

Multivariate Probabilistic Forecasting of Intraday Electricity Prices using Normalizing Flows

Eike Cramer^{a,b}, Dirk Witthaut^{c,d}, Alexander Mitsos^{e,a,b}, Manuel Dahmen^{a,*}

^a Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Energy Systems Engineering (IEK-10), Jülich 52428, Germany

^b RWTH Aachen University, Process Systems Engineering (AVT.SVT), Aachen 52074, Germany

^c Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Systems Analysis and Technology Evaluation (IEK-STE), Jülich 52428, Germany

^d Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany

^e JARA-ENERGY, Jülich 52425, Germany

Electricity is traded on various markets with different time horizons and regulations. Short-term intraday trading becomes increasingly important due to the higher penetration of renewables. In Germany, the intraday electricity price typically fluctuates around the day-ahead price of the European Power Exchange (EPEX) spot markets in a distinct hourly pattern. This work proposes a probabilistic modeling approach that models the intraday price difference to the day-ahead contracts. The model captures the emerging hourly pattern by considering the four 15 min intervals in each day-ahead price interval as a four-dimensional joint probability distribution. The resulting nontrivial, multivariate price difference distribution is learned using a normalizing flow, i.e., a deep generative model that combines conditional multivariate density estimation and probabilistic regression. Furthermore, this work discusses the influence of different external impact factors based on literature insights and impact analysis using explainable artificial intelligence (XAI). The normalizing flow is compared to an informed selection of historical data and probabilistic forecasts using a Gaussian copula and a Gaussian regression model. Among the different models, the normalizing flow identifies the trends with the highest accuracy and has the narrowest prediction intervals. Both the XAI analysis and the empirical experiments highlight that the immediate history of the price difference realization and the increments of the day-ahead price have the most substantial impact on the price difference.

Keywords: Electricity price forecasting; probabilistic forecasting; deep learning; multivariate modeling

1 Introduction

The liberalization of the European electricity markets has introduced the auction-based day-ahead market and the continuously traded intraday market (Mayer and Trück, 2018). The hourly day-ahead auction market is cleared on the day before delivery. Afterward, the continuous 15 min interval intraday markets aim to reduce the balance requirements by enabling short-term changes to delivery and consumption commitments and react to the ramping of supply and demand (Ocker and Ehrhart, 2017; European Power Exchange, 2021; Bublitz et al., 2019). The increasing penetration of renewable electricity sources with low marginal cost (Sensfuß et al., 2008) causes intraday prices to become more volatile (Viehmann, 2017; Shinde and Amelin, 2019). Thus, electricity price forecasting (EPF) (Weron, 2014) for the intraday market becomes increasingly difficult (Lago et al., 2021; Jedrzejewski et al., 2022a) as the forecasts have to account for the inherent uncertainty of intermittent

renewable electricity supply and human interaction on the continuous market (Mayer and Trück, 2018; Bublitz et al., 2019). Still, accurate forecasts of intraday prices can lead to significant financial gains (Serafin et al., 2022).

A promising approach to address the uncertainty in EPF uses probabilistic forecasting models that predict a distribution of possible realizations instead of a single point forecast (Zhang et al., 2003; Nowotarski and Weron, 2018). The quantitative knowledge of the uncertainty allows the market participants to develop different strategies considering possible outcomes and to derive more effective bidding strategies (Nowotarski and Weron, 2018). Established approaches for probabilistic forecasting include prediction intervals (Zhang et al., 2003; Serinaldi, 2011; Jónsson et al., 2014; Uniejewski and Weron, 2021) and density forecasts (Huurman et al., 2012; Panagiotelis and Smith, 2008; Andrade et al., 2017; Narajewski and Ziel, 2020b). Other works use ensemble forecasts (Narajewski and Ziel, 2020b) or

*Manuel Dahmen, Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Energy Systems Engineering (IEK-10), Jülich 52428, Germany
E-mail: m.dahmen@fz-juelich.de

propose combinations of deterministic and probabilistic approaches (Marcjasz et al., 2020). There is also an increasing trend towards machine learning methods (Jedrzejewski et al., 2022b).

Most probabilistic EPF papers utilize univariate modeling approaches (Nowotarski and Weron, 2018). Here, the models predict single time steps in a step-by-step fashion and consider the correlation between time steps via the autoregressive components. Alternatively, multivariate models predict sequences in multi-step forecasting (Panagiotelis and Smith, 2008; Ziel and Weron, 2018). For instance, Panagiotelis and Smith (2008) use a multivariate student-t distribution model for intraday prices in Australia. In a study on deterministic forecasting models, Ziel and Weron (2018) report on the advantages of using a multivariate approach for day-ahead EPF. Most of the literature on probabilistic EPF only considers the day-ahead markets (Lago et al., 2021). A notable exception is presented by Andrade et al. (2017), who perform probabilistic forecasting for both day-ahead and intraday markets.

The literature on intraday EPF generally addresses the intraday prices as a stand-alone prediction problem without considering the relation to the day-ahead prices (Pape et al., 2016; Kremer et al., 2021). However, separating the intraday EPF from the day-ahead prices ignores the connection between the two markets, where the intraday market mainly serves to make adjustments to the fixed day-ahead contracts (Koch and Hirth, 2019). Consequently, the intraday price follows the day-ahead price trend. Furthermore, the adjustments react to the ramping of supply and demand, which leads to steadily increasing or decreasing prices in the four intraday trading intervals of each day-ahead trading interval (Kiesel and Paraschiv, 2017). Thus, there is a distinct hourly fluctuation of the intraday prices around the day-ahead prices that averages to the day-ahead price (Han et al., 2022).

This work proposes a modeling approach for intraday electricity prices that exploits the special relationship between the two electricity markets. Assuming that the intraday EPF is performed short-term, i.e., after the day-ahead price is set, the intraday price is modeled via the difference between the two market prices. Furthermore, the hourly increasing or decreasing patterns of the four 15 min intervals of each day-ahead price interval are captured using a multivariate forecasting scheme. Hence, the forecasting problem needs to predict the four-dimensional price difference vector

probability distribution which is defined by heavy tails, high correlations between the dimensions, and conditional dependence on external impact factors. Thus, the price difference vector distribution cannot be modeled via standard probability distribution models. Instead, this work uses normalizing flows (Papamakarios et al., 2021), a non-parametric, i.e., fully data-driven, distribution model, to learn the probability distribution of the price difference vector. Normalizing flows learn complex distributions without a priori assumptions about the data and can include external impact factors for probabilistic regression (Winkler et al., 2019; Rasul et al., 2021). For scenario generation of other energy time series, such as renewable electricity generation and electricity demand, normalizing flows have already shown promising results (Dumas et al., 2022; Cramer et al., 2022a,b; Arpogaus et al., 2021). The forecasting performance of the normalizing flow is compared to an informed selection of historical data, a Gaussian copula, and a multivariate Gaussian regression.

Both day-ahead and intraday prices are heavily influenced by external impact factors such as renewable electricity generation, import and export, and demand (Trebbien et al., 2022). However, the literature is scarce on evaluations on the impact factors for the price *differences* analyzed and modeled in this study. Thus, this work investigates the impact of different external factors on the realization of the price difference vector using explainable artificial intelligence (XAI). XAI is a branch of machine-learning research aiming to design human-understandable models and post-modeling explanations for black-box models (Gunning, 2017; Tjoa and Guan, 2020). An important subbranch of post-modeling explanations in XAI considers the predictability of the desired labels or values given the considered input features (Lundberg et al., 2020). In particular, this work uses SHAP-values (SHapley Additive exPlanations), which are a typical approach to quantify the impact of each input feature on the output of the machine-learning model. In previous works, SHAP-values have been successfully applied to analyze the predictability of power grid frequencies (Kruse et al., 2020), to reveal drivers and risks for power grid frequency stability (Kruse et al., 2021), predict the electricity mix in the African grid (Alova et al., 2021), and to gain insight into solar power forecasting (Kuzlu et al., 2020).

The remainder of this work is organized as follows: Section 2 introduces the four-dimensional

price difference modeling approach and highlights the advantages of multivariate modeling over the traditional univariate approach. Section 3 introduces the different multivariate probabilistic models. Section 4 discusses relevant external input factors and performs the feature selection using SHAP-values analysis. Section 5 applies the different distribution modeling approaches to learn the four-dimensional conditional distribution and discusses the impact of each external input factor via an empirical study. Finally, Section 6 concludes this work.

2 Electricity price analysis

This Section first describes the used data sets. Second, an approach for multivariate modeling of intraday electricity prices is proposed. Finally, the Section compares a univariate and a multivariate sampling approach for the price difference vector using assessment scores for multivariate probabilistic forecasting.

2.1 Data description

The majority of electricity is traded on the auction-based day-ahead market, while the intraday market is primarily used to make short-term adjustments to the previously submitted bids (Kiesel and Paraschiv, 2017; Koch and Hirth, 2019; Han et al., 2022). The intraday contracts are negotiated between the producers and the consumers on different time horizons and for different periods. Thus, different prices occur for the same intraday trading interval (Mayer and Trück, 2018; Ocker and Ehrhart, 2017). As an indicator of the intraday electricity prices, the market operators release indices of volume-weighted averages (European Power Exchange, 2021). This work uses the ID₃ index, i.e., the index reflecting the volume-weighted average price of the last three hours before delivery, as it is the most commonly used indicator (European Power Exchange, 2021; Bublitz et al., 2019). The data set considered in this work is the day-ahead and ID₃ price time series recorded over the years of 2018 and 2019 provided by the Fraunhofer-Institut für Solare Energiesysteme ISE (2022). The day-ahead forecasts and actual values for the renewable electricity production and the loads are taken from ENTSO-E Transparency Platform (2022). All renewable electricity time series are recorded in 15 min intervals. Note that the electricity price data from 2020 and 2021 shows atypical behavior due to the COVID-19 pandemic (Michał Narajewski, 2020;

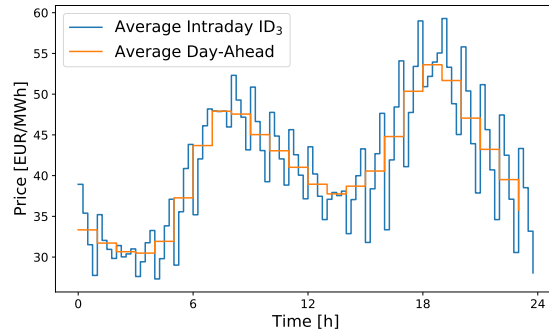


Fig. 1. Average daily profiles of day-ahead and ID₃ price trends of 2018 and 2019. EPEX spot price data from Fraunhofer-Institut für Solare Energiesysteme ISE (2022).

Badesa et al., 2021) and is, therefore, not included.

2.2 Hourly price difference vectors

The volume of electricity traded on the day-ahead and the intraday markets exemplifies the relationship between the two markets. In 2019, the volume of day-ahead trades was 501.6 TWh, and the intraday trades amounted to 83.2 TWh (European Power Exchange (EPEX SPOT), 2020). These numbers highlight how traders primarily trade on the day-ahead market and then adjust their bids on the intraday market. Intraday trades address ramping of supply and demand within the day-ahead trading intervals and compensate for forecast errors of demand and renewable electricity supply (Kiesel and Paraschiv, 2017; Spodniak et al., 2021). In particular, ramping of large suppliers and consumers leads to a distinct pattern of increasing or decreasing intraday prices within each day-ahead trading interval (Kiesel and Paraschiv, 2017). For instance, a scheduled ramp-up of electricity supply results in an undersupply in the first 15 minutes and an oversupply in the last 15 minutes of the respective hour, which leads to higher prices in the first 15 minutes and lower prices in the last 15 minutes, respectively. Thus, the hour-wise constant day-ahead contracts lead to hourly patterns in the 15 min intraday prices (Kiesel and Paraschiv, 2017; Han et al., 2022).

Figure 1 shows the average daily day-ahead and intraday prices in 2018 and 2019, i.e., each step represents the average price interval over the two years. As indicated by Kiesel and Paraschiv (2017), the average ID₃ price within each hour is close to the respective day-ahead price. Furthermore, the ID₃ price either increases or decreases within the

respective day-ahead trading intervals. These observations lead to the following conclusions about the relationship of the two price time series: First, the ID_3 index fluctuates around the day-ahead auction price. Second, Figure 1 highlights that the distinct hourly patterns of either increasing or decreasing steps appear systematically. Exceptions of this pattern only appear at local minima and maxima of the day-ahead price. In summary, the relationship between the day-ahead and the intraday markets as primary and adjustment markets leads to a distinct hourly pattern of intraday price fluctuations around the day-ahead prices.

