Understanding electricity prices beyond the merit order principle using explainable AI

Julius Trebbien [a,b,*], Leonardo Rydin Gorjão [c], Aaron Praktiknjo [d,f], Benjamin Schäfer [e], Dirk Witthaut [a,b]

[a] *Forschungszentrum Jülich, Institute for Energy and Climate Research – Systems Analysis and Technology Evaluation (IEK-STE), 52428 Jülich, Germany*
[b] *University of Cologne, Institute for Theoretical Physics, Zülpicher Str. 77, 50937 Cologne, Germany*
[c] *Faculty of Science and Technology, Norwegian University of Life Sciences, 1432 As, Norway*
[d] *Chair for Energy System Economics, Institute for Future Energy Consumer Needs and Behavior (FCN), E.ON Energy Research Center, RWTH Aachen University, Mathieustr. 10, 52074 Aachen, Germany*
[e] *Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany*
[f] *JARA-ENERGY, 52074 Aachen, Germany*

GRAPHICAL ABSTRACT
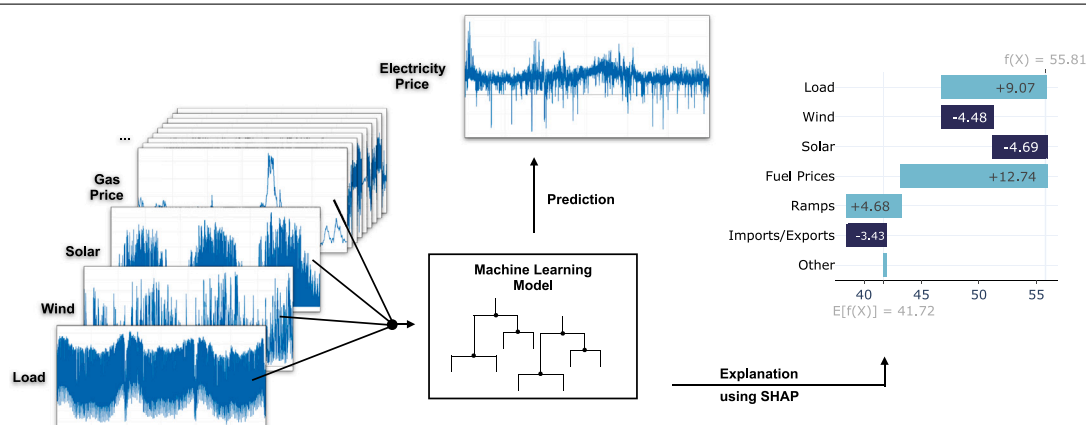
HIGHLIGHTS

- Explainable ML model outperforms benchmark model based on the merit order principle.
- SHAP analysis reveals which features affect electricity prices beyond residual load.
- Load, wind and solar generation are key features, but dependencies differ slightly.
- Model quantifies the impact of fuel prices as well as generation ramps.

ARTICLE INFO

ABSTRACT

Electricity prices in liberalized markets are determined by the supply and demand for electric power, which are in turn driven by various external influences that vary strongly in time. In perfect competition, the merit order principle describes that dispatchable power plants enter the market in the order of their marginal costs to meet

* Corresponding author at: Forschungszentrum Jülich, Institute for Energy and Climate Research – Systems Analysis and Technology Evaluation (IEK-STE), 52428 Jülich, Germany.
E-mail addresses: j.trebbien@fz-juelich.de (J. Trebbien), d.witthaut@fz-juelich.de (D. Witthaut).

the residual load, i.e. the difference of load and renewable generation. Various market models are based on this principle when attempting to predict electricity prices, yet the principle is fraught with assumptions and simplifications and thus is limited in accurately predicting prices. In this article, we present an explainable machine learning model for the electricity prices on the German day-ahead market which foregoes of the aforementioned assumptions of the merit order principle. Our model is designed for an ex-post analysis of prices and builds on various external features. Using SHapley Additive exPlanation (SHAP) values we disentangle the role of the different features and quantify their importance from empiric data, and therein circumvent the limitations inherent to the merit order principle. We show that load, wind and solar generation are the central external features driving prices, as expected, wherein wind generation affects prices more than solar generation. Similarly, fuel prices also highly affect prices, and do so in a nontrivial manner. Moreover, large generation ramps are correlated with high prices due to the limited flexibility of nuclear and lignite plants. Overall, we offer a model that describes the influence of the main drivers of electricity prices in Germany, taking us a step beyond the limited merit order principle in explaining the drivers of electricity prices and their relation to each other.

1. Introduction

The reliable supply of electric power is vital for modern societies [1, 2]. A stable operation of the electric power system requires that power generation and load are always balanced [3]. Electricity markets are central to coordinate generation and demand; from long-term future contract to short-term spot trading. This coordination is becoming increasingly challenging due to the ongoing energy transition [4,5]. Generation from renewable sources such as wind and solar power is determined by the weather and thus highly volatile [6], making also the electricity price highly volatile [7].

European electricity markets were liberalized starting in the 1990s [8]. Before, generation, transmission, and distribution were typically integrated in a single company holding a regional monopoly. In liberalized markets, different suppliers compete which should lead to an improved efficiency and lower costs [8]. Today, several markets exist in Europe which enable electricity trading on different time horizons [9]. In particular, the day-ahead markets of the European Power Exchange (EPEX SPOT) cover 13 countries and reached a market volume of more than 500 TWh in 2019 [9]. Recently, European energy markets were heavily disturbed by the Russian invasion of Ukraine causing an intensified research interest (see, e.g. [10,11]).

The functioning and design of electricity markets is a central topic in energy economics [12,13]. Mechanistic models, such as agent-based models, start from first principles and derive statements that can be verified or falsified on data [14,15]. Data-based approaches act complementary, i.e., they start from data and make no assumptions of underlying mechanisms. General rules that govern the system are inferred, providing valuable inputs to improve mechanistic models. In recent years, machine learning has become an important method in electricity market research. The vast majority of studies focuses on the forecasting of electricity prices [16]. However, modern developments in interpretable machine learning enable much richer applications. In particular, they allow for in-depth scientific insights from large heterogeneous data sets [17].

In this article, we pursue the novel road of offering explanations for the dependencies and correlations within the energy markets and forego focusing on forecasting performance. We establish a machine learning model for the electricity prices on the German day-ahead spot market. We apply SHapley Additive exPlanations (SHAP) [18] to explain the model and provide insights into which factors determine the market price. We use an extended data set to identify driving factors which are commonly neglected in elementary studies or machine learning models [16]. The model substantially outperforms the more commonly employed merit order principle, revealing additional details into the function of the market. For instance, the model quantifies the impact of fossil fuel prices and load ramps, as well as nonlinear interactions of different features, and thus takes us one step further in accurately understanding the drivers of the electricity prices.

The remaining article is structured as follows. In Section 2, we provide an introduction to the German electricity market, market models based on the merit order principle and review previous works in the field. In Section 3, we discuss how we obtained and processed data, how the machine learning model is trained and interpreted. We then continue in Section 4 to analyze the results of the machine learning model, in particular demonstrating how load, wind and solar generation but also fuel prices critically influences electricity prices. We close in Section 5 with a discussion.

2. Background and literature

2.1. The German electricity markets

Electricity is traded on many different markets until it is delivered at a time t_0. Basic trading is done in the forward, day-ahead and intraday markets, while long-term non-public contracts between electricity producers and consumers via over-the-counter agreements are also possible. Here, we give a short overview of the German market structure focusing on day-ahead markets. Notably, Germany shares its bidding zone with Luxembourg and also shared it with Austria until 1st of October 2018. In the following, we will refer to this shared bidding zone as *German markets* for the sake of simplicity.

To ensure that power generation matches the load, electricity is traded until a day before delivery in the day-ahead market or even minutes before delivery in the intraday market. This makes adaptation to sudden events like weather changes or unavailability of power plants possible and therefore reduces the need of expensive control reserves [19]. Since the intraday markets are smaller in volume and open after the day-ahead markets are closed, they are mainly driven by the day-ahead prices and react to short-term system changes like weather forecast error. This leaves the day-ahead markets as the main driver for general electricity prices reacting directly to availability of generation and load capacity.

