



Predicting Day-Ahead Electricity Market Prices through the Integration of Macroeconomic Factors and Machine Learning Techniques

Adela Bâra¹ · Simona-Vasilica Oprea¹

Received: 23 August 2023 / Accepted: 5 December 2023
© The Author(s) 2024

Abstract

Several events in the last years changed to some extent the common understanding of the electricity day-ahead market (DAM). The shape of the electricity price curve has been altered as some factors that underpinned the electricity price forecast (EPF) lost their importance and new influential factors emerged. In this paper, we aim to showcase the changes in EPF, understand the effects of uncertainties and propose a forecasting method using machine learning (ML) algorithms to cope with random events such as COVID-19 pandemic and the conflict in Black Sea region. By adjusting the training period according to the standard deviation that reflects the price volatility, feature engineering and by using two regressors for weighing the results, significant improvements in the performance of the EPF are achieved. One of the contributions of the proposed method consists in adjusting the training period considering the price variation. Thus, we introduce a rule-based approach given an empirical observation that for days with a higher growth in prices the training interval should be shortened, capturing the sharp variations of prices. The results of several cutting-edge ML algorithms represent the input for a predictive meta-model to obtain the best forecasting solution. The input dataset spans from Jan. 2019 to Aug. 2022, testing the proposed EPF method for both stable and more tumultuous intervals and proving its robustness. This analysis provides decision makers with an understanding of the price trends and suggests measures to combat spikes. Numerical findings indicate that on average mean absolute error (MAE) improved by 48% and root mean squared error (RMSE) improved by 44% compared to the baseline model (without feature engineering/adjusting training). When the output of the ML algorithms is weighted using the proposed meta-model, MAE further improved by 2.3% in 2020 and 5.14% in 2022. Less errors are recorded in stable years like 2019 and 2020 (MAE = 6.71, RMSE = 14.67) compared to 2021 and 2022 (MAE = 9.45, RMSE = 20.64).

Keywords Black swan · Machine learning · Feature engineering · Electricity price forecast · Day-ahead market

List of Symbols

h	Time of the historical records used for training and testing (hours)	RMSE, R^2 , MAPE, MAE	Evaluation metrics of the ML models
hf	Time of the forecasted values (hours)	EIQ^h	Traded quantity of electricity on DAM (MWh)
d, m	Day, month corresponding to the training interval	EIP^h	Hourly electricity price on DAM (RON/MWh)
T	Time interval for the training and testing of the ML algorithms (hours)	$GasP^d$ $GasQ^d$ $InfRO^m$	Daily gas price on DAM (RON/MWh) Traded quantity of gas on DAM (MWh) Monthly inflation for Romania
		$InfEU^m$ $Robor^m$ $OilP^d$	Monthly inflation for EU ROBOR index Daily oil price (\$/barrel)

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

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

LevelTM ^d , LevelBR ^d , LevelTL ^d	Danube water levels in 3 regions (Turnu Magurele, Braila, Tulcea) (cm)	$\hat{y}_k^h, k \in \{\text{RF, LGBR, HGBR, XGB, VR}\}$	Output of the ML models (forecasted values of electricity price)
EIC ^h	Hourly electricity consumption (MW)	$\widehat{\text{EIP}}^h$	Forecast of the hourly electricity price (RON/MWh)
EIG ^h	Hourly electricity generation (MW)	θ_k	Weights used to adjust the output of the ML models to obtain $\widehat{\text{EIP}}^h$
CG ^h	Hourly coal generation (MW)		
HG ^h	Hourly hydro generation (MW)		
OGG ^h	Hourly oil and gas generation (MW)		
NG ^h	Hourly nuclear generation (MW)		
WG ^h	Hourly wind generation (MW)		
PVG ^h	Hourly photovoltaic generation (MW)		
BG ^h	Hourly biomass generation (MW)		
Ex ^h	Hourly exchange (MW)		
EUETS ^d	Daily price per tonne of CO ₂ (EUR/tCO ₂)		
W ^d	Weekday, W ^d = 0.6		
μEIP^{d-1}	Mean price of the previous day (RON/MWh)		
EIPmax ^{d-1} , EIPmin ^{d-1}	Maximum and minimum price of the previous day (RON/MWh)		
$\sigma\text{EIP}^{d-1}, \sigma^2\text{EIP}^{d-1}$	Standard deviation and variance of the price of the previous day		
$\widetilde{\text{EIP}}^{d-1}$	Median of the price of the previous day (RON/MWh)		
EIPrange ^{d-1}	Range price of the previous day (RON/MWh)		
EIPratio ^{d-1}	Ratio price index of the previous day		
$\mu\text{EIP}^{h3}, \mu\text{EIP}^{h7}$	Averages of the previous hourly prices for 3 and 7 consecutive days		
$\Delta\text{EIP}^{h1}, \Delta\text{EIP}^{h2}, \Delta\text{EIP}^{h3}$	Hourly variation of the prices for 3 consecutive days		
X ^h	Input variable of the ML algorithms (array)		
y ^h	Actual target variable of the ML algorithms (RON/MWh)		

1 Introduction

Starting from 1996, the European Union (EU) by Directive 96/92/EC progressively enabled the competition in the energy market, aiming to create one single integrated internal European electricity market, to reduce prices, increase the security of supply, and enhance the life standards for the European citizens. This effort has started more than 20 years ago and is still ongoing [27]. It assumed the translations of the EU directives and regulations into the national legislation [16]. Therefore, the structure of the electricity market is relatively similar in the European state members. In this context, the performance of the EPF is important for economic activities, energy traders, suppliers, producers, researchers and policy makers as the electricity traded on the DAM has currently a predominant share (43%) in the total electricity consumption. Thus, the DAM prices tremendously influence on the economy as they have a significant impact on the final electricity prices [23]. Furthermore, the EPF have become essential for energy suppliers to optimize their energy procurement strategies [33].

Differences between long- and mid-term bilaterally negotiated contracts and short-term energy requirements are settled through DAM and intraday markets. However, even in more stable intervals, the EPF has been a challenge for researchers as it depends on numerous complex drivers; some influences are difficult to be modeled and sharp fluctuations of electricity prices also known as spikes often occur [20, 38]. The share of RES is increasing in the European state members, although its generation is more volatile, and administrators rely on short-term forecast and prefer to trade on DAM rather than to sign forward contracts [12, 31]. On the other hand, random events have tremendous side effects that influence electricity prices. Therefore, most of the EPF methods are not robust enough to cope with random events or black swans such as those recently experienced.

The first event, the COVID-19 pandemic disease shook the world business and was something that the modern

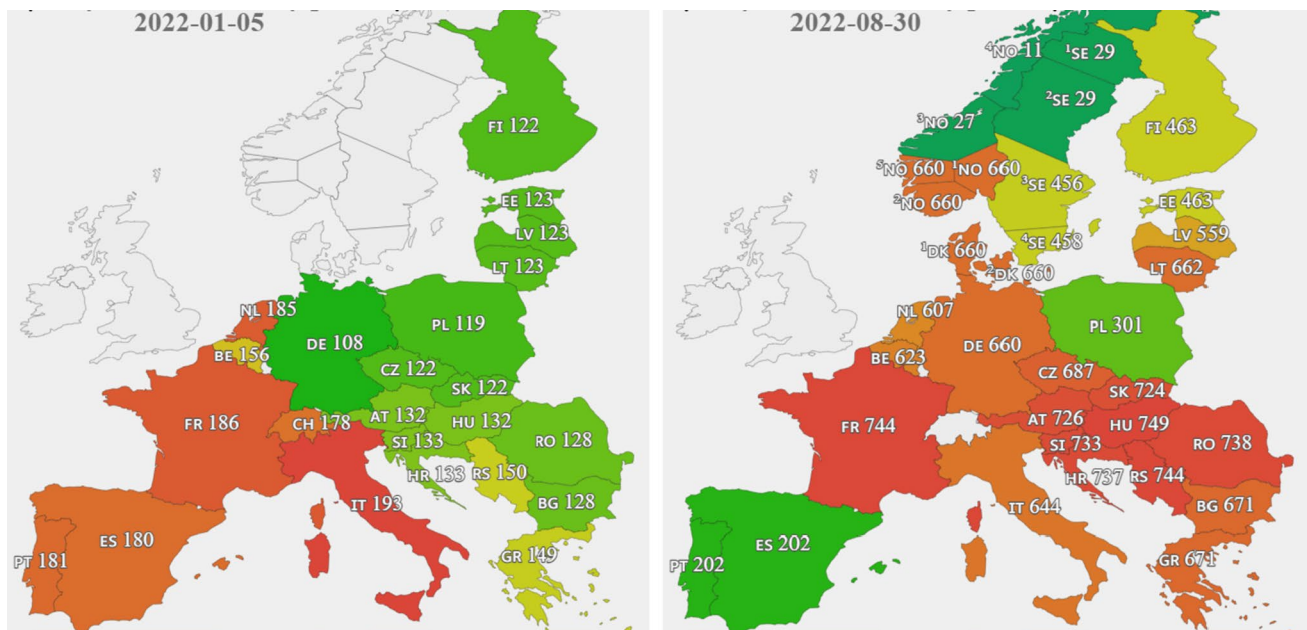


Fig. 1 Average price variations on DAM on 5th of January and 30th of August in 2022 (<https://euenergy.live/>)

citizens have not experienced yet. Moreover, the constraints and lockdowns baffled and slowed down the economic activities slightly decreasing the electricity prices. It started to spread in March 2020 and soon affected the entire world. From the electricity consumption point of view, between 2019 and 2020, it rather stalled at almost the same level and the prices slowly went down until 2021, when after the lockdowns, the activities were almost fully resumed, and the electricity requirements increased. In 2021, the electricity prices on the DAM were more than doubled compared with 2020 and, in 2022 vs. 2021, they doubled and even tripled on average at certain time intervals. Such soaring prices were also influenced by excessive draughts in recent years and higher inflation after the lockdowns, when the economic activities were interrupted or drastically reduced. Thus, the offer of goods and services was reduced in lockdowns, but afterwards the request increased abruptly leading to higher inflation and interest rates. Furthermore, some electricity producers encountered higher prices for emission allowances via the European Union Emissions Trading Scheme (EU ETS) leading to more expensive electricity especially generated by coal-based power plants. These facts created the snowball effect that amplified the electricity prices.

The second event, the military conflict that started in Ukraine due to Russian invasion that eventually has been transformed into a war with millions of refugees and casualties was also rather unexpected. In modern times, when business prevails and artificial intelligence should guide and improve our living standards, nobody expected conflicts in Europe. It also led to sanctions and energy sabotage

that underlined the dependency of Europe from the energy resources of Russia especially gas and oil, on which Europe unfortunately relied. This conflict increased the gas and oil prices to unexpected levels, acting as a cold shower on the most European day-ahead markets. Furthermore, the measures taken in the European countries increased the speed of installing PV systems and transforming consumers into prosumers, as well as efficiency measures such as replacing light bulbs, installing heat pumps and so on. However, this stress on the electricity price inflated more on the prices of goods and services and the interest rate, exacerbating the side effects of the first event. Huge variations of average prices were encountered on the DAMs of the European countries. In Fig. 1, two snapshots of the average electricity prices from 5th of Jan. and 30th of Aug. 2022 are presented.

Noticeable changes took place in the DAM, the energy crisis hit this market altogether: prices go up and down, the curve level is 3–4 times higher than in 2020, it is more humped, and the spikes are sharper and more frequent. The effects of these events and the evolution of prices on the DAMs motivated us to understand the trends and propose a forecast method to cope with random events. To obtain a robust forecasting method able to handle random events, an economical understanding of electricity markets, combined with feature engineering, and stand-out machine learning algorithms are necessary.

The proposed forecasting method relies on a couple of ML algorithms, such as: random forest, light gradient boosting, histogram-based gradient boosting, extreme gradient boosting and voting regressors. The results of the ML

algorithms are weighted, and the training interval is adjusted per forecasting interval considering the standard deviation of the prices. Furthermore, its strengths relies in collecting numerous fundamental features extracted from various sources such as: National Statistics Institute, Electricity Market Operator, Romanian Commodities Exchange, Transmission System Operator, reliable websites for commodities historical prices, Romanian National Bank for projections, Eurostat, etc. and feature engineering or derivate features extracted from the raw ones.

An interesting finding is related to the shape of electricity price curve and the importance of the features. Some of them lost or gained importance depending on the forecast interval as we train and test the algorithms for a longer interval from January 2019 to August 2022. The novelty of our method relies in adjusting the training period considering the price variation in the past days. Thus, we build a rule-based approach based on an empirical observation that for days with a higher increase in prices the training interval should be shortened capturing the sharp increase and speculative effect of the price soaring. The results of several state-of-the-art ML algorithms are input for a regressor (that could be linear regression or gradient descent) to obtain the best solution or the weighed forecast. This approach is robust and provides accurate results for both more stable intervals and for those with much higher price variation.

The remainder of this paper is structured as follows: in Sect. 2, an investigation of similar EPF methods is performed and in Sect. 3, input data are analyzed to understand the current general context. Section 4 focuses on the proposed EPF method, Sect. 5 is dedicated to the results and comparison, whereas Sect. 6 is dedicated to the discussions and interpretation of the results. Conclusion and implications for DAM are presented in Sect. 7.

2 Literature Review

As the world transitions to cleaner and more sustainable energy sources to combat climate change, there are risks associated with this transition [17] including economic and employment disruptions in fossil fuel-dependent regions, technological challenges, and policy uncertainties. Considering the importance of the EPF, researchers struggled to perform accurate EPF. There is a large palette of methods implemented to obtain the EPF: from statistical and hybrid methods [13, 32] to deep learning [26, 35]. However, in comparison with other research studies, that consider shorter intervals that do not cope with uncertainties, the proposed EPF solution is robust and efficient in both stable and tumultuous context of the electricity market. The drawback of shorter intervals in which EPF solutions are tested is that

they are unable to effectively predict prices outside these intervals.

To identify the factors that influence the electricity price on DAM, several hypothesis are tested [14] with data from 2007 to 2019. In terms of random events, this interval is characterized by the financial crisis that started in 2008 and continued in the following years. The influence of the cost of active energy, utilization cost of transmission and distribution capacity, losses in transmission and distribution systems, supplier and market operator fees, fee for stimulating RES and electricity generated by RES on the electricity price on DAM in Montenegro are tested using a regression model. Only the cost of active energy had a significant impact on the price of electricity. Furthermore, the impact of distribution capacity utilization on the price of electricity was significant.