The proposed modeling-approach for the ID_3 price index aims to capture the observed patterns. Since the intraday market is used for adjustments to the day-ahead market, the difference between the two price time series represents the characteristics of the intraday market. Assuming that the prediction takes place after the day-ahead auction has been settled and the day-ahead prices are known, the ID_3 price index can be described via the sum of the day-ahead price and the difference between the two price time series. Furthermore, the hourly pattern of fluctuations is captured by modeling the four steps of price differences in each day-ahead trading interval as a four-dimensional vector. For a given hour t the price difference vector ΔID_3 then reads:

$$\Delta ID_3(t) = \begin{bmatrix} ID_3^{00}(t) - DA(t) \\ ID_3^{15}(t) - DA(t) \\ ID_3^{30}(t) - DA(t) \\ ID_3^{45}(t) - DA(t) \end{bmatrix} \quad (1)$$

The four dimensions each describe the difference between intraday ($ID_3^{XX}(t)$) and day-ahead ($DA(t)$) price for a 15 min interval in the hour t . The superscripts 00, 15, 30, and 45 indicate the starting minutes of the four intraday trading intervals.

Note that while the day-ahead market sets the prices for all time intervals simultaneously, the intraday market allows for trading up to five minutes prior to the delivery, and the prices for each trading interval are, thus, set independently. Intuitively, a step-by-step forecasting approach, where each step is informed by the previous, could be considered more appropriate than the chosen multivariate approach. However, the multivariate modeling approach is specifically designed to consider the correlation between the forecast dimensions, i.e., the four quarter-hour intervals, and there are no limitations to the practical application of the multivariate modeling approach.

Tab. 1. Pearson correlation within the 4D- ΔID_3 distribution (see Equation (2)). Price data from Fraunhofer-Institut für Solare Energiesysteme ISE (2022).

	ΔID_3^{00}	ΔID_3^{15}	ΔID_3^{30}	ΔID_3^{45}
ΔID_3^{00}	1.00	0.77	0.41	0.09
ΔID_3^{15}	0.77	1.00	0.74	0.46
ΔID_3^{30}	0.41	0.74	1.00	0.83
ΔID_3^{45}	0.09	0.46	0.83	1.00

2.3 Analysis of the price difference vector distribution

To showcase the advantage of the multivariate modeling approach, this Subsection first highlights the strong correlation between the four price difference time steps in each day-ahead trading interval by showing the Pearson correlation and then compares a univariate and a multivariate sampling approach to highlight how the multivariate approach better captures the correlations between the time steps.

Table 1 shows the Pearson correlation of the four dimensions in the joint distribution of price differences, i.e., the autocorrelations of the respective time steps. The correlations between neighboring time steps are high and, thus, highlight the strong correlation between the time steps within each day-ahead trading interval. In fact, it appears that each interval is predetermined by the previous dimension, which motivates the multivariate modeling approach.

To confirm that the multivariate approach captures the correlation of neighboring time steps better than the classical univariate approach, two different approaches of univariate and multivariate selection of historical data are compared using scoring metrics for multivariate probabilistic forecasts. Figure 1 shows a strong dependency of the average realization on the hour of the day. The multivariate selection is based on this observation and randomly selects historical realizations of the hourly price difference intervals that occurred during the same hour of the day. Similarly, the univariate selection approach randomly selects values that occurred during the same 15 min interval and combines the selections into four-dimensional samples. Both approaches randomly select price differences from January 2018 to June 2019. July 2019 is set aside as a test set.

Unlike point forecasts, probabilistic forecasts cannot be evaluated by residual metrics such as the

mean squared error or the mean absolute error. Instead, probabilistic forecasts are typically evaluated in terms of reliability and sharpness (Nowotarski and Weron, 2018). Here, reliability describes whether the realization lies within certain prediction intervals. Sharpness analyzes the tightness of the prediction intervals, i.e., how close the lower and upper boundaries of the prediction intervals are together (Nowotarski and Weron, 2018). A good probabilistic forecast should be both reliable and sharp to avoid missing the realization and avoid large prediction intervals that carry little information.

To evaluate the different sampling approaches, this work uses the energy score (Gneiting and Raftery, 2007; Pinson and Girard, 2012) and the variogram score (Scheuerer and Hamill, 2015). The energy score (Gneiting et al., 2008; Pinson and Girard, 2012) is a widely applied metric for multistep probabilistic forecasts. It assesses how close the prediction samples are to the realizations and how diverse the samples are:

$$ES = \frac{1}{N} \sum_{s=1}^N \|\mathbf{x} - \hat{\mathbf{x}}_s\|_2 - \frac{1}{2N^2} \sum_{s=1}^N \sum_{s'=1}^N \|\hat{\mathbf{x}}_s - \hat{\mathbf{x}}_{s'}\|_2$$

Here, $\mathbf{x}$ is the realized ID₃ price-vector, i.e., the actual realization from the test set, $\hat{\mathbf{x}}_s$ are sample vectors drawn from the predicted distributions, N is the number of samples, and $\|\cdot\|_2$ is the 2-norm. The variogram score (Scheuerer and Hamill, 2015) aims to assess whether a multivariate forecast correctly describes the correlation between the individual time steps:

$$VS = \frac{1}{N} \sum_{t=1}^{\hat{T}} \sum_{t'=1}^{\hat{T}} \left(|x_t - x_{t'}|^\gamma - \frac{1}{N} \sum_{s=1}^N |\hat{x}_t^s - \hat{x}_{t'}^s|^\gamma \right)^2$$

Here, x_t are the realized ID₃ price time steps, $\hat{x}_t$ are time steps of samples from the predicted ID₃ distribution, $\hat{T} = 4$ is the dimension of x , i.e., the number of 15 min intervals in the multivariate prediction, N is the number of samples, γ is the variogram order (typically $\gamma = 0.5$ (Scheuerer and Hamill, 2015)), and $|\cdot|$ denotes the absolute value. Both energy score and variogram score are negatively oriented metrics, i.e., lower values indicate better performance (Scheuerer and Hamill, 2015; Pinson and Girard, 2012).

Figure 2 shows box plots (Waskom, 2021) of the energy score and the variogram score distributions of the univariate and the multivariate selection approaches applied to each day-ahead trading interval of the test month, respectively. The multivariate

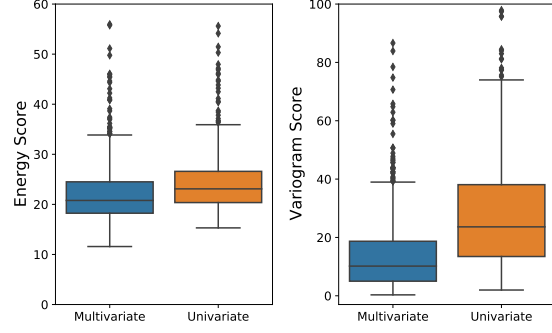


Fig. 2. Box plots (Waskom, 2021) of the energy score (Gneiting and Raftery, 2007; Pinson and Girard, 2012) and variogram score (Scheuerer and Hamill, 2015) distributions of historical data selected using univariate and multivariate approaches. Price data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022) is used and the evaluation is performed on July 2019 as the test set.

approach yields lower values for energy scores and variogram scores and, thus, indicates better agreement with the realizations. In particular, the variogram score suggests that the univariate selection fails to represent the correct correlation between the time steps whereas the multivariate approach captures the hourly pattern. In conclusion, the results in Figure 2 confirm that the multivariate approach is better suited to describe the price difference distribution than a univariate approach. In the following, the term historical selection refers to the multivariate selection of historical data.

3 Multivariate probabilistic forecasting of electricity price differences

The realizations of the hourly price difference vector in Equation (1) follow an unknown four-dimensional joint distribution:

$$\Delta \mathbf{ID}_3 \sim p_{\Delta \mathbf{ID}_3}(\Delta \mathbf{ID}_3) \quad (2)$$

Here, $p_{\Delta \mathbf{ID}_3}(\Delta \mathbf{ID}_3)$ is the probability density function (PDF) of the vector-valued random variable $\Delta \mathbf{ID}_3$. The realizations of the price difference distribution are assumed to be influenced by known external factors. Therefore, the proposed modeling approach includes these external factors as input

features $\mathbf{y}$ to form a conditional distribution:

$$\Delta \mathbf{ID}_3 \sim p_{\Delta \mathbf{ID}_3 | Y}(\Delta \mathbf{ID}_3 | \mathbf{y}) \quad (3)$$

Here, $p_{\Delta \mathbf{ID}_3 | Y}(\Delta \mathbf{ID}_3 | \mathbf{y})$ is the conditional PDF of $\Delta \mathbf{ID}_3$ given the input feature $\mathbf{y}$. Forecasting this conditional multivariate probability distribution poses a difficult modeling problem. The probability distribution is of unknown shape and electricity prices are known to exhibit heavy tails (Han et al., 2022). Furthermore, the relationship with the conditional input factors is unknown and likely nonlinear. Thus, the proposed probabilistic forecasting approach requires a flexible distribution model that can seamlessly include external input features with potentially nonlinear effects.

To model the conditional probability distribution of price differences described by Equation (3), this work uses the following four approaches: (1) conditional normalizing flows (Winkler et al., 2019), (2) the informed selection of historical data presented in Section 2, (3) Gaussian copulas with quantile regression (Pinson et al., 2009), and (4) multivariate Gaussian regression (Dillon et al., 2017). The selection of models aims to provide a variety of naïve, established, and recently published approaches.

Normalizing flows are non-parametric distribution models that model high-dimensional complex distributions as an invertible neural network transformation T , with inverse T^{-1} , of a multivariate standard Gaussian $\phi(\mathbf{z})$ (Dinh et al., 2017; Papamakarios et al., 2021). Utilizing the change of variables formula, normalizing flows describe the conditional PDF $p_{X|Y}(\mathbf{x}|\mathbf{y})$ of a multivariate random variable X with realizations $\mathbf{x}$ and input features $\mathbf{y}$ explicitly (Papamakarios et al., 2021; Winkler et al., 2019):

$$p_{X|Y}(\mathbf{x}|\mathbf{y}) = \phi(T^{-1}(\mathbf{x}, \mathbf{y})) |\det \mathbf{J}_{T^{-1}}(\mathbf{x}, \mathbf{y})| \quad (4)$$

Here, $\mathbf{J}_{T^{-1}}$ is the Jacobian of the inverse transformation. Normalizing flows make no assumptions about the data structure and distribution. Thus, by using sufficiently expressive transformations, normalizing flows can model any distribution (Papamakarios et al., 2021). To sample from a normalizing flow, first, samples are drawn from the multivariate Gaussian, i.e., $\hat{\mathbf{z}}_i \sim \phi(\mathbf{z})$. Second, the samples and the conditional inputs are transformed using the forward form of the invertible neural network:

$$\hat{\mathbf{x}}_i = T(\hat{\mathbf{z}}_i, \hat{\mathbf{y}}) \quad \forall i = 1, 2, \dots, N$$

Here, $\hat{\mathbf{y}}$ are the external input features at the respective time, which are assumed to be given. Figure 3 shows a sketch of the conditional normalizing

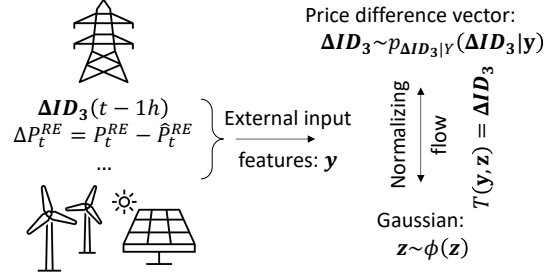


Fig. 3. Multivariate probabilistic forecasting scheme. The external input features are collected in a feature vector $\mathbf{y}$ and used to condition the transformation of Gaussian samples via the normalizing flow.

flow modeling scheme. This work implements the invertible neural network using the real non-volume preserving transformation (RealNVP) (Dinh et al., 2017).