Day-ahead trading is possible via different exchanges. In Germany the three operating exchanges are EPEX Spot SE, Nord Pool EMCO AS and EXAA AG [20]. Trading is possible from 10:00 until 12:00 the day before delivery at t_0, where all exchanges are coupled with the pan European Single Day-Ahead Coupling (SDAC) to create one coupled market clearing prices (MCP) for all bidding zones [21]. The MCP is calculated using an algorithm called EUPHEMIA, by using the bids and offers from all exchanges and the network constraints provided by the responsible TSOs.

In a simple and uncoupled market, electricity producers place their offer, which is a specific amount of electricity offered for a specific price for a given time period. Oppositely, electricity consumers place their bid, representing a specific amount of electricity they want to consume for a specific price. Once the order book closes, meaning no bids and offers can be placed anymore, the market clearing price is determined as the highest offer that finds a matching bid in each respective time window. The market clearing price is paid by every
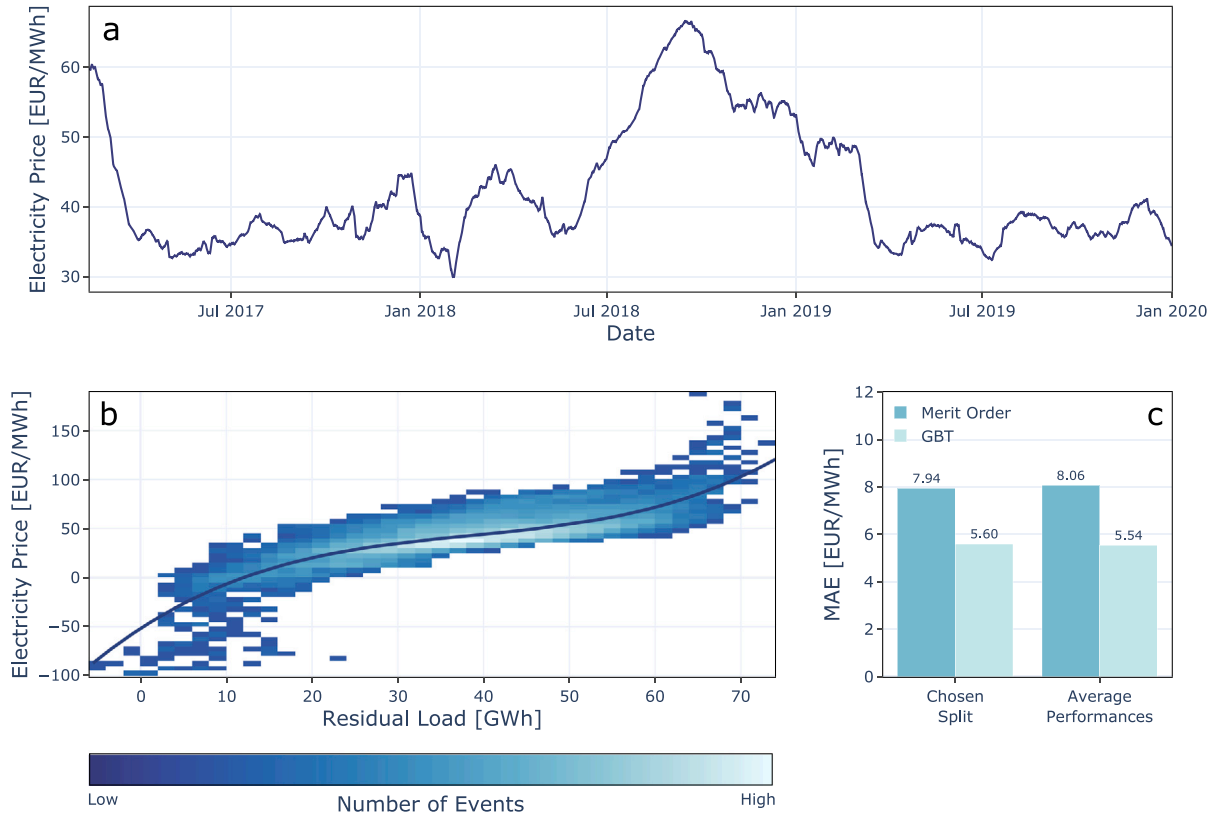
Fig. 1. Explainable Machine Learning for day-ahead electricity prices. (a) Electricity price time series from the German day-ahead EPEX spot market from January 2017 to 2020. (b) In a single-feature benchmark model based on the merit order principle, prices are a function of the residual load, i.e. the difference of load and non-dispatchable renewable generation. The colormap shows a 2D histogram of the raw data, the line is a 3rd order polynomial fit. (c) Performance of the benchmark model and a gradient boosted tree (GBT) model, measured by the mean absolute error (MAE) on the test set. The GBT model outperforms the single-feature benchmark model and reveals more information about the electricity market.

market participant [7]. Looking at the European markets, the SDAC enables cross-border trading, where offers and bids can be matched over different bidding zones, while also respecting network constraints.

Energy in the day-ahead markets is mainly traded for all 24 h of the next day separately while there exist some smaller trading time frames and also specific blocks of important hours e.g. which represent base- and peak-load, depending on the operating exchange.

2.2. The merit order principle and a single-feature price model

Models based on the merit order principle provide a first approximation for the outcome of the day-ahead electricity market. This class of models is frequently used to describe the impact of renewable energy on electricity markets in a mechanistic way [22–24]. We briefly review this approach, which will later serve as a benchmark model for more advanced machine learning models.

The electricity demand or load L is mostly inelastic in the short term [25]. Demand side management aims to increase the short-term flexibility and elasticity of demand, such that it can be adapted to the availability of renewable power generation. However, the progress of demand side management in the German market is very limited as private consumers and small enterprises typically experience no price signals [26]. Furthermore, day-ahead trading relies only on load forecasts, not the actual load. Load forecasts in the ENTSO-E region are calculated on the historic load profile on similar days [27] and generally assume "inelasticity of the load to price" [28]. Hence, we assume that L depends only on the time t, but not on the price p for the time being.

In perfect competition, the demand is satisfied by generating units according to their marginal costs. All units with marginal costs below the market price p can realize positive contribution margins and are thus "on", all others are "off". Hence one can obtain an approximate

view of the market outcome by sorting all generating units according to their estimated marginal costs. Renewable power plants, in particular wind and solar power, have high investment costs but almost vanishing variable costs. As marginal costs are dependent on the variable costs only, the marginal costs of these two renewable power sources are usually neglected. Furthermore, renewable plants are prioritized in the German market according to national regulations ("Erneuerbare Energien Gesetz"). Hence, we can assume that the total renewable generation G_{ren} is independent of the market price p. However, renewable generation depends on the weather and thus varies strongly with time t. If the availability of the dispatchable power plants varies little in time, we can assume that the generation G_{dis} depends only on the price p. In the German market, nuclear and lignite power plants have the smallest marginal costs and thus contribute first. Notably, the marginal costs of individual power plants are not known exactly and must be estimated.

We can now formulate the condition of market equilibrium. Supply and demand match if

$$L(t) = G_{\text{ren}}(t) + G_{\text{dis}}(p(t)). \tag{1}$$

Solving for the price yields

$$p(t) = G_{\text{dis}}^{-1}\left[L(t) - G_{\text{ren}}(t)\right], \tag{2}$$

where G_{dis}^{-1} denotes the inverse function. That is, the market price is a function of the difference of load and renewable generation, which is commonly referred to as the residual load.

Fig. 1b shows the price in the German day-ahead market as a function of the residual load. Data has been collected for 3 years and is displayed as a 2D histogram. We observe that the assumption (2) provides a reasonable approximation of the actual market behavior — the price generally increases with the residual load. We fit a third-order polynomial to the data, which had the best performance among

all polynomials up to the ninth order on the test set. The third-order polynomial fit will serve as a benchmark model in the following. We find that the data scatters quite strongly around this fit, as various effects are not taken into account in this approximate treatment.

