

Regular article

Short- and long-term forecasting of electricity prices using embedding of calendar information in neural networks

Andreas Wagner^{a,b,*}, Enislay Ramentol^a, Florian Schirra^a, Hendrik Michaeli^b^a Fraunhofer Institute for Industrial Mathematics ITWM, Department for Financial Mathematics, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany^b Karlsruhe University of Applied Sciences, Moltkestraße 30, 76133 Karlsruhe, Germany

ARTICLE INFO

Keywords:

Machine learning
Neural networks
Embedding
Electricity market
Spot price
Forecasting
Price-forward curve
Renewables

ABSTRACT

Electricity prices strongly depend on seasonality of different time scales, therefore any forecasting of electricity prices has to account for it. Neural networks have proven successful in short-term price-forecasting, but complicated architectures like LSTM are used to integrate the seasonal behavior. This paper shows that simple neural network architectures like DNNs with an embedding layer for seasonality information can generate a competitive forecast. The embedding-based processing of calendar information additionally opens up new applications for neural networks in electricity trading, such as the generation of price forward curves. Besides the theoretical foundation, this paper also provides an empirical multi-year study on the German electricity market for both applications and derives economical insights from the embedding layer. The study shows that in short-term price-forecasting the mean absolute error of the proposed neural networks with an embedding layer is better than the LSTM and time-series benchmark models and even slightly better as our best benchmark model with a sophisticated hyperparameter optimization. The results are supported by a statistical analysis using Friedman and Holm's tests.

1. Introduction

Forecasting electricity prices is an important task in the trading process of energy utilities. This paper focuses on short-term price forecasting in the day-ahead market and long-term forecasting of the price profile. Typical applications for short-term forecasting are proprietary trading and short-term dispatching of power plants. The long-term price profile, on the other hand, is needed to generate (usually hourly) price-forward curves from observed futures market prices. Both forecasting tasks face a very strong dependency on calendar information, i.e. season (as a proxy for expected temperature levels), day of the week, and hour. This is visible in Fig. 1, which shows a typical time series of EPEX day-ahead electricity prices for Germany. This paper shows how calendar information can be used in neural-network-based forecasting models to significantly improve the forecasting quality compared to the default of using dummy variables. Our main contribution is to prove the use of embeddings as a very successful way to represent calendar information.

* Corresponding author at: Fraunhofer Institute for Industrial Mathematics ITWM, Department for Financial Mathematics, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany.

E-mail address: andreas.wagner@h-ka.de (A. Wagner).

<https://doi.org/10.1016/j.jcomm.2022.100246>

Received 27 July 2020; Received in revised form 17 January 2022; Accepted 31 January 2022

Available online 20 February 2022

2405-8513/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

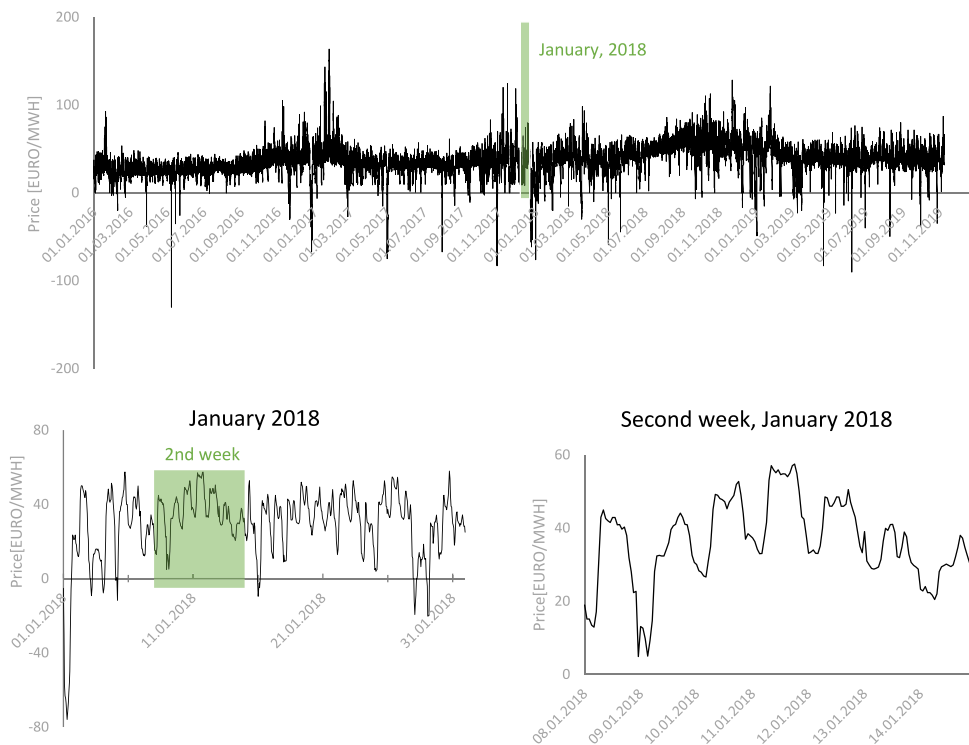


Fig. 1. Hourly EPEX day-ahead electricity prices for Germany on different time scales.

