/stable/python/python\\_api.html#xgboost.XGBRegressor](https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.XGBRegressor)

<sup>14</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html>

The ML algorithms are trained on the  $T$  interval and provide for the next 24 h an individual forecast denoted by  $\widehat{y}_i^{hf}, \forall hf \in d + 1, i \in \{RF, LGBR, HGBR, XGB, VR\}$ . These estimations are further weighted with a set of weights ( $\omega_i$ ) and used as input for another weighting ML algorithm that adjusts  $\widehat{y}_i^{hf}$  of the initial five ML models in order to forecast the hourly electricity price for the next 24 h and several days:

$$\widehat{P}_{RO}^{hf} = \sum_i \omega_i \times \widehat{y}_i^{hf} \tag{15}$$

The ML algorithm determines the weights using Linear Regression (LR) or Decision Tree (DT) regressor by minimizing the sum of the squared errors during the training interval ( $T$ ) using Eq. (16):

$$\omega_i = \operatorname{argmin}_{\omega_i} \sum_{h \in T} \left( P_{RO}^h - \sum_i \omega_i \times \widehat{y}_i^h \right)^2 \tag{16}$$

### 4.2.2 Rotation of the Variables

The strong correlation among DAM prices in the Eastern-European region indicates a potential to rotate variables and rely on their reciprocal influence. Therefore, the dependent variable could be any of the prices in the region. For example, in our case study, the dependent variable was the Romanian price ( $P_{RO}^h$ ), and the other prices (exogenous) were shifted to the previous day ( $P_{HU}^{h-24}, P_{SR}^{h-24}, P_{BG}^{h-24}, P_{CZ}^{h-24}, P_{SK}^{h-24}$ ). But we may consider estimating the electricity price on DAM in Hungary ( $P_{HU}^h$ ). Therefore, it will be kept for day  $d (\forall h \in d)$  and the rest of the prices are shifted to day  $d-1 (\forall h - 24 \in d - 1)$ . The new input is transformed as in Eq. (17):

$$X^h = [Q_{RO}^{h-24}, Q_{HU}^{h-24}, P_{RO}^{h-24}, Q_{SR}^{h-24}, P_{SR}^{h-24}, Q_{BG}^{h-24}, P_{BG}^{h-24}, P_{SK}^{h-24}, P_{CZ}^{h-24}, Flow_{RO-HU}^{h-24}, Flow_{HU-RO}^{h-24}, W^d, h] \tag{17}$$

Replacing  $P_{RO}^h$  with  $P_{HU}^h$  in Eqs. (2–11), the final input  $X^h$  becomes:

$$X^h = [X^h, P_{HU}^{h-24}, P_{HU}^{h-48}, P_{HU}^{h-72}, P_{HU}^{h-96}, P_{HU}^{h-120}, P_{HU}^{h-144}, P_{HU}^{h-168}, P_{range}^{d-1}, P_{ratio}^{d-1}] \tag{18}$$

$$X^h = [X^h, \mu P^{h_3}, \mu P^{h_7}, \Delta P^{h_1}, \Delta P^{h_2}, \Delta P^{h_3}] \tag{19}$$

The target of the ML algorithms also changes and becomes  $y^h = P_{HU}^h$ .

### 4.2.3 Evaluation of the Accuracy of the Models

To evaluate the accuracy of the ML models, the following metrics are calculated for the training and testing process: Root-Mean Squared Error (RMSE), coefficient of determination ( $R^2$ ), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE):

$$RMSE = \frac{1}{T} \sum_{h=1}^T (y^h - \hat{y}_i^h)^2 \quad (20)$$

$$R^2 = 1 - \frac{\sum_{h=1}^T (y^h - \hat{y}_i^h)^2}{\sum_{h=1}^T (y^h - \bar{y}^h)^2} \quad (21)$$

$$MAPE = \frac{1}{T} \sum_{h=1}^T \frac{|y^h - \hat{y}_i^h|}{|y^h|} \times 100\% \quad (22)$$

$$MAE = \frac{1}{T} \sum_{h=1}^T |y^h - \hat{y}_i^h| \quad (23)$$

## 5 Results

### 5.1 Romanian DAM Price Forecasting Case Study

After splitting the data set as usual into training (80%) and testing (20%), the algorithms provided the results that are graphically presented for four consecutive days in March (from 16 to 19) for each year 2019–2022 as in Figs. 6, 7, 8, and 9. It is worth noting that the proposed method (the blue curve *RO\_price\_PF*) is compared with the individual (*P1–P5*) and an average forecast of the five ML algorithms (the red curve *RO\_price\_F*). *P1–P5* represent the predictions obtained with the five ML algorithms.

The prediction is very good; both spikes and valleys are approximated with an individual estimation of the ML algorithms and also with the average and weighted forecast. Very small deviations from the actual curves were observed for days in March 2019, but on average the estimation was performant. Furthermore, in 2020, the prediction for the same four consecutive days is accurate.

Small deviations were recorded on 16 of March 2020. In 2021, for four consecutive days (but also for the rest of the year), the results are promising, and the prediction is almost perfect. The EPF has not encountered obstacles in estimating spikes and valleys. Higher deviations between actual prices and predictions were recorded in 2022 for two days out of the six consecutive days (19, 20 March), but on average the estimation is good.

One can notice that each of the five algorithms provides the best results at certain intervals. Moreover, it is difficult to follow and chose one of the five curves that are the predicted prices obtained with the ML algorithms. Thus, two approaches to mediate the results are considered: average and weighting the results of the five algorithms. In Figs. 10, 11, 12, and 13, three curves are depicted: the actual prices—the