2.3. Previous work on electricity price prediction and modeling

Understanding and forecasting electricity prices is an important research question in energy science. Therefore, day-ahead electricity prices have been studied from many different perspectives using a variety of methods and models [16,29]. Data based approaches range from autoregressive models, where predictions of future prices are computed from the historic time series alone [30,31], to regression models including many different external features. Which features are important for electricity price predictions is still a matter of current research. For instance, the importance of weather forecasts has been highlighted in [32,33]. Most recently, more complex machine learning models have been proposed for electricity price forecasting [16], using many different approaches such as convolutional neural networks (CNNs) [34] or recurrent neural networks (RNNs) [35,36].

Many machine learning approaches are based on black-box models, which limits scientific insights [17,37] and may induce security risks in critical sectors [38,39]. A promising alternative is provided by methods from eXplainable artificial intelligence (XAI), including inherently transparent models as well as post-hoc model explanations [40]. The field of XAI has gained a strong interest in energy systems analysis in recent years [41], in particular for applications to power system operation and stability. For instance, XAI was used in transient stability assessment [42], the identification of risks for frequency stability [43] or load and renewable generation forecast [44]. Furthermore, XAI has been used to analyze factors that determine the success of large power system infrastructure projects [45]. Applications of XAI methods for electricity markets are still in its infancy. A recent study by Tschora et al. [46] mainly focused on the identification of relevant features in forecasting models.

Overall, machine learning methods have been almost exclusively used for electricity price forecasting, not for explanations. However, understanding which factors determine electricity prices is an important question in energy science, including for instance the impact of fossil fuel prices [11].

In contrast to data-based methods, model-based approaches try to predict prices from fundamental economic considerations. There exist several mechanistic models that have been developed to simulate electricity markets and explain the emerging electricity prices. For instance, agent-based models have been developed to simulate market processes using a variety of external features from many areas and explain the emerging electricity prices [15,47]. Additionally, optimization models are routinely used to determine the dispatch and the market price [3,12]. Since simulation and optimization models are inherently transparent they provide an explanation of electricity prices from economic principles. However, they are limited to the mechanisms and interactions explicitly included by the modeler. XAI models can complement this approach by identifying key features, their dependencies and interactions, which in turn could inspire model extensions.

3. Methods

We develop an explainable machine learning model to understand German day-ahead electricity prices beyond the merit order effect introduced in Section 2.2. Our focus lies on predicting electricity prices given all other feature values at that point in time. That is, we transparently model and analyze the electricity market and the feature-price-interactions, which would not be fully possible in a forecasting setting.

For any further technical details about the methods the full project code is available on GitHub [48]. This includes data preparation, hyperparameter optimization and model explanation.

3.1. Data

As our prediction target, we use the hourly day-ahead electricity prices for Germany, which we collect from the ENTSO-E transparency platform [27]. Since European day-ahead market prices are coupled via the SDAC explained in Section 2.1, we get only one price for all exchanges.

It is important to note that Germany shares its bidding zone with Luxembourg and also shared it with Austria until the 1st of October 2018 [49]. Throughout the article, *German electricity prices* denote the price in the bidding zone of the given time period. Prices before and after the change of the bidding zones are joined together to create one continuous time series (see Fig. 1a).

As inputs for our prediction model, we use power system features and fuel prices. Power system features are collected from the ENTSO-E transparency platform [27] and fuel prices are collected from ARIVA.DE AG [50]. A full list of the features can be seen in Fig. 2.

Power system features include day-ahead forecasts of load, solar generation, wind generation, the day-ahead total generation and imports and exports. The features are aggregated for the four control areas of 50Hertz Transmission, Amprion, TenneT and TransnetBW. Wind generation is aggregated from wind on- and offshore generation. Total generation corresponds to the total scheduled generation in the day-ahead market. Import and export is aggregated from the cross-border flows between Germany and the neighboring bidding zones, where a positive (negative) value corresponds to more energy imports (exports). We also supplement the power system features with ramps for each feature, which are calculated using the formula $\text{ramp}(t) = f(t) - f(t-1)$ where $f(t)$ denotes the feature at a point t in time.

Fuel prices include oil prices and natural gas prices. Because both features have a daily time resolution we create a linear interpolation to get an hourly time resolution matching the time sampling of the model. Coal prices vary only very little during the considered time span. We note that we exclude CO_2 prices, although they affect electricity prices in the long-term. During the considered time period, CO_2 prices have increased almost monotonically allowing the ML model to memorize the train set leading to immediate overfitting. In particular, we tested models including the CO_2 price and found a high generalization error.

We use 3 years of data from the years 2017, 2018 and 2019 in order to get enough data for training and evaluation. Hourly data points with missing values for any feature are dropped to prevent fitting corrupted data.

3.2. Model

To model the German electricity price, we use Gradient Boosted Trees (GBTs) on our input data consisting of power system and fuel price features. GBTs offer complex non-linear models which we need in order to get more precise predictions of the electricity price than the benchmark model based on a common approximation of the merit order principle [51]. We use the LightGBM framework for our implementation in order to achieve a fast model training [52].

While single decision or regression trees are interpretable by reporting their decision path, ensemble methods, such as GBTs, trade a higher performance for a harder-to-interpret model. Still, using methods such as SHAP enables us to get a detailed explanation of the GBTs which we explain in detail in Section 3.3.

For the training process, we split our data into a training (48%), validation (32%) and test (20%) set. The reason for the unusual size of the validation set is that overfitting represents a serious issue for the given dataset. By doubling the validation set and splitting the set into four separate validation sets, we were able to reduce overfitting to an acceptable level. Additionally, reducing the train set did not cause any performance loss. The four validation sets are used for evaluation of the performance after each training epoch, where the training is stopped if the performance of one of the validation sets is not improving
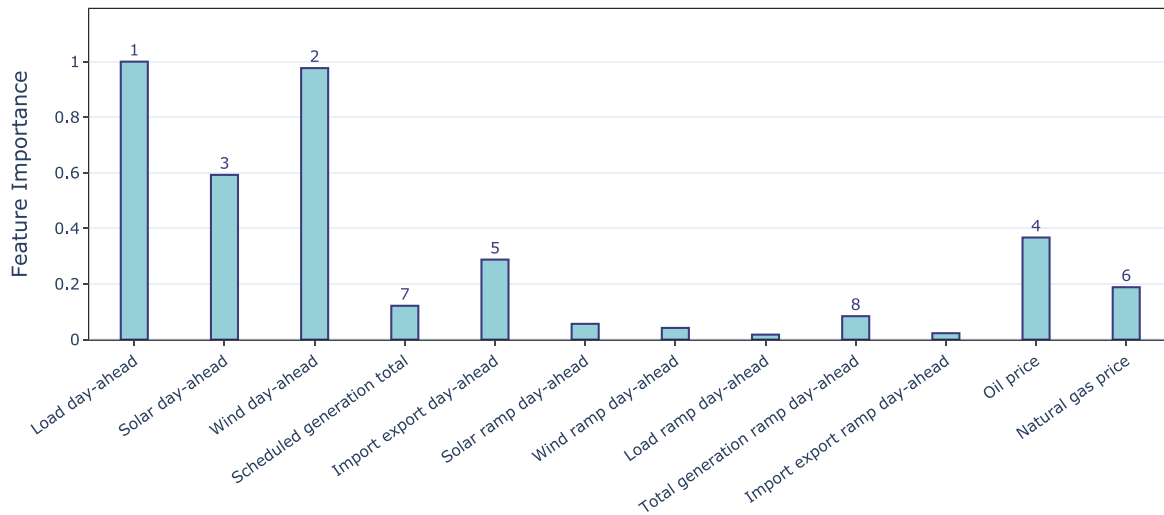
Fig. 2. Feature importance in the GBT model for the day-ahead electricity prices. Feature importances are computed from SHapley Additive exPlanations and normalized to one (see text for details). Features contributing to the residual load are most important as expected, but fuel prices also rank high.

for a predefined number of epochs. Since we focus on explaining our model instead of forecasting electricity prices as mentioned above, we shuffle our data before splitting. We use a weekly shuffle, where we only shuffle the dataset by weeks instead of hours before splitting. This gives us a more general model because we reduce the amount of similar data points in the training, validation and test set. We use the L^2 loss for the training process and the corresponding MAE score for evaluating the performance of the models.