The Gaussian copula is a well-established method to describe multivariate distributions with nontrivial correlations between the individual dimensions. The basic approach relies on transforming samples from a uniform distribution via an estimated inverse cumulative distribution function (CDF). In forecasting applications, the inverse CDF is estimated via an interpolation of quantile forecasts from quantile regression. For a detailed introduction to the combination with quantile regression, we refer the interested reader to Pinson et al. (2009). The present work uses the quantile regression implemented in the ‘statsmodels’ library (Seabold and Perktold, 2010).

Finally, the multivariate Gaussian regression uses neural networks to predict the mean vector μ_X and the covariance matrix Σ_X of the multivariate Gaussian as a function of the conditional inputs $\mathbf{y}$ (Dillon et al., 2017):

$$p_{X|Y}(\mathbf{x}|\mathbf{y}) = \mathcal{N}_{X|Y}(\mathbf{x}; \mu_X(\mathbf{y}), \Sigma_X(\mathbf{y})) \quad (5)$$

As the covariance matrix is symmetric, the neural network only needs to predict a lower-triangular version of $\Sigma_X(\mathbf{y})$ (Dillon et al., 2017). The predicted mean and covariance can then be used to sample from the Gaussian distribution parameterized by the predicted values.

In all models, the electricity price difference vector $\Delta \mathbf{ID}_3$ takes the role of the random variable X , and the external factors are the conditional input features Y . For every realization of the external factors $\mathbf{y}$, the models allow sampling from the electricity price difference vector distribution

$\Delta \hat{\mathbf{ID}}_3 \sim p_{\Delta \mathbf{ID}_3}$ as described for the normalizing flow above.

4 Input feature selection

The realizations of both the day-ahead and the intraday prices are influenced by external factors such as the residual loads (Kiesel and Paraschiv, 2017). However, the driving factors for the electricity price difference vector are likely to be different compared to those for the absolute day-ahead and intraday prices. In particular, the difference alleviates the impact of some external factors and amplifies the impact of others. Hence, the selection of input features should not be based on the standard inputs for day-ahead or intraday price forecasting known in the literature.

This Section aims to find a highly informative set of external factors that should be included as input features to the models in Section 3. First, Section 4.1 derives a set of possible input feature based on literature impressions and an analysis of the price time series. Second, Section 4.2 uses explainable artificial intelligence (XAI) to narrow down the full set to a highly informative subset of input features.

4.1 Possible input features

The day-ahead prices are determined by the intersection of supply and demand (Pape et al., 2016; Gürtler and Paulsen, 2018) and influenced by the merit-order curve (Kremer et al., 2021). The intraday markets are used to adjust the previously submitted bids to address short-term changes in supply and demand. Thus, the impact of the forecasted total electricity demand and the forecasted total renewable electricity supply is already reflected in the day-ahead prices and has little influence on the difference between day-ahead prices and intraday prices (Kiesel and Paraschiv, 2017; Spodniak et al., 2021; Han et al., 2022). The differences are mostly impacted by the forecasting errors in the residual load, i.e., for load and renewable electricity generation (Kiesel and Paraschiv, 2017; Han et al., 2022). For the absolute intraday prices, Kalukov and Ziel (Kulakov and Ziel, 2019) show that the impact of forecast errors is nonlinear. Notably, the actual renewable forecast errors are only known a posteriori. Hence, these values are neither known to intraday traders, nor can they be applied in actual forecasts of the intraday prices. However, hour-ahead renewable forecast are available (Sweeney et al., 2020)

and can be used to correct day-ahead forecasts, thus affecting the intraday electricity prices. Unfortunately, these hour-ahead renewable forecasts are not publicly available. Hence, we use the a posteriori renewable forecasting errors as a proxy in our analysis:

$$\Delta P_t^{RE} = P_t^{RE} - \hat{P}_t^{RE} \quad RE \in \{\text{Solar, Wind, Load}\}$$

Here, P_t^{RE} is the actual, $\hat{P}_t^{RE}$ is the day-ahead forecast.

Another important external factor for the fluctuation of the intraday prices is the ramping of renewables and load (Koch and Hirth, 2019). For instance, solar electricity generation ramps up in the morning and ramps down in the evening, which leads to large changes in solar power supply within single day-ahead trading intervals (Koch and Hirth, 2019). Wolff and Feuerriegel (2017) and Ocker and Ehrhart (2017) find that the ratio of conventional and renewable electricity impacts the realizations of both day-ahead and intraday prices. To investigate the impact of the share of conventional electricity, the prices of oil and natural gas are considered as input factors.

Han et al. (2022) find that the increments of the day-ahead price, i.e., the difference between one hour and the next, have a substantial impact on the direction (increasing or decreasing) and the intensity of the hourly fluctuation. This observation is confirmed by the average daily price profiles shown in Figure 1. For increasing or decreasing day-ahead prices, the hourly fluctuation shows increasing or decreasing patterns, respectively. Furthermore, larger increments appear to induce more significant fluctuations and vice versa. For local minimum and maximum peak hours in the day-ahead price, Figure 1 shows no distinct trend of the fluctuation.

Figure 1 also indicates that the realizations of the price differences depend on the time of the day. Thus, the list of possible input features includes a trigonometric encoding of the hour of the day (Moon et al., 2019).

Narajewski and Ziel (2020a) and Han et al. (2022) state that the day-ahead markets in Germany are *weak-form efficient*, i.e., the recent history of day-ahead prices does not inform on future realizations, whereas for intraday prices, there is an advantage from including the most recent value in the prediction. To evaluate the impact of prior realizations on the price difference, Figure 4 shows the Pearson autocorrelation function (ACF) of the day-ahead price, the intraday price, and the difference between

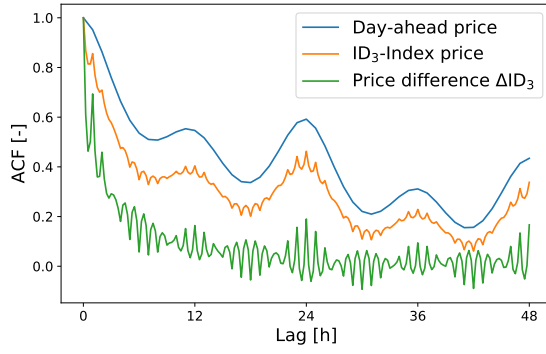


Fig. 4. Pearson autocorrelation (ACF) of day-ahead price (“Day-ahead price”), ID_3 index (“ ID_3 -Index price”), and difference between the two time series (“Price difference ΔID_3 ”) with time lag of 0 to 48 h. EPEX spot price data in 2018 and 2019 from [Fraunhofer-Institut für Solare Energiesysteme ISE \(2022\)](#).

the two time series for time lags up to 48 h. The ACF of the price difference drops quickly to values that are significantly lower compared to those for the day-ahead ACF and the intraday ACF that persist at high levels over 0.3 over longer time spans. After the rapid drop-off, the price difference ACF fluctuates at a low level under 0.2 indicating negligible impact for forecasting tasks. Yet, the first few hours still show high ACF values. Thus, the two previous hours are included as possible features. The supplementary material of this work shows further analysis of the time dependency on longer scales. Notably, the price difference time series does not show any significant trends over the days of the week or the months of the year over the considered dataset.

In the following, the possible input features are considered in three groups: The forecast error features, the ramping features, and the price and time features. The considered feature labels and their descriptions are listed in Table 2, 3, and 4, respectively.

4.2 Input feature selection via explainable AI

To improve usability and to avoid overfitting of the forecasting models, the set of input features should be narrowed down to a highly informative subset. While relevant external impact factors for day-ahead and intraday prices are well studied within the literature ([Wolff and Feuerriegel, 2017](#); [Treb-](#)

Tab. 2. Price and time feature labels. The variable t denotes the considered time interval in hours.

Description	Labels
Day Ahead Auction	DA
Day Ahead Increments back	ΔDA^-
Day Ahead Increments ahead	ΔDA^+
cosine time encoding	$\cos(t)$
sine time encoding	$\sin(t)$
$DA - ID_3$ at t-1 h (1 st interval)	$\Delta ID_3^{00}(t-1h)$
$DA - ID_3$ at t-1 h (2 nd interval)	$\Delta ID_3^{15}(t-1h)$
$DA - ID_3$ at t-1 h (3 rd interval)	$\Delta ID_3^{30}(t-1h)$
$DA - ID_3$ at t-1 h (4 th interval)	$\Delta ID_3^{45}(t-1h)$
$DA - ID_3$ at t-2 h (1 st interval)	$\Delta ID_3^{00}(t-2h)$
$DA - ID_3$ at t-2 h (2 nd interval)	$\Delta ID_3^{15}(t-2h)$
$DA - ID_3$ at t-2 h (3 rd interval)	$\Delta ID_3^{30}(t-2h)$
$DA - ID_3$ at t-2 h (4 th interval)	$\Delta ID_3^{45}(t-2h)$

Tab. 3. Forecast error feature labels.

Description	Labels
Solar Error [MW] (1 st interval)	Solar Error 00
Solar Error [MW] (2 nd interval)	Solar Error 15
Solar Error [MW] (3 rd interval)	Solar Error 30
Solar Error [MW] (4 th interval)	Solar Error 45
Load Error [MW] (1 st interval)	Load Error 00
Load Error [MW] (2 nd interval)	Load Error 15
Load Error [MW] (3 rd interval)	Load Error 30
Load Error [MW] (4 th interval)	Load Error 45
Wind Error [MW] (1 st interval)	Wind Error 00
Wind Error [MW] (2 nd interval)	Wind Error 15
Wind Error [MW] (3 rd interval)	Wind Error 30
Wind Error [MW] (4 th interval)	Wind Error 45

Tab. 4. Day-ahead ramping feature labels.

Description	Labels
Solar	DA solar ramp
Wind	DA wind ramp
Load	DA load ramp
Total power generation	DA total gen ramp
Import/export	DA import/export ramp

bien et al., 2022), there is no evaluation of impact factors for the difference between the two prices. The previous Section lists a number of possible input features. However, the true effect on the realization of the price differences remains unknown. This work uses XAI to find the most impactful input features. In particular, the explanatory power of SHAP-values is used to quantify the impact of each input feature and, thus, make informed decisions for the feature set selection.

Given a model $f(\mathbf{y})$ with feature vector $\mathbf{y}$, the SHAP-value approach (Lundberg et al., 2020) decomposes each model prediction into the contributions of the individual features:

$$f(\mathbf{y}) = \varphi_0 + \sum_i \varphi_i(y_i; f), \quad (6)$$

where $\varphi_0 = \mathbb{E}_Y(f(\mathbf{y}))$ is a constant offset and φ_i are the SHAP-values of the respective input features $y_i \in \mathbf{y}$. The SHAP-value decomposition ensures consistent interpretability, and the individual SHAP-values quantify the contribution of each input feature towards the model output value. The SHAP value method is designed specifically to quantify the impact of each input feature considering all possible (nonlinear) combinatorial effects represented in the model $f(\mathbf{y})$. Hence, SHAP values are not restricted to simple marginal feature perturbations under the ceteris paribus assumption. They capture the contribution of a given feature for all possible coalitions of other features in a consistent way, such that there is no possibility for missing combinatorial effects of different features changing the model outputs. Thus, the SHAP value method evaluates how each feature impacts the effects of the other features on the model output. In the supplementary material, we provide more information on the SHAP-value method. For a detailed introduction, the reader is referred to the original work by Lundberg et al. (2020).