Our paper focuses on the following research questions:

- How can calendar information be included in neural-network-based price forecasting?
- How does an approach based on embeddings perform in short-term price forecasting and long-term profile forecasting?
- Which economic insights can be gained from the embedding layer?

We base our empirical work on the EPEX German day-ahead electricity market and use only publicly available data.

The following Section 2 gives a literature review on electricity price forecasting (EPF), followed by Section 3 which introduces our approach based on neural networks and in particular embeddings, as well as their application to calendar information. Empirical results are presented in Section 4. Embeddings can be used to check the plausibility of the results, which is presented in Section 5. Finally Section 6 concludes.

2. Applications and related literature on electricity price forecasting

This paper distinguishes between two applications of EPF:

- Short-term price forecasting: the aim is to forecast the (usually hourly) prices for the next day as closely as possible
- Long-term profile forecasting: the aim is to generate a time series of hourly prices for several years into the future; in this application, the relation between the prices should be as realistic as possible (e.g. the relative behavior of prices on a Sunday compared to prices on a weekday).

We provide more details on the applications and existing literature for both cases.

Short-term price forecasting

In short-term forecasting, we want to predict the 24 hourly prices of the next day, one day-ahead. A good forecast is needed for trading in the day-ahead market and in decision support concerning power-plant dispatch, the scheduling of an industrial plant, or trading in alternative markets like secondary control or other auxiliary services. The prediction of electricity prices has been widely studied by the research community in areas such as financial mathematics and machine learning.

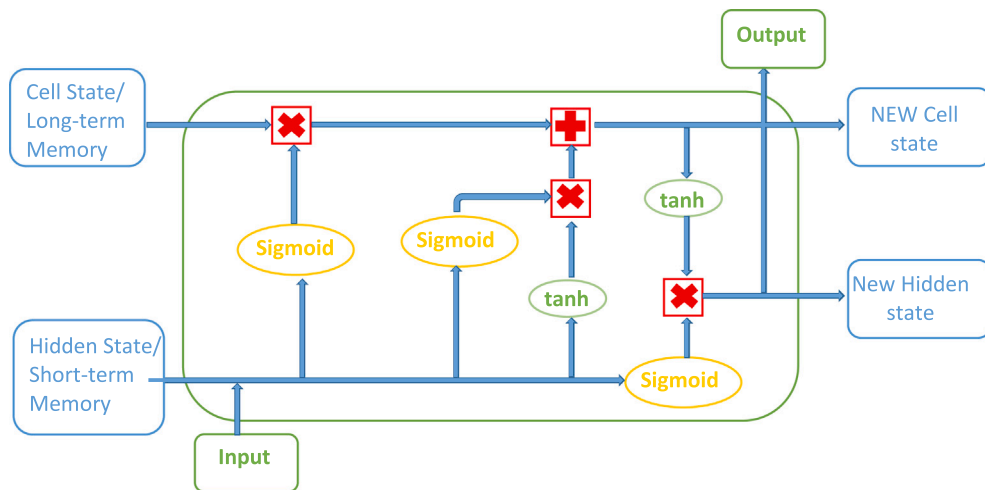


Fig. 2. Internal operation of the LSTM cell.

Overview. Aggarwal et al. (2009) gives an early overview including 47 papers published between 1997 and 2006, with topics ranging from game-theoretic to time series and machine learning models. Weron (2014) provides an extensive overview including game-theoretic, fundamental, reduced-form, statistical, and machine learning models.

One distinguishes univariate models (same model for each hour) and multivariate models (separate models for each hour) and Ziel and Weron (2018) show that there is no clear preference in empirical results. The modeling methods range from time-series approaches as in Ugurlu et al. (2018) and Narajewski and Ziel (2020), dynamic regression and transfer functions (Nogales et al., 2002), wavelet transformation followed by an ARIMA model (Conejo et al., 2005) and weighted nearest neighbor techniques (Troncoso et al., 2007).