We use a random search to find the best hyperparameters where we evaluate the performance of the fully trained models on the unseen test set. We choose the model with best performance on the test set.

For specific analysis tasks, we need to ensure the consistency of our models. We achieve this by analyzing the 10 best models of our random search for 10 different weekly random splits.

3.3. Model explanation

The model is interpreted using SHapley Additive exPlanation (SHAP) values after training [18,53]. SHAP values quantify the positive or negative impact of each feature relative to the overall base value. In our case, SHAP values show which features lead to higher or lower price predictions. SHAP values are unique among all feature attribution methods as they fulfill certain desirable properties for model explanation [53,54]. They provide a mathematically consistent attribution of a prediction to the individual features and thus avoid inconsistencies present in other methods.

For example, *local accuracy* guarantees that SHAP values sum up to the model prediction, meaning all feature contributions plus a base value [54]. Overall for the n feature values $x_1, \ldots, x_n$ the output of the model $f(x_1, \ldots, x_n)$ can be written as

$$f(x_1, \ldots, x_n) = \phi_0 + \sum_{j=1}^{n} \phi_j(x_1, \ldots, x_n),$$

where ϕ_0 corresponds to the base value of the model predictions and $\{\phi_j\}_{j=1}^{n}$ are the SHAP values of the corresponding features.

From the *local* SHAP values we quantify a *global* feature importance by averaging over all SHAP values for each feature and normalizing them by the highest value (see Fig. 2). We also use SHAP dependency plots (see Fig. 3a–c), which give detailed insights into the feature contributions and utilize SHAP interaction plots to get insights about feature interactions inside the model (see Fig. 4 and 5).

Before we proceed, we briefly discuss how SHAP values help to interpret complex energy systems. In general, analysis of the raw data for these complex systems [55] is problematic because features are

Table 1
Summary of the performance measures on the merit order based benchmark and GBT model. The GBT model outperforms the merit order model for every metric up to a relative increase of 31.1% for the MAE.

	MAE	SMAPE	R^2-score
Merit order	8.06	23.25	0.66
GBT	5.53	17.82	0.8

highly correlated and hence many data-analysis tools will yield spurious results. For example, wind and solar power have opposite seasonal dependencies, and solar and load have opposite daily dependencies. Correlations in the raw data may thus be largely due (anti-) correlations with another feature. This problem is well known in the literature and different methods were discussed to mitigate the problem, see e.g. [56]. Using machine learning methods, such as GBTs, coupled with explainability approaches, such as SHAP, allow us to disentangle these interactions. In this way, we obtain a mathematically consistent attribution of the prediction to the individual features [57,58]. Nevertheless, our analysis is bound to a specific model, which will only reach a finite accuracy. The remaining effects cannot be explained (yet).

4. Results

4.1. Model performance

The developed machine learning model is capable of predicting day-ahead electricity prices with an average performance of MAE = 5.54 (Fig. 1c). Roughly speaking, the model explains the price with an error of 5.54 EUR/MWh on average. The performance is substantially better than for the benchmark model based on a common approximation of the merit order principle reaching only MAE = 8.06. Our model outperforms the benchmark model by 2.52 EUR/MWh in absolute values, which is a relative increase in performance of 31.3%. We also evaluated both models with the SMAPE and R^2-score for further validation of the results (Table 1). In summary, all three metrics are within a reasonable range for predicting electricity prices. Our model outperforms the benchmark model for all three metrics. We conclude that the machine learning model captures several market effects which are neglected in the single-feature benchmark model described in Section 2.2. We will now discuss these effects in detail, interpreting the machine learning model with the SHAP framework.
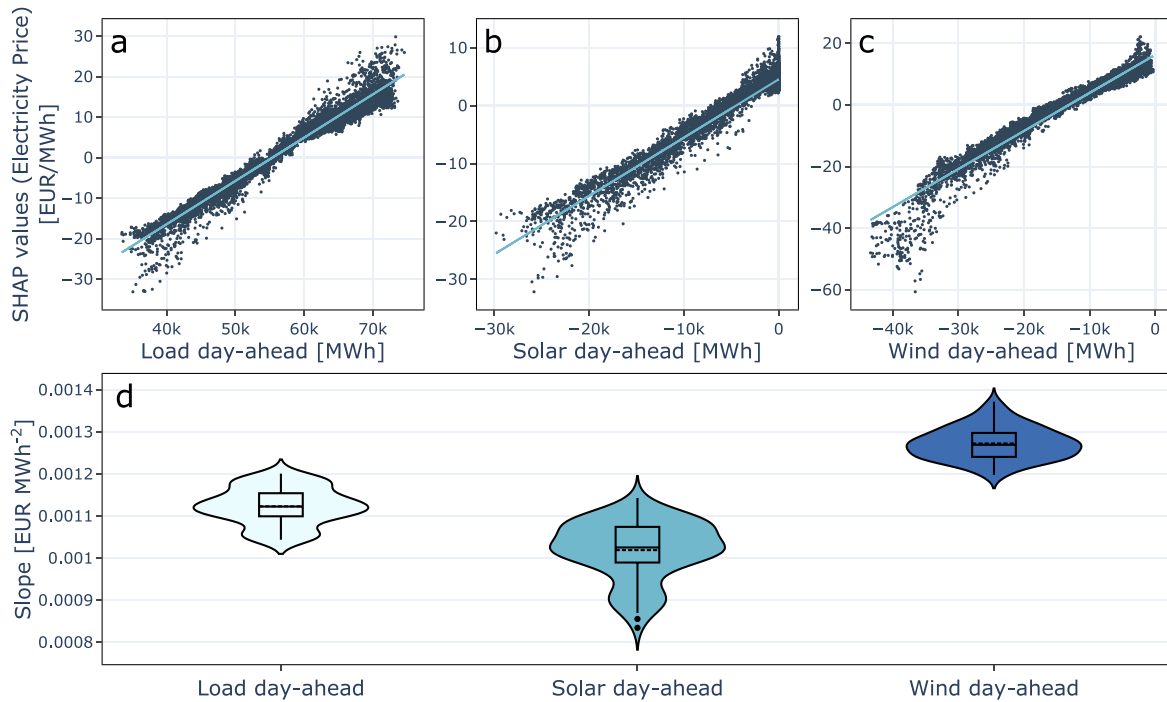
Fig. 3. Impact of the residual load features in the GBT model. (a)–(c) Dependency plots for the residual load features load, solar generation and wind generation (measured by mean absolute SHAP values). The light blue line is a linear fit. (d) Slope of the linear fits in the dependency plots. Violin plots shows the results for the ten best models for ten different data splits. The residual load features have slightly different linear relation to the electricity price, which is not captured by the benchmark model based on the merit order principle. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Features affecting the electricity prices

The developed machine learning model takes into account a variety of different features beyond the residual load. The SHAP values provide a consistent measure of the feature importance and thus reveal which factors have the strongest influences on the prices. Cumulative feature importance is shown in Fig. 2.

As expected, the main driver of the electricity prices is given by the residual load. More precisely, the three residual load features (load, wind and solar generation) are also the most important features in the machine learning model. The dependency of these features will be discussed in more detail in Section 4.3.

Fuel prices rank at position 4 and 6, with oil prices being more important than gas prices. This dependency is not surprising as fuel prices directly affect the variable costs of the respective power plants. However, the precise interpretation of this finding is less clear and will be further discussed in Section 4.6.

Prices are obviously related to cross-border trading. The import-export balance is the fifth most important feature and will be discussed in detail in Section 4.4. The total generation and its ramp rank at position 7 and 8. The generation ramp is particularly interesting as it reveals the influence of previous time steps, see Section 4.5 for details.

4.3. The role of wind, solar and load

The residual load features, i.e. load, solar generation and wind generation, are the most important features for the machine learning model (Fig. 2), in agreement with the benchmark model based on the merit order principle explained in Section 2.2. We take a more detailed look at the contribution of the residual load features by analyzing the corresponding partial dependency plots and interaction plots.