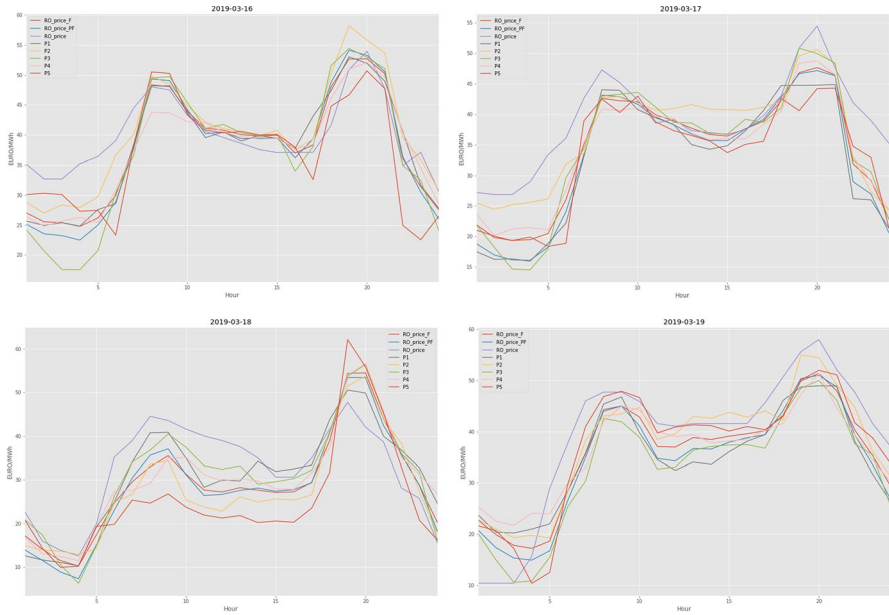


Fig. 6 Hourly RO price forecast in Euro/MWh for 4 consecutive days in March 2019

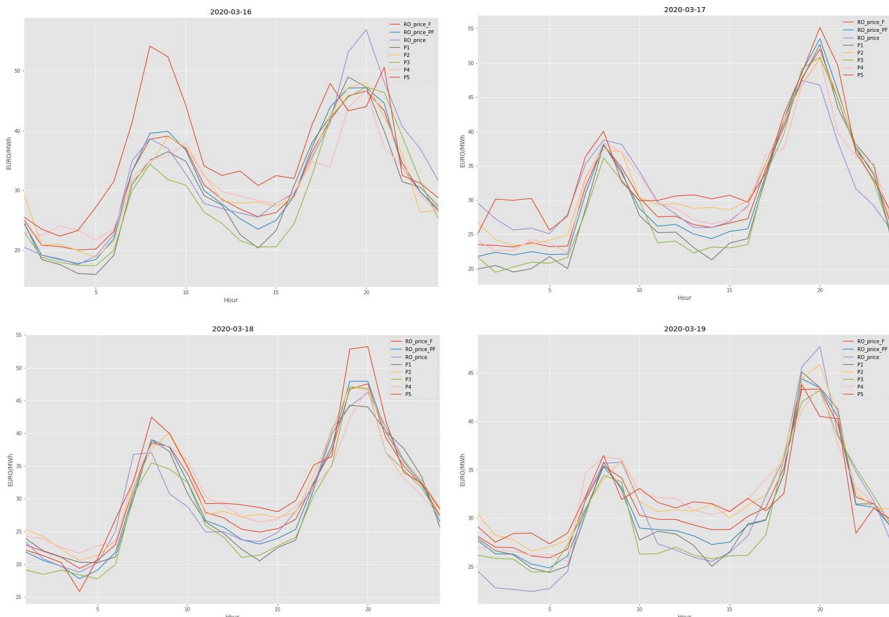


Fig. 7 Hourly RO price forecast in Euro/MWh for 4 consecutive days in March 2020

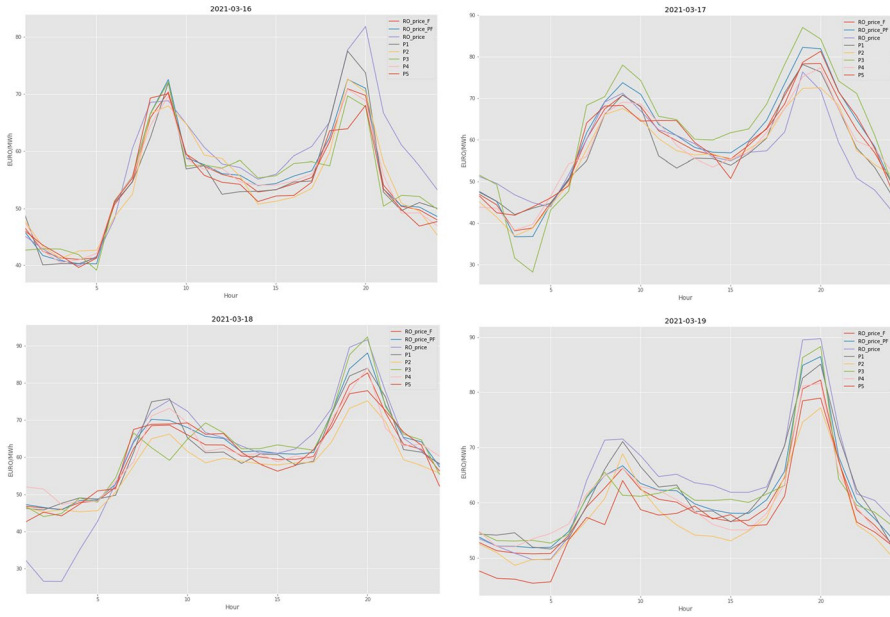


Fig. 8 Hourly RO price forecast in Euro/MWh for 4 consecutive days in March 2021