Machine learning. There are many applications of machine learning methods in electricity price forecasting. Amjady (2006) compares the performance of a fuzzy neural network with one hidden layer to ARIMA, wavelet-ARIMA, multilayer perceptron, and radial basis function network models for the Spanish market. Chen et al. (2012) use a neural network with one hidden layer on Australian data. On the same market, Mosbah and El-Hawary (2016) use a multilayer neural network focusing on forecasting the next month, but their analysis is based on data from 2005 only. As in this study, Keles et al. (2016) use neural networks to forecast prices on the EPEX German/Austrian power market and show that the machine-learning approach performs better than a competitive time-series model like seasonal ARIMA. In recent years, following the rapid progress of deep learning, more sophisticated variants of neural networks have become popular in EPF (Lago et al., 2018; Zhu et al., 2018; Brusafferri et al., 2019; Kuo and Huang, 2018; Marcjasz et al., 2019). Lago et al. (2018) compare different neural networks and show, using a Diebold–Mariano test, that deep feed-forward, GRU (gated recurrent unit) and LSTM (long-short-term memory) networks perform best on Belgian market data. Note that the *Diebold–Mariano test* refers to the test introduced in Diebold and Mariano (1995). Lago et al. (2021) also propose machine-learning models for EPF (which we will use later in our study) and introduce a methodology to benchmark EPF-models. In the literature there is evidence that a LSTM approach tends to be a competitive neural network setup in EPF, which can be ascertained from various recent studies (see below after the details on LSTM neural networks). Therefore we explain LSTM in more detail and also add it as a benchmark in our study.

LSTM in EPF. The **Long short-term memory (LSTM)**, proposed in Hochreiter and Schmidhuber (1997), is a deep learning framework that has proven successful with time series problems due to its versatility and great efficiency at *remembering* information from the time-series history in the long and short term. LSTMs are a particular variant of Recurrent Neural Networks (RNN). The LSTM networks solved the *short-term memory problem* suffered by their predecessors, RNNs. For each time step, the LSTM cell takes three different inputs: the current input data, the short-term memory from the previous cell, and the long-term memory. The short-term memory is also known as the hidden state, and long-term memory is generally referred to as the cell state. Fig. 2 shows the internal architecture of a LSTM cell. Another relevant characteristic of the LSTM cell is the use of gates to regulate the information to be kept or discarded at each time step. These gates are known as the Input Gate, the Forget Gate, and the Output Gate.

In the following, we summarize the literature on the application of LSTM in EPF. Zhu et al. (2018) use hourly price data from the New England and PJM day-ahead market, and train the model with different input lengths, forecasting horizons and data sizes. The experimental study shows good results in comparison with Support Vector Machines (SVM) and Decision Trees (DT). The LSTM network designed in Zhu et al. (2018) uses the previous prices in a certain time window of length L (lookback window, $L \in [4, 8, 12, 24]$) as features. The authors achieved the best results using $L = 24$, which shows that the price depends heavily on all prices of the previous 24 h. Jiang and Hu (2018) present a study for day-ahead electricity price forecasting using LSTM on the Australian market in the Victoria region and the Singapore market. They use not only historical prices as features in the model but

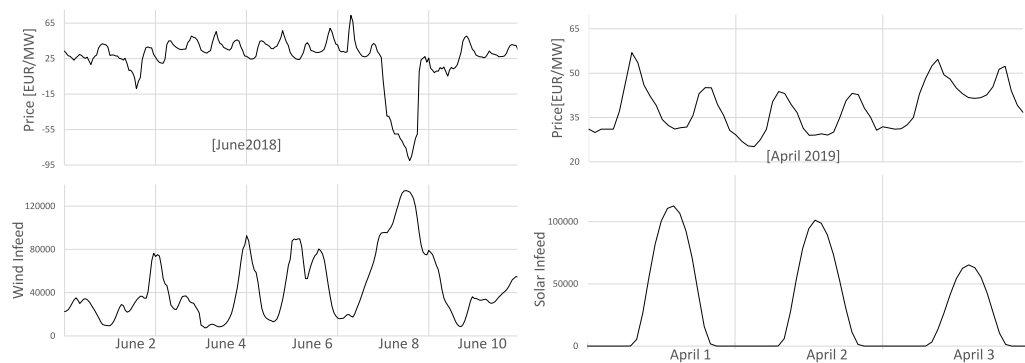


Fig. 3. Influence of wind and photovoltaic infeed on prices shown for examples in June 2018 and April 2019.