SHAP-values provide local information for each element of the data set. A global measure of the importance of each feature can be obtained by taking the average of the absolute values:

$$FI_i = \sum_{t \in \chi_{\text{test}}} |\varphi_i(y(t); f)| \quad (7)$$

Here, FI_i is the feature importance of the i -th input feature, $|\cdot|$ denotes the absolute value operator, and χ_{test} is the test set. Notably, SHAP-values do not inform on the quality of the prediction itself. For a detailed introduction to SHAP-values and XAI, the reader is referred to Lundberg et al. (2020) and to

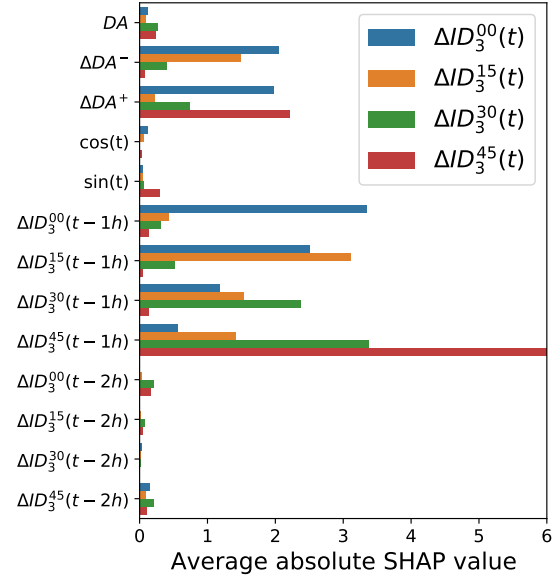


Fig. 5. SHAP-values (Lundberg et al., 2020) of the price difference vector for the price and time features. DA : day-ahead price, $\Delta DA^{+/-}$: day-ahead increments ahead and back, $\sin(t)/\cos(t)$: trigonometric hour encodings, and $\Delta ID_3^{X,X}$: price difference at given hour and 15 min interval.

Kruse et al. (2021) for an example application to power grid frequencies.

The SHAP value concept is designed for deterministic regression models. Thus, SHAP-values cannot be computed for the probabilistic models in Section 3. Instead, this work uses a tree regressor model, which is a popular choice to compute SHAP-values (Lundberg et al., 2020). In particular, gradient boosted trees have been shown to offer high levels of interpretability (Lundberg et al., 2020; Kruse et al., 2021). The gradient boosted trees are trained on 80% of the days from 2018 and 2019 and the remaining 20% of days are set aside as a test set. The test set is then used to score the accuracy of the trained models and to compute the SHAP-values. As SHAP-values can only be computed for scalar regressors, each of the four dimensions of the price difference vector is considered separately. For the SHAP-value analysis, this work uses the ‘SHAP’ library (Lundberg and Lee, 2017) and the gradient boosted tree implementation in the machine-learning library scikit-learn (Pedregosa et al., 2011). For the test set, the regressor returns R^2 of 74%, 60%, 61%, and 63% for each of the four price differences, respectively.

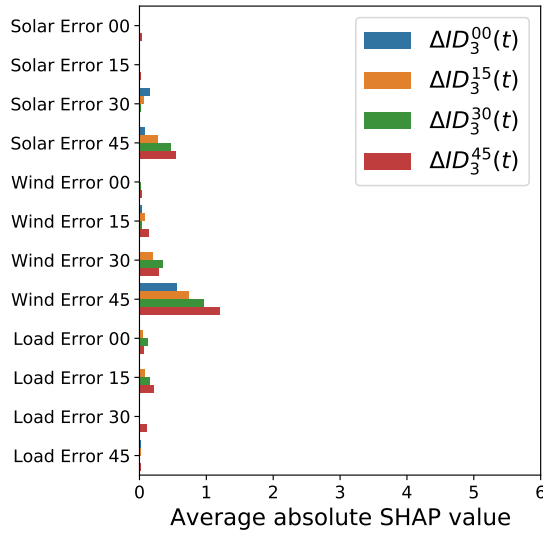


Fig. 6. SHAP-values (Lundberg et al., 2020) of the price difference vector for the forecast error features. Solar, Wind, and Load: forecast errors at given hour and 15 min interval.

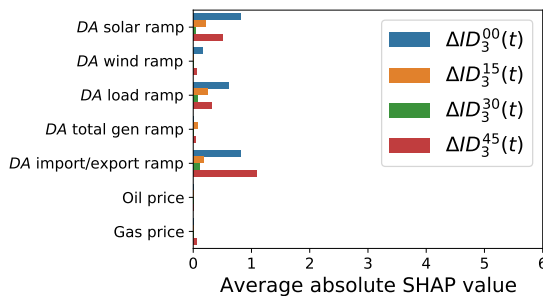


Fig. 7. SHAP-values (Lundberg et al., 2020) of the price difference vector for the day-ahead ramping features. Day-ahead ramps for solar, wind, load, total power generation, and import and export balances.

The SHAP-values are computed using a single regressor for all input features. Figure 5, Figure 6, and Figure 7 show the feature importance defined in Equation (7) for price and time features, the forecast error features, and the ramping features, respectively. The SHAP-values imply that there are few highly impactful features and many negligible features. In particular, the realizations of the price difference vector in the previous hour and the day-ahead increments show the most significant impact. The previous hour shows the highest SHAP-values for the same 15 min interval, respectively. Meanwhile, the absolute day-ahead price, the trigonometric time encoding, the fossil fuel prices, and the second to last price difference vector realizations also show negligible impact. Of the forecast errors, only the renewables forecast errors in the last 15 minute interval show considerable impact. Note that using hour-ahead forecasts may change the feature importance for the forecast errors. However, the small difference between the hour-ahead forecasts and the realizations combined with the low feature importance of the proxy indicate that considering the hour-ahead forecast is unlikely to cause significant changes. The ramping features mostly impact the first and the fourth interval in the price difference vector. Thus, the ramps mostly impact the orientation of the increasing or decreasing pattern in the data. Notably, the overall impact of the ramping features is minor.

In summary, the SHAP-value analysis indicates that the day-ahead increments and the realization of the previous hour are the most informative as input features to the models in Section 3. The analysis further suggests that the performance of models trained to predict the price difference is independent of most external impact factors. For instance, oil and gas prices have negligible feature importance for the price differences. Intraday trading decisions are made in situations of immediate requirement rather than a planning situation. Thus, the differences to the day-ahead prices are not affected by fuel prices. Still, electricity markets are continuously evolving, and the SHAP results in this study do not necessarily translate to other times.

For an analysis of the Pearson correlation of the impact features, see the supplementary material of this work.

5 Numerical experiments

This Section applies the different probabilistic modeling approaches to predict the price difference vec-

tor distribution. Each model is trained on the data set from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022), and the month of July 2019 is used as a test set. The $24\text{ h} \times 31\text{-day}$ test set results in 744 test intervals. Note that July 1st, 2019, was a Monday. To preprocess the data and avoid complications from the heavy-tailed distributions (Han et al., 2022), the data is transformed using the probability integral transform discussed in Uniejewski et al. (2017), which renders the data in a Gaussian form. The normalizing flow and the Gaussian regression are trained via log-likelihood maximization over 500 epochs with a batch size of 128 using the python-based machine-learning library TensorFlow version 2.8.0 (Abadi and Agarwal, 2015). The Gaussian copula uses the quantile regression in the ‘statsmodels’ python library (Seabold and Perktold, 2010). For details on the implementation of the models, see the supplementary material. One hundred samples are drawn for each hour in the test set and used to compute the different metrics outlined in Section 2. Sections 5.1 and 5.2 use the full set of input features derived in Section 4.1. Section 5.3 performs an empirical investigation of the impact of the different groups of input features.

5.1 Prediction intervals

This Section analyzes the forecasted distributions by investigating the 50% and 90% prediction intervals. The intervals are estimated by computing the respective quantiles of the drawn samples for each time step and then adding the day-ahead price of the given hour to reverse the difference in Equation (1).

Figure 8 shows the predicted mean, the 50%, and 90% prediction intervals, and the intraday price realization for the first four days in the test month of July 2019. The Figure shows the results estimated from probabilistic forecasts from the normalizing flow (top), the selection of historical data (second from top), the Gaussian copula (third from top), and the Gaussian regression (bottom). For the presented test days, the normalizing flow prediction intervals enclose the realizations for most of the time steps, and the forecasted mean identifies the realized trends well, with very few exceptions. The selection of historical data reflects the trends moderately well and portrays wider prediction intervals compared to the normalizing flow. The Gaussian copula and the Gaussian regression often fail to identify the correct trends and do not fit the realizations as tightly as the normalizing flow.

Tab. 5. Percentage of realizations within 50% and 90% prediction intervals (PI).

	PI 50%	PI 90%
Normalizing Flow	59.4%	93.5%
Historical	62.1%	95.5%
Copula	61.6%	96.0%
Gaussian	54.9%	90.2%

In particular, the Gaussian copula results in extensive prediction intervals that show vast peaks for the 90% prediction interval. Compared to the other approaches, the normalizing flow performs significantly better in identifying the trends and keeping narrow prediction intervals. Overall, the normalizing flow yields the sharpest prediction intervals, while the Gaussian copula shows low sharpness.

On the first day shown in Figure 8, there is a significant peak in the ID₃ price realization. This type of extreme event may occur due to unforeseen incidences in the markets, such as unplanned changes in the operation of large consumers or suppliers. Notably, the normalizing flow is the only approach that captures this peak. The supplementary material shows the prediction intervals for other peak days in the test month that confirm the normalizing flows’ ability to capture the price peaks.

Table 5 shows the percentage of time steps in the test set that lie within the 50% and 90% prediction interval, respectively. All approaches show conservative results for both prediction intervals, i.e., the actual percentage of realizations within the prediction intervals is higher than the selected probability. Thus, all approaches yield reliable prediction intervals. The supplementary material also lists the percentage of realizations within the prediction intervals for each of the four time steps of the price difference vector. The results show that the multivariate approach is able to predict all four dimensions well and that there is no deterioration of the prediction accuracy for the later time steps.

In summary, the normalizing flow shows the best results with sharp prediction interval presented in Figure 8 and reliable results in Table 5. The selection of historical data shows surprisingly good results but has poor sharpness in the prediction intervals due to the missing conditional input information. The Gaussian copula results in prediction intervals with poor sharpness and extreme peaks that are likely a result of a poor quantile regression of the tails of the price difference distribution. By overestimating the tails in the inverse CDF, the

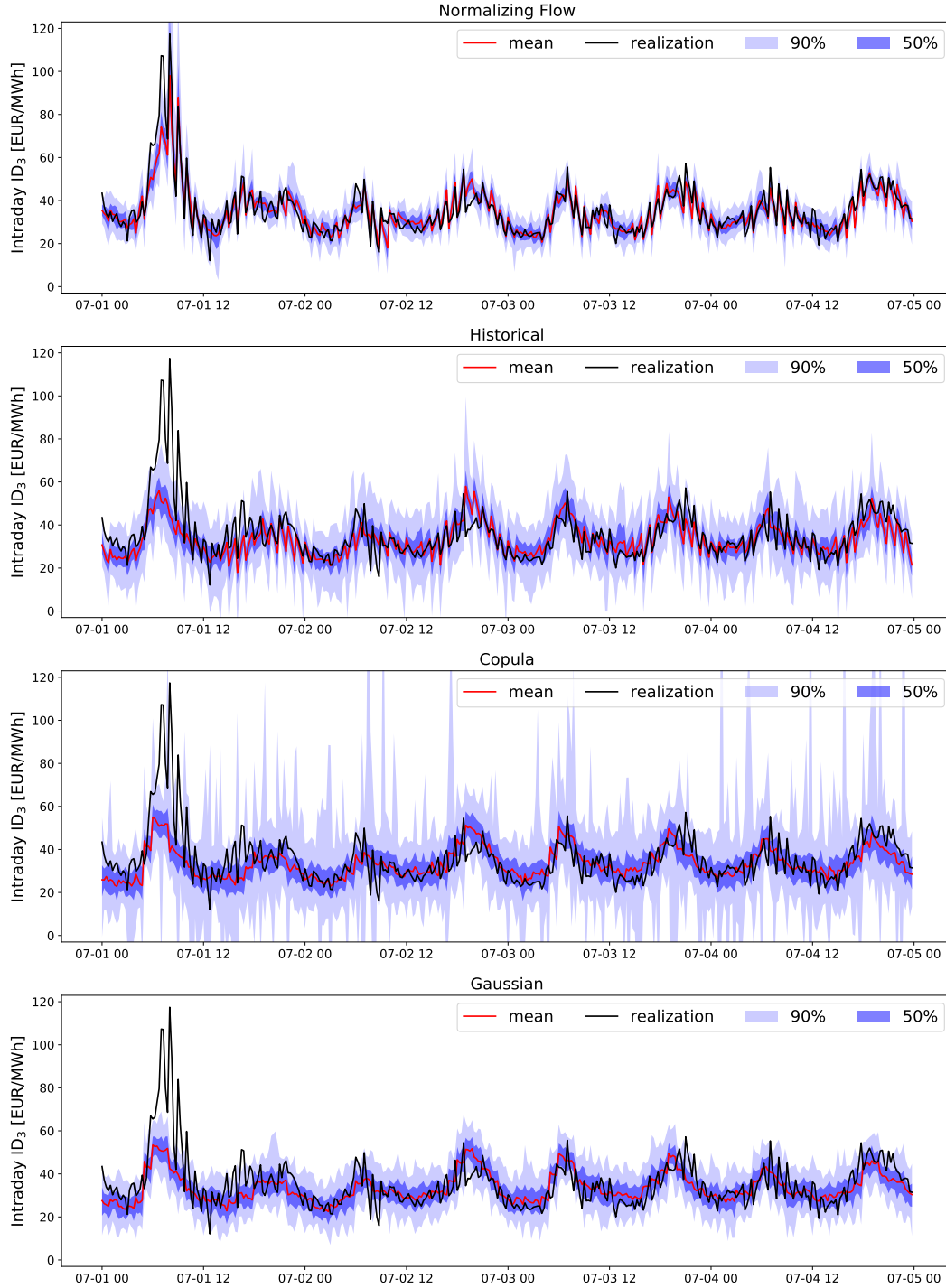


Fig. 8. Comparison of probabilistic forecasts by the normalizing flow, the informed selection of historical data, The Gaussian copula, and the Gaussian regression model to the actual realization (black line) for the first five days in July 2019. The probabilistic forecasts are visualized in terms of the median (red line) and the 90% and 50% quantiles (shaded areas), respectively. Note the truncated y-axis for the Gaussian copula. Training data from January 2018 to June 2019 ([Fraunhofer-Institut für Solare Energiesysteme ISE, 2022](#)). Results generated with all conditional inputs derived in Section 4.1.