Aiming to forecast the baseline and spikes in the electricity price, decision trees are involved [15]. One classification tree is envisioned to predict the occurrence of spikes and two regression trees are applied for EPF. First, the price is classified as spike or normal, then the price variation is calculated. Feature selection is performed to choose the features with the highest explanatory capacity. As input, hourly electricity prices on different bidding zones, load, transmission capacity, cross-border physical flows, wind and PV forecast are collected to perform EPF. A price value is identified as spike if it is outside a band of plus/minus 2 times standard deviation of the price series. Econometric time-series models for EPF, such as autoregressive ARMA, ARIMA, VAR, GARCH and ARIMAX, singular spectrum analysis [29] are usually outperformed by the machine learning algorithms [22, 25, 34] artificial neural networks are considered to perform EPF focusing on the selection and processing of the input variables [24], such as gas prices, CO₂ certificate prices, renewable feed-in, cost of fuel and consumption. It is interesting that in the recent past, electricity prices were depended on the weekdays and weekend days, holidays, seasons, weekly and annual cycles, but they are not significant anymore as random events are stronger drivers. RES volatility also changed the pattern of the DAM, increasing its share and introducing more price variation. [24] focused on preparing 55 fundamental variables and a combination of multiple configurations of the ANN model tested in with various input parameters, finding the best setting of the model. They compared ANN with (S)ARIMA results using RMSE and MAD. However, [24] did not provide forecast at different moments in time (in various years, for instance) and the capacity of the model to handle random events was not tested.

As many recent studies [18, 19, 28] included RES generation into the main EPF driver along with total load, weather data (temperatures in 2 cities in Germany), and intraday prices, investigating several years, from October 2015 to September 2019, but no significant random events took place. They used autoregressive (AR) models with

exogenous variables for day-ahead and intraday prices. However, the lack of the random events (numerous fundamental variables are ignored as they were not significant at that time such as gas/oil price, emission price and price index) and the fame of the AR models do not recommend this approach as a robust method. Additionally, more recent research is focused on the RES optimization and integration into the power systems, energy communities and microgrids [2, 3, 8, 30], prediction [2, 6–8], air quality [4], conventional resources optimization [6, 7, 9], water quality [1], reducing computation [21] and data processing [5].

Zhang et al. [36] proposed an integrated model using the improved empirical mode decomposition, ARMA with exogenous terms (demand for Spanish and Australian markets), exponential GARCH and adaptive network-based fuzzy inference system. They considered 672–1200 observations for training and 24–168 observations for testing and compared the results with wavelet transform combined with ARMA and back propagation neural network, kernel extreme learning machine, and least squares support vector machine. Ziel and Steinert [37] proposed a model, namely X-model, using the sale and purchase curves of the electricity exchange or bidding structure of the auction data. They processed the data, reduced its dimension and considered lasso-based estimation methods to obtain the day-ahead EPF for the German and Austrian markets. Usually, the individual bids and offers are not open-source data. Another drawback of the model is that it was not tested on long-term considering severe changes generated by random events.

More recent studies regarding electricity price predictions were provided in [10, 11]. In this paper, the dataset interval is generous and recent. It refers to an S–E European country (Romania), whereas most previous studies focused on Western and Northern European markets (such as UK, Norway, Germany, Austria, France, etc.). The authors identified a research gap as the S–E European countries were less investigated. Moreover, most of the studies do not include recent events (such as: COVID-19 pandemic and the conflict in the Black Sea region) and their impact on electricity prices. Another interesting aspect refers to the novelty of the dataset that was entirely built for this research. It can be used in further studies to analyze the storage requirements to replace conventional power plants or to identify the required power system flexibility that can come from consumers or business partners.

3 Input Data Depiction

In this section, the input dataset is described. Numerous time series are considered in the EPF to cope with the uncertainties brought by random events. We collected several datasets for January 2019–August 2022 from various open sources

and merged these time series on a daily and hourly basis. Different time resolutions were encountered: trimestral, monthly, daily hourly, and 10-min resolution. As the target is the electricity price that is extracted from the Romanian Market Operator (MO) on an hourly basis, all datasets were adjusted to be in line with this resolution. Furthermore, the traded electricity on DAM was extracted from MO.¹ The data related to the Romanian power system operational state (total consumption and generation, wind, hydro, PV, oil and gas, biomass, coal and nuclear generation and exchange on the overhead tie-lines) were extracted from the Transmission System Operator website.² Furthermore, gas and oil prices, gas quantity traded in Romania (Romanian Commodities Exchange³), price index or inflation in Romania (National Statistics Institute) and at the European Union level (Eurostat⁴), interest rate,⁵ Danube water level⁷ (measured at three points) and emission allowance⁸ were scrapped from open data sources. The initial data processing flow is briefly depicted in Fig. 2.

Out of the total 43 features, 21 are raw features (electricity quantity on DAM, gas price on DAM, oil price, etc.), 22 are engineered features (by aggregation six features: minimum, mean, variance, standard deviation, maximum, median, and 16 are derivatives: range, max/mean, previous prices 24 h, previous prices 48 h, previous prices 72 h, previous prices 96 h, previous prices 120 h, previous prices 144 h, previous prices 168 h, previous 3 days prices mean, previous 7 days prices mean, hourly variations of the previous prices for the last 3 days, encoded weekday from 0 to 6, hour from 1 to 24). The electricity price on DAM is the target that is present in the training dataset and removed from the testing datasets. The raw features extracted from open data sources from 2019 to 2022 are depicted in Appendix 1, Tables 4, 5, 6 and 7, showing the average basic statistics. It is worth mentioning that the electricity price standard deviation varied from 85 in 2020 to 707 in 2022. In 2019, the electricity price on DAM (or target) is directly correlated with traded electricity (60%) on DAM and total consumption (64%), and inversely correlated with Danube water levels (–30%) as in

¹ https://www.opcom.ro/pp/grafice_ip/raportPIPsiVolumTranzactiонат.php?lang=ro

² https://www.transelectrica.ro/widget/web/tel/sen-grafic/-/SENGrafic_WAR_SENGraficportlet

³ <https://www.brm.ro/piata-spot-gn/>

⁴ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Inflation_in_the_euro_area

⁵ <https://www.bnr.ro/Proiectii-BNR-22694-mobile.aspx>

⁶ <http://statistici.insse.ro/shop/?page=ipc1>

⁷ <https://www.cotele-dunarii.ro/>

⁸ <https://www.investing.com/commodities/carbon-emissions-historical-data>

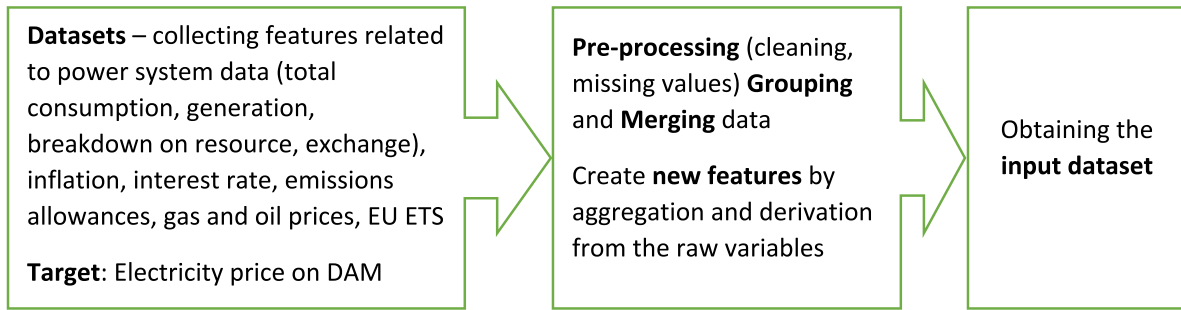


Fig. 2 Data processing initial flow

Table 1 Pearson correlation coefficients

Correlation coefficient	EIP_{2019}^h	EIP_{2020}^h	EIP_{2021}^h	EIP_{2022}^h
El_quantity (EIQ^h)	0.596587	0.310967	− 0.13812	0.168683
Gas_price_DAM ($GasP^d$)	0.125493	0.441562	0.82858	0.757485
Gas_quantity_DAM ($GasQ^d$)	− 0.0297	− 0.1866	0.486406	0.334951
Inflation_RO ($InfRO^m$)	− 0.3424	− 0.05039	0.789006	0.329347
Inflation_EU ($InfEU^m$)	− 0.16297	− 0.07235	0.786641	0.454506
ROBOR 3 M ($Robor^m$)	− 0.31901	− 0.07501	0.704966	0.586826
Oil_price ($OilP^d$)	− 0.32725	0.428199	0.593899	0.042344
Level_Turnu Magurele ($LevelTM^d$)	− 0.31371	− 0.11808	− 0.55886	− 0.44648
Level_BR ($LevelBR^d$)	− 0.27231	− 0.07297	− 0.57319	− 0.55958
Level_TL ($LevelTL^d$)	− 0.26438	− 0.09949	− 0.5755	− 0.54291
Consumption (EIC^h)	0.641719	0.765722	0.322032	0.193566
Generation (EIG^h)	0.234691	0.535417	− 0.03781	− 0.1943
Coal_gen (CG^h)	0.413524	0.574996	0.132546	0.228913
Oil&Gas_gen (OGG^h)	0.368002	0.566714	0.414415	0.048759
Hydro_gen (HG^h)	0.025771	0.401482	− 0.33788	− 0.0179
Nuclear_gen (NG^h)	0.043153	0.074555	0.196464	− 0.01421
Wind_gen (WG^h)	− 0.17033	− 0.17981	− 0.00498	− 0.3163
PV_gen (PVG^h)	0.073434	− 0.11813	− 0.09232	− 0.00558
Biomass_gen (BG^h)	0.246113	0.351802	− 0.07165	− 0.21031
Exchange (Ex^h)	0.281236	0.355652	0.306741	0.242605
Price_EUETS ($EUETS^d$)	0.081138	0.447834	0.707506	0.004049
Weekday (W^d)	0.100084	0.085415	0.028807	0.038833
Hour (h)	0.338499	0.289218	0.181837	0.211503

Table 1. In 2020, the electricity price on DAM is directly correlated with total consumption (77%), coal and oil & gas generation (57%). Furthermore, it was medium correlated with the price of emissions (44%). In 2021, the electricity price on DAM is directly correlated with gas price on DAM (83%), inflation (79%), price of emissions (71%), interest rate—ROBOR 3 M (70%), oil price (59%) and inversely correlated with Danube water levels (− 30%) and inversely correlated with Danube water levels (− 57%). As in Table 1, in 2022, the electricity price on DAM is directly correlated with gas price on DAM (76%), interest rate (59%)

and inversely correlated with Danube water levels (− 52% on average) and inversely correlated with wind generation (− 32%). Several more comprehensive correlations are presented as heatmaps in Appendix B.

On average, the hourly consumption evolution is presented in Fig. 3a, its variations are small in amplitude, whereas the hourly average price variation is depicted in Fig. 3b. It can be noticed that both level and shape of the electricity price have changed in time. The price soared to unprecedented levels in 2022. The electricity price shapes in 2019 and 2020 are almost flat, but in 2021 and 2022 they

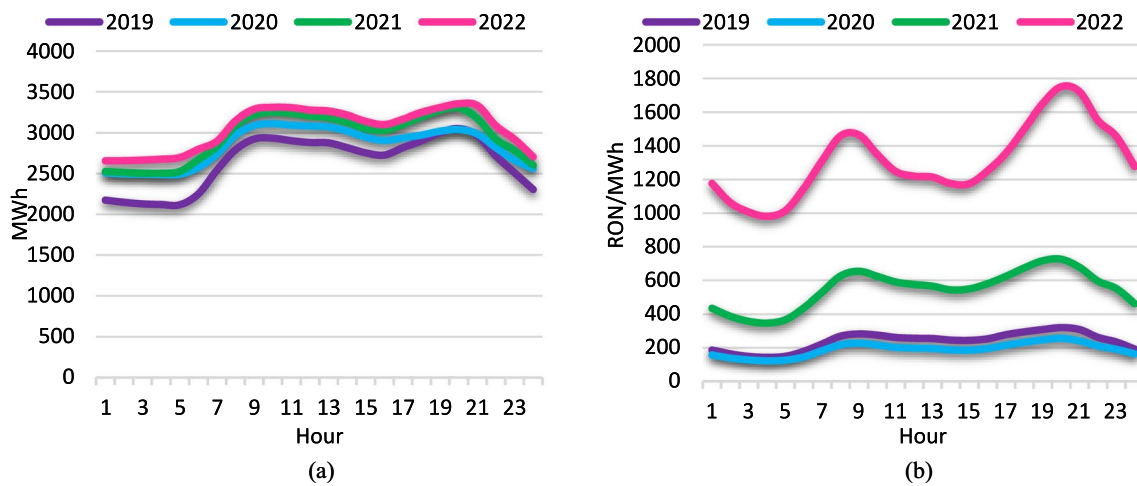


Fig. 3 Average hourly consumption (a) and average hourly prices (b)

show two humps that are similar with the daily load curves, showing morning/evening peaks and night valleys. This characteristic is related to the very expensive generating units that are able to satisfy the request bids at peaks. Also, there is a speculative effect of the ever-increasing prices, thus new features are built to overcome and predict these spikes. Producers tend to ask for more on DAM only because they know the trend and try to maximize its advantage. The producers also know the hydrologic conditions, RES forecast and the planned outages of the nuclear units, anticipating the level of scarcity and asking for higher prices. Another aspect is related to the auctioning price mechanism that favors the producers with cheaper fuel. The uniform price mechanism offers the price of the most expensive unit of energy to all producers with accepted offers regardless of their prices. This means the same price is paid for all producers, but this increases the burden on the supplying side especially during energy scarcity. The pay-as-bid pricing mechanism could be more appropriate for this market.

In Fig. 4, the average electricity traded and prices on DAM are depicted in combo graphs, year by year: from 2019 to 2022. The price scales are similar in 2019 and 2020, but in the next two years, tremendously soaring prices are encountered.

The average monthly prices for electricity and gas on DAM in 2019–2022 are displayed in Fig. 5. It is noticeable that in the last two years 2021 and 2022, the electricity price is highly correlated with gas price. The increase is constant from May 2021 to December 2021. Then, the increase is even sharper in 2022.

The monthly electricity traded on DAM and the share of electricity traded on DAM in total consumption are shown in Fig. 6. Monthly electricity traded on DAM varied according to Fig. 6a. The monthly electricity traded is higher in July and August in 2022, whereas in January 2020 was the

highest electricity traded on DAM. However, the differences are not big, but the last two months in 2022 show an increase that is directly correlated with the increasing prices. The share of electricity traded on DAM in the total consumption varied between 38.5% in 2019 and 46.6% in 2022 (as in Fig. 6b).

The evolution of electricity price and quantity on DAM and gas price on DAM are also represented in Fig. 7. There is an obviously directly correlated increase in the prices starting from mid-2021, when lockdowns and restrictions were eliminated. Then, the increase is amplified by the second random event—conflict in Ukraine.

The variations of the electricity prices and quantities on DAM are showcased in Fig. 8 using boxplots. They show an accelerated increase in the electricity price from 2020 to August 2022. In 2022, values above 2900 RON/MWh are outliers, whereas 50% of the prices varied between 849 and 1645 RON/MWh. The price variations is more evident in 2021 and 2022, showing more extreme values and even outliers that were less frequent in the previous years.