The benchmark model assumes that the price depends only on the residual load, hence the three features enter in an equal way up to a sign. Analyzing the respective partial dependency plots in Fig. 3a–c, we observe a similar dependency as expected, but also some subtle differences. For the detailed analysis of the differences and the observed scattering, we simplify the comparison of the three features by multiplying the renewable generations by −1.

The dependency on load, wind and solar generation is approximately linear (Fig. 3a–c), hence we use a linear fit for a quantitative analysis. We create linear fits for the partial dependency plots for the 10 best models after random search of 10 different random weekly shuffled splits. Fig. 3d shows the slope of the linear fits on the dependency plots for the 100 different models as violin plots. The different models seem to be consistent with their dependencies since the violin plots show a clear distribution around the mean value of the slope. Only the violin plot for solar generation shows some outliers.

The benchmark model based on the merit order principle assumes an equal contribution of all residual load features, but the machine learning model reveals some subtle differences. The slopes of the dependency on load and solar are rather similar, with solar being slightly smaller. In contrast, the slope of the dependency on wind is notably larger. The smaller influence of solar generation may be due to the fact that solar generation is more distributed in the German power grid, with a large fraction installed directly at the consumers. This could lead to solar generation acting as a negative load in the power grid, which would naturally cause load and solar generation to have a similar impact on the electricity price. In contrast, the generation of wind is more concentrated in the north of Germany, especially in the case of off-shore wind generation. This could lead to wind generation acting more like other power plants in the energy system, making the presence or absence of wind generation more important for the electricity price.

A further reason for the different role of wind and solar may lie in their respective market rules. While small scale PV installations typically rely on fixed feed-in tariffs, wind turbines are incentivized to sell their power according to the wholesale electricity market prices ('Direktvermarktung', see [59]).

The small scattering visible in the dependency plots can be explained by feature interactions. Neglecting all interactions in the dependency plot in Fig. 4a–c second column, scattering becomes smaller
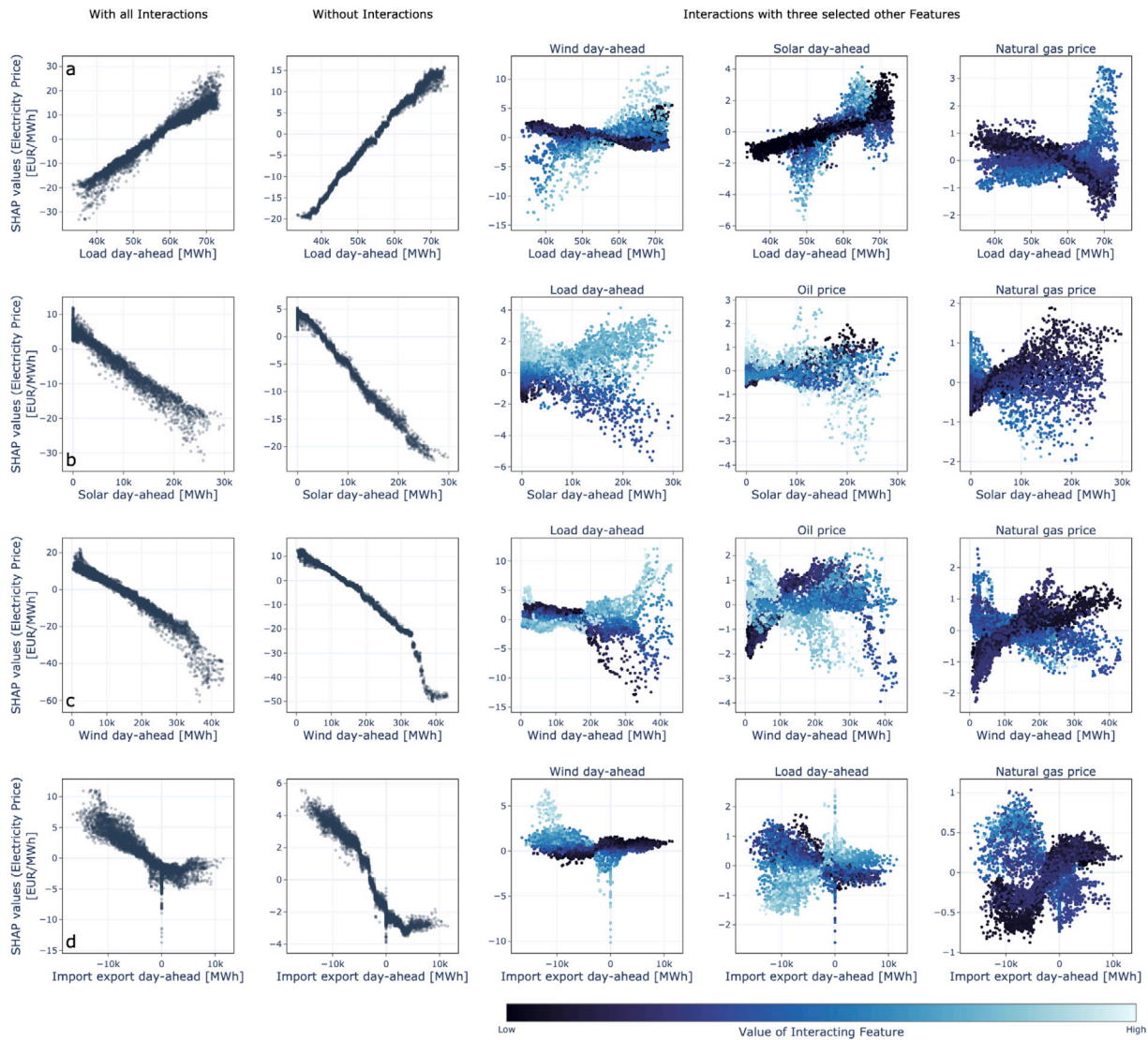
Fig. 4. Dependencies and interactions for (a) the load, (b) solar generation, (c) wind generation and (d) cumulative import or export, respectively. The first column shows the full SHAP dependency plot, the second column the SHAP dependency plot without any interactions. The third to fifth column show the SHAP interactions of three selected features, where the color indicates the value of the interacting feature. Fuel prices play an important role when interacting with other features in the GBT model. Further details are given in the text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and the dependency of the residual load features is even clearer. The dependency on load and solar generation is approximately linear, while the dependency on wind shows a clear non-linearity for large in-feeds above 30 GW. In this case, flexible power plants such as natural gas or oil plants have already left the market. Then, market equilibrium requires either an increase in exports, an increase of the load or a reduction of generation from mostly inflexible power plants such as lignite or nuclear. It appears plausible that these three mechanisms are comparatively inelastic such that the price decreases rapidly.

Looking at three of the most important interacting features, we can attribute most of the scattering present in the normal partial dependency plots.

Load has its strongest interactions with the renewable generation wind and solar. Both interactions are similar and enhance the dependency on load. This is reasonable since these features combined serve as a good approximation for the residual load, which is again already a good predictor for the electricity price. We can also see a strong interaction with the gas price, especially for high gas prices and high load. This is because more gas power plants are active for higher load, which then leads to higher electricity prices if gas prices are also high.

Wind and solar generation have a similar interaction structure, in particular there is a strong interaction with the load but also with fuel prices. For the load, interactions are particularly strong for high renewable generation. If the load is small, we recover the situation discussed above for the case of high wind generation, where market equilibrium requires the adaption of comparatively inelastic participants and thus entails strong price signals which may even include negative prices. This effect is largely compensated if the load is also high, leading to a strong increase of the price.

We also see strong interaction with the fuel prices, both amplifying the dependency for low wind or solar generation and reducing it for high wind or solar generation. These interactions originate from the fact that more fuel-dependent power plants are active if renewable generation is low and therefore the electricity price is more dependent on fuel prices in this case.