also external variables, like holidays, day of the week, hour of the day, weather conditions, oil prices and demand. Their LSTM model predicts only the next hour, so the whole day (24 h) is forecasted in a recursive manner. In Bano et al. (2019) the authors study the New York Independent System Operator (NY-ISO).¹ They compare several machine-learning-techniques using one year of hourly data (2016–2017). In their experimental study, the authors combine Multilayer Perceptrons (MLP), LSTMs, SVMs and Logistic Regression (LR) models with two feature selection methods. They conclude that LSTM networks perform better than MLPs for the EPF problem. A novel approach based on Gated Recurrent Units (GRU) is introduced in Ugurlu et al. (2018) to EPF in the Turkish day-ahead market. The authors compare their proposal with seven methods based on neural networks, including the well-known LSTM and the Convolutional Neural Network (CNN). They use a rolling window of three historical years to predict one day ahead. In their experimental study, the LSTM approach outperforms all selected state-of-the-art methods with a Mean Absolute Forecast Error (MAE) of 5.36.

As shown in our literature review we expect being LSTMs the most promising machine learning approach in EPF, while they perform very well on electricity markets around the world. However, according to Lago et al. (2021), it is hardly possible to define a generally advantageous state-of-the-art EPF method. They claim that there is no sufficient evidence for LSTMs to be more accurate than other methods. To facilitate the comparison between different EPF approaches they also introduce two very well studied, accurate and open source benchmark models (see section below). We take both models from Lago et al. (2021) as benchmarks.

Features. In the literature, the features (also dependent variables or model input) used vary from price data only (Weron et al., 2005) to a whole range of fundamental data like demand, commodity prices, renewable infeed, weather data, etc. Jiang and Hu (2018). Some studies, e.g. Ziel and Steinert (2016) and Schnürch and Wagner (2020), use deep price information from order-book data. To keep our study as concise as possible, we only use the most dominant features. This is calendar information, which plays a major role in the structure of electricity prices as shown in Section 1. Moreover, due to the high share of renewables in the German market, we also use forecasted infeed from wind and photovoltaic (solar radiation) in some of our short-term models. The important role of wind and photovoltaic in the German market has been proven in many studies. The first work that recognized the need to integrate wind and photovoltaic power into models of the German power market is (Wagner, 2014), which also proved their strong impact on the day-ahead electricity price. Using multivariate regression methods, various authors have quantified the influence renewable infeed has on price (Cludius et al., 2014; Würzburg et al., 2013). Due to the regulation, higher renewable infeed generally leads to lower market prices. This relation is shown for two exemplary periods in Fig. 3.

Long-term profile forecasting

The second application to which we apply our proposed forecasting approach is the long-term profile forecasting. The need for a long-term profile forecast is due to the fact that for weeks, months, or even years ahead, no hourly price information can be obtained from the market until the day before delivery. However, for many applications in electricity trading, future price expectations are needed on an hourly granularity. Typical applications include power plant dispatch (hourly price expectations serve as an input for deterministic or stochastic optimization), pricing of contracts for delivery like full-service contracts, production planning in industry, or even as a seasonality component in stochastic price models, as in Hinderks and Wagner (2020), for use in risk-management or the valuation of contracts of delivery with options.

The futures, as traded on electricity markets, always have a delivery period of years, months, or other intervals. The market participants break them down into hourly prices using historically observed day-ahead market prices (which are in hourly granularity). The resulting price curve is called (*hourly*) *price forward curve* (HPFC). A good overview of the concept of HPFC's and multiple approaches for their construction is given in Sæthør (2018). In general, the generation of an HPFC can be separated into two steps:

¹ NY-ISO provides electricity to different countries like United States, Canada and Israel.

Step 1 Long-Term Profile: Generation of an hourly profile using historical hourly prices

Step 2 Absence of Arbitrage: Transformation of the curve according to market quotes for futures to achieve an arbitrage-free HPFC

The larger part of the existing literature is on the absence of arbitrage, which often focuses on methods to smoothen the curve. Two prominent approaches to transform the long-term profile into an HPFC are proposed in Fleten and Lemming (2003) and Benth et al. (2007). Fleten and Lemming (2003) calculate the values of the HPFC directly by minimizing the distance to the long-term profile simultaneously to optimize the smoothness of the HPFC. The absence of arbitrage is ensured by constraints in the optimization. In Benth et al. (2007), on the other hand, the HPFC is not directly constructed. Instead they calculate a correction term consisting of multiple polynomial splines. This provides a smoothing function which adds up with the long-term profile to an HPFC. Especially for the latter method there are multiple extensions. In Sæthre (2018) the polynomial splines are substituted by trigonometric splines. Caldana et al. (2017) introduces a second correction term so that there is one term for base load futures and one for peak load futures. Our model, which we outline in the upcoming sections, can be used to generate a long-term profile. Again, we benchmark it against popular approaches from the literature, which we detail in the following.