In comparison with 2022, in 2020, 50% of the prices varied between 173 and 289 RON/MWh. Impressive variations took place in the European market for emission allowances, as in Fig. 9a. The price of emitting a tone of carbon (in the EU ETS) attained a minimum value in March 2020 (16 €/tCO₂) and then it rose from 33 €/tCO₂ (1st Jan. 2021) to a record high of 89 €/tCO₂ (8th Dec. 2021). The barrier of 40 €/tCO₂ was broken in March 2021 due to cold temperatures. In May 2021, prices climbed above 56 €/tCO₂ for the first time and in Aug. 2022, it continued to head higher, above 98 €/tCO₂. The monthly variation in Fig. 9b shows a constant increase in 2021.

Three of the Danube water levels are represented in Fig. 10. The water levels were daily recorded in Turnu Magurele, Braila and Tulcea points. They reflect the precipitation level in different years and the potential that some important hydro-power

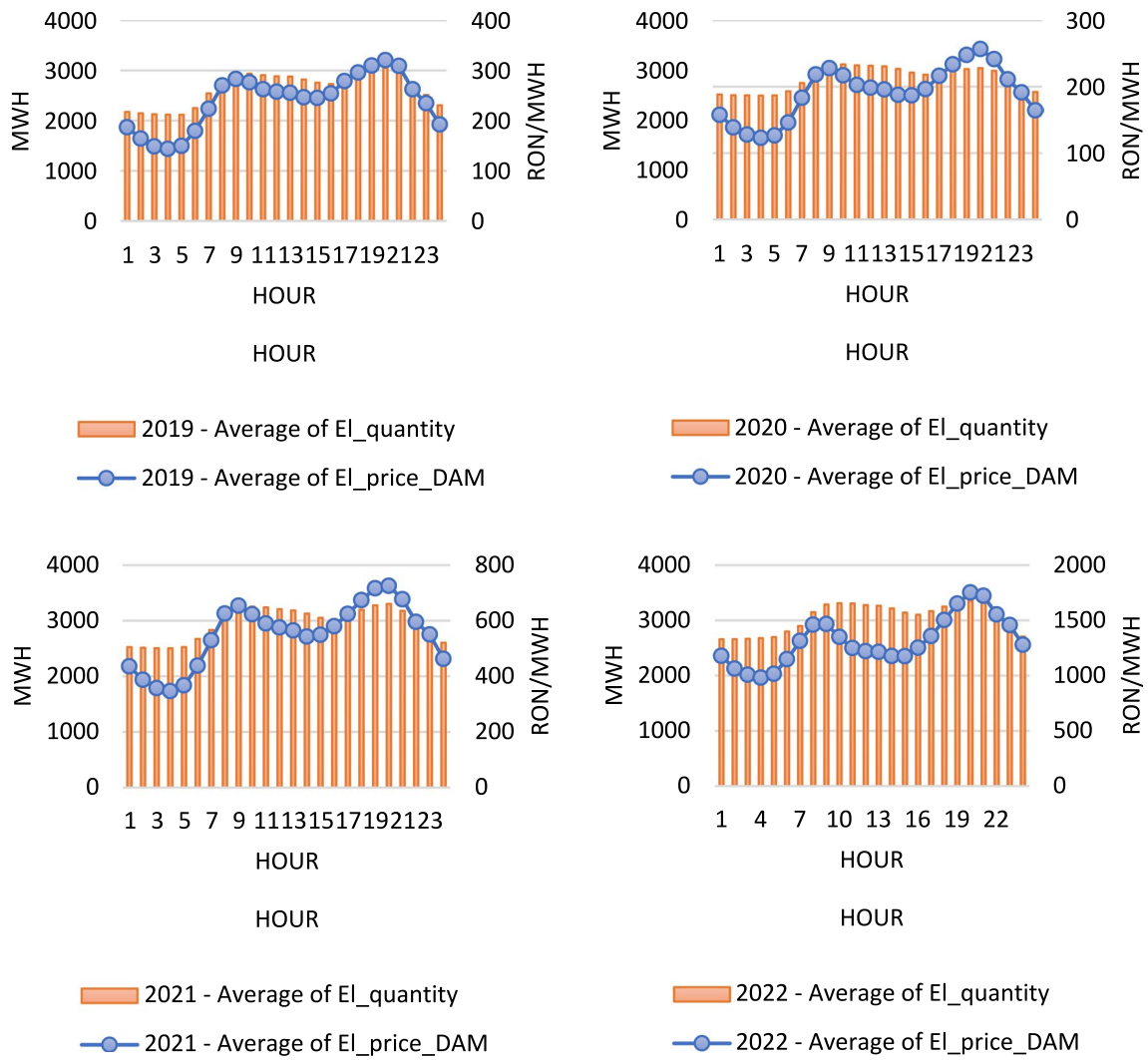


Fig. 4 Average electricity traded and prices on DAM 2019–2022

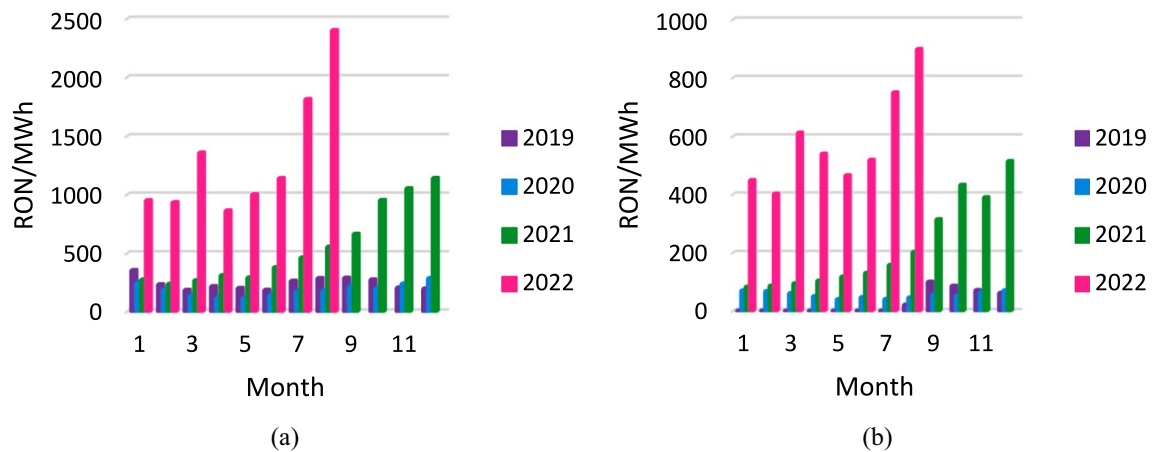


Fig. 5 Average electricity price (a) and gas price (b) on DAM

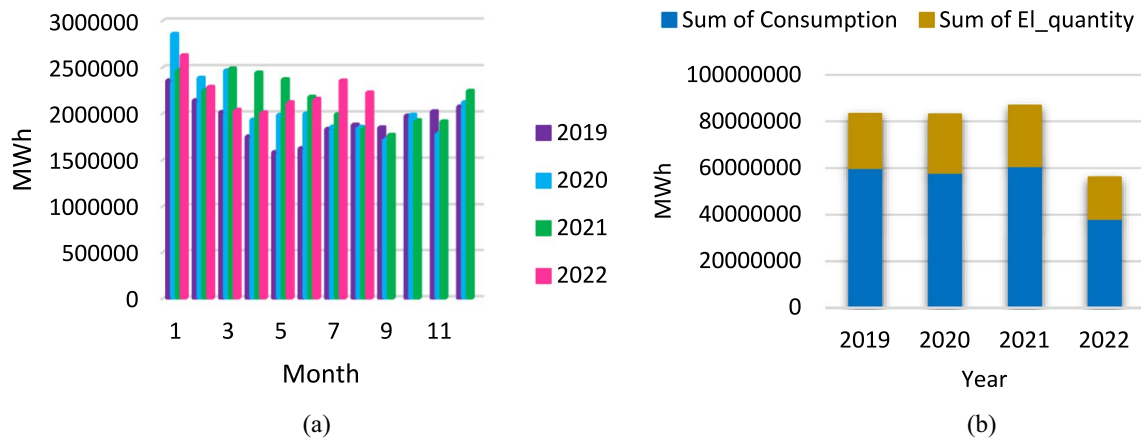


Fig. 6 Monthly electricity traded on DAM (a) and electricity traded on DAM in total consumption (b)

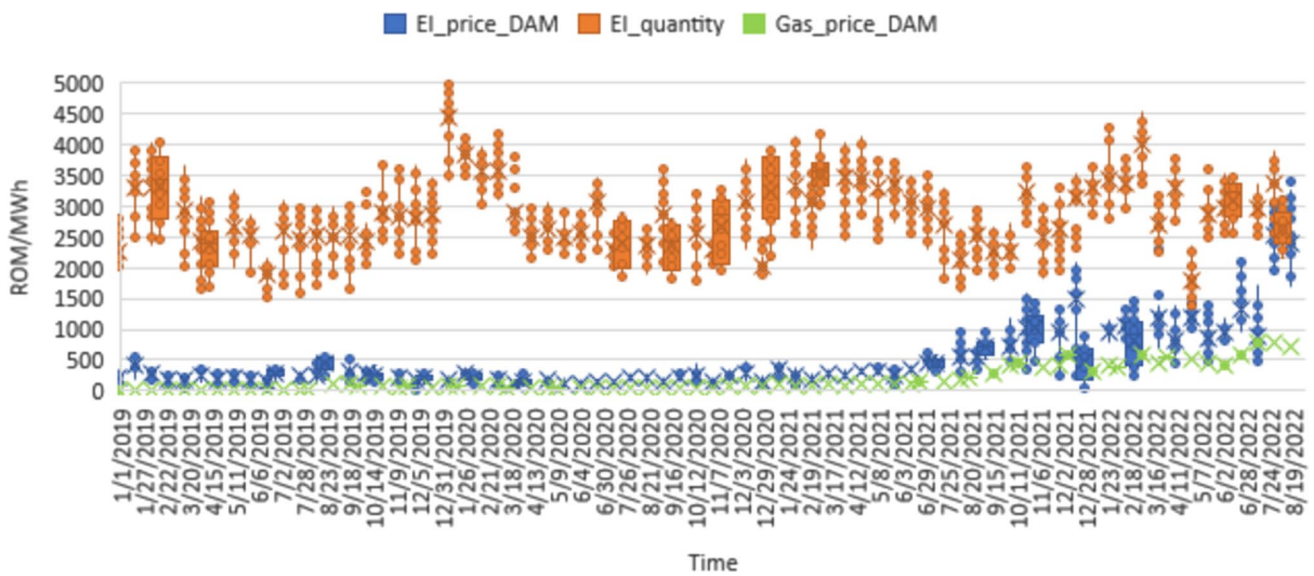


Fig. 7 Variation of electricity price, traded quantity and gas price from 2019 to 2022

plants (like Portile de Fier) had to generate electricity. There is a clear decrease of the water levels starting from the end of 2021 that accentuated the drought and diminished the hydro generation. This issue led to a smaller hydro capacity and a higher market concentration.

The levels are highly correlated and especially in 2022, the water levels were extremely low indicating a weak hydrology and less potential for hydro-power plants depending on the Danube river.

4 Methodology and Research Hypotheses

The EPF methodology consists in two major steps that first prepare and configure the input of the ML models, then trains the algorithms, compute the final forecast, and

evaluate the accuracy of the models. For training the ML models, a rule-based approach is built based on an empirical observation that for a month with a higher deviation in prices, the training interval should be shortened capturing the sharp increase and speculative effect of the soaring price. After training, the output of the ML algorithms is the input for a regressor model in order to increase the individual accuracy and obtained a weighed forecast of the electricity prices.

4.1 Configure the Input for the ML Algorithms

The input of the ML algorithms is initially composed by the 22 raw features (including the target) that represents the recorded values of the previous day:

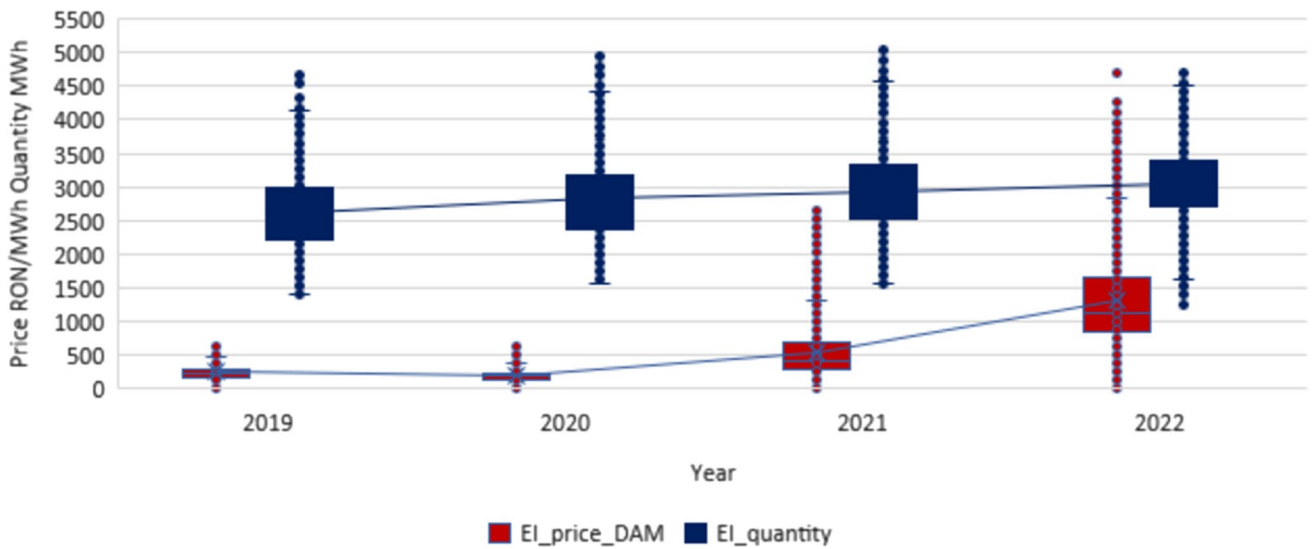


Fig. 8 Prices and quantities on DAM distribution in 2019–2022

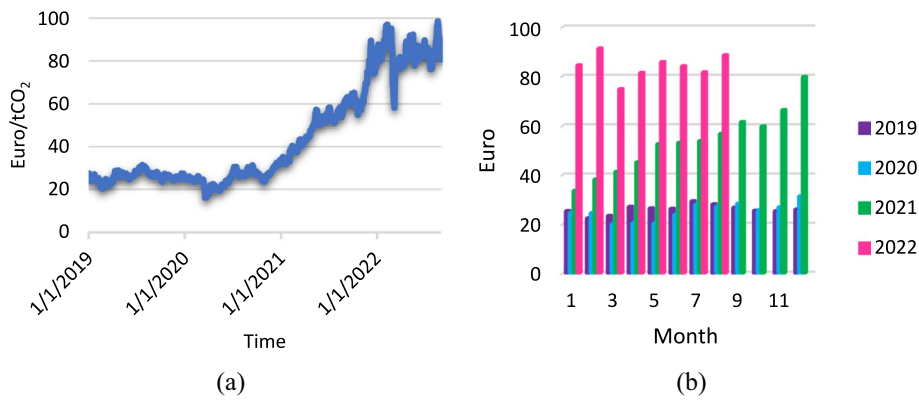


Fig. 9 Evolution of emissions trading (a) and average monthly price of emissions (b)