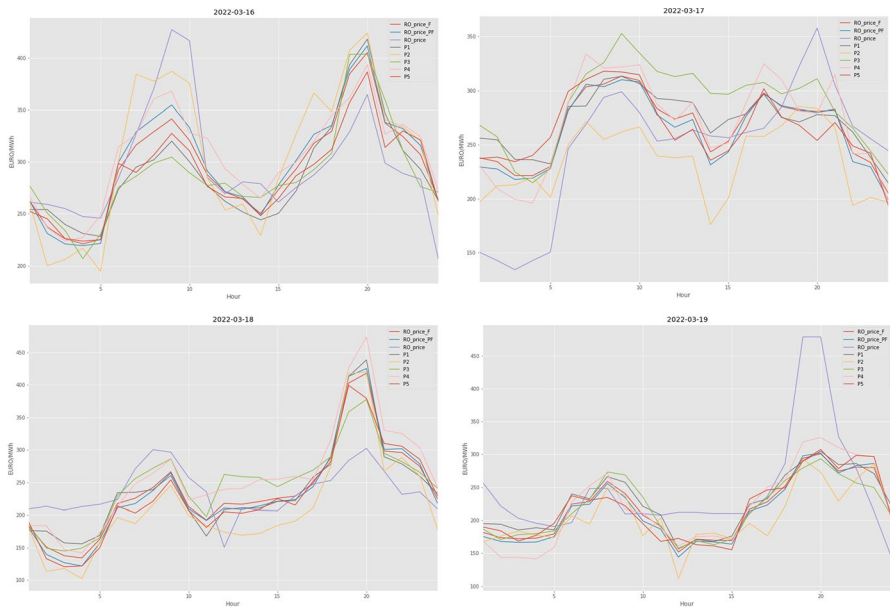


Fig. 9 Hourly RO price forecast in Euro/MWh for 4 consecutive days in March 2022

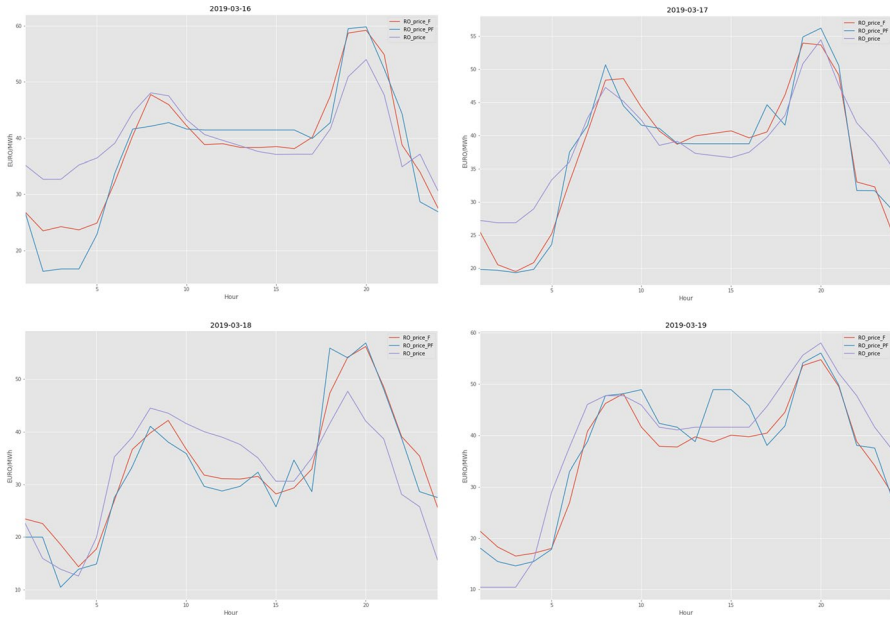


Fig. 10 Average (red) and weighted results of ML algorithms (blue) for 4 days in March 2019 with DT. (Color figure online)

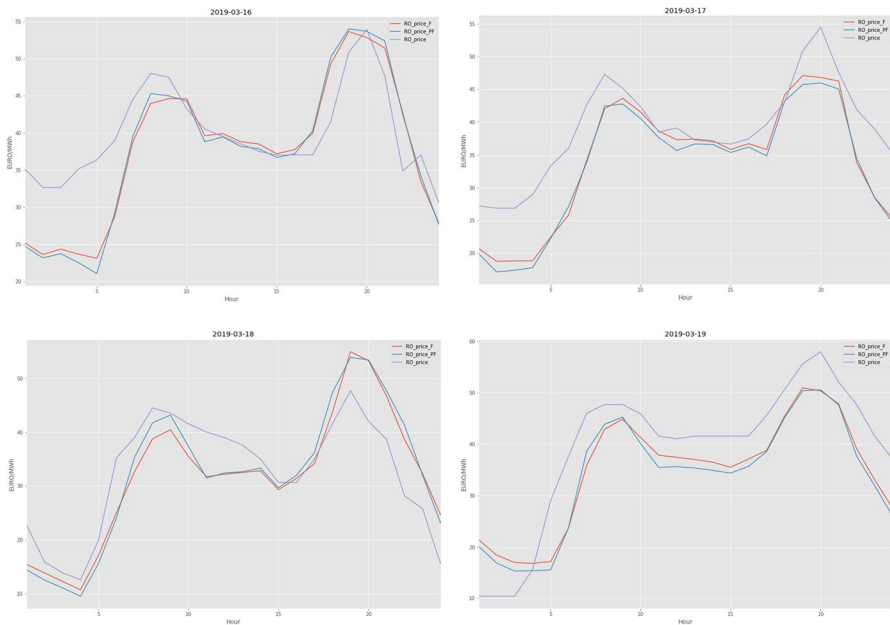
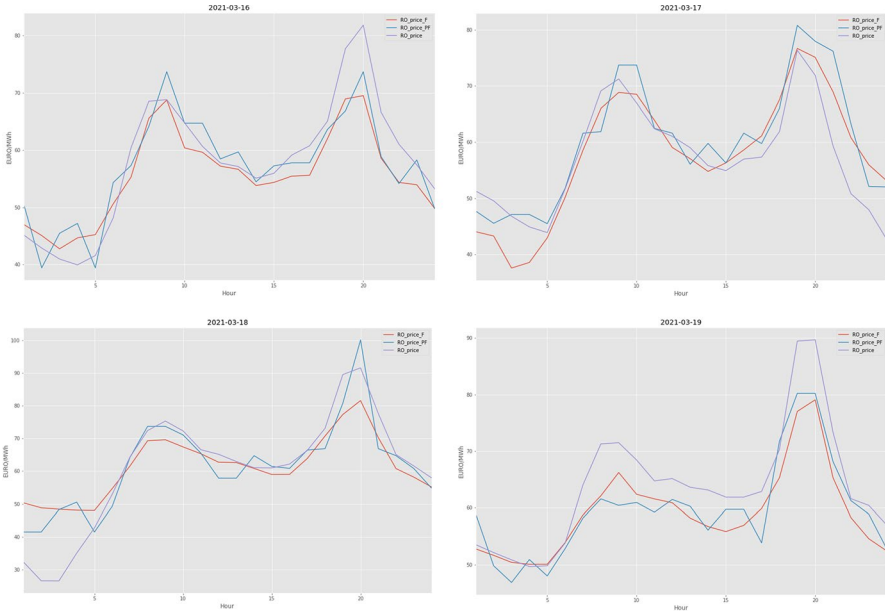
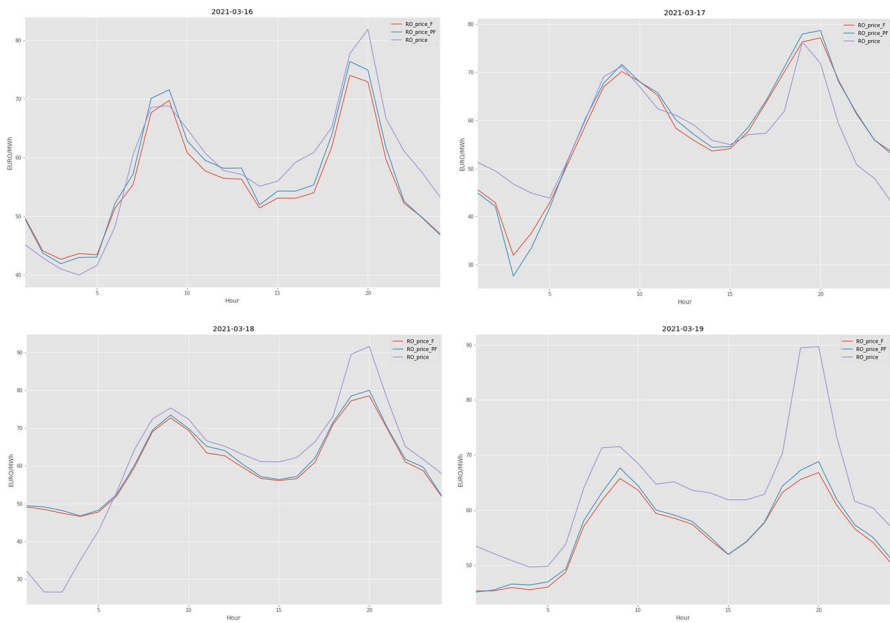