Gaussian copula samples a higher share of outliers than the actual distribution. The high reliability indicated in Table 5 comes at the cost of poor sharpness. Finally, the Gaussian regression predicts 14 values for the mean vector and a lower-triangular covariance matrix (Dillon et al., 2017). Likely, this number is too high to achieve good fits in the regression model, and instead, the training regresses to constant outputs.

5.2 Energy and variogram score

Figure 9 shows the energy score for each model and each hour of the test month (left), as well as box plots of the overall energy score distributions (right). The Gaussian regression performs only marginally better than the historical selection. As shown in Figure 8, the Gaussian regression struggles to identify the correct trends, and the historical selection shows wider prediction intervals, i.e., both approaches show weaknesses that lead to an increase in the energy score in different ways. The Gaussian copula shows significantly worse results than the other forecasting models. As already observed in Figure 8, the Gaussian copula results in extensive prediction intervals and fails to identify the intraday price patterns. Thus, the large energy scores in Figure 9 confirm the qualitative observations from Figure 8. Finally, the normalizing flow shows the lowest energy score, i.e., it provides the most reliable predictions while maintaining sharp prediction intervals with a diverse set of samples.

In general, the energy scores over time in the case of all methods show rare incidences of extreme price peaks, e.g., on the first day and at the beginning of the last quarter of the month. Notably, however, the normalizing flow shows the best approximation of these extreme events, e.g., for the price peak on the first day of the test month shown in Figure 8 and the other peak days shown in the supplementary material.

Next, the variogram score is used to compare whether the different models can identify the correlation between the dimensions. Figure 10 shows the variogram score for each model and each hour of the test month (left), as well as box plots of the overall distributions (right). Note that the y-axis of the right of Figure 10 is truncated to allow for better visualization of the distributions. Similar to the energy score results, the normalizing flow shows the best results, while the Gaussian regression performs approximately equal to the historical selection. The Gaussian copula appears to miss large parts of the correlation and shows the worst variogram scores.

The flexibility of the normalizing flow allowed the model to learn the correlation between the four time steps in the price difference vector better than any other considered approach. In particular, the days with extreme price peaks result in consistently low variogram scores for the normalizing flow, while all other models result in spiking variogram scores. In the supplementary material, we show the evaluations of energy score and variogram score over a longer period from July to December 2019. The results confirm the observations described here.

In conclusion, the flexibility of the normalizing flow can describe the non-trivial distribution of the considered price differences, as highlighted by both the energy scores and the variogram scores for the test month. Furthermore, the combination with external impact factors leads to a significant improvement over the historical selection. Meanwhile, the Gaussian copula results in erroneous trends and inflated prediction intervals, and the Gaussian regression shows no advantage compared to the selection of historical data.

5.3 Significance of external factors

Up to this point, it remains unclear which external features contribute to the performance of the normalizing flow predictions. The SHAP-value analysis in Section 4.2 has provided first insights, albeit for a different deterministic prediction model. This Subsection now compares the performance of the normalizing flow using different sets of input features. Figure 11 shows box plots of the energy score and variogram score distributions over the test month for samples generated using the normalizing flow with the different groups of input features in comparison to the selection of historical data. Note the truncated y-axis to highlight the majority parts of the distributions.

The energy score results confirm the observations from the SHAP-value analysis in Section 4.2. The combination of the previous realization of the price difference vector and the day-ahead increments (“ $\Delta ID_3(t-1h)$ & ΔDA ”) results in the largest improvement. The forecast errors (“Forecast errors”) lead to a moderate improvement, while the ramping features (“Ramping”) only result in a minor improvement of the energy score compared to the historical selection (“Historical”). Notably, the combination of all external factors (“All”) achieves an energy score marginally below the combination of the previous realization of the price difference vector and the day-ahead increments.

The variogram score confirms the observations

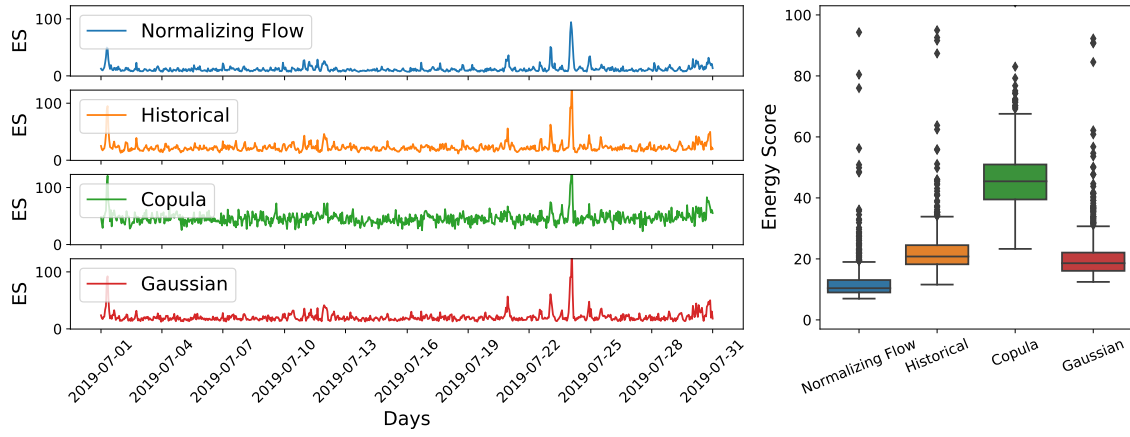


Fig. 9. Energy score (Gneiting et al., 2008; Pinson and Girard, 2012) of the normalizing flow (“Normalizing Flow”), the selection of historical data (“Historical”), the Gaussian copula (“Copula”), and the Gaussian regression (“Gaussian”) for each hour in test month of July 2019 (left) and box plot (Waskom, 2021) of the overall distribution for each model (right). Training data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022). Conditional inputs as derived in Section 4.1.

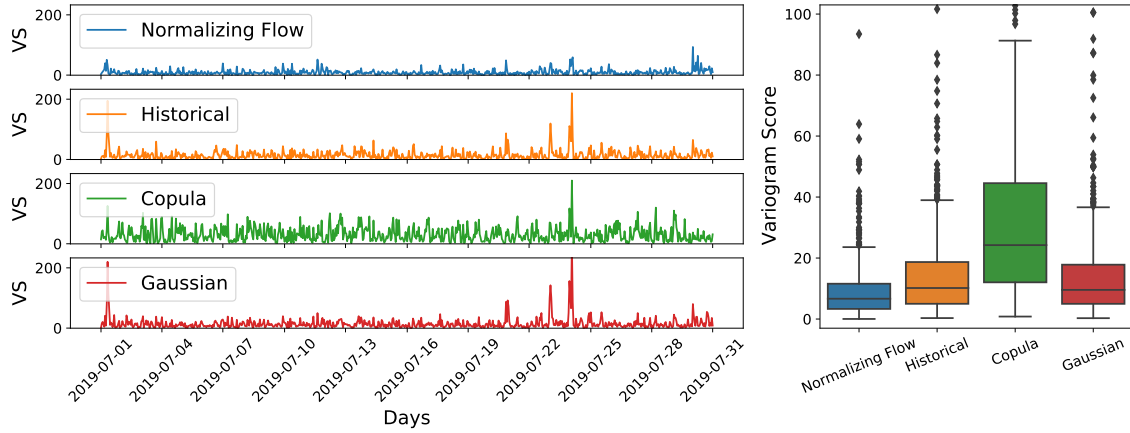


Fig. 10. Variogram score (Scheuerer and Hamill, 2015) of the normalizing flow (“Normalizing Flow”), the selection of historical data (“Historical”), the Gaussian copula (“Copula”), and the Gaussian regression (“Gaussian”) for each hour in test month of July 2019 (left) and box plot (Waskom, 2021) of the overall distribution for each model (right). Training data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022). Conditional inputs as derived in Section 4.1.

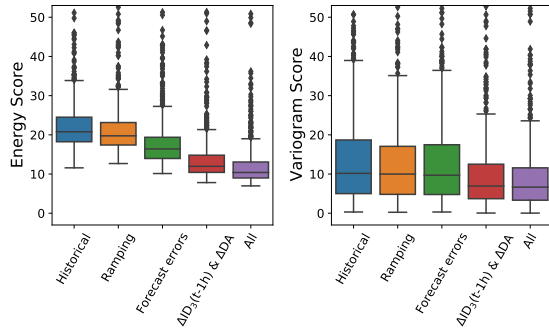


Fig. 11. Energy score (Gneiting et al., 2008; Pinson and Girard, 2012) and variogram score (Scheuerer and Hamill, 2015) of normalizing flow predictions using different sets of external factors. Note the truncated y-axis. Training data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022) and July 2019 as the test set.

from the energy score. Except for the combination of all external factors (“All”) and the combination of the previous realization of the price difference vector and the day-ahead increments (“ $\Delta ID_3(t-1h) \& \Delta DA$ ”), all inputs lead to approximately equal results compared to the informed historical selection.

In conclusion, the combination of the previous realization of the price difference vector and the day-ahead increments has by far the strongest impact on the performance of the probabilistic forecasts and appears to be sufficient to achieve high-quality prediction results. However, the normalizing flow can still extract additional information from the remaining external factors.

Narajewski and Ziel (2020a) and Han et al. (2022) argue that the intraday markets are not efficient in the sense of the efficient-market hypothesis as there are strong relations between the current and previous realizations. The results of this work confirm this observation.

5.4 Discussion

The analysis in this Section shows promising results that indicate valuable density forecasts of the price difference vector by the normalizing flow using appropriate external input features. Both the Gaussian copula and the Gaussian regression show weaknesses and fail to improve on the results of the informed historical selection suggesting they are not suitable prediction methods. Figure 11 shows that only the knowledge of the realizations in the pre-

vious hour and the day-ahead increments offers a significant advantage over using the informed selection of historical data. Hence, lead times of multiple hours rescind the advantage of the normalizing flow over the informed historical selection. In conclusion, the normalizing flow yields high-quality density forecasts for short lead times. If the available information is limited, the selection of historical price differences based on the hour of the day appears to be sufficient.

6 Conclusion

This work proposes a modeling approach for probabilistic forecasting of intraday electricity prices in the German EPEX spot market. The novel approach captures the strong relation between day-ahead and intraday prices by assuming fixed day-ahead prices and predicting the difference between the two price time series. Furthermore, the model captures the hourly patterns of fluctuation of the intraday prices around the day-ahead prices by considering each day-ahead trading interval as a four-dimensional joint distribution.