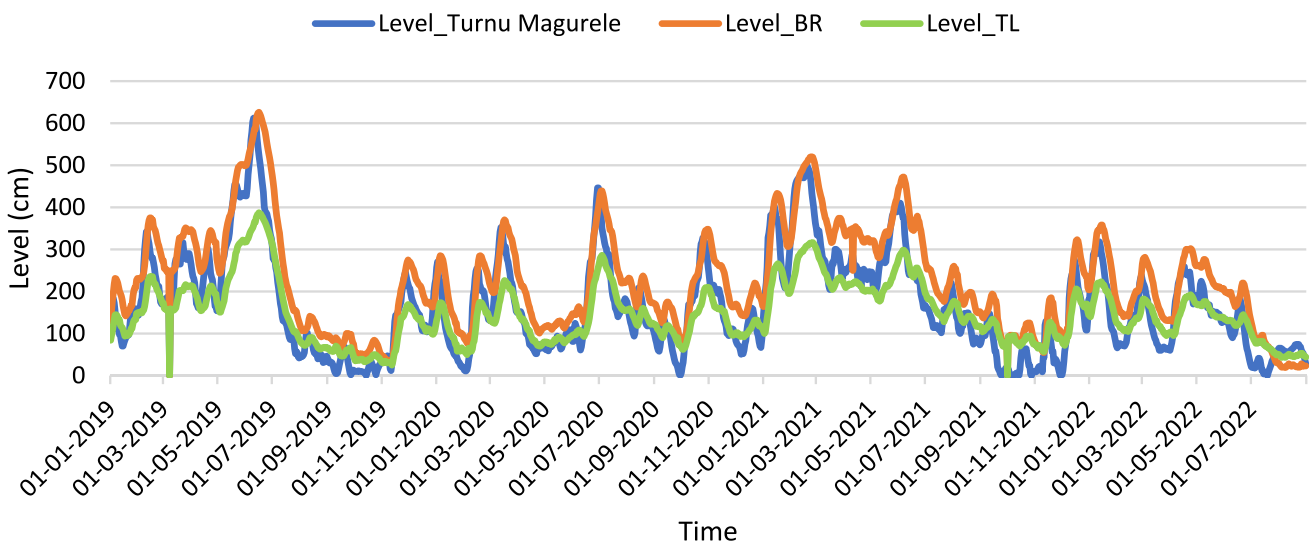


Fig. 10 Three Danube water levels

$$\begin{aligned}
 X^h = & [\text{EI}Q^{h-24}, \text{EI}P^{h-24}, \text{Gas}P^{d-1}, \text{Gas}Q^{d-1}, \\
 & \text{InfRO}^m, \text{InfEU}^m, \text{Robor}^m, \text{Oil}P^{d-1}, \\
 & \text{LevelTM}^{d-1}, \text{LevelBR}^{d-1}, \text{LevelTL}^{d-1}, \text{EIC}^{h-24}, \text{EIG}^{h-24}, \\
 & \text{CG}^{h-24}, \text{HG}^{h-24}, \text{OGG}^{h-24}, \text{NG}^{h-24}, \text{WG}^{h-24}, \\
 & \text{PVG}^{h-24}, \text{BG}^{h-24}, \text{Ex}^{h-24}, \text{EUETS}^{d-1}].
 \end{aligned}
 \tag{1}$$

The input is filled with 22 more engineered features obtained by aggregating the previous day’s hourly prices and by determining the range and variations of the current prices versus previous days prices.

4.1.1 Aggregated Features

The following variables are calculated for the previous day: minimum and maximum price, mean price, variance, standard deviation and median of the previous day prices.

$$\text{EIPmin}^{d-1} = \min_{h|h \in d-1} \text{EIP}^h; \text{EIPmax}^{d-1} = \max_{h|h \in d-1} \text{EIP}^h,
 \tag{2}$$

$$\mu\text{EIP}^{d-1} = \frac{1}{24} \sum_{h|h \in d-1} \text{EIP}^h,
 \tag{3}$$

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

$$\widetilde{\text{EIP}}^{d-1} = \text{median}_{h|h \in d-1} \text{EIP}^h.
 \tag{5}$$

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

$$X^h = [X^h, \text{EIPmin}^{d-1}, \text{EIPmax}^{d-1}, \mu\text{EIP}^{d-1}, \sigma^2\text{EIP}^{d-1}, \sigma\text{EIP}^{d-1}, \widetilde{\text{EIP}}^{d-1}].
 \tag{6}$$

4.1.2 Derivate Features

Based on the previous day aggregated values, in this step, 16 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:

$$\text{EI}Prange^{d-1} = \text{EIPmax}^{d-1} - \text{EIPmin}^{d-1}.
 \tag{7}$$

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

$$\text{EI}Pratio^{d-1} = \frac{\text{EIPmax}^{d-1}}{\mu\text{EIP}^{d-1}}.
 \tag{8}$$

Previous hourly prices with a lag of 24–168 h: $\text{EI}Q^{h-24}, \text{EI}Q^{h-48}, \text{EI}Q^{h-72}, \text{EI}Q^{h-96}, \text{EI}Q^{h-120}, \text{EI}Q^{h-144}, \text{EI}Q^{h-168}$. The average of the previous hourly prices for 3–7 consecutive days are calculated using Eqs. (9), (10):

$$\mu\text{EIP}^{h3} = \frac{\sum_{i \in \{24,48,72\}} \text{EI}Q^{h-i}}{3},
 \tag{9}$$

$$\mu\text{EIP}^{h7} = \frac{\sum_{i \in \{24,48,72,96,120,144,168\}} \text{EI}Q^{h-i}}{7}.
 \tag{10}$$

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

$$\Delta\text{EIP}^{h1} = 100 - \frac{\text{EI}Q^{h-24}}{\text{EI}Q^{h-48}} \times 100,
 \tag{11}$$

$$\Delta\text{EIP}^{h2} = 100 - \frac{\text{EI}Q^{h-48}}{\text{EI}Q^{h-72}} \times 100,
 \tag{12}$$

$$\Delta\text{EIP}^{h3} = 100 - \frac{\text{EI}Q^{h-72}}{\text{EI}Q^{h-96}} \times 100.
 \tag{13}$$

The weekday (W^d) and the hour (h) are also added to the input variable to capture the time variations. Finally, the input of the ML algorithms is completed with the above calculated features:

$$\begin{aligned}
 X^h = & [X^h, \text{EI}Prange^{d-1}, \text{EI}Pratio^{d-1}, \text{EI}Q^{h-24}, \text{EI}Q^{h-48}, \text{EI}Q^{h-72}, \\
 & \text{EI}Q^{h-96}, \text{EI}Q^{h-120}, \text{EI}Q^{h-144}, \text{EI}Q^{h-168}, \\
 & \mu\text{EIP}^{h3}, \mu\text{EIP}^{h7}, \Delta\text{EIP}^{h1}, \Delta\text{EIP}^{h2}, \Delta\text{EIP}^{h3}, W^d, h].
 \end{aligned}
 \tag{14}$$

The actual electricity hourly prices constitute the target variable ($y^h = \text{EIP}^h$) of the ML algorithms that are trained and applied to provide their estimations or output (\widehat{y}_m^h).

4.2 Training and Evaluation of the Ensemble ML Algorithms

Five ML algorithms are trained: Random Forest (RF) from bagging ensemble machine learning algorithms category, Light Gradient Boosting Regressor (LGBR), Histogram-Based Gradient Boosting Regressor (HGBR), eXtreme Gradient Boosting (XBR) from boosting category and voting regressors (VR) from voting category.

4.2.1 Setting the Training Interval

The training process reveals that the results depend on the training interval. Thus, for periods with small fluctuations a longer training interval is suitable for ML training, while

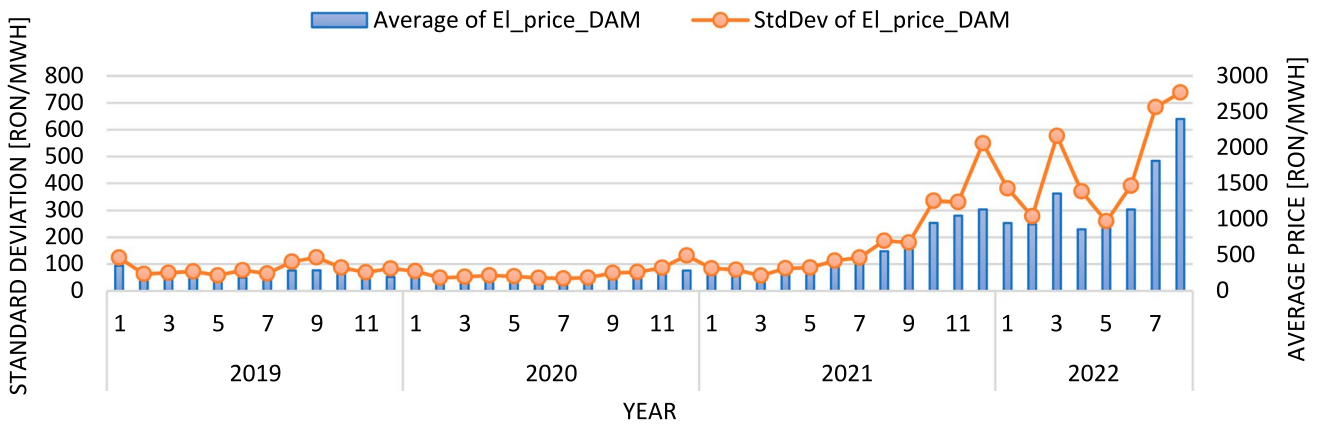


Fig. 11 Monthly standard deviation during 2019 and 2022

for periods with higher fluctuations it is recommended to shorten the training interval. Therefore, the training interval is adapted considering the monthly standard deviation of the price. The greater the deviation, the shorter the training interval. As it can be noticed from Fig. 11, the deviations are higher during 2021–2022. This is also evident from the basic statistics performed at the yearly level and shown in Appendix A (variable EI_price_DAM).

Based on the results obtained during the training process, the following rules can be applied to the training interval (T):

$$T = \begin{cases} 60, & \text{if } \sigma EIP^{m-1} < 100 \\ 45, & \text{if } \sigma EIP^{m-1} \geq 100 \text{ and } \sigma EIP^{m-1} < 200 \\ 30, & \text{if } \sigma EIP^{m-1} \geq 200 \text{ and } \sigma EIP^{m-1} < 300 \\ 15, & \text{if } \sigma EIP^{m-1} \geq 300 \end{cases} \quad (15)$$

where σEIP^{m-1} represents the previous month standard deviation of electricity price. The ML models are trained and tested on the pair (X^h, y^h) and provide their output \hat{y}_k^h that is evaluated according to Eqs. (18–20) and used in the next step to train a regressor model and obtain the final prediction of electricity price.

4.2.2 Weighted Forecast of Electricity Price

The estimations provided by the ML algorithms are used as input for a regressor model to obtain a weighed forecast that increases the accuracy of the individual models. The final forecast of the hourly electricity prices for the next day ($hf = 1, 24$) is obtained using Eq. (16) by adjusting the individual estimations of the ML models (y_k^{hf}) with a set of weights (θ_k):

$$\widehat{EIP}^{hf} = \sum_k \theta_k \times y_k^{hf} \quad (16)$$

The weights are calculated with a regressor model (e.g., using linear regression or gradient descent) by minimizing the difference between the actual electricity prices and their estimation during the training interval (T). Therefore, the regressor model minimizes the sum of squared errors using Eq. (17). Figure 24 (Appendix C) represents the steps of the proposed methodology to obtain the forecast of the hourly electricity prices.

$$\begin{aligned} \theta_k &= \operatorname{argmin}_{\theta_k} \sum_{h \in T} \left(EIP^h - \widehat{EIP}^h \right)^2 \\ &= \operatorname{argmin}_{\theta_k} \sum_{h \in T} \left(EIP^h - \sum_k \theta_k \times \hat{y}_k^h \right)^2 \end{aligned} \quad (17)$$

4.2.3 Assessing 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):

$$RMSE = \frac{1}{T} \sum_{h=1}^T \left(y^h - \hat{y}_k^h \right)^2, \quad (18)$$

$$R^2 = 1 - \frac{\sum_{h=1}^T \left(y^h - \hat{y}_k^h \right)^2}{\sum_{h=1}^T \left(y^h - \bar{y}^h \right)^2}, \quad (19)$$

$$MAPE = \frac{1}{T} \sum_{h=1}^T \frac{|y^h - \hat{y}_k^h|}{|y^h|} \times 100\%, \quad (20)$$

Table 2 Results for 2019 and 2020

Metric/Year	2019			2020		
	Baseline No feature engineering	5 ML algorithms	Weighted forecast	Baseline No feature engineering	5 ML algorithms	Weighted forecast
R^2	0.73	0.96	0.97	0.75	0.97	0.98
MAE	15.18	7.82	7.58	14.64	6.87	6.71
RMSE	29.34	16.14	15.85	27.67	15.01	14.67
MAPE	12.98	4.06	3.67	11.76	3.89	3.37

Table 3 Results for 2021 and 2022

Metric/year	2021			2022		
	Baseline No feature engineering	5 ML algorithms	Weighted forecast	Baseline No feature engineering	5 ML algorithms	Weighted forecast
R^2	0.68	0.93	0.95	0.65	0.92	0.93
MAE	19.35	9.14	8.67	23.88	11.03	10.23
RMSE	40.12	18.20	17.39	43.56	24.12	23.9
MAPE	26.94	5.05	4.76	29.18	6.43	5.88

$$\text{MAE} = \frac{1}{T} \sum_{h=1}^T |y^h - \hat{y}_k^h|. \quad (21)$$

For the weighted forecast, the same equations are used by replacing y_k^h with $\widehat{\text{EIP}}^h$.

5 Results and Assessment

The training interval varies according to Eq. (15) and the forecasting horizon is 7 days although the focus is on the day-ahead forecast (or the electricity prices for the next 24 h) that is required for DAM. The comparison is provided in Tables 2 and 3 between the proposed method (weighted forecast), the average results of the five ML algorithms and the baseline approach with fixed training interval and no feature engineering. The results for the baseline approach are obtained by calculating the yearly average of the estimations of the five ML algorithms without feature engineering.

Several graphical results are extracted and showcased in Figs. 12, 13, 14, 15 and 16. Results are grouped by several consecutive days in each year (2019–2022). The forecast is performed for the next 7 days, but the results are reliable for only for the first days: the more distant the time horizon, the greater the prediction errors. For 2019–2021, the forecast is reliable for the next 4–5 consecutive days, whereas for 2022, only the electricity prices for the next 2 consecutive days are accurately predicted. P1–P5 represent the estimations (\hat{y}_k^h) of the five ML algorithms described in Sect. 4.2; El_price_DAM (orange) represents the actual

electricity price (y^h); El_price_DAM_F (blue) represents the average of the estimations obtained by the five ML algorithms (average of the \hat{y}_k^h values). For simulations performed for 2019 and 2020, the results are showcased in Figs. 12 and 13.