Fig. 11 Average (red) and weighted results of ML algorithms (blue) for 4 days in March 2019 with LR. (Color figure online)



**Fig. 12** Average (red) and weighted results of the ML algorithms (blue) for 4 days in March 2021 with DT. (Color figure online)



**Fig. 13** Average (red) and weighted results of ML algorithms (blue) for 4 days in March 2021 with LR. (Color figure online)



**Table 2** Performance indicators for 2019—August 2022

Metric/Year	2019		2020		2021		2022	
	Baseline	EPF	Baseline	EPF	Baseline	EPF	Baseline	EPF
R <sup>2</sup>	0.88	0.98	0.89	0.98	0.74	0.95	0.71	0.93
MAE	6.89	2.58	7.24	2.71	8.34	5.25	10.83	6.03
RMSE	14.32	3.82	13.22	3.67	17.12	4.55	21.65	24.12
MAPE	10.98	3.11	8.98	3.37	13.94	5.05	17.01	6.89

purple curve, the average prediction—the red curve and the weighted curve—the blue curve. Out of the interval 2019–2022, two years are chosen: 2020—as more stable and 2021 as more turbulent, when the prices started to increase.

In Figs. 10, 11, 12, and 13, the average (red) and weighted results of the ML algorithms (blue) for the four consecutive days (from 16 to 19) in March 2020 and the same 4 consecutive days in March 2021, using Decision Tree (DT) regressor and Linear Regression (LR) for weighting the results of the individual ML algorithms (using Eqs. 15, 16) are graphically showcased. The DT regressor performs better than LR that is smoother, and it is chosen to calculate the performance indicators in Table 2. We tested the proposed method for the entire interval, from January 2019 to August 2022. To benchmark, we compare our method with the classic approach using ML algorithms (baseline), but without feature engineering, no adjustment of the training interval (fixed interval of 45 days) and no weighting of the individual estimations of the ML algorithms.

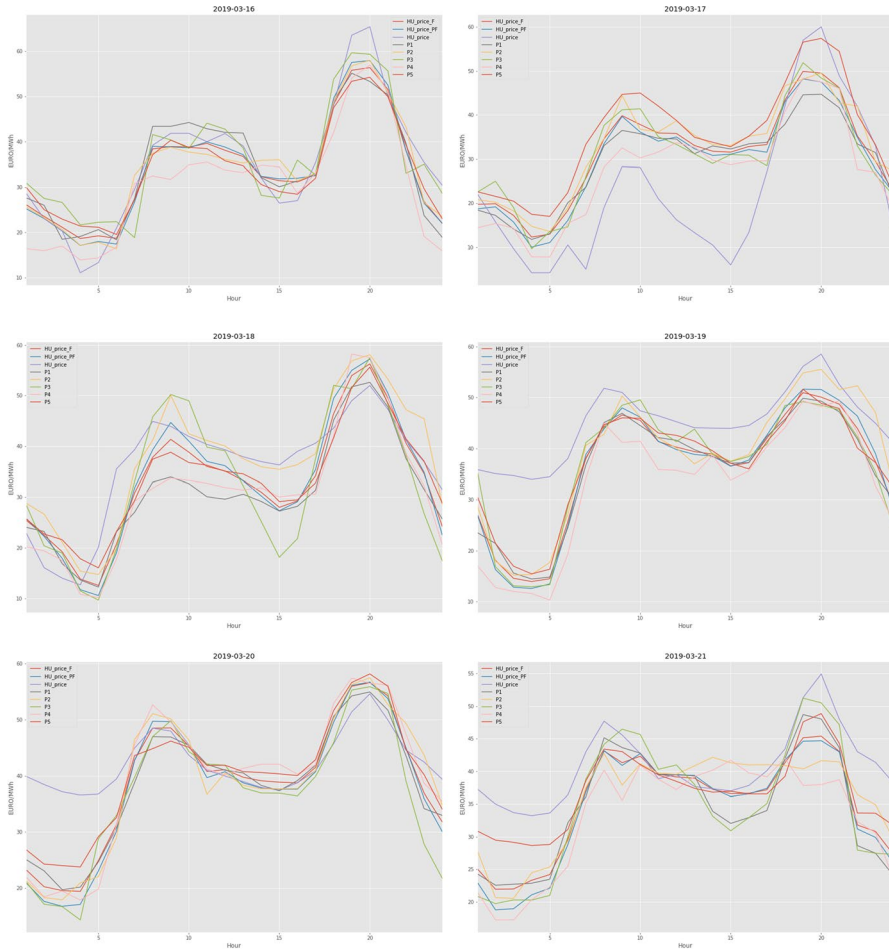
On average, all metrics improved with EPF. MAE enhances by 62.5% in 2019 and 2020, respectively by 37 and 44% in 2021 and 2022, whereas MAPE improves by 72 and 62% in 2019 and 2020, respectively by 64 and 59% in 2021 and 2022. It is obvious that the EPF is more efficient for peaceful intervals, but by adjusting the training intervals based on standard deviation of the prices in the previous month as in Eqs. (14)–(16) and inserting new features that consider price variations and high spikes, EPF proved efficient for more hectic years like 2021 and 2022.