Normalizing flow, a non-parametric distribution model, is used to learn the proposed four-dimensional distribution of price differences and includes external input factors to form a conditional density forecast. The proposed model yields a complete end-to-end approach for multivariate density estimation and probabilistic regression that does not require any a priori assumptions about the data.

Compared to a selection of historical data and other multivariate probabilistic models, namely the Gaussian copula and multivariate Gaussian regression, the normalizing flow returns the most reliable and sharpest predictions. Using methods from XAI and careful performance analysis, the role of different input features has been analyzed in detail. In summary, the normalizing flow best identifies the trends of the intraday prices and shows the narrowest prediction intervals.

Declaration of Competing Interest

We have no conflict of interest.

Acknowledgements

We gratefully acknowledge Johannes Kruse (FZ-Jülich IEK-STE), Julius Trebbien (FZ-Jülich IEK-STE), and Leonard Rydin Gorjão (Department of Computer Science, OsloMet) for the discussion about the concepts of explainable AI and the importance of different impact factors. This work was performed as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE) and received funding from the Helmholtz Association of German Research Centres.

Nomenclature

Symbol	Description
$f(\mathbf{y})$	Deterministic forecasting model
FI	Feature importance
N	Number of samples
P_t^{RE}	Power values $RE \in \{\text{Solar, Wind, Load}\}$
$\hat{P}_t^{RE}$	Forecast of renewable electricity
ΔP_t^{RE}	Forecast error of renewable electricity
T	Invertible neural network
$\mathbf{J}_{T^{-1}}$	Jacobian of T^{-1}
$\hat{T}$	Day-ahead trading interval $\hat{T} = 4$
x_t	Value at time step t
$\hat{x}_t$	Sample at time step t
$\mathbf{x}, \mathbf{y}$	Data vector
$\hat{\mathbf{z}}$	Gaussian sample
$\hat{\mathbf{x}}, \hat{\mathbf{y}}$	Sample of data vector
$\mathbf{y}$	External input factors
ID_3	Intraday price
$\text{ID}_3^{00}, \text{ID}_3^{15}, \text{ID}_3^{30}, \text{ID}_3^{45}$	Quarter hour steps of intraday prices
DA	Day-ahead price
ΔID_3	4D vector of price difference
ϕ	Multivariate standard Gaussian
φ	SHAP value
$\mathcal{N}_{X Y}(\mathbf{x}; \boldsymbol{\mu}_X(\mathbf{y}), \boldsymbol{\Sigma}_X(\mathbf{y}))$	Conditional Gaussian distribution
$p_{X Y}(\mathbf{x} \mathbf{y})$	Conditional PDF of X given Y
$p_{\Delta \text{ID}_3 Y}(\Delta \text{ID}_3 \mathbf{y})$	Conditional PDF of ΔID_3 given Y
$\mathcal{N}_{X Y}$	Gaussian distribution with moments predicted via inputs Y
$\boldsymbol{\mu}_X(\mathbf{y})$	Neural network mean value predictor
$\boldsymbol{\Sigma}_X(\mathbf{y})$	Neural network covariance matrix value predictor
γ	Variogram order (typically $\gamma = 0.5$)

Bibliography

- Abadi, M. and Agarwal, A. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/>. Accessed on 08-08-2021.
- Alova, G., Trotter, P. A., and Money, A. (2021). A machine-learning approach to predicting Africa’s electricity mix based on planned power plants and their chances of success. *Nature Energy*, 6(2):158–166.
- Andrade, J. R., Filipe, J., Reis, M., and Bessa, R. J. (2017). Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model. *Sustainability*, 9(11):1990.
- Arpogaus, M., Voß, M., Sick, B., Nigge-Uricher, M., and Dürr, O. (2021). Probabilistic short-term low-voltage load forecasting using bernstein-polynomial normalizing flows. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, virtual.
- Badesa, L., Strbac, G., Magill, M., and Stojkovska, B. (2021). Ancillary services in Great Britain during the COVID-19 lockdown: A glimpse of the carbon-free future. *Applied Energy*, 285:116500.
- Bublitz, A., Keles, D., Zimmermann, F., Fraunholz, C., and Fichtner, W. (2019). A survey on electricity market design: Insights from theory and real-world implementations of capacity remuneration mechanisms. *Energy Economics*, 80:1059–1078.
- Cramer, E., Mitsos, A., Tempone, R., and Dahmen, M. (2022a). Principal component density estimation for scenario generation using normalizing flows. *Data-Centric Engineering*, 3:e7.
- Cramer, E., Paeleke, L., Mitsos, A., and Dahmen, M. (2022b). Normalizing flow-based day-ahead wind power scenario generation for profitable and reliable delivery commitments by wind farm operators. *Computers & Chemical Engineering*, 166:107923.
- Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., and Saurous, R. A. (2017). Tensorflow distributions. *arXiv preprint arXiv:1711.10604*.
- Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using Real NVP. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.
- Dumas, J., Wehenkel, A., Lanaspese, D., Cornélusse, B., and Sutura, A. (2022). A deep generative model for probabilistic energy forecasting in power systems: Normalizing flows. *Applied Energy*, 305:117871.
- ENTSO-E Transparency Platform (2022). ENTSO-E Transparency Platform. <https://transparency.entsoe.eu/>. Accessed: 2022-03-01.
- European Power Exchange (2021). EPEX SPOT documentation. <http://www.epexspot.com/en/extras/download-center/documentation>. Accessed: 2021-05-30.
- European Power Exchange (EPEX SPOT) (2020). Annual report 2019. https://www.epexspot.com/sites/default/files/download_center_files/Epex-spot-2019_200703_Planche.pdf. Accessed: 2021-05-30.
- Fraunhofer-Institut für Solare Energiesysteme ISE (2022). Energy Charts. <https://www.energy-charts.info>. Accessed: 2022-03-01.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.
- Gneiting, T., Stanberry, L. I., Gruit, E. P., Held, L., and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, 17(2):211–235.
- Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA), and Web*, 2(2):1.
- Gürtler, M. and Paulsen, T. (2018). The effect of wind and solar power forecasts on day-ahead and intraday electricity prices in Germany. *Energy Economics*, 75:150–162.
- Han, C., Hilger, H., Mix, E., Böttcher, P. C., Reyers, M., Beck, C., Witthaut, D., and Gorjão, L. R. (2022). Complexity and persistence of price time series of the european electricity spot market. *PRX Energy*, 1(1):013002.

- Huurman, C., Ravazzolo, F., and Zhou, C. (2012). The power of weather. *Computational Statistics & Data Analysis*, 56(11):3793–3807.
- Jedrzejewski, A., Lago, J., Marcjasz, G., and Weron, R. (2022a). Electricity price forecasting: The dawn of machine learning. *IEEE Power and Energy Magazine*, 20(3):24–31.
- Jedrzejewski, A., Lago, J., Marcjasz, G., and Weron, R. (2022b). Electricity price forecasting: The dawn of machine learning. *IEEE Power and Energy Magazine*, 20(3):24–31.
- Jónsson, T., Pinson, P., Madsen, H., and Nielsen, H. A. (2014). Predictive densities for day-ahead electricity prices using time-adaptive quantile regression. *Energies*, 7(9):5523–5547.
- Kiesel, R. and Paraschiv, F. (2017). Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, 64:77–90.
- Koch, C. and Hirth, L. (2019). Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing germany’s electricity system. *Renewable and Sustainable Energy Reviews*, 113:109275.
- Kremer, M., Kiesel, R., and Paraschiv, F. (2021). An econometric model for intraday electricity trading. *Philosophical Transactions of the Royal Society A*, 379(2202):20190624.
- Kruse, J., Schäfer, B., and Witthaut, D. (2020). Predictability of power grid frequency. *IEEE Access*, 8:149435–149446.
- Kruse, J., Schäfer, B., and Witthaut, D. (2021). Revealing drivers and risks for power grid frequency stability with explainable AI. *Patterns*, 2(11):100365.
- Kulakov, S. and Ziel, F. (2019). The impact of renewable energy forecasts on intraday electricity prices. *arXiv preprint arXiv:1903.09641*.
- Kuzlu, M., Cali, U., Sharma, V., and Güler, Ö. (2020). Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. *IEEE Access*, 8:187814–187823.
- Lago, J., Marcjasz, G., De Schutter, B., and Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1):56–67.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.
- Marcjasz, G., Uniejewski, B., and Weron, R. (2020). Probabilistic electricity price forecasting with narx networks: Combine point or probabilistic forecasts? *International Journal of Forecasting*, 36(2):466–479.
- Mayer, K. and Trück, S. (2018). Electricity markets around the world. *Journal of Commodity Markets*, 9:77–100.
- Michał Narajewski, F. Z. (2020). Changes in electricity demand pattern in europe due to COVID-19 shutdowns. *IAEE Energy Forum*, pages 44–47.
- Moon, J., Park, S., Rho, S., and Hwang, E. (2019). A comparative analysis of artificial neural network architectures for building energy consumption forecasting. *International Journal of Distributed Sensor Networks*, 15(9):1550147719877616.
- Narajewski, M. and Ziel, F. (2020a). Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, 19:100107.
- Narajewski, M. and Ziel, F. (2020b). Ensemble forecasting for intraday electricity prices: Simulating trajectories. *Applied Energy*, 279:115801.
- Nowotarski, J. and Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568.
- Ocker, F. and Ehrhart, K.-M. (2017). The “German Paradox” in the balancing power markets. *Renewable and Sustainable Energy Reviews*, 67:892–898.
- Panagiotelis, A. and Smith, M. (2008). Bayesian density forecasting of intraday electricity prices

- using multivariate skew t distributions. *International Journal of Forecasting*, 24(4):710–727.
- Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64.
- Pape, C., Hagemann, S., and Weber, C. (2016). Are fundamentals enough? explaining price variations in the german day-ahead and intraday power market. *Energy Economics*, 54:376–387.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pinson, P. and Girard, R. (2012). Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96:12–20.
- Pinson, P., Madsen, H., Nielsen, H. A., Papaefthymiou, G., and Klöckl, B. (2009). From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 12(1):51–62.
- Rasul, K., Sheikh, A.-S., Schuster, I., Bergmann, U. M., and Vollgraf, R. (2021). Multivariate probabilistic time series forecasting via conditioned normalizing flows. In *International Conference on Learning Representations*.
- Scheuerer, M. and Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334.
- Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, pages 92–96, Austin, TX.
- Sensfuß, F., Ragwitz, M., and Genoese, M. (2008). The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in germany. *Energy policy*, 36(8):3086–3094.
- Serafin, T., Marcjasz, G., and Weron, R. (2022). Trading on short-term path forecasts of intraday electricity prices. *Energy Economics*, 112:106125.
- Serinaldi, F. (2011). Distributional modeling and short-term forecasting of electricity prices by generalized additive models for location, scale and shape. *Energy Economics*, 33(6):1216–1226.
- Shinde, P. and Amelin, M. (2019). A literature review of intraday electricity markets and prices. In *2019 IEEE Milan PowerTech*, pages 1–6.
- Spodniak, P., Ollikka, K., and Honkapuro, S. (2021). The impact of wind power and electricity demand on the relevance of different short-term electricity markets: The nordic case. *Applied Energy*, 283:116063.
- Sweeney, C., Bessa, R. J., Browell, J., and Pinson, P. (2020). The future of forecasting for renewable energy. *Wiley Interdisciplinary Reviews: Energy and Environment*, 9(2):e365.
- Tjoa, E. and Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural networks and Learning Systems*, 32(11):4793–4813.
- Trebbien, J., Gorjão, L. R., Praktiknjo, A., Schäfer, B., and Witthaut, D. (2022). Understanding electricity prices beyond the merit order principle using explainable ai. *arXiv preprint arXiv:2212.04805*.
- Uniejewski, B. and Weron, R. (2021). Regularized quantile regression averaging for probabilistic electricity price forecasting. *Energy Economics*, 95:105121.
- Uniejewski, B., Weron, R., and Ziel, F. (2017). Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems*, 33(2):2219–2229.
- Viehmann, J. (2017). State of the German short-term power market. *Zeitschrift für Energiewirtschaft*, 41(2):87–103.
- Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081.