For simulations performed for 2021, the results are showcased in Fig. 14.

For simulations performed for 2022, the results are showcased in Figs. 15 and 16.

For stable years such as 2019 and 2020, the forecast was more accurate, the model providing very good results. Additionally, for more disruptive years (2021 and 2022), the forecast provided reliable results by adjusting the training interval and extracting features that enhance the proposed model.

The weighted forecast ($\widehat{\text{EIP}}^h$) is obtained using two regressors (linear regression and gradient descent) that calculate the weights (θ_k) and adjust the predictions obtained with the five ML algorithms. El_price_DAM_PF represents the weighted forecast ($\widehat{\text{EIP}}^h$) that is obtained either by linear regression or gradient descent (as in Figs. 17, 18, 19). The results are provided for days in 2019 and 2022, showing that the proposed method proved to be robust for calmer and more thrilling intervals. Our approach relies on adjusting training intervals based on the monthly standard deviation, multiple fundamental features from various open data sources, feature engineering, several standout ML algorithms and combining the individual estimations of the ML algorithms with a set of weights. The regressor model (either linear regression or gradient descent) minimizes the sum of squared errors. For simulations performed for 2019 (known

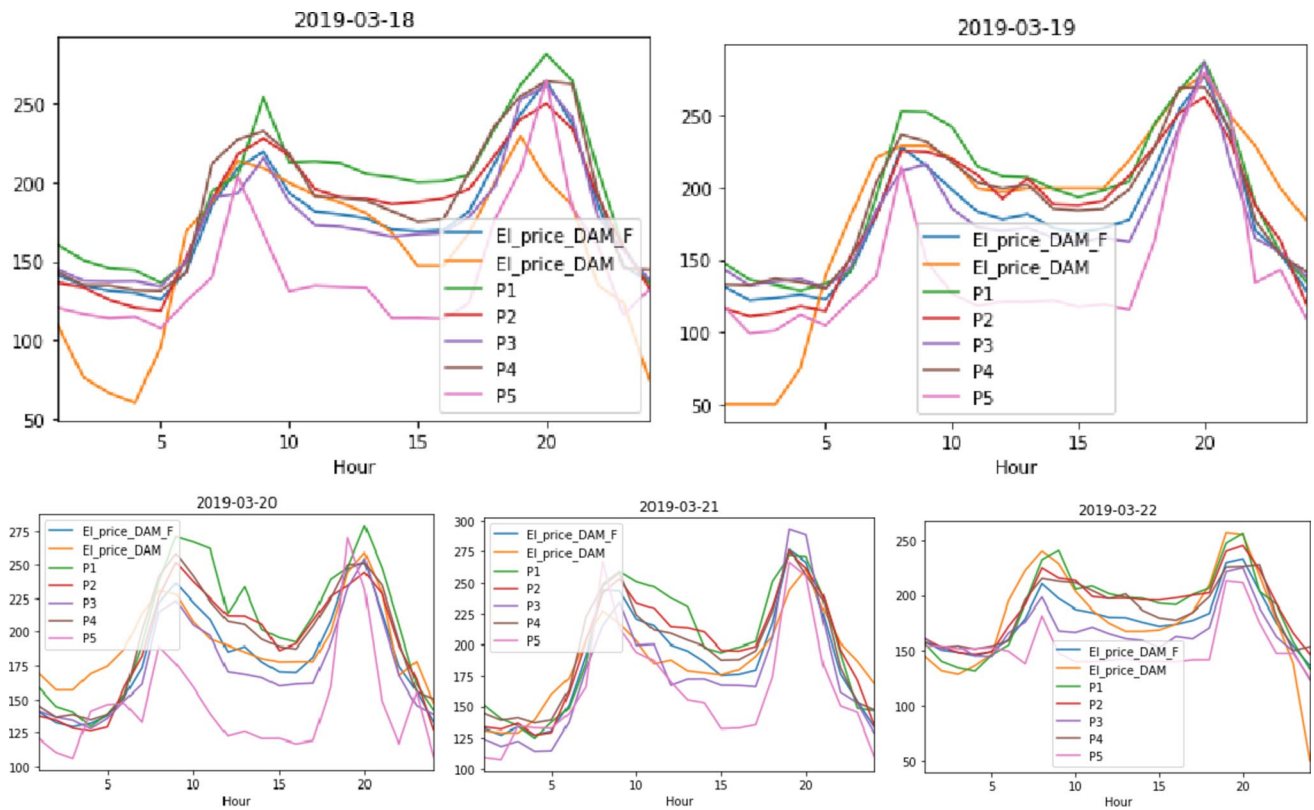


Fig. 12 Forecast of the hourly electricity price on DAM (RON/MWh) for 5 days in March 2019

as a stable year from the price fluctuations point of view) using the proposed methodology, the results are showcased in Figs. 17 and 18.

For simulations performed for 2022 (known as a disruptive year from the price fluctuations point of view) using the proposed methodology, the results are showcased in Fig. 19a, b. Thus, based on the results, the proposed EPF approach is able to cope with random events such as the COVID-19 pandemic, lockdowns and conflict in Ukraine.

6 Discussion

In this paper, we investigated a newly built data set with numerous variables that spans from January 2019 to August 2022. It entirely covers the interval before and after COVID-19 pandemic and the emergence of the conflict in the Black Sea region. For interpreting the results, we grouped them on the two types of years: more stable 2019 and 2020 and more disruptive from the price fluctuations point of view: 2021 and 2022. The numerical results are presented in Tables 2 and 3. Numerical results indicate that on average MAE improved by 48% and RMSE improved by 44% over

the entire interval compared to the baseline model (without feature engineering).

When the output of the ML algorithms is weighted using the proposed meta-model, MAE further improved by 2.3% in 2020 and 5.14% in 2022. Less errors are recorded in stable years like 2019 and 2020 (MAE = 6.71, RMSE = 14.67) compared to 2021 and 2022 (MAE = 9.45, RMSE = 20.64). MAPE and coefficient of determination follow similar trends. The latter significantly improved from 0.73 to 0.97 in 2019, from 0.75 to 0.98 in 2020, from 0.68 to 0.95 in 2021 and from 0.65 to 0.93 in 2022.

When the results of the ML algorithm were averaged, MAPE decreased from 12.98 to 4.06 compared to the baseline model in 2019, from 11.76 to 3.89 in 2020, from 26.94 to 5.05 in 2021 and from 29.18 to 6.43. It further decreased by around 1 unit when the proposed model was applied. The lowest error values were recorded in the more stable years: 2019 and 2020. However, good results were obtained for 2021 and 2022 when the price fluctuations were much higher.

To graphically visualize the prediction results, they were depicted in several figures showcased in the previous section. Several consecutive days (in March and April) were chosen to show the hourly price variation and the

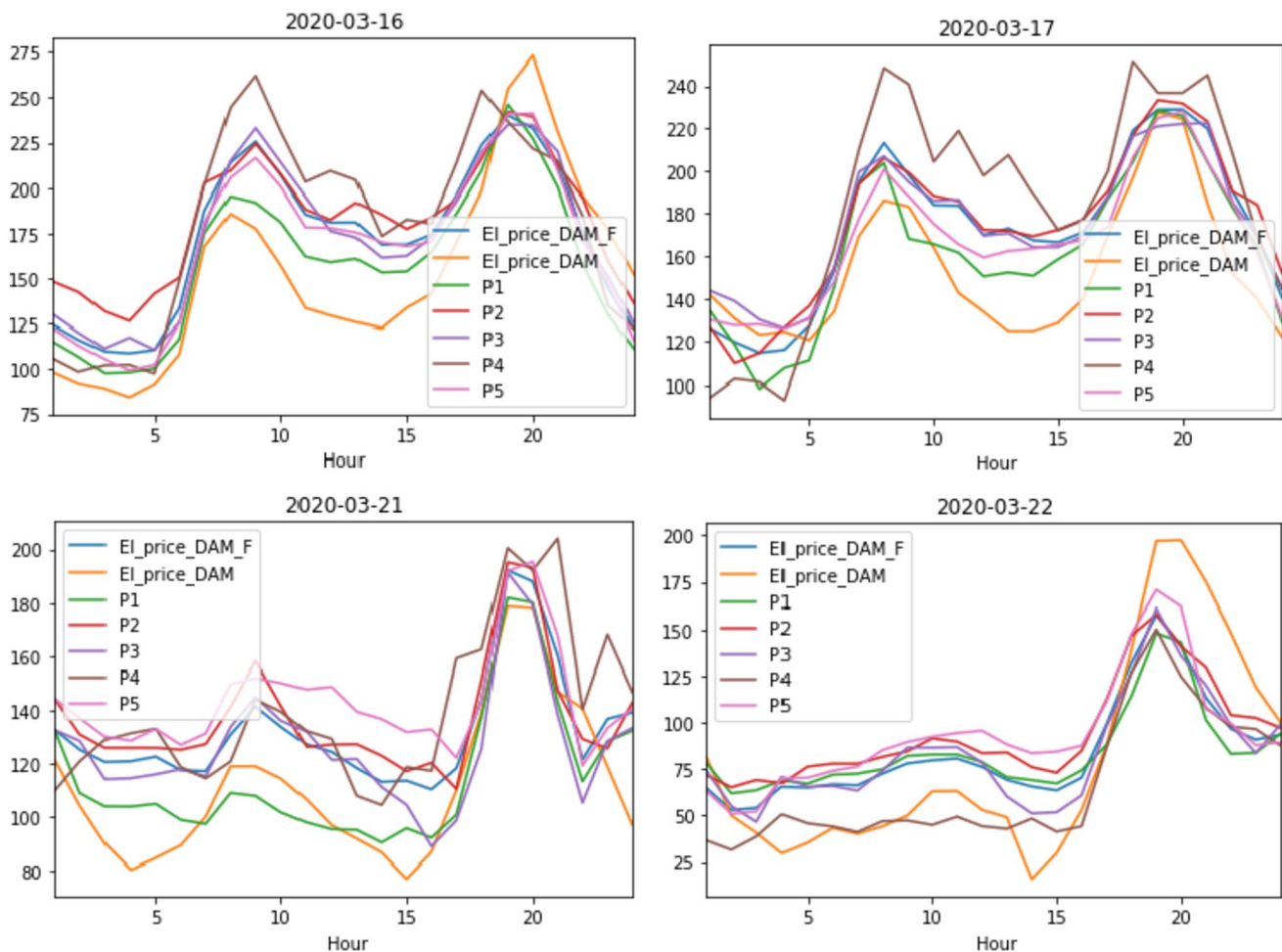


Fig. 13 Forecast of the hourly electricity price on DAM (RON/MWh) for 4 days in March 2020

prediction performed with the individual ML algorithms and with the proposed model in which their results are combined. One can notice that the forecasting lines follow the price for both stable and disruptive intervals, indicating a good predictive capacity of the proposed model. The linear regressor meta-model provides better results. Additionally, the average output of the ML models (blue line) proved to be a good estimation.

7 Conclusion

To cope with black swans, a robust method for EPF is required, in which more fundamental features and feature engineering should be considered. The training interval requires adjustment, the more price varies the shorter the training. They are adjusted based on the price variations to obtain the best performance. Aggregated features and features generated by shifting the interval from 24 to 168 h are important especially for the last 2 years. Finally, two

approaches are taken into account to combine the results of the five standout ML algorithms—random forest, light gradient boosting, histogram-based gradient boosting, extreme gradient boosting and voting regressors. Thus, the results are weighted using two regressors (linear regression and gradient descent). The algorithms learn from different features depending on the forecasting interval.

From the exploratory data analysis, we noticed that the most important features in 2019 lost their importance and other features become significant in the coming years marked by black swans or random events, such as COVID-19 and conflict in Ukraine. Some features lost importance (e.g., total consumption in 2021 and 2022 is less important) and other emerged from uncertainties (e.g., gas price, price index, interest rate, emissions allowance, water level) brought by the random events or extreme climate conditions. In simulations, we also added the electricity prices and quantities traded on other DAMs from neighboring countries (Bulgaria, Serbia, Hungary). Moreover, the capacity flows on the

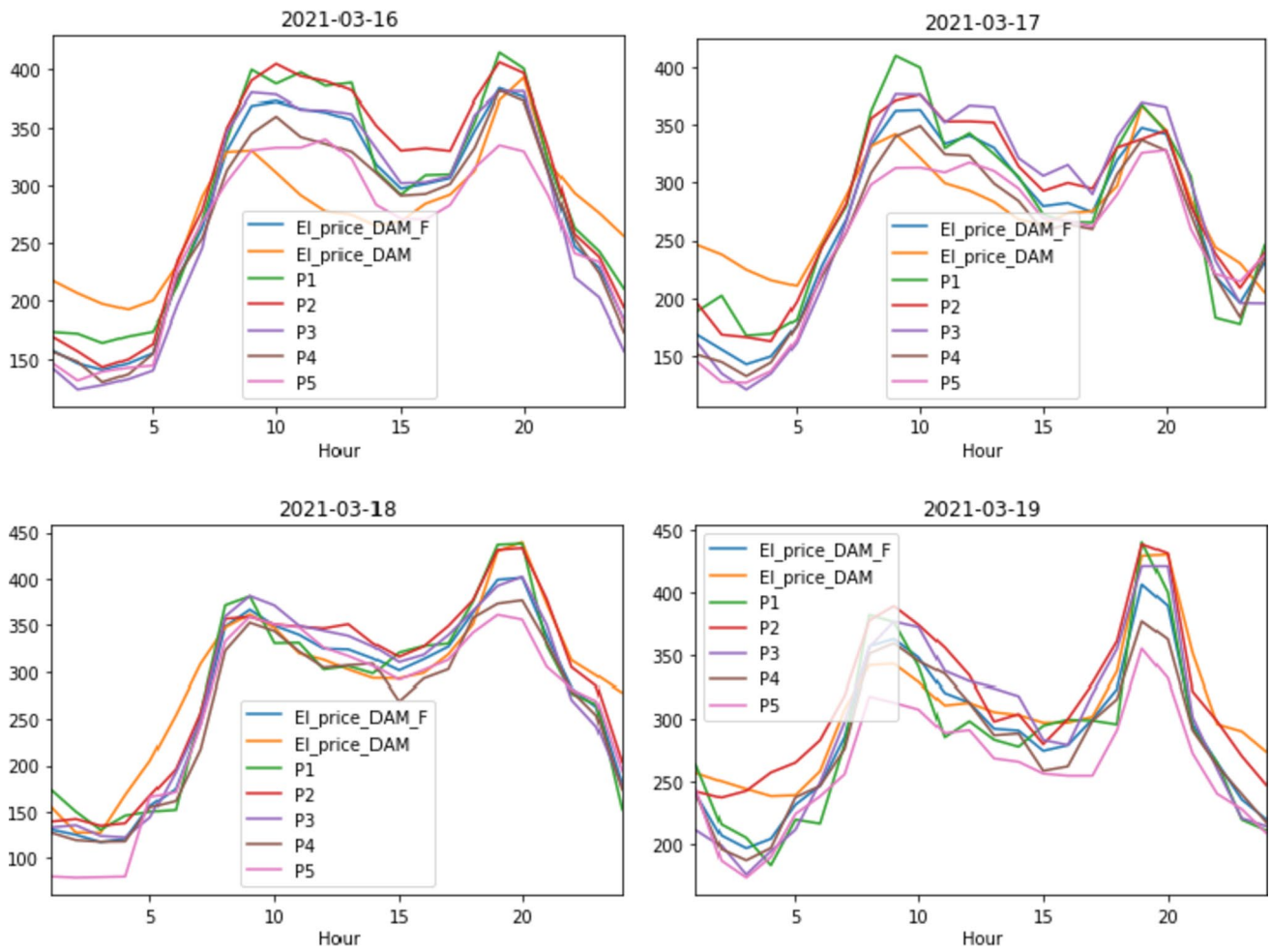


Fig. 14 Forecast of the hourly electricity price on DAM (RON/MWh) for 4 days in March 2021