Summarizing, the residual load features have an overall strong interaction with fuel prices, mostly due to more dispatchable generation being active for specific values of the residual load features. Fuel prices not being included in the single-feature benchmark model could explain its lower performance compared to the GBT model.
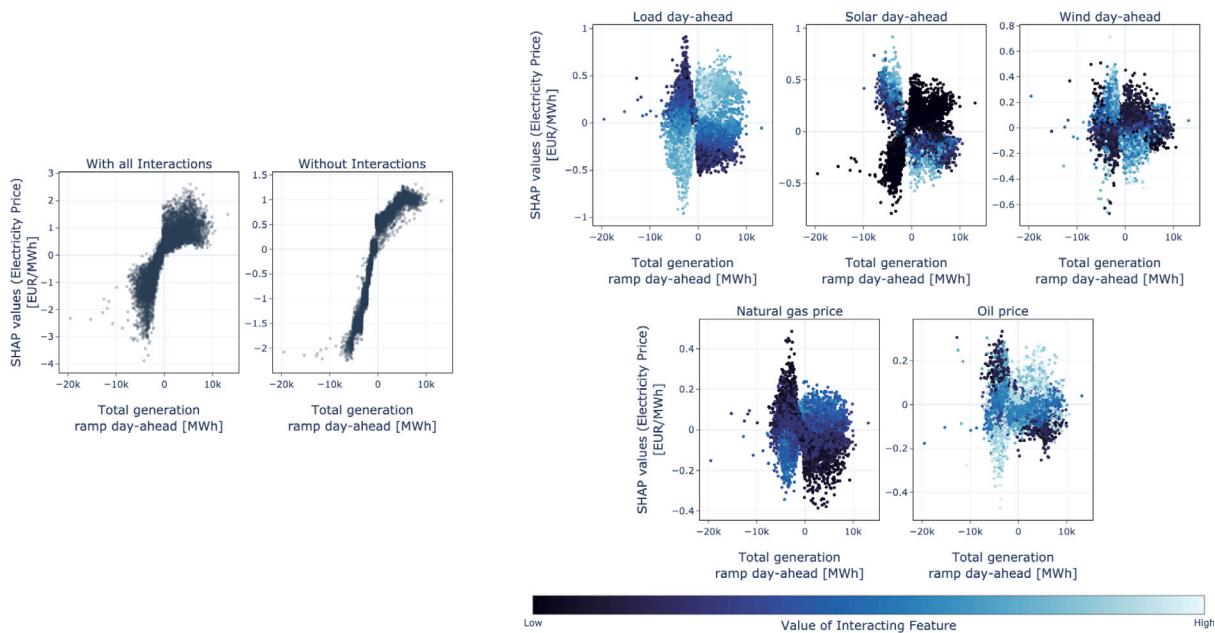
Fig. 5. Effect of generation ramps and limited generation flexibility on electricity prices. The first column shows the full SHAP dependency plot of the total generation ramp, the second column the SHAP dependency plot without any interactions. Further panels show the SHAP feature interactions for five selected interacting features. The color indicates the value of the interacting feature. Ramps affect/alter the electricity price prediction by up to 10%, depending on factors as load and renewable generation, but also on fuel prices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.4. Prices and cross-border trading

Electricity prices and trades are intimately related. In the machine learning model, the import-export balance ranks at fifth place in terms of the feature importance (Fig. 2). Different mechanisms can contribute to this dependency. On the one hand, high prices foster imports from other countries. On the other hand, imports provide a further source of electricity and should thus lead to lower prices. These two interactions can be interpreted as opposite causal relations, where either the price or the import-export balance is acting as the driver.

However, we stress that a causal interpretation is not that straightforward. The prices in different bidding zones and the cross-border trades are not determined sequentially, but simultaneously via the EUPHEMIA algorithm [21], see also Section 2.1.

The SHAP dependency plot reveals a negative correlation between the import balance and the day-ahead price (Fig. 4d). That is, imports are typically related to lower prices, while exports are related to higher prices. To understand this correlation, we consider one specific market situation. Assume that there is a high wind power generation in Germany. Typically, there are many low price offers in the German bidding zone, leading to a low market clearing price. However, if there is a strong demand in a neighboring country, additional offers will be accepted for exports, leading to an increase of the market clearing price. Hence, it appears as if the exports drive the market price, but in fact both are driven by a common cause: the total supply and demand in the two neighboring countries combined.

The strongest feature interactions are found for wind power generation and load. The interaction is opposite, which is comprehensible because the two features enter the residual load with opposite sign. We find that an increase of the residual load reduces the observed dependency, while a decrease of the residual load increases it.

Since SHAP values reveal only correlations of features and targets, it is difficult to reach a comprehensive causal interpretation. Still, based on our results, we formulate the following hypothesis. A high demand from a neighboring country generally leads to exports and to an increase of prices. But if the domestic demand (the residual load) is also high, there are no cheap offers left in the order book that

would allow for exports. Hence, the dependency of exports and prices diminishes.

Notably, there is large scatter to lower prices in the case of a vanishing import-export balance. This might be due to a temporary reduction in the transmission capacity preventing exports, or corrupted data (see [60] for a discussion of the data quality of the ENTSO-E transparency platform).

Finally, we also see an interesting interaction of the import-export balance with gas prices. Lower gas prices tend to reduce the overall contribution of import-export while higher prices amplify this contribution. This could indicate that at low gas prices, local gas generation reduces the dependency of a country to exchange power with neighboring countries and hence the price is influenced less by its imports and exports. In the opposite case of high gas prices, countries will be more willing to exchange energy and the effect of cross-border flows on prices increases.

4.5. Impact of ramps

The machine learning model reveals a weak dependency of the price and the power generation ramps (Fig. 2). Hence, the market outcome in a certain hour is affected by previous hours. The partial dependency plot (Fig. 5 left) indicates a positive correlation between the price and the total generation ramp. Hence, prices tend to be higher if generation is ramped up and lower if generation is ramped down.

This finding can be attributed to a limited flexibility of conventional power plants, in particular nuclear and lignite plants [61]. First, technical limits exist for the ramping speed and the minimum generation in partial load. Second, ramping and cycling induce additional costs, for instance, due to wear and tear of the power plant. Hence, there is an incentive to limit generation ramps which can affect the bidding on the market. In case of a decreasing total generation, operators may bid at a lower price to remain in the market and avoid ramping downwards. Similarly, in case of an increasing total generation, operators may bid at a higher price. As a consequence, the price increases with the total generation ramp.