For the generation of the long-term profile, there are two common approaches in the literature, see Kiesel et al. (2019) and Hinderks and Wagner (2020). Up to the daily granularity we use two of the approaches – namely the dummy median and the dummy sinusoidal approaches, from Hinderks and Wagner (2020) – as benchmark models. To model the profile of hourly prices we base our approach on Caldana et al. (2017) and Blöchlinger (2008), which use dummy variables with different clusters for days with a similar hourly structure.

Dummy Variables are indicator variables combined with a certain value. They are used in cases where a state is either present or not, e.g. the month of a given date either is January or not. In the literature, one commonly distinguishes between four groups of dummy variables: quarters, months, day types² and hours. They build consecutively on each other to form the long-term forecast. The hourly dummy variables are clustered firstly by the quarter of the considered day and then by the day type. Hence we have 20 different hourly structures. The formula for the dummy-variable-based forecast is then:

$$LTF(t)_{Dummy} = \sum_{i=1}^4 D_i^q(t)c_i^q + \sum_{i=1}^{12} D_i^m(t)c_i^m + \sum_{i=1}^5 D_i^d(t)c_i^d + \sum_{i=1}^4 \sum_{j=1}^5 \sum_{k=1}^{24} D_{ijk}^h(t)c_{ijk}^h \quad (1)$$

Sinusoids Since the dummy variable approach yields profiles with jumps at every time step we introduce a second approach based on trigonometric functions. This tends to produce smoother curves in contrast to the dummy variables. However, it has the drawback that the periodicity ignores irregularly occurring events, such as Easter. Therefore, it is usually combined with the dummy approach such that sinusoids are used for quarterly and monthly variations, while the weekly and daily profiles are modeled by dummy variables. This leads to the following formula for this approach:

$$LTF(t)_{Sinusoidal} = a_0 + a_1 \sin\left(\frac{2\pi t}{8760}\right) + b_1 \cos\left(\frac{2\pi t}{8760}\right) + \sum_{i=1}^5 D_i^d(t)c_i^d + \sum_{i=1}^4 \sum_{j=1}^5 \sum_{k=1}^{24} D_{ijk}^h(t)c_{ijk}^h \quad (2)$$

Note that for calibration the day-ahead prices are deseasonalized with the yearly median price. This is not an issue for the application, as the expected yearly average price is observed from traded Year-Futures. The parameters of the dummy variables are robustly calculated using the median. For the sinusoids, the parameters are calculated via a least-squares approach. The results are shown in Section 4.

3. Embeddings for calendar information and proposed neural network

Dense Neural Networks (DNN) are fully connected networks. Each neuron in a layer receives an input from all the neurons present in the previous layer as outlined in Fig. 4. Despite the fact that this neural network architecture is quite old (McCulloch and Pitts, 1943; Hopfield, 1982), they have gained great popularity in recent years due to the evolution of Deep Learning (DL). The introduction of dense layers in neural networks has brought about a considerable improvement in their performance (Huang et al., 2017). Another contribution of DL is word embedding (Bengio et al., 2003), which almost 20 years after its creation has taken natural language processing (NLP) to levels never before reached. Word embedding is one of the most fascinating areas in DL at the moment and draws the attention of a huge community of researchers (Bian et al., 2014; Deng and Liu, 2018; Khabiri et al., 2019). Algorithms like Word2vec or Doc2vec have allowed natural language analysis that, until a few years ago, was just a futuristic dream.

A word embedding $W: words \rightarrow \mathbb{R}^n$ is a parameterized function mapping words to high-dimensional vectors (Chen and Song, 2017). Typically, the function is a lookup table, parameterized by a matrix, θ , with one row for each word: $W_\theta(w_n) = \theta_n$. W is initialized with a random vector for each word. During the training, it learns meaningful vectors in order to perform some task. The resulting embedding vectors can be interpreted as semantic features, and can be used to understand similarities or differences between words. The distance between vectors describes their semantic similarity. Using the numeric/semantic vectors, it is possible

² We use 1. Mondays, 2. Tuesdays, Wednesdays and Thursdays, 3. Fridays, 4. Saturdays, Partial Holidays and Bridge Days and 5. Sundays and Public Holidays.