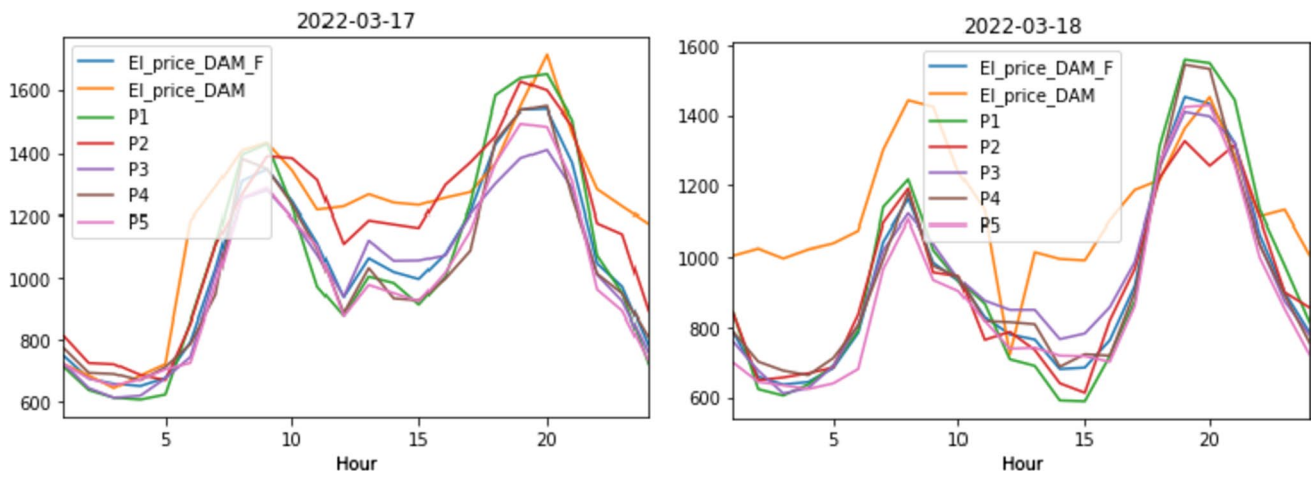


Fig. 15 Forecast of the hourly electricity price on DAM (RON/MWh) for 2 days in March 2022

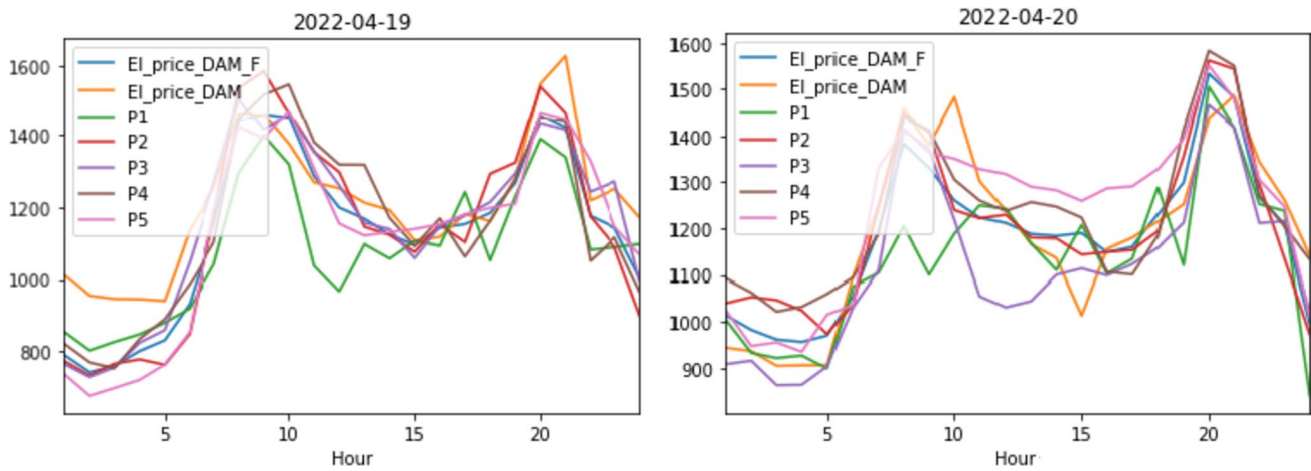


Fig. 16 Forecast of the hourly electricity price on DAM (RON/MWh) for 2 days in April 2022

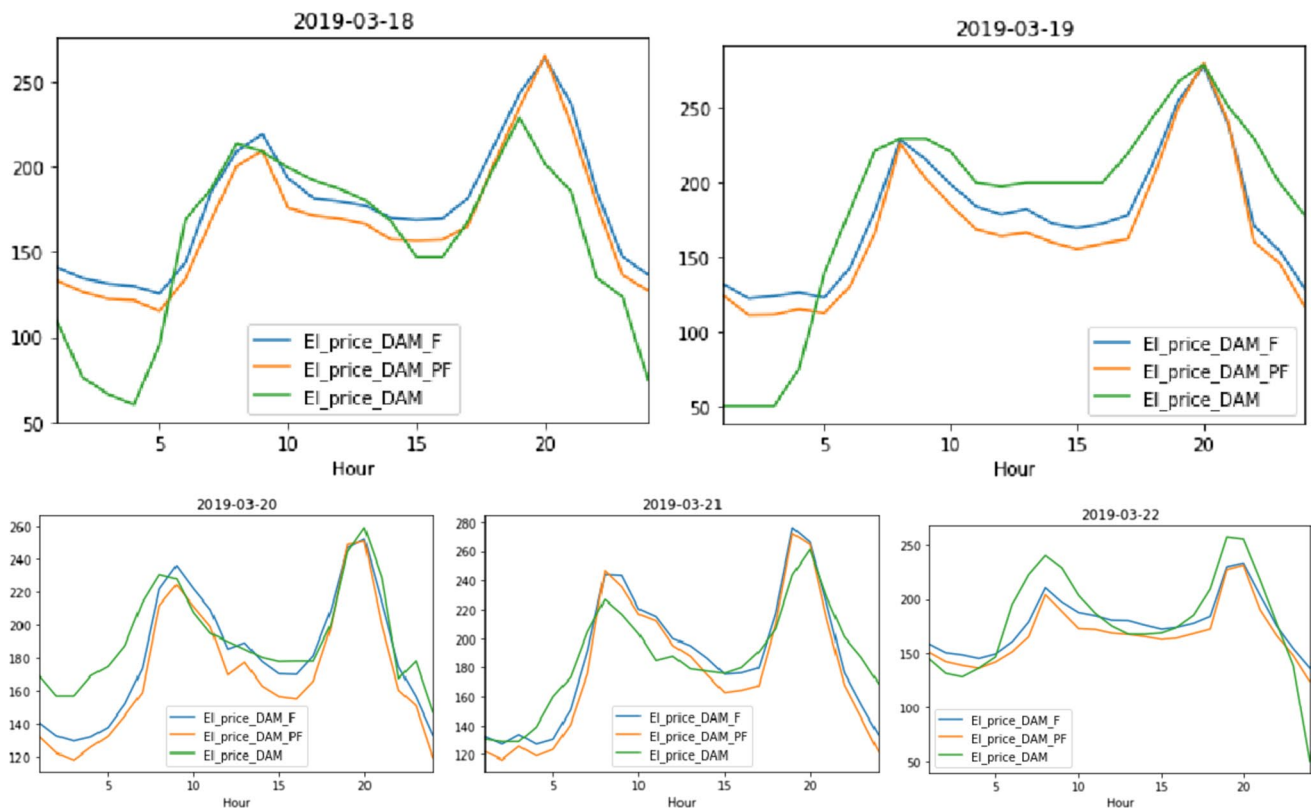


Fig. 17 Forecast of the hourly electricity price on DAM (RON/MWh). Weighted the output of the five ML algorithm using a linear regressor for 5 days in March 2019

interconnection with Hungary was added in the model as fundamental features, but these did not improve the performance of the results and they were removed from the input dataset.

From investigating the evolution of the variables between January 2019 and August 2022, we found out that

the shape of the electricity price showed sharper humps in the last 2 years proving that the most expensive generators are used at peaks. RES producers and even other producers with a more predictable output tend to trade more on DAM. The share of electricity traded on DAM

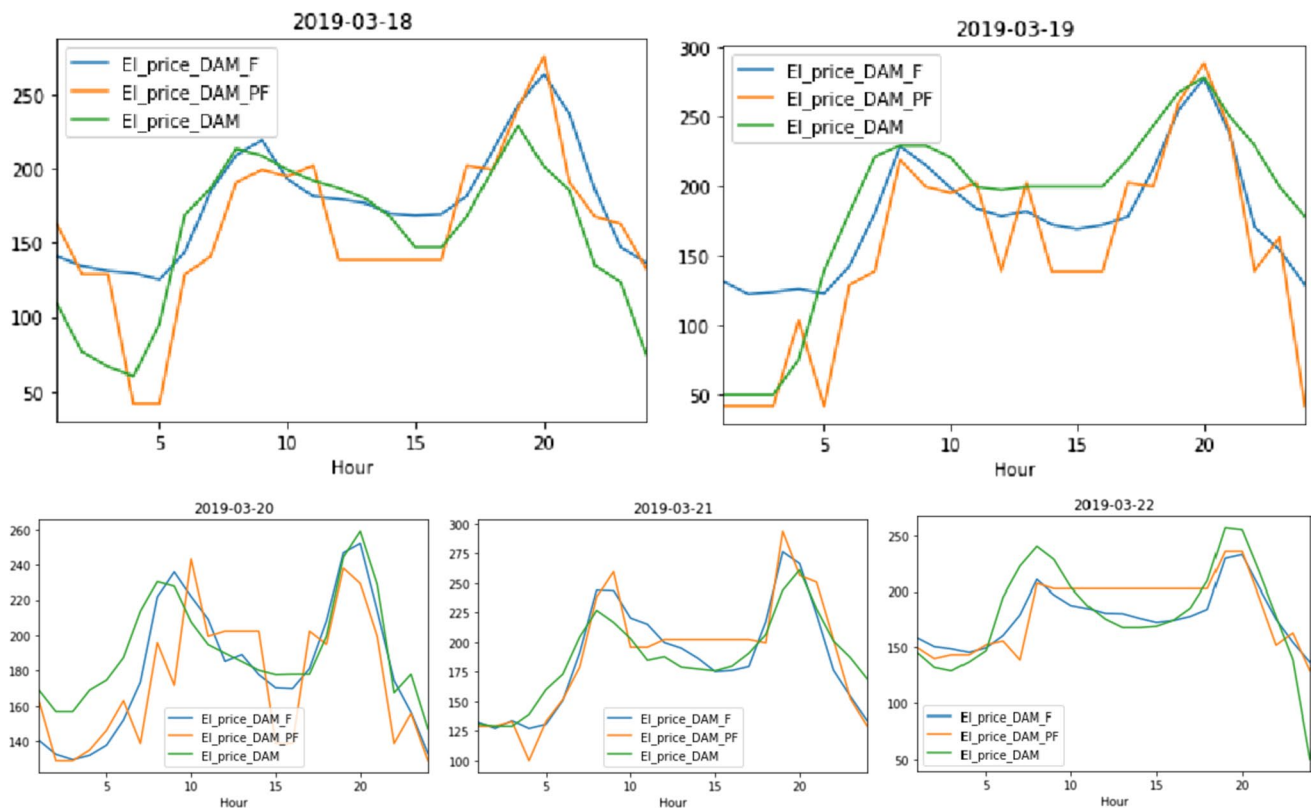


Fig. 18 Forecast of the hourly electricity price on DAM (RON/MWh). Weighted the output of the five ML algorithm using gradient descent for 5 days in March 2019

in the total consumption increased at 46% in 2022 (in the interval January–August). The humps of the electricity price shape indicate that the uniform price mechanism may not be appropriate for DAM in the context of severe drought, higher gas and oil prices, higher price index and increased consumption level. The most expensive price for each MWh will be paid to all generators regardless of their costs, but the burden on the electricity suppliers and final consumers is tremendous. Alternative price mechanisms should be investigated and applied for DAM, such as more popular pay-as-bid, or less speculative generalized second-price or Vickrey–Clark–Glove pricing mechanisms.

In Romania, governmental measures to protect consumers consider capping the price of the electricity to the households to 0.68 RON/kWh if the consumption is less or equal to 100 kWh, 0.8 RON/kWh if the consumption is between 101 and 255 kWh. For those with higher consumption, the market price that can be five times higher will be applied. Some offers are as high as 4 RON/kWh. However, for this measure, the budgetary impact is 1 billion RON/month up to August 2023. One of the drawbacks of this measure is that there is no incentive for consumers to reduce the actual consumption since the

thresholds are set for 2021. One suggestion is to consider the current consumption and better stimulate consumption reduction.

To reduce the burden on the governmental budget, alternatives should be further analyzed: such as more incentives for PV panel installations at the households' and small and medium business's side. Since December 2021, the greenhouse program was stopped for verifications of files due to the lack of administrative capacity. This capacity should be increased, and the process speed up. Alternatively, the flexibility of consumers to adjust or reduce consumption should be moreover investigated. What if the consumers' flexibility is cheaper than the financial effort of capping the price? Therefore, to combat spikes, as future work, the flexibility of consumers and the storage facilities potential will be investigated.

One of the limitations is related to the data extraction as in the current paper, the data were scraped from various sources. When the page source or libraries/drivers change over time, the scraper has to be updated to further extract data. Therefore, an API or a download option could provide more reliability in terms of data extraction. Another limitation is related to the variables that are selected for Romania in the current paper, but for other country, more or less variables could be considered, thus the variables may depend on the region.