The role of generation ramps depends on several other factors. The SHAP dependency plot is strongly scattered which can be attributed to
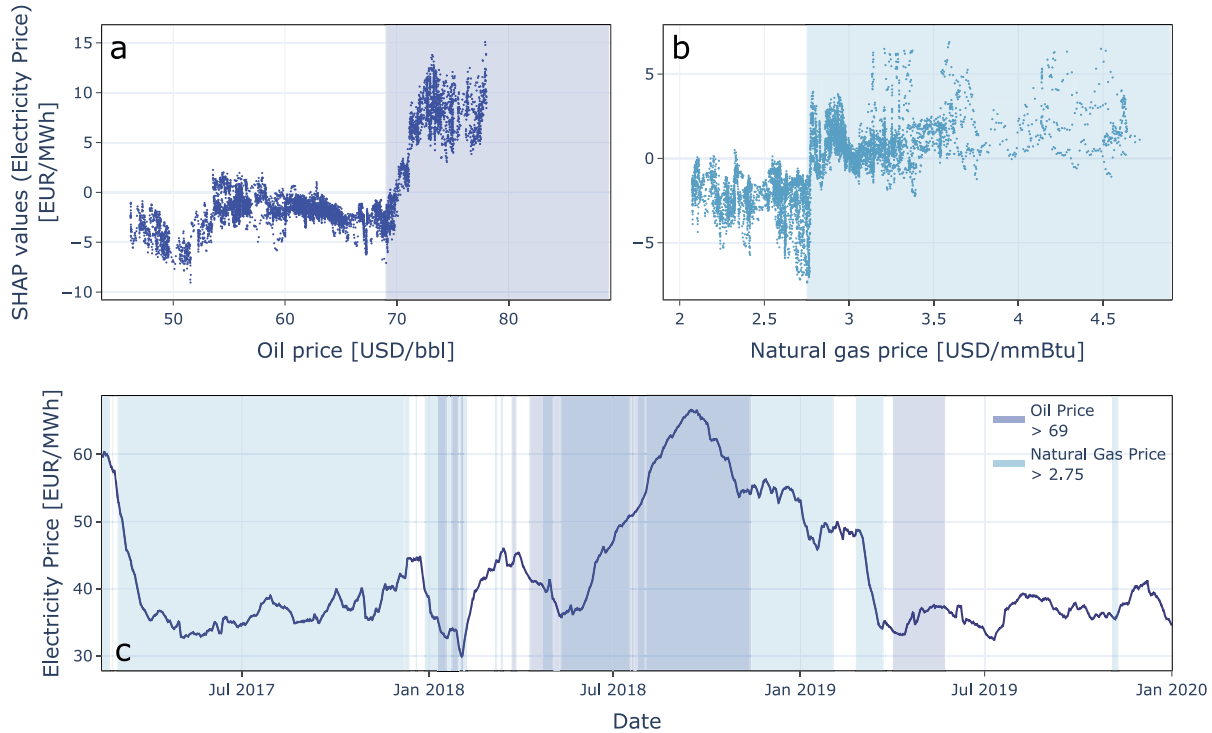
Fig. 6. Impact of oil and gas prices on the electricity price predictions. (a)–(b) SHAP dependency plot of the oil price (gas price), where the colored area marks a change for the dependency in the GBT model. (c) Electricity price time series from the German day-ahead EPEX spot market from January 2017 to 2020. Dark (light) blue areas mark the time periods with oil (gas) prices above the dependency threshold. We observe a clear change in the dependency on fuel prices in the model, but it cannot be asserted whether this is a causal relation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the presence of feature interaction. A high value of the load amplifies the impact of generation ramps. If the load is high, more conventional generation is needed in general, such that ramping limits and costs are more important. Vice versa, high values of wind and solar generation mitigate the impact of generation ramps as less conventional plants are needed. Furthermore, high oil and gas prices amplify the impact of generation ramps, too. Oil and gas power plants typically have a higher flexibility than nuclear or coal power plants. Higher fuel prices penalize these plants, such that nuclear or coal power plants may have to contribute more strongly to the ramping process.

4.6. Impact of fuel prices: Correlation or causality?

Looking at the feature importance in Fig. 2 we note that the machine learning model is critically dependent on fuel prices. The oil price is the 4th most important feature, i.e. the most important feature after the residual load features, while the gas price is the 6th most important. We analyze the dependency on oil and gas prices in detail.

The dependency plots for the fuel prices in Fig. 6a–b are strongly nonlinear, with an almost step-like behavior. The dependency of the electricity price on the oil price is approximately constant below a threshold of 69 [USD/bbl]. Above this threshold, the dependency increases until it saturates. While the change in dependency is almost linear for oil prices, the dependency for gas prices shows a step-like behavior, with a threshold at 2.75 [USD/mmBTu], albeit with a stronger scattering. A causal interpretation is comprehensible as an increase in fuel prices leads to an increase in the operational costs of the respective power plants and thus to offers at higher prices.

For further analysis, we focus on the time series of the electricity price in Fig. 6c, highlighting the time periods with oil and gas prices above the dependency threshold respectively. High oil prices seem to be correlated with high electricity prices, especially for the maximum of electricity prices at the end of 2018. In contrast, in the middle of 2019, electricity prices stay low while oil prices are above the threshold. For

gas, the dependency is even less clear. Lower prices seem to be a proxy for low electricity prices in 2019. Meanwhile, high gas prices do not display a clear correlation with overall electricity prices in the time series. We further note that a delayed relation could also be possible, if power plants purchase fuel well before the usage.

In general, it is difficult to pinpoint the relation of fuel prices to the electricity price beyond statistical correlations. Although we find a strong change in the dependency on the fuel prices in the machine learning model, it is still possible that the machine learning model is using the fuel prices to remember specific time periods where electricity prices are higher or lower than expected from the other features. Nevertheless, we find reasonably strong interactions of the residual load features with fuel prices, as discussed in Section 4.3. This points to a causal interaction, but a confounding effect is also possible. Overall, fuel prices are correlated with electricity prices which could be one of the reasons the machine learning model outperforms the benchmark model.

5. Discussion and conclusion

Summarizing, we have developed a machine learning model based on gradient boosted trees and demonstrated how it accurately estimates electricity prices, outperforming a single-feature benchmark model based on a common approximation of the merit order principle. Using SHAP to interpret our black-box model, we obtained deeper insights into the characteristics of the day-ahead market. SHAP values quantify how the price depends on the input features and thus reveals drivers beyond the benchmark model. Our analysis confirmed that high load leads to high prices, while large shares of wind or solar generation reduces prices. Furthermore, the model quantified the role of fuel prices, imports and exports, as well as load and generation ramps.

We saw that the SHAP analysis of our model is limited when it comes to a deeper causal interpretation of feature impacts. Nevertheless, the SHAP values provide detailed insights into the working

of the model by revealing how and which features are mostly used and by quantifying dependencies and interactions. Only by including domain knowledge of market mechanisms and power systems we can hypothesize to causal relations. Estimating causal models directly from data is in principle possible, e.g. using Causal SHAP [62] or causal representation learning [63] but requires more explicit assumptions about the underlying causal structure than we wanted to employ in this first exploratory study.

Concluding, we have demonstrated the usefulness of XAI models to analyze electricity price dynamics in the German market. These insights may contribute to the improvement of mechanistic models of electricity markets as well as data-driven forecasting models by identifying the relevant features to be included [64]. For instance, our results suggest slightly different roles of wind and solar power, while they enter the residual load equally. Furthermore, the SHAP analysis quantifies the role of generation ramps, which are subject to strong feature interactions.

There remain many open questions and starting points for further research. It would be interesting to investigate how XAI price models differ between electricity markets in different countries. Similarly, XAI models may also be used to compare markets at different times to quantify changes. For instance, the phase-out of nuclear power in Germany or changes of the regulatory framework should impact the dependencies of prices and features. Furthermore, XAI methods may also be used to analyze the impact of exceptional events such as the energy crisis after the Russian invasion of Ukraine once sufficient data is available.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

References

[1] van der Vleuten E, Lagendijk V. Transnational infrastructure vulnerability: The historical shaping of the. Energy Policy 2010;38:2042–52.

[2] Praktiknjo A. The value of lost load for sectoral load shedding measures: The German case with 51 sectors. Energies 2016;9:116.

[3] Wood AJ, Wollenberg BF, Sheblé GB. Power generation, operation, and control. 3rd ed.. New Jersey: John Wiley & Sons, Hoboken; 2013.

[4] Witthaut D, Hellmann F, Kurths J, Kettemann S, Meyer-Ortmanns H, Timme M. Collective nonlinear dynamics and self-organization in decentralized power grids. Rev Modern Phys 2022;94:015005.

[5] Milano F, Dörfler F, Hug G, Hill DJ, Verbič G. Foundations and challenges of low-inertia systems (invited paper). In: 2018 Power systems computation conference. p. 1–25.

[6] Staffell I, Pfenninger S. The increasing impact of weather on electricity supply and demand. Energy 2018;145:65–78.

[7] Han C, Hilger H, Mix E, Bøttcher PC, Reyers M, Beck C, et al. Complexity and persistence of price time series of the European electricity spot market. PRX Energy 2022;1:013002.

[8] Jamasb T, Pollitt M. Electricity market reform in the European Union: Review of progress toward liberalization & integration. Energy J 2005;26:11–42.

[9] European power exchange (EPEX SPOT). Annual Report 2019, 2019, https://www.epexspot.com/sites/default/files/sites/catalogue/catalogue. (Accessed 28 October 2022).

[10] Osička J, Černoch F. European energy politics after Ukraine: The road ahead. Energy Res Soc Sci 2022;91:102757.

[11] Zakeri B, Staffell I, Dodds P, Grubb M, Ekins P, Jääskeläinen J, et al. Energy transitions in Europe–Role of natural gas in electricity prices. 2022, Available at SSRN.

[12] Stoft S. Power system economics: designing markets for electricity. Vol. 468. Wiley–IEEE Press; 2002, online.