The following performance indicators are obtained (as in Table 2) revealing that the proposed method (EPF) is superior to the classic approach or baseline.

## 5.2 Hungarian DAM Price Forecasting and Other Case Studies

Rotating the variables, we estimate the electricity prices on DAM in Hungary. Prices in Hungary are kept for day  $d$  and the previous prices in neighboring countries are shifted to  $d-1$ . Therefore, the exogenous variables are the prices for the previous day that are known on day  $d$ .

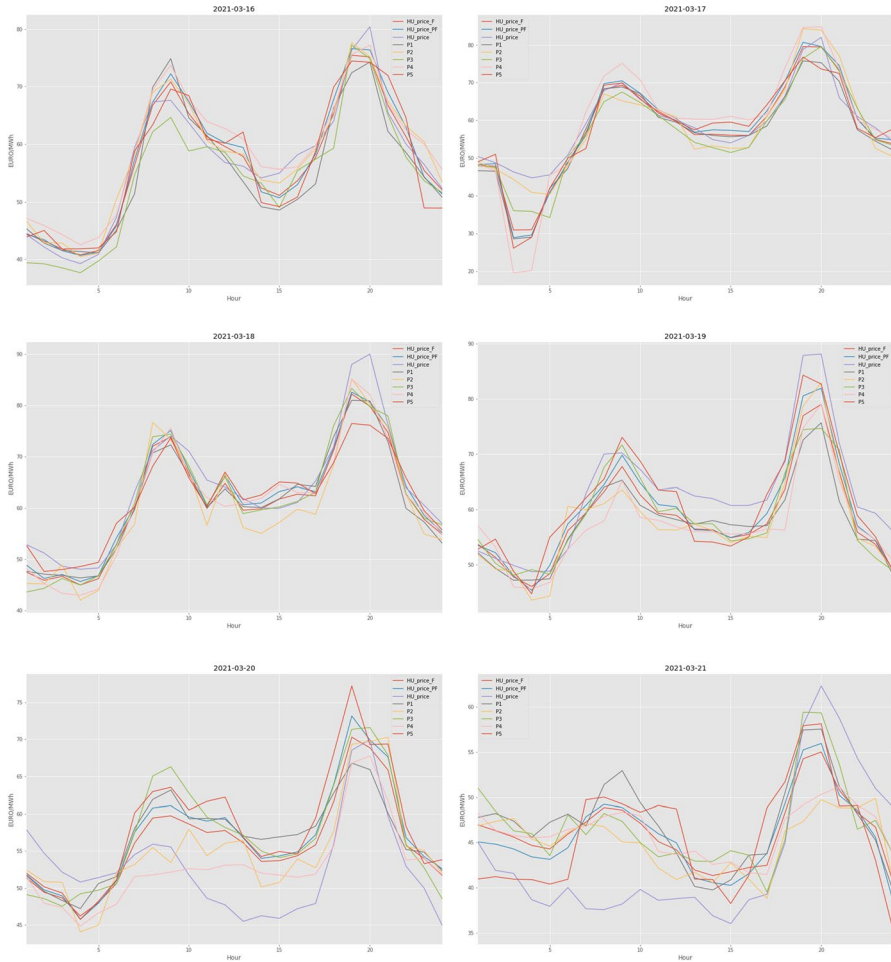
The graphical results are presented in Figs. 14 and 15 for six consecutive days in March 2019 and 2021. 2019 is chosen as a quieter interval, whereas 2021 is a more turbulent one. The training intervals are adjusted considering Eq. (14) and monthly standard deviations as in Fig. 16.



**Fig. 14** HU hourly price forecast in Euro/MWh for 6 consecutive days in March 2019

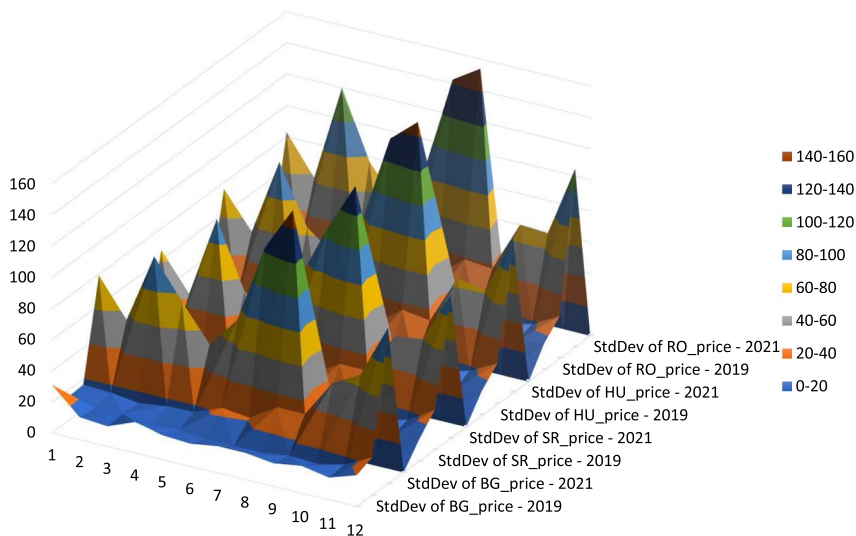
In March 2019, except the prediction for Hungary on 17-03-2019 from the analyzed interval, the rest of the five predictions are very closed to the actual values. On 17-03-2019, the morning peak and afternoon valley were lower than the predicted curves, but the evening peak overlaps the predictions. In March 2021, except the last two days from the analyzed interval, the predictions for Hungary are very close to the actual values. On 20-03-2022 and 21-03-2021, a similar issue occurred: the morning peak noon valley are slightly under the prediction curves, whereas the evening peak is much closer.

From Fig. 16, one can notice numerous similarities among the price standard deviations in the four analyzed countries. The highest standard deviations were recorded in the summer months (July and August) in 2022 when they went up to 160. A higher standard deviation was also recorded closer to the winter holidays in 2021.