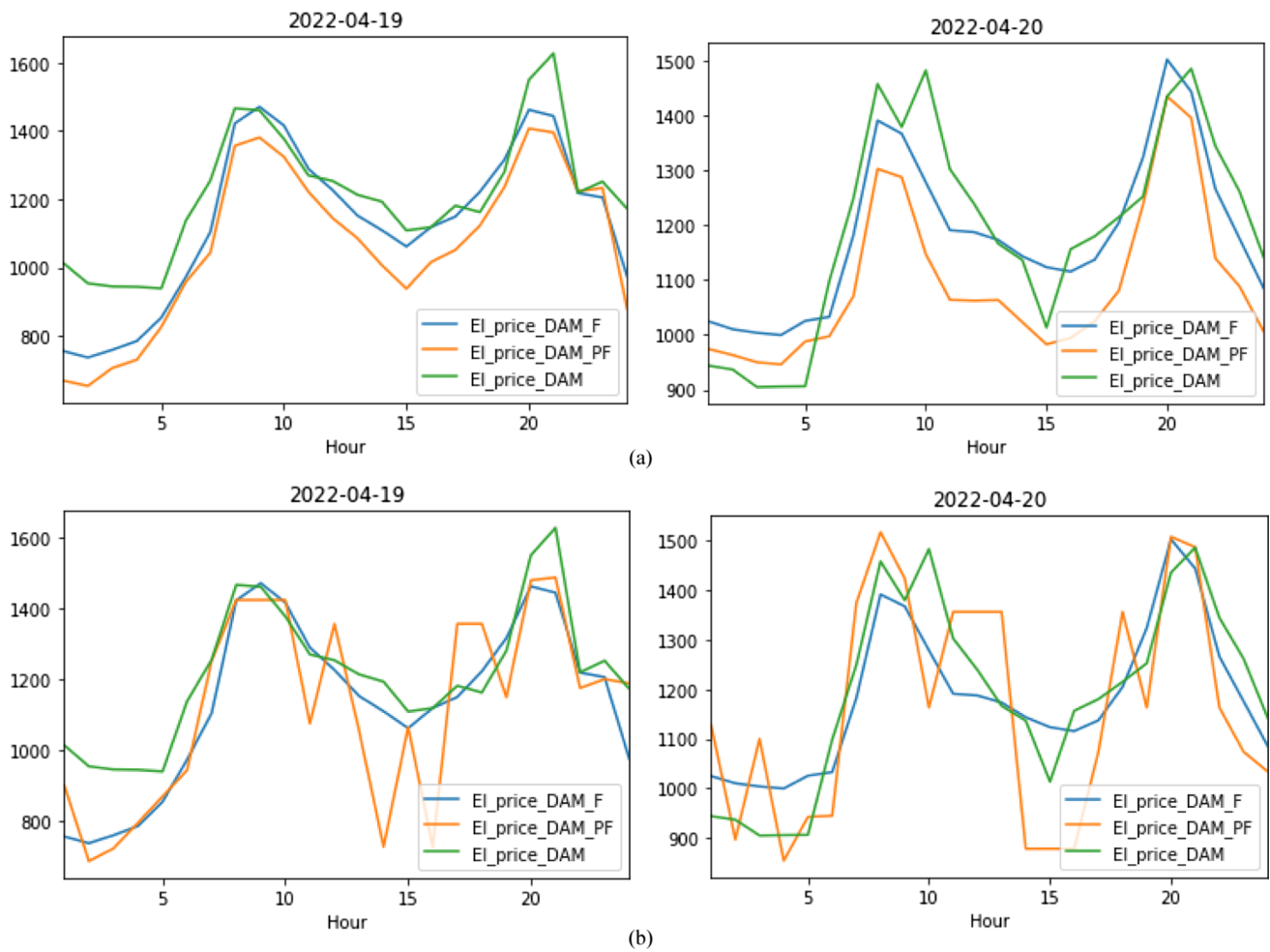


Fig. 19 Forecast of the hourly electricity price on DAM (RON/MWh). **a** weighted the output of the five ML algorithm using a linear regressor; **b** weighted the output of the five ML algorithm using gradient descent for 2 days in April 2022

Appendix A

See Table 4, 5, 6, 7.

Table 4 Electricity, gas, oil prices and quantities

Stat	El_price_DAM	El_quantity	Gas_price_DAM	Gas_quantity_DAM	Oil_price
Mean	505.53	2853.92	188.33	528,681.48	67.65
Standard error	3.07	3.29	1.29	9385.76	0.13
Median	278.00	2828.40	74.98	11,837.00	65.16
Mode	229.00	2686.00	0.10	376,800.00	101.13
Standard deviation	550.73	589.39	231.75	1,682,435.86	23.64
Sample variance	303,307.08	347,386.07	53,707.14	2,830,590,421,508.72	558.82
Kurtosis	7.33	- 0.23	2.30	43.22	0.20
Skewness	2.45	0.28	1.64	5.64	0.48
Range	4698.99	3808.80	1207.80	19,879,969.00	124.06
Minimum	0.01	1244.10	0.07	47.00	9.12
Maximum	4699.00	5052.90	1207.87	19,880,016.00	133.18
Sum	16,243,770.74	91,702,079.90	6,051,482.78	16,987,593,470.77	2,173,669.25

Table 5 EU Inflation, emission allowance and Danube water levels

Stat	Price_EUETS	Inflation_EU	Level_Turnu Magurele	Level_BR	Level_TL
Mean	43.94	2.47	166.05	225.57	148.44
Standard error	0.13	0.02	0.66	0.68	0.40
Median	29.78	1.30	147.00	209.00	139.00
Mode	25.68	- 0.30	60.00	198.00	97.00
Standard deviation	23.50	2.73	118.80	121.89	72.19
Sample variance	552.08	7.45	14,114.59	14,857.45	5210.95
Kurtosis	- 0.85	0.18	0.62	0.19	0.32
Skewness	0.82	1.21	0.91	0.66	0.74
Range	82.31	9.20	612.00	626.00	387.00
Minimum	16.12	- 0.30	0.00	0.00	0.00
Maximum	98.43	8.90	612.00	626.00	387.00
Sum	1,411,751.46	79,498.80	5,335,623.00	7,247,944.00	4,769,568.00

Table 6 Total consumption, generation, exchange level, Romanian inflation and interest rate

Stat	Consumption	Generation	Exchange	Inflation_RO	ROBOR 3 M
Mean	6763.46	6528.09	607.06	6.10	2.97
Standard error	5.78	5.61	2.33	0.02	0.01
Median	6750.67	6484.83	543.67	4.74	2.94
Mode	6324.17	6864.83	93.33	15.18	3.05
Standard deviation	1035.10	1006.10	417.31	3.73	1.44
Sample variance	1,071,425.68	1,012,240.24	174,150.58	13.93	2.06
Kurtosis	- 0.66	- 0.23	0.15	0.81	3.36
Skewness	0.15	0.17	0.73	1.42	1.81
Range	5536.50	6479.33	2373.90	12.71	6.65
Minimum	4087.17	3323.50	12.43	2.47	1.49
Maximum	9623.67	9802.83	2386.33	15.18	8.14
Sum	217,215,133.35	209,656,026.94	19,496,441.30	195,856.70	95,458.06

Table 7 Breakdown on generation source

Stat	Coal_gen	Oil&Gas_gen	Hydro_gen	Nuclear_gen	Wind_gen	PV_gen	Biomass_gen
Mean	1293.24	1153.65	1809.13	1281.78	774.87	157.52	58.45
Standard error	1.75	2.39	3.87	1.40	3.77	1.27	0.08
Median	1287.17	1189.58	1711.58	1385.17	563.75	4.67	60.50
Mode	1186.17	790.00	1717.67	1423.67	12.00	0.00	64.00
Standard deviation	313.05	429.06	694.43	250.34	675.76	227.18	13.52
Sample variance	98,002.50	184,090.81	482,231.97	62,670.43	456,646.64	51,611.55	182.91
Kurtosis	0.42	- 0.82	0.45	1.72	- 0.06	0.21	- 0.35
Skewness	- 0.04	- 0.05	0.77	- 1.89	0.94	1.25	- 0.36
Range	2227.79	1926.00	4183.00	822.83	2809.33	868.00	152.83
Minimum	279.50	211.67	160.33	611.17	0.83	0.00	19.00
Maximum	2507.29	2137.67	4343.33	1434.00	2810.17	868.00	171.83
Sum	41,533,665.85	37,050,666.48	58,102,159.89	41,165,521.76	24,885,754.12	5,059,049.75	1,877,297.05

Appendix B

See Figs. 20, 21, 22 and 23.

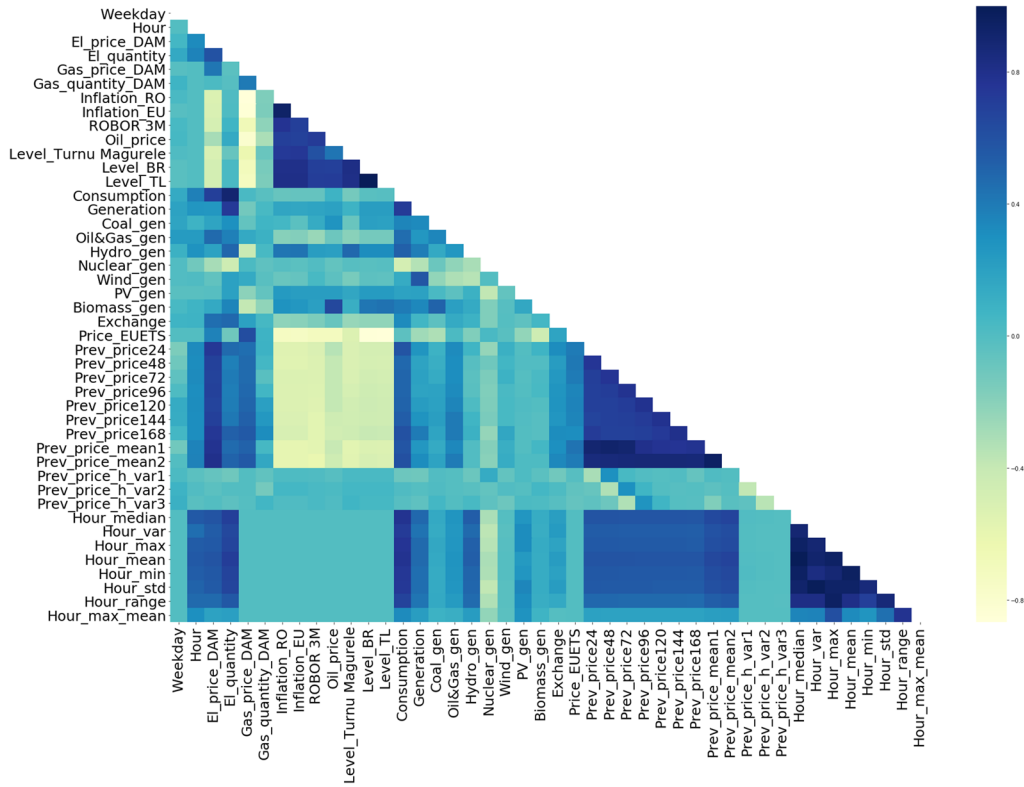


Fig. 20 Pearson correlation coefficients in 2019

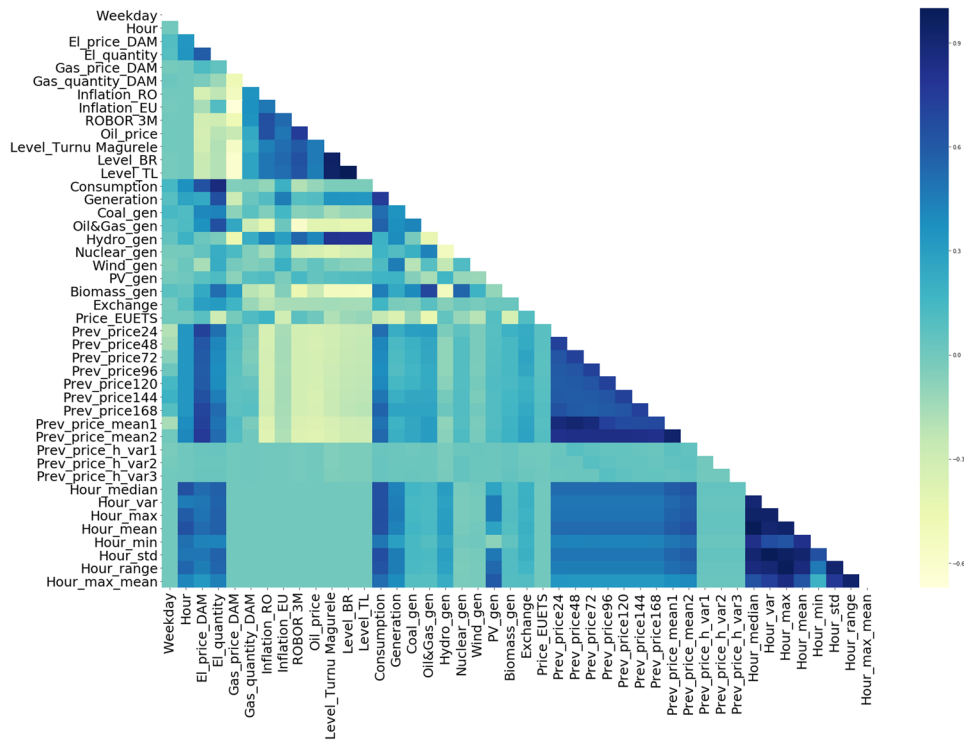


Fig. 21 Pearson correlation coefficients in 2020

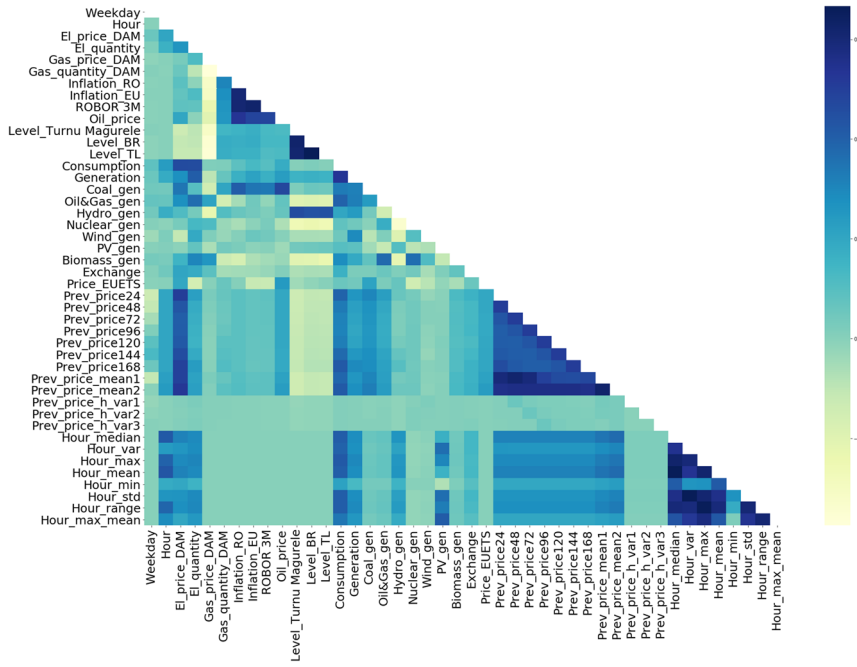


Fig. 22 Pearson correlation coefficients in 2021

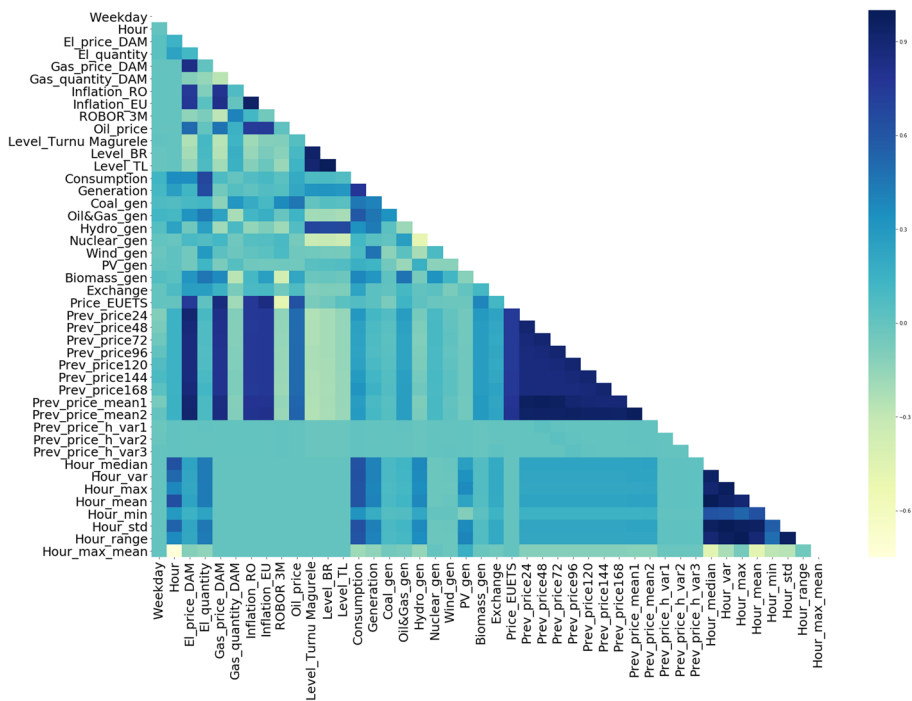


Fig. 23 Pearson correlation coefficients in 2022

Appendix C

See Fig. 24.