[13] Bublitz A, Keles D, Zimmermann F, Fraunholz C, Fichtner W. A survey on electricity market design: Insights from theory and real-world implementations of capacity remuneration mechanisms. Energy Econ 2019;80:1059–78.

[14] Hansen P, Liu X, Morrison GM. Agent-based modelling and socio-technical energy transitions: A systematic literature review. Energy Res Soc Sci 2019;49:41–52.

[15] Reeg M, Hauser W, Wassermann S, Kast T, Klann U, Nienhaus K, et al. AMIRIS: An agent-based simulation model for the analysis of different support schemes and their effects on actors involved in the integration of renewable energies into energy markets. In: 2012 23rd International workshop on database and expert systems applications. p. 339–44.

[16] Lago J, Marcjasz G, Schutter BD, Weron R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Appl Energy 2021;293:116983.

[17] Roscher R, Bohn B, Duarte MF, Garcke J. Explainable machine learning for scientific insights and discoveries. IEEE Access 2020;8:42200–16.

[18] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2:56–67.

[19] Kruse J, Schäfer B, Witthaut D. Secondary control activation analysed and predicted with explainable AI. Electr Power Syst Res 2022;212:108489.

[20] European union agency for the cooperation of energy regulators (ACER), List of currently designated NEMOs. 2021, https://www.acer.europa.eu/electricity/market-rules/capacity-allocation-and-congestion-management/implementation/designation-of-nemos.

[21] Kühling J, Knapp P, Kollmann D, Nöcker T, Wamback A. 8th Energy sector report of the monopolies commission. Technical Report, Monopolkommission, 2021, https://www.monopolkommission.de/images/PDF/SG/chapter-3-competition-between-electricity-exchanges.pdf.

[22] Sensfuß F, Ragwitz M, Genoese M. The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. Energy Policy 2008;36:3086–94.

[23] Nicolosi M, Fürsch M. The impact of an increasing share of res-e on the conventional power market-the example of Germany. Z Energwirtsch 2009;33:246–54.

[24] Huber M, von Roon S. Modeling spot market pricing with the residual load. In: Enerday: 5th conference on energy economics and technology, Dresden.

[25] Lijesen MG. The real-time price elasticity of electricity. Energy Econ 2007;29:249–58.

[26] Hanny L, Wagner J, Buhl HU, Heffron R, Körner M-F, Schöpf M, et al. On the progress in flexibility and grid charges in light of the energy transition: The case of Germany. Energy Policy 2022;165:112882.

[27] ENTSO-E. ENTSO-E transparency platform. 2022, https://transparency.entsoe.eu/.

[28] ENTSO-E. Enhanced load forecasting. 2023, https://www.entsoe.eu/Technopedia/techsheets/enhanced-load-forecasting.

[29] Praktiknjo A, Erdmann G. Renewable electricity and backup capacities: An (un-)resolvable problem? Energy J 2016;37:89–106.

[30] Contreras J, Espinola R, Nogales F, Conejo A. ARIMA models to predict next-day electricity prices. IEEE Trans Power Syst 2003;18:1014–20.

[31] Conejo A, Plazas M, Espinola R, Molina A. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans Power Syst 2005;20:1035–42.

[32] Sgarlato R, Ziel F. The role of weather predictions in electricity price forecasting beyond the day-ahead horizon. IEEE Trans Power Syst 2022. (Early Access).

[33] Goodarzi S, Perera HN, Bunn D. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. Energy Policy 2019;134:110827.

[34] Khan ZA, Fareed S, Anwar M, Naeem A, Gul H, Arif A, et al. Short term electricity price forecasting through convolutional neural network (CNN). In: Barolli L, Amato F, Moscato F, Enokido T, Takizawa M, editors. Web, artificial intelligence and network applications, advances in intelligent systems and computing. Cham: Springer International Publishing; 2020, p. 1181–8.

[35] Li W, Becker DM. 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;237:121543.

[36] Iwabuchi K, Kato K, Watari D, Taniguchi I, Catthoor F, Shirazi E, et al. Flexible electricity price forecasting by switching mother wavelets based on wavelet transform and long short-term memory. Energy AI 2022;10:100192.

[37] Adadi A, Berrada M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access 2018;6:52138–60.

[38] Ahmad T, Zhang D, Huang C, Zhang H, Dai N, Song Y, et al. Artificial intelligence in sustainable energy industry: Status quo, challenges and opportunities. J Clean Prod 2021;289:125834.

[39] Cremer JL, Konstantelos I, Strbac G. From optimization-based machine learning to interpretable security rules for operation. IEEE Trans Power Syst 2019;34:3826–36.

[40] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115.

[41] Machlev R, Heistrene L, Perl M, Levy KY, Belikov J, Mannor S, et al. Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. Energy AI 2022;9:100169.

[42] Chen M, Liu Q, Chen S, Liu Y, Zhang C-H, Liu R. XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. IEEE Access 2019;7:13149–58.

[43] Kruse J, Schäfer B, Witthaut D. Revealing drivers and risks for power grid frequency stability with explainable AI. Patterns 2021;2:100365.

[44] Mitrentsis G, Lens H. An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting. Appl Energy 2022;309:118473.

[45] Alova G, Trotter PA, Money A. A machine-learning approach to predicting Africa's electricity mix based on planned power plants and their chances of success. Nat Energy 2021;6:158–66.

[46] Tschora L, Pierre E, Plantevit M, Robardet C. Electricity price forecasting on the day-ahead market using machine learning. Appl Energy 2022;313:118752.

[47] Nitsch F, Deissenroth-Uhrig M, Schimeczek C, Bertsch V. Economic evaluation of battery storage systems bidding on day-ahead and automatic frequency restoration reserves markets. Appl Energy 2021;298:117267.

[48] Trebbien J, Witthaut D, Schäfer B. Github repository will be made available upon publication. 2022.

[49] EEX. Vorbereitung der Trennung der Stromgebotszone zwischen Deutschland und Österreich. 2018, https://www.eex.com/fileadmin/EEX/Downloads/Products/Power/Phelix-DE/20180904-customer-information---bidding-zone-split-data.pdf.

[50] ARIVA.DE AG, ARIVA.DE. 2022, https://www.ariva.de/.

[51] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery; 2016, p. 785–94.

[52] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017;30.

[53] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30:4768–77.

[54] Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. 2019, arXiv preprint arXiv:1802.03888.

[55] Hülsmann J, Steinke F. Explaining complex energy systems: A challenge. In: Proceedings of the 34th conference on neural information processing systems. p. 6–12, Virtual Conference.

[56] Hirth L. What caused the drop in european electricity prices? a factor decomposition analysis. Energy J 2018;39.

[57] Chen H, Janizek JD, Lundberg S, Lee S-I. True to the model or true to the data? 2020, arXiv preprint arXiv:2006.16234.

[58] Hamilton RI, Stiasny J, Ahmad T, Chevalier S, Nellikkath R, Murzakhanov I, et al. Interpretable machine learning for power systems: establishing confidence in shapley additive explanations. 2022, arXiv preprint arXiv:2209.05793.

[59] German Bundestag, Gesetz für den Ausbau erneuerbarer Energien (Erneuerbare Energien-Gesetz – EEG 2021). 2014, http://www.gesetze-im-internet.de/eeg_2014/EEG_2021.pdf.

[60] Hirth L, Mühlenpfordt J, Bulkeley M. The ENTSO-E Transparency Platform–A review of Europe's most ambitious electricity data platform. Appl Energy 2018;225:1054–67.

[61] Van den Bergh K, Delarue E. Cycling of conventional power plants: Technical limits and actual costs. Energy Convers Manage 2015;97:70–7.

[62] Heskes T, Sijben E, Bucur IG, Claassen T. Causal shapley values: Exploiting causal knowledge to explain individual predictions of complex models. Adv Neural Inf Process Syst 2020;33:4778–89.

[63] Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, et al. Toward causal representation learning. Proc IEEE 2021;109:612–34.

[64] Cramer E, Witthaut D, Mitsos A, Dahmen M. Multivariate probabilistic forecasting of intraday electricity prices using normalizing flows. 2022, arXiv preprint arXiv:2205.13826.