- Winkler, C., Worrall, D., Hoogeboom, E., and Welling, M. (2019). Learning likelihoods with conditional normalizing flows. *arXiv preprint arXiv:1912.00042*.
- Wolff, G. and Feuerriegel, S. (2017). Short-term dynamics of day-ahead and intraday electricity prices. *International Journal of Energy Sector Management*, 11:557–573.
- Zhang, L., Luh, P. B., and Kasiviswanathan, K. (2003). Energy clearing price prediction and confidence interval estimation with cascaded neural networks. *IEEE Transactions on Power Systems*, 18(1):99–105.
- Ziel, F. and Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420.

Multivariate Probabilistic Forecasting of Intraday Electricity Prices using Normalizing Flows

Eike Cramer^{a,b}, Dirk Witthaut^{c,d}, Alexander Mitsos^{e,a,b}, Manuel Dahmen^{a,*}

^a Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Energy Systems Engineering (IEK-10), Jülich 52425, Germany

^b RWTH Aachen University, Process Systems Engineering (AVT.SVT), Aachen 52074, Germany

^c Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Systems Analysis and Technology Evaluation (IEK-STE), Jülich 52428, Germany

^d Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany

^e JARA-ENERGY, Jülich 52425, Germany

1 Input feature selection

Table 1, Table 2, and Table 3 list the Pearson correlation of all considered external factors with the 4 dimensions of the price difference vector ΔID_3 . The Pearson correlation indicates the highest dependency on the immediate history of the price difference vector.

2 Time dependency

Figure 1 shows the dependency of the price difference vector on the hour of the day, the day of the week, and the month of the year, respectively. Notably, only the hour of the day appears to have a significant impact on the realizations. This also indicates a similar behavior of the price differences for weekdays and weekends.

3 Shapley Additive explanations

For black-box machine learning models, the quantitative relationship between input features and model outputs is often unknown. Here, explainable artificial intelligence (XAI) offers methods to estimate the quantitative relationship between features and outputs. A simple approach estimates the impact of different features using variations under the ceteris paribus assumption, i.e., by varying a single feature while keeping the remaining features constant. However, such one-at-a-time variations ignore possible combinatorial effects, which often lead to nonlinear behavior in the outputs. Shapley Additive explanations (SHAP) values are designed to describe the impacts of the input features on the model output including nonlinear combinatorial effects (Lundberg and Lee, 2017). Thus, SHAP

values give quantitative insight into the underlying systematic behavior of the system and can, e.g., be used to reveal the driving factors of a stochastic process like the power grid frequency (Kruse et al., 2021) or electricity prices (Trebbien et al., 2022).

SHAP values are computed by combining the marginal feature perturbations, i.e., the impact of including or excluding the feature from the feature set, for *all* possible coalitions and combinations of features. Notably, this evaluation of all possible combinations is computationally prohibitively expensive for large feature sets and arbitrary models. However, there exist efficient algorithms that compute SHAP values for tree-based models in polynomial time (Lundberg and Lee, 2017; Lundberg et al., 2020). For more information on the computation of SHAP values, the reader is referred to the original publications (Lundberg and Lee, 2017; Lundberg et al., 2020) and the documentation of the SHAP python package (Lundberg and Lee, 2017).

SHAP values are based on the idea to express a model output $f(\mathbf{y})$ using an additive feature attribution (Lundberg et al., 2020):

$$f(\mathbf{y}) \approx \varphi_0 + \sum_i \varphi_i(y_i; f) \quad (1)$$

Here, φ_0 is the null output, and $\varphi_i(y_i; f)$ are the SHAP values for each input feature i . The null output is computed as the expected model output over the full dataset:

$$\varphi_0 = \frac{1}{N} \sum_t f(\mathbf{y}_t)$$

Here, N is the number of points in the dataset. For tree-based models, like the ones used for the SHAP analysis in the main paper, the SHAP algorithm provides exact local accuracy, i.e., the approximate additive feature attribution in Equation (1) is met exactly. The formulation via the additive feature attribution also ensures the *missingness* in the

*Manuel Dahmen, Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Energy Systems Engineering (IEK-10), Jülich 52425, Germany
E-mail: m.dahmen@fz-juelich.de

Tab. 1. Pearson correlation between the price difference dimensions ΔID_3 and the price and time features. 00, ..., 45 indicate the for 15 min intervals in each day-ahead trading interval. Price data from [Fraunhofer-Institut für Solare Energiesysteme ISE \(2022\)](#) and renewables data from [ENTSO-E \(ENTSO-E Transparency Platform, 2022\)](#).

External factors	ΔID_3^{00}	ΔID_3^{15}	ΔID_3^{30}	ΔID_3^{45}
DA	-0.02	-0.04	-0.05	-0.04
ΔDA^-	-0.44	-0.16	0.14	0.31
ΔDA^+	0.38	0.10	-0.22	-0.43
$\cos(t)$	0.11	0.03	-0.04	-0.10
$\sin(t)$	0.01	0.00	0.02	-0.01
$\Delta ID_3^{00}(t - 1h)$	0.72	0.55	0.28	0.05
$\Delta ID_3^{15}(t - 1h)$	0.70	0.69	0.51	0.31
$\Delta ID_3^{30}(t - 1h)$	0.49	0.67	0.69	0.56
$\Delta ID_3^{45}(t - 1h)$	0.24	0.53	0.70	0.67
$\Delta ID_3^{00}(t - 2h)$	0.46	0.37	0.22	0.09
$\Delta ID_3^{15}(t - 2h)$	0.51	0.50	0.38	0.24
$\Delta ID_3^{30}(t - 2h)$	0.44	0.52	0.48	0.36
$\Delta ID_3^{45}(t - 2h)$	0.31	0.45	0.48	0.42

Tab. 2. Pearson correlation between the price difference dimensions ΔID_3 and the forecast errors. 00, ..., 45 indicate the for 15 min intervals in each day-ahead trading interval. Price data from [Fraunhofer-Institut für Solare Energiesysteme ISE \(2022\)](#) and renewables data from [ENTSO-E \(ENTSO-E Transparency Platform, 2022\)](#).

External factors	ΔID_3^{00}	ΔID_3^{15}	ΔID_3^{30}	ΔID_3^{45}
Solar Error 00	-0.20	-0.23	-0.21	-0.15
Solar Error 15	-0.20	-0.24	-0.23	-0.17
Solar Error 30	-0.20	-0.24	-0.24	-0.18
Solar Error 45	-0.20	-0.24	-0.24	-0.19
Wind Error 00	-0.28	-0.36	-0.38	-0.32
Wind Error 15	-0.28	-0.36	-0.39	-0.33
Wind Error 30	-0.28	-0.37	-0.39	-0.33
Wind Error 45	-0.28	-0.36	-0.39	-0.34
Load Error 00	0.12	0.11	0.09	0.04
Load Error 15	0.12	0.11	0.09	0.04
Load Error 30	0.12	0.11	0.09	0.04
Load Error 45	0.12	0.11	0.08	0.03

Tab. 3. Pearson correlation between the price difference dimensions ΔID_3 and the ramping features. 00, ..., 45 indicate the for 15 min intervals in each day-ahead trading interval. Price data from [Fraunhofer-Institut für Solare Energiesysteme ISE \(2022\)](#) and renewables data from [ENTSO-E \(ENTSO-E Transparency Platform, 2022\)](#).

External factors	ΔID_3^{00}	ΔID_3^{15}	ΔID_3^{30}	ΔID_3^{45}
DA solar ramp	0.23	0.08	-0.07	-0.21
DA wind ramp	0.05	-0.03	-0.10	-0.14
DA load ramp	-0.27	-0.09	0.12	0.25
DA total gen ramp	-0.02	-0.01	0.03	-0.01
DA import/export ramp	-0.27	-0.09	0.09	0.26
Oil price	-0.01	0.01	0.01	0.00
Gas price	-0.02	-0.01	-0.02	-0.03

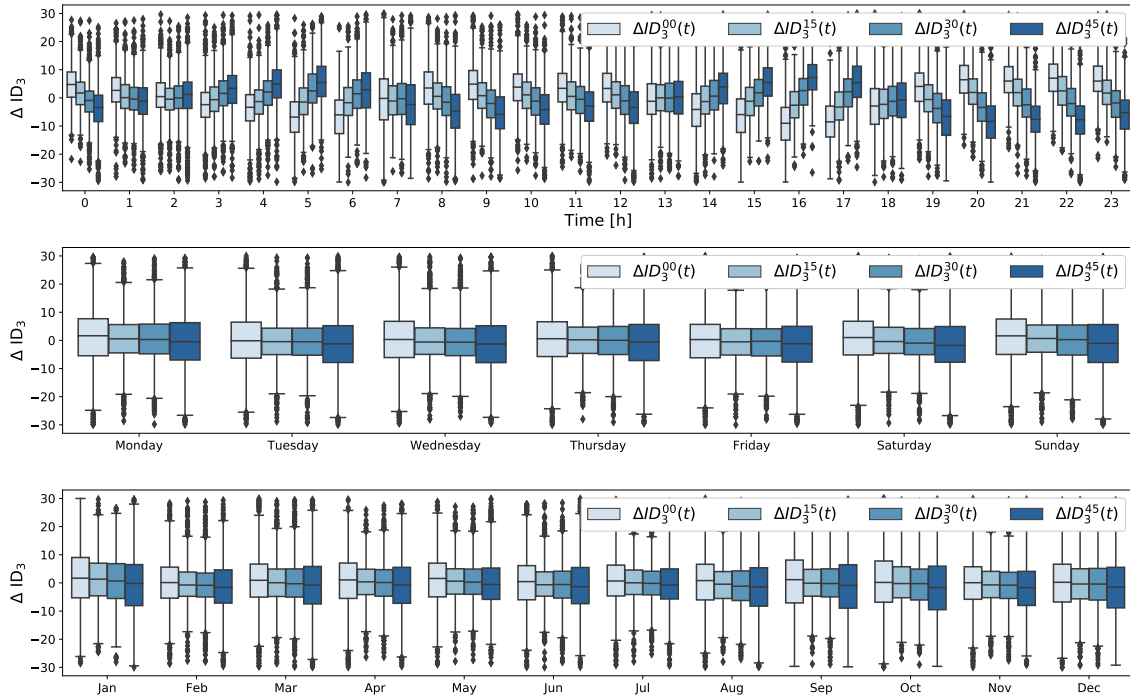


Fig. 1. Box plots of distribution of price difference as function of hour of the day, day of the week, and month of the year.

model such that, if a SHAP value is zero, the model output does not change. Furthermore, the SHAP value decomposition is consistent, i.e., a change to an input feature leads to a change in the corresponding SHAP value (Lundberg and Lee, 2017).

4 Additional analysis of Winkler score

The Winkler score (Winkler, 1972; Gneiting and Raftery, 2007) is a well-established test to investigate reliability and sharpness:

$$\text{WS} = \begin{cases} \delta_t & x_t \in [\hat{L}_t^\alpha, \hat{U}_t^\alpha] \\ \delta_t + \frac{2}{1-\alpha} (\hat{U}_t^\alpha - x_t) & x_t \geq \hat{U}_t^\alpha \\ \delta_t + \frac{2}{1-\alpha} (x_t - \hat{L}_t^\alpha) & x_t \leq \hat{L}_t^\alpha \end{cases}$$

Here, x_t is the time step, $\hat{L}_t^\alpha$ and $\hat{U}_t^\alpha$ are the lower and upper bound of the confidence interval with probability α , respectively. The term δ_t indicates the width of the predicted confidence interval:

$$\delta_t = \hat{U}_t^\alpha - \hat{L}_t^\alpha$$

The Winkler score is a negatively oriented score, i.e., lower values indicate better performance.

Figure 2 shows the distributions of Winkler scores over the full month of July 2019 for each of the probabilistic models and over α values between 0.0 and 0.9. Overall, the normalizing flow shows the best results.

5 ANN structures

The Gaussian regression uses a fully-connected neural network with one hidden layer, 32 neurons, and ‘tanh’ activation. The output layer has 14 neurons to describe the four-dimensional mean and the lower triangular covariance matrix implemented in TensorFlow-probability (Dillon et al., 2017). The neuron that output the mean have no activation and the neurons that output the covariance matrix use ‘softmax’ activation to ensure strict positivity.

The normalizing flow uses the affine coupling layer implementation in (Dinh et al., 2017). The full model uses four affine coupling layers with fully connected conditioner models. The conditioner models have two hidden layers with ‘ReLU’ activation and four neurons per layer.