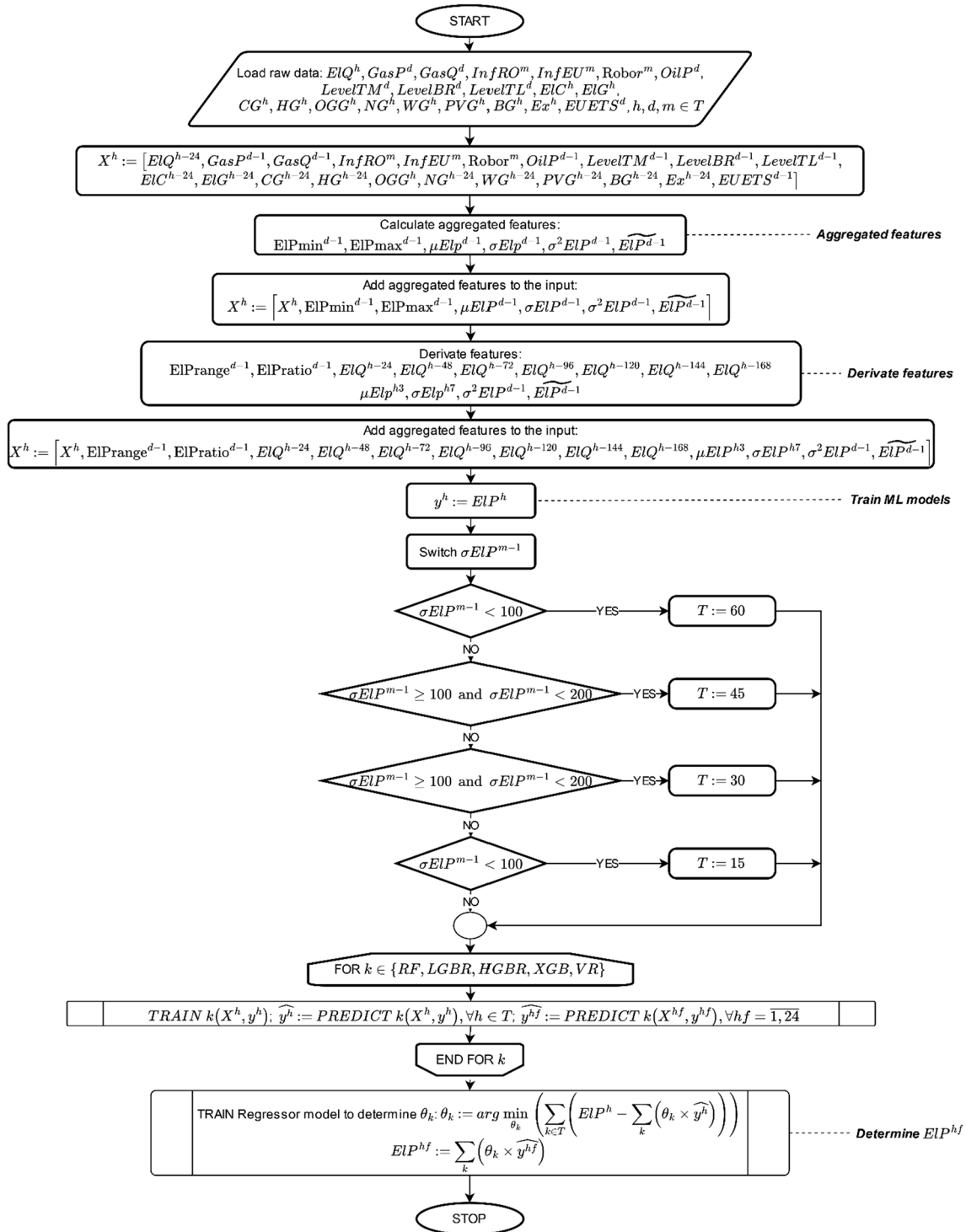


Fig. 24 Flowchart of the proposed methodology to obtain the EPF

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

Author contributions AB: conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing—original draft, writing—review and editing, visualization, and supervision. SVO: conceptualization, validation, formal analysis, investigation, writing—original draft, writing—review and editing, visualization, and project administration.

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

Availability of data and materials Data will be made available on request.

Declarations

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al-Janabi, S., Al-Barmani, Z.: Intelligent multi-level analytics of soft computing approach to predict water quality index (IM12CP-WQI). *Soft. Comput.* (2023). <https://doi.org/10.1007/s00500-023-07953-z>
- Al-Janabi, S., Al-Janabi, Z.: Development of deep learning method for predicting DC power based on renewable solar energy and multi-parameters function. *Neural Comput. Appl.* (2023). <https://doi.org/10.1007/s00521-023-08480-6>
- Al-Janabi, S., Alkaim, A.: A novel optimization algorithm (Lion-AYAD) to find optimal DNA protein synthesis. *Egypt. Inform. J.* (2022). <https://doi.org/10.1016/j.eij.2022.01.004>
- Al-Janabi, S., Alkaim, A., Al-Janabi, E., Aljeboree, A., Mustafa, M.: Intelligent forecaster of concentrations (PM_{2.5}, PM₁₀, NO₂, CO, O₃, SO₂) caused air pollution (IFCsAP). *Neural Comput. Appl. Comput. Appl.* (2021). <https://doi.org/10.1007/s00521-021-06067-7>
- Al-Janabi, S., Alkaim, A.F.: A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation. *Soft. Comput.* (2020). <https://doi.org/10.1007/s00500-019-03972-x>
- Al-Janabi, S., Alkaim, A.F., Adel, Z.: An innovative synthesis of deep learning techniques (DCapsNet & DCOM) for generation electrical renewable energy from wind energy. *Soft. Comput.* (2020). <https://doi.org/10.1007/s00500-020-04905-9>
- Al-Janabi, S., Mohammad, M., Al-Sultan, A.: A new method for prediction of air pollution based on intelligent computation. *Soft. Comput.* (2020). <https://doi.org/10.1007/s00500-019-04495-1>
- Al-Janabi, Z.K., Al-Janabi, S.: An efficient predictor of renewable energy based on deep learning technique (DGBM) and multi-objectives optimization function (2023). <https://doi.org/10.1109/iiccit55816.2022.10010380>
- Alkaim, A.F., & Al-Janabi, S.: Multi objectives optimization to gas flaring reduction from oil production. In: *Lecture Notes in Networks and Systems* (2020) https://doi.org/10.1007/978-3-030-23672-4_10
- Băra, A., Oprea, S.-V., Georgescu, I.A.: Understanding electricity price evolution—day-ahead market competitiveness in Romania. *J. Bus. Econ. Manag.* **24**(2), 221–244 (2023)
- Băra, A., Oprea, S.-V., Tudorică, B.G.: From the East-European regional day-ahead markets to a global electricity market. *Comput. Econ. Econ.* (2023). <https://doi.org/10.1007/s10614-023-10416-0>
- Bashir, N., Irwin, D., Shenoy, P.: A probabilistic approach to committing solar energy in day-ahead electricity markets. *Sustain. Comput. Inform. Syst.* (2021). <https://doi.org/10.1016/j.suscom.2020.100477>
- Chai, S., Xu, Z., Jia, Y.: Conditional density forecast of electricity price based on ensemble ELM and Logistic EMOS. *IEEE Trans. Smart Grid* (2019). <https://doi.org/10.1109/TSG.2018.2817284>
- Dragasevic, Z., Milovic, N., Djuricic, V., Backovic, T.: Analyzing the factors influencing the formation of the price of electricity in the deregulated markets of developing countries. *Energy Rep.* (2021). <https://doi.org/10.1016/j.egy.2021.07.046>
- Fragkioudaki, A., Marinakis, A., Cherkaoui, R.: Forecasting price spikes in European day-ahead electricity markets using decision trees. In: *International conference on the European energy market, EEM* (2015) <https://doi.org/10.1109/EEM.2015.7216672>
- Gencer, B., Larsen, E.R., van Ackere, A.: Understanding the coevolution of electricity markets and regulation. *Energy Policy* (2020). <https://doi.org/10.1016/j.enpol.2020.111585>
- Georgescu, I.: Possibility theory and the risk. *Stud. Fuzz. Soft Comput.* (2012). https://doi.org/10.1007/978-3-642-24740-8_1
- Gianfreda, A., Ravazzolo, F., Rossini, L.: Comparing the forecasting performances of linear models for electricity prices with high RES penetration. *Int. J. Forecast.* (2020). <https://doi.org/10.1016/j.ijforecast.2019.11.002>
- Gürtler, M., Paulsen, T.: The effect of wind and solar power forecasts on day-ahead and intraday electricity prices in Germany. *Energy Econ.* (2018). <https://doi.org/10.1016/j.eneco.2018.07.006>
- Hu, X., Jaraitė, J., Kažukauskas, A.: The effects of wind power on electricity markets: a case study of the Swedish intraday market. *Energy Econ.* (2021). <https://doi.org/10.1016/j.eneco.2021.105159>
- Kadhuim, Z.A., Al-Janabi, S.: Intelligent deep analysis of DNA sequences based on FFGM to enhancement the performance and reduce the computation. *Egypt. Inform. J.* (2023). <https://doi.org/10.1016/j.eij.2023.02.004>
- Karakatsani, N.V., Bunn, D.W.: Fundamental and behavioural drivers of electricity price volatility. *Stud. Nonlinear Dyn. Econ.* (2010). <https://doi.org/10.2202/1558-3708.1657>
- Kath, C., Ziel, F.: The value of forecasts: quantifying the economic gains of accurate quarter-hourly electricity price forecasts. *Energy Econ.* (2018). <https://doi.org/10.1016/j.eneco.2018.10.005>
- Keles, D., Scelle, J., Paraschiv, F., Fichtner, W.: Extended forecast methods for day-ahead electricity spot prices applying

- artificial neural networks. *Appl. Energy* (2016). <https://doi.org/10.1016/j.apenergy.2015.09.087>
25. Lago, J., De Ridder, F., De Schutter, B.: Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. *Appl. Energy* (2018). <https://doi.org/10.1016/j.apenergy.2018.02.069>
 26. Li, W., Becker, D.M.: Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy* (2021). <https://doi.org/10.1016/j.energy.2021.121543>
 38. Li, X., Li, Z., & Guan, X.: Does the investor attention matter in predicting the cny exchange rate? evidence from the asymmetric perspective. *Econ. Comput. Econ. Cybern. Stud. Res.* (2022). <https://doi.org/10.24818/18423264/56.1.22.19>
 27. Liu, J., Wang, J., Cardinal, J.: Evolution and reform of UK electricity market. *Renew. Sustain. Energy Rev.* (2022). <https://doi.org/10.1016/j.rser.2022.112317>
 28. Maciejowska, K., Nitka, W., Weron, T.: Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices. *Energy Econ.* (2021). <https://doi.org/10.1016/j.eneco.2021.105273>
 29. Miranian, A., Abdollahzade, M., Hassani, H.: Day-ahead electricity price analysis and forecasting by singular spectrum analysis. *IET Gener. Transm. Distrib.* (2013). <https://doi.org/10.1049/iet-gtd.2012.0263>
 30. Mohammed, G.S., Al-Janabi, S.: An innovative synthesis of optimization techniques (FDIRE-GSK) for generation electrical renewable energy from natural resources. *Results Eng.* (2022). <https://doi.org/10.1016/j.rineng.2022.100637>
 31. Sánchez de la Nieta, A.A., Contreras, J.: Quantifying the effect of renewable generation on day-ahead electricity market prices: The Spanish case. *Energy Econ.* (2020). <https://doi.org/10.1016/j.eneco.2020.104841>
 32. Shafie-Khah, M., Moghaddam, M.P., Sheikh-El-Eslami, M.K.: Price forecasting of day-ahead electricity markets using a hybrid forecast method. *Energy Convers. Manag.* (2011). <https://doi.org/10.1016/j.enconman.2010.10.047>
 33. Spodniak, P., Ollikka, K., Honkapuro, S.: The impact of wind power and electricity demand on the relevance of different short-term electricity markets: the Nordic case. *Appl. Energy* (2021). <https://doi.org/10.1016/j.apenergy.2020.116063>
 34. Weron, R.: Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int. J. Forecast.* (2014). <https://doi.org/10.1016/j.ijforecast.2014.08.008>
 35. Zhang, C., Li, R., Shi, H., Li, F.: Deep learning for day-ahead electricity price forecasting. *IET Smart Grid* (2020). <https://doi.org/10.1049/iet-stg.2019.0258>
 36. Zhang, J.L., Zhang, Y.J., Li, D.Z., Tan, Z.F., Ji, J.F.: Forecasting day-ahead electricity prices using a new integrated model. *Int. J. Electr. Power Energy Syst.* (2019). <https://doi.org/10.1016/j.ijepes.2018.08.025>
 37. Ziel, F., Steinert, R.: Electricity price forecasting using sale and purchase curves: the X-Model. *Energy Econ.* (2016). <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.