**Fig. 15** HU hourly price forecast in Euro/MWh for 6 consecutive days in March 2021

The training interval is variable as in Eq. (14), and the forecasting horizon spans to six days, providing the estimation at hourly resolution for short- and mid-run intervals. As resulted from Figs. 14 and 15, excellent results are obtained for Hungary DAM prices. On average, the performance of the estimation is better with 1.5–2% compared with the results obtained in Table 2. Additionally, the Serbian and Bulgarian DAM forecasting price case studies are analyzed. In case of Serbian DAM, the performance metrics are lower by 2–3% in comparison with the Romanian DAM forecasting prices case. The errors are reasonable and far from the baseline case, but they are probably generated because the Serbian market is not operating in coupling mode.



**Fig. 16** Monthly standard deviations in Romania, Hungary, Serbia, and Bulgaria 2019–2022

## 6 Conclusions

Numerous changes took place in the last three years and a half. The most unexpected ones were COVID-19 pandemic and war in Ukraine, events that competed to increase prices on the electricity markets in the European countries. The traditional features that impacted prices lost their influence, but other features such as prices in neighboring countries became more correlated. Lockdowns influenced economies and business sector whereas the conflict in the Black Sea region had great influence on the prices of other energy resources such as gas and oil. The prices for certificates paid for CO<sub>2</sub> emissions were increasing to unprecedented levels and the drought in Europe also led to expensive generation based on coal, gas and oil. The prices for commodities also increased due to the rapid economy recovery and higher demand occasioned by the war necessities. In this context, the conventional tools for prediction based on deterministic methods and classical features are not useful due to the fact that the dynamics of the economies drastically changed.

This paper proposes a machine learning-based approach to EPF. The main novelty elements are as follows:

- The paper comparatively analyzes the effects of applying five different ML technologies to EPF. These methods are compared and analyzed from various points of view, including sensitivity.
- According to the authors' knowledge, this is the first paper that scrutinizes the Eastern European market for a longer time span, a span that includes the period of the pandemic and the conflict in Ukraine, with the significant market fluctuations induced by these two significant events.

This paper provides interesting insights into the correlation between electricity prices in Eastern-European neighboring countries. Despite some markets being coupled (Romania, Hungary, Slovakia, Czech, Bulgaria) and others not (Romania, Serbia), the prices for electricity in these countries are strongly correlated. This aspect is important because any electricity price can be considered as a target or dependent variable and the others as exogenous. Simulations were carried out for Romanian or Hungarian electricity prices as targets for the entire interval between January 2019 and August 2022, and very good results were obtained in both cases. This is especially relevant because most research studies and scientific papers focus on Nordic or Western European countries or Australia. Furthermore, previous studies analyzed historical data up to March 2020, while our study goes further and approaches the more feverish intervals starting from 2019 to August 2022. The results indicate that the proposed method is robust for both the stable intervals and the more turbulent intervals characterized by random events or by the so-called black swans (such as the COVID-19 pandemic and the conflict in Ukraine).

Moreover, this paper proposes a robust method for predicting electricity prices, since the price on the DAM has been increasing tremendously and showing more fluctuations year by year starting from 2019. The proposed method proved to accurately estimate hourly prices for both stable years (like 2019 and 2020) and more turbulent ones (2021 and 2022). Traditional variables do not ensure the quality of the prediction, especially in intervals with higher fluctuations. Therefore, the prices on DAMs in neighboring countries and flows on the tie-lines are considered as input variables.

We noticed that there is a very strong direct correlation among electricity prices on DAM in the neighboring countries, therefore, we checked the hypothesis that one price series can be the target at a time and proposed the rotating variables concept. The first case study aims to predict the electricity price on the Romanian DAM using the prices recorded on the previous day ( $d-1$ ) in adjacent countries (Hungary, Bulgaria, and Serbia). We also considered the flows on the line with Hungary, as the two countries are integrated, and even more regional prices from the neighboring countries of Hungary (Slovakia and the Czech Republic). Apart from that, feature engineering, which consists of aggregated features and derivatives, substantially increased the performance of the estimation.

The prices fluctuations have to be carefully analyzed as they influence the training interval. In this sense, we noticed that the best training interval can be obtained considering the price standard deviation of the previous month. The standard deviation of electricity prices in Romania varied from 17 in 2020 to 145 in 2022, indicating large fluctuations, especially in 2021 and 2022. Consequently, we proposed to adjust the training intervals, considering an empirical observation, from 20 to 75 days based on the previous month's deviation. The higher the deviations, the shorter the training interval. The concept of adjusting the training interval stems from the fact that in more disruptive intervals when the deviation exceeds a certain threshold, it is better to shorten the training interval to capture the speculative effect of random events on the DAM. Therefore, the training interval is variable, and the forecasting horizon spans to six days, providing performant estimation at an hourly resolution for short- and mid-run intervals.

Also, a rotation of the variables (i.e., prices on DAMs) was performed as a second case study. The initial target variable, or the electricity price on DAM in Romania, was replaced by the price in Hungary, and the exogenous variables (including Romanian hourly prices) were shifted accordingly to the previous day's known values. The results proved that the electricity price forecast on the Hungarian market is also reliable. Therefore, we tested the proposed method for several years, from January 2019 to August 2022, and obtained very good results in terms of MAE, MAPE, and other performance indicators for both DAMs in Romania and Hungary. The results in the Romanian DAM were compared to the no feature engineering and fixed training interval case (baseline). The EPF proved to be more efficient not only for more peaceful intervals but even so, on average, the MAE diminished by 37% in 2021 and 44% in 2022.

To enlarge the sensitivity analyses, the hourly prices in Serbia are also considered the target for prediction whereas the other prices are shifted by 24 h. We perform the price forecast for Serbian DAM to check the hypothesis that the variables can be rotated with reliable results. The performance for Serbian DAM was 2–3% lower in terms of MAE, but the results are better than the baseline and somehow expected as the Serbian DAM is not operating in coupling mode with the analyzed countries. Also, prediction for Bulgaria is performed and the results in terms of accuracy are 0.5% lower than accuracy obtained for Romanian price prediction. However, the best prediction on average was obtained for Hungary using the proposed adjusted training interval and building new features to capture the spikes.

**Acknowledgements** This work was supported by a grant from the Ministry of Research, Innovation and Digitization, CNCS- UEFISCDI, project number PN-III-P4-PCE-2021–0334, within PNCDI III.

**Funding** This work was supported by a grant from the Ministry of Research, Innovation and Digitization, CNCS- UEFISCDI, project number PN-III-P4-PCE-2021–0334, within PNCDI III.

## Declarations

**Conflict of interest** The authors declare that there is no conflict of interest.

## References

- Adolfson, J. F., Kuik, F., Lis, E. M., & Schuler, T. (2022). The impact of the war in Ukraine on euro area energy markets. *European Central Bank Economic Bulletin*. [https://www.ecb.europa.eu/pub/economic-bulletin/focus/2022/html/ecb.ebbox202204\\_01~68ef3c3dc6.en.html](https://www.ecb.europa.eu/pub/economic-bulletin/focus/2022/html/ecb.ebbox202204_01~68ef3c3dc6.en.html)
- Alkawaz, A. N., Abdellatif, A., Kanesan, J., Khairuddin, A. S. M., & Gheni, H. M. (2022). Day-ahead electricity price forecasting based on hybrid regression model. *IEEE Access*, *10*, 108021–108033. <https://doi.org/10.1109/ACCESS.2022.3213081>
- Bernal, B., Molero, J. C., & Perez De Gracia, F. (2019). Impact of fossil fuel prices on electricity prices in Mexico. *Journal of Economic Studies*, *46*(2), 356–371. <https://doi.org/10.1108/JES-07-2017-0198/FULL/PDF>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brownlee, J. (2020a). How to develop a random forest ensemble in python. *Machine Learning Mastery*. <https://machinelearningmastery.com/random-forest-ensemble-in-python/>

- Brownlee, J. (2020b). How to develop voting ensembles with python. *Machine Learning Mastery*. <https://machinelearningmastery.com/voting-ensembles-with-python/>
- Busse, J., & Rieck, J. (2022). Mid-term energy cost-oriented flow shop scheduling: Integration of electricity price forecasts, modeling, and solution procedures. *Computers and Industrial Engineering*. <https://doi.org/10.1016/j.cie.2021.107810>
- Butler, N. (2022). The impact of the Ukraine war on global energy markets. *Centre for European Reform*. <https://www.cer.eu/insights/impact-ukraine-war-global-energy-markets>
- Castelli, M., Groznic, A., & Popović, A. (2020). Forecasting electricity prices: A machine learning approach. *Algorithms*. <https://doi.org/10.3390/A13050119>
- dmlc XGBoost. (2022). *XGBoost Documentation*. Xgboost.Readthedocs.Io. <https://xgboost.readthedocs.io/en/stable/>
- Dragasevic, Z., Milovic, N., Djuricic, V., & Backovic, T. (2021). Analyzing the factors influencing the formation of the price of electricity in the deregulated markets of developing countries. *Energy Reports*. <https://doi.org/10.1016/j.egy.2021.07.046>
- Edmond, C. (2022). How much energy does the EU import from Russia? *World Economic Forum*. <https://www.weforum.org/agenda/2022/03/eu-energy-russia-oil-gas-import>
- Elliot, L. (2022). Ukraine war 'will mean high food and energy prices for three years.' *The Guardian*. <https://www.theguardian.com/business/2022/apr/26/ukraine-war-food-energy-prices-world-bank>
- European Commission. (2022). Gas and electricity market reports (in Energy, Data and analysis, Market analysis). [https://energy.ec.europa.eu/data-and-analysis/market-analysis\\_en](https://energy.ec.europa.eu/data-and-analysis/market-analysis_en)
- Eurostat. (2020a). From where do we import energy? (in Infographs, Energy). <https://ec.europa.eu/eurostat/cache/infographs/energy/bloc-2c.html>
- Eurostat. (2020b). Where does our energy come from? (in Infographs, Energy). <https://ec.europa.eu/eurostat/cache/infographs/energy/bloc-2a.html>
- Fragkioudaki, A., Marinakis, A., & Cherkaoui, R. (2015). Forecasting price spikes in European day-ahead electricity markets using decision trees. In *International Conference on the European Energy Market, EEM*. <https://doi.org/10.1109/EEM.2015.7216672>
- Gabrielli, P., Wüthrich, M., Blume, S., & Sansavini, G. (2022). Data-driven modeling for long-term electricity price forecasting. *Energy*. <https://doi.org/10.1016/j.energy.2022.123107>
- GeeksforGeeks. (2022). *Random forest regression in python*. GeeksforGeeks. <https://www.geeksforgeeks.org/random-forest-regression-in-python/>
- Great Learning. (2022). *Random forest algorithm in machine learning: An overview*. My Great Learning. <https://www.mygreatlearning.com/blog/random-forest-algorithm/>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3149–3157.
- Kuik, F., Adolfsen, J. F., Lis, E. M., & Meyler, A. (2022). Energy price developments in and out of the COVID-19 pandemic - from commodity prices to consumer prices. In *European Central Bank Economic Bulletin*. [https://www.ecb.europa.eu/pub/economic-bulletin/articles/2022/html/ecb.ebart.202204\\_01~7b32d31b29.en.html](https://www.ecb.europa.eu/pub/economic-bulletin/articles/2022/html/ecb.ebart.202204_01~7b32d31b29.en.html)
- Hubicka, K., Marcjasz, G., & Weron, R. (2019). A note on averaging day-ahead electricity price forecasts across calibration windows. *IEEE Transactions on Sustainable Energy*. <https://doi.org/10.1109/TSTE.2018.2869557>
- Irfan, A. S. M., Bhuiyan, N. H., Hasan, M., & Khan, M. M. (2021). Performance analysis of machine learning techniques for wind speed prediction. In: *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. <https://doi.org/10.1109/ICCCNT51525.2021.9579564>
- Keles, D., Scelle, J., Paraschiv, F., & Fichtner, W. (2016). Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2015.09.087>
- Kolaczowski, M. (2022). How does the war in Ukraine affect oil prices? *World Economic Forum*.
- Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2021.116983>
- Lam, L. H., Ilea, V., & Bovo, C. (2018). European day-ahead electricity market coupling: Discussion, modeling, and case study. *Electric Power Systems Research*. <https://doi.org/10.1016/j.epr.2017.10.003>

- Lehna, M., Scheller, F., & Herwartz, H. (2022). Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. *Energy Economics*. <https://doi.org/10.1016/j.eneco.2021.105742>
- Li, R., & Leung, G. C. K. (2021). The relationship between energy prices, economic growth and renewable energy consumption: Evidence from Europe. *Energy Reports*, 7, 1712–1719. <https://doi.org/10.1016/J.EGYR.2021.03.030>
- Li, W., & Becker, D. M. (2021). Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy*. <https://doi.org/10.1016/j.energy.2021.121543>
- Liu, L., Bai, F., Su, C., Ma, C., Yan, R., Li, H., Sun, Q., & Wennersten, R. (2022). Forecasting the occurrence of extreme electricity prices using a multivariate logistic regression model. *Energy*. <https://doi.org/10.1016/j.energy.2022.123417>
- Lu, X., Qiu, J., Lei, G., & Zhu, J. (2022). Scenarios modelling for forecasting day-ahead electricity prices: Case studies in Australia. *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2021.118296>
- Maciejowska, K., Nitka, W., & Weron, T. (2021). Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. *Energy Economics*. <https://doi.org/10.1016/j.eneco.2021.105273>
- Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S., & Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for short-term load forecasting. *Energy*, 214, 118874. <https://doi.org/10.1016/J.ENERGY.2020.118874>
- Mayer, A. (2022). Fossil fuel dependence and energy insecurity. *Energy, Sustainability and Society*, 12(1), 1–13. <https://doi.org/10.1186/S13705-022-00353-5/TABLES/4>
- Menon, S. (2022). War and gas: What Russia's war on Ukraine means for energy prices and the climate. *Environmental Defense Fund*. <https://www.edf.org/article/war-ukraine-driving-gas-prices>
- Mondal, A. (2021). *Complete guide on how to Use LightGBM in Python*. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/08/complete-guide-on-how-to-use-lightgbm-in-python/>
- Naumzik, C., & Feuerriegel, S. (2021). Forecasting electricity prices with machine learning: Predictor sensitivity. *International Journal of Energy Sector Management*, 15(1), 157–172. <https://doi.org/10.1108/IJESM-01-2020-0001/FULL/PDF>
- Nvidia. (2022). *XGBoost*. Nvidia.Com. <https://www.nvidia.com/en-us/glossary/data-science/xgboost/>
- Punmiya, R., & Choe, S. (2019). Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Transactions on Smart Grid*, 10(2), 2326–2329.
- Schnürch, S., & Wagner, A. (2020). Electricity price forecasting with neural networks on EPEX order books. *Applied Mathematical Finance*, 27(3), 189–206. <https://doi.org/10.1080/1350486X.2020.1805337>
- SciKit. (2022). *Ensemble methods*. <https://Scikit-Learn.Org/>. <https://scikit-learn.org/stable/modules/ensemble.html?highlight=xgb#forests-of-randomized-trees>
- SciKit Learn. (2022a). *Histogram-Based Gradient Boosting*. SciKit Learn. <https://scikit-learn.org/stable/modules/ensemble.html?highlight=histgradientboostingclassifier#histogram-based-gradient-boosting>
- SciKit Learn. (2022b). *Voting Regressor*. SciKit Learn. <https://scikit-learn.org/stable/modules/ensemble.html#voting-regressor>
- SciKit Learn. (2022c). *VotingRegressor*. SciKit Learn. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html>
- Solibakke, P. B. (2022). Step-ahead spot price densities using daily synchronously reported prices and wind forecasts. *Journal of Forecasting*. <https://doi.org/10.1002/for.2759>
- Spodniak, P., Ollikka, K., & Honkapuro, S. (2021). The impact of wind power and electricity demand on the relevance of different short-term electricity markets: The Nordic case. *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2020.116063>
- Supply Agency of the European Atomic Energy Community. (2021). *ESA Annual Report 2021*. [https://euratom-supply.ec.europa.eu/publications/esa-annual-reports\\_en](https://euratom-supply.ec.europa.eu/publications/esa-annual-reports_en)
- Thompson, H. (2022a). What does the war in Ukraine mean for the geopolitics of energy prices? *Economics Observatory*. <https://www.economicsobservatory.com/what-does-the-war-in-ukraine-mean-for-the-geopolitics-of-energy-prices>
- Thompson, H. (2022b). The geopolitics of fossil fuels and renewables reshape the world. *Nature*, 603, 7901.
- Tian, J., Yu, L., Xue, R., Zhuang, S., & Shan, Y. (2022). Global low-carbon energy transition in the post-COVID-19 era. *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2021.118205>



- Tolefson, J. (2022). What the war in Ukraine means for energy, climate and food. *Nature*. <https://www.nature.com/articles/d41586-022-00969-9>
- Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. *Applied Energy*. <https://doi.org/10.1016/j.apenergy.2022.118752>
- Van den Bergh, K., Boury, J., & Delarue, E. (2016). The flow-based market coupling in central western Europe: Concepts and definitions. *Electricity Journal*. <https://doi.org/10.1016/j.tej.2015.12.004>
- Wang, R., Lu, S., & Li, Q. (2019). Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustainable Cities and Society*, 49, 101623. <https://doi.org/10.1016/J.SCS.2019.101623>
- Zhang, J. L., Zhang, Y. J., Li, D. Z., Tan, Z. F., & Ji, J. F. (2019). Forecasting day-ahead electricity prices using a new integrated model. *International Journal of Electrical Power and Energy Systems*. <https://doi.org/10.1016/j.ijepes.2018.08.025>
- Zhang, T., Tang, Z., Wu, J., Du, X., & Chen, K. (2022). Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. *Electric Power Systems Research*. <https://doi.org/10.1016/j.epsr.2021.107762>
- Ziel, F., & Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Economics*. <https://doi.org/10.1016/j.eneco.2016.08.008>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Adela Băra<sup>1</sup> · Simona-Vasilica Oprea<sup>1</sup> · Bogdan George Tudorică<sup>2</sup>

✉ Simona-Vasilica Oprea  
simona.oprea@csie.ase.ro

Adela Băra  
bara.adela@ie.ase.ro

Bogdan George Tudorică  
btudorica@upg-ploiesti.ro

<sup>1</sup> Department of Economic Informatics and Cybernetics, Bucharest University of Economic Studies, No. 6 Piața Romană, Bucharest, Romania

<sup>2</sup> Department of Cybernetics, Economic Informatics, Finance and Accountancy, Petroleum-Gas University of Ploiești, No. 39 București Blvd., Ploiești, Romania