Tab. 4. Percentage of realizations within 50% and 90% prediction intervals (PI) for each quarter hour. Results for normalizing flow over test interval from July 1st, 2019 to December 31st, 2019.

	ΔID_3^{00}	ΔID_3^{15}	ΔID_3^{30}	ΔID_3^{45}
PI 50%	0.54	0.53	0.54	0.62
PI 90%	0.91	0.90	0.91	0.93

6 Prediction intervals on extreme peaks

Figure 3 shows the predicted mean, the prediction intervals of 50% and 90%, and the realization between July 21st and July 26th of 2019. Similar to the prediction intervals shown in the main part of the paper, the results confirm that the normalizing flow results in the narrowest prediction intervals and is the only approach that captures the extreme price peaks.

7 Quarter-hourly resolution of prediction interval accuracy

Table 4 shows the share of realizations within the prediction intervals for each of the four dimensions of the distribution predicted by the normalizing flow. The results show no deterioration of performance for the later time steps. Thus, any concern about limitations of the multivariate model being applied to the step-by-step realizing intraday price are theoretical rather than practical.

8 Energy and variogram score

Figure 4 shows the energy score and Figure 5 shows the variogram score for each model and each hour from July to December 2019 (left), as well as box plots of the overall distributions (right). Note that there are some peaks in the results of all methods in both the energy score and the variogram score, e.g., in September or December. Still, the distribution of the values for energy score and variogram score for the extended test periods confirm the observations from the main paper.

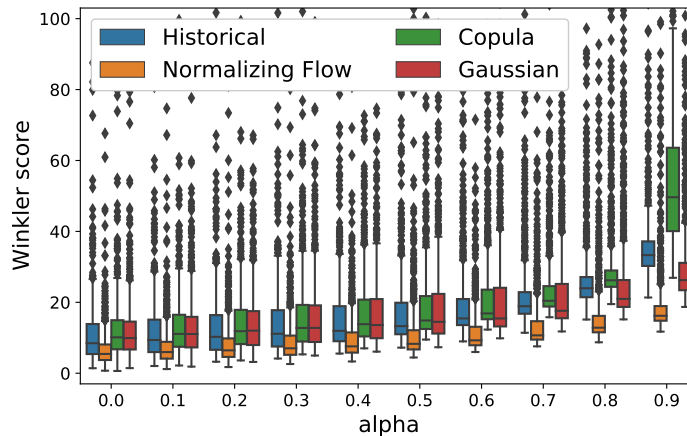


Fig. 2. Winkler score (Winkler, 1972; Gneiting and Raftery, 2007) over confidence interval width α for the different models and the historical data. Training data from January 2018 to June 2019. Full distribution of scores over the test month of July 2019.

Bibliography

- Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., and Saurous, R. A. (2017). Tensorflow distributions. *arXiv preprint arXiv:1711.10604*.
- Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). Density estimation using Real NVP. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. Open-Review.net.
- ENTSO-E Transparency Platform (2022). ENTSO-E Transparency Platform. <https://transparency.entsoe.eu/>. Accessed: 2022-03-01.
- Fraunhofer-Institut für Solare Energiesysteme ISE (2022). Energy Charts. <https://www.energy-charts.info>. Accessed: 2022-03-01.
- Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.
- Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, 17(2):211–235.
- Kruse, J., Schäfer, B., and Witthaut, D. (2021). Revealing drivers and risks for power grid frequency stability with explainable AI. *Patterns*, 2(11):100365.
- Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, 2(1):56–67.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, page 4768–4777, Red Hook, NY, USA. Curran Associates Inc.
- Pinson, P. and Girard, R. (2012). Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96:12–20.
- Scheuerer, M. and Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334.
- Trebbien, J., Gorjão, L. R., Praktikno, A., Schäfer, B., and Witthaut, D. (2022). Understanding electricity prices beyond the merit order principle using explainable ai. *arXiv preprint arXiv:2212.04805*.

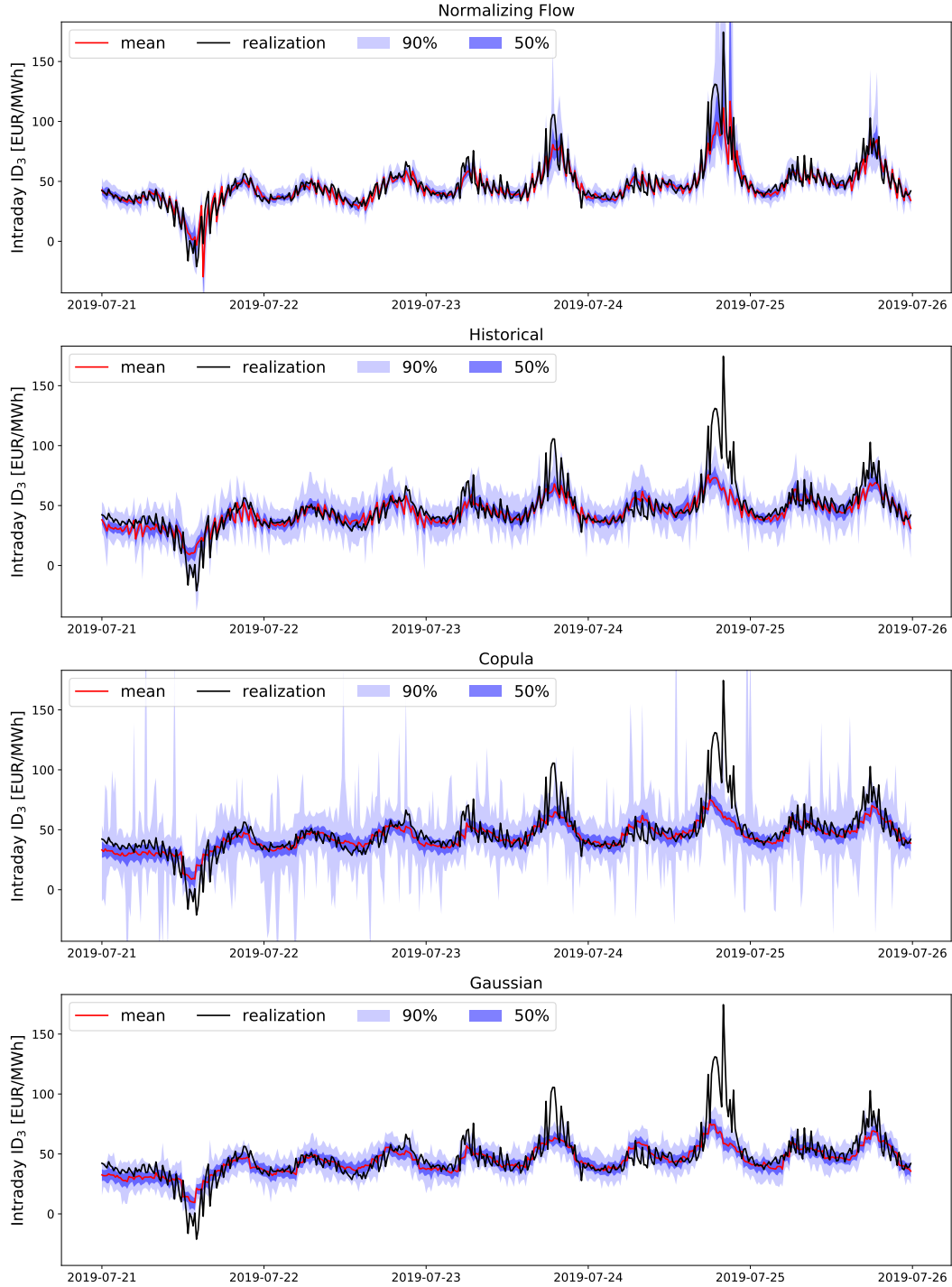


Fig. 3. Predicted mean, prediction intervals of 50% and 90%, and realization between July 21st and July 26th of 2019 estimated from the probabilistic forecasts from the normalizing flow (top), the selection of historical data (second from top), the Gaussian copula (third from top), and the Gaussian regression (bottom). Note the truncated y-axis for the Gaussian copula. Training data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022). Results generated with all conditional inputs.

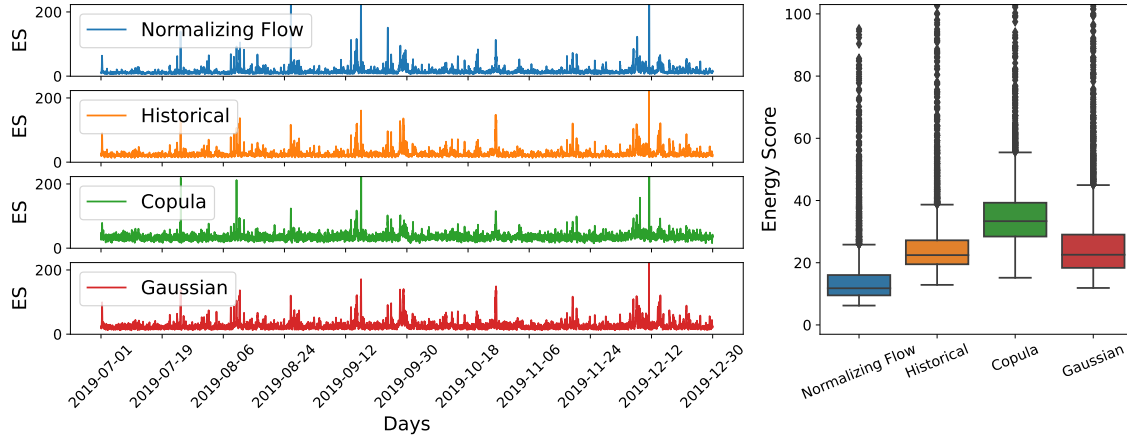


Fig. 4. Energy score (Gneiting et al., 2008; Pinson and Girard, 2012) of the normalizing flow (“Normalizing Flow”), the selection of historical data (“Historical”), the Gaussian copula (“Copula”), and the Gaussian regression (“Gaussian”) for each hour from July to December 2019 (left) and box plot (Waskom, 2021) of the overall distribution for each model (right). Training data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022).

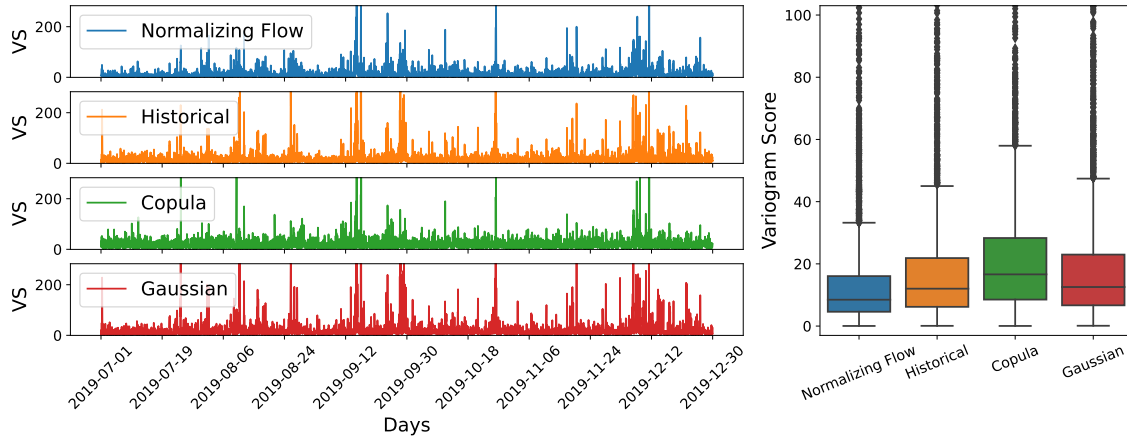


Fig. 5. Variogram score (Scheuerer and Hamill, 2015) of the normalizing flow (“Normalizing Flow”), the selection of historical data (“Historical”), the Gaussian copula (“Copula”), and the Gaussian regression (“Gaussian”) for each hour from July to December 2019 (left) and box plot (Waskom, 2021) of the overall distribution for each model (right). Training data from January 2018 to June 2019 (Fraunhofer-Institut für Solare Energiesysteme ISE, 2022).

- Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021.
- Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. *Journal of the American Statistical Association*, 67(337):187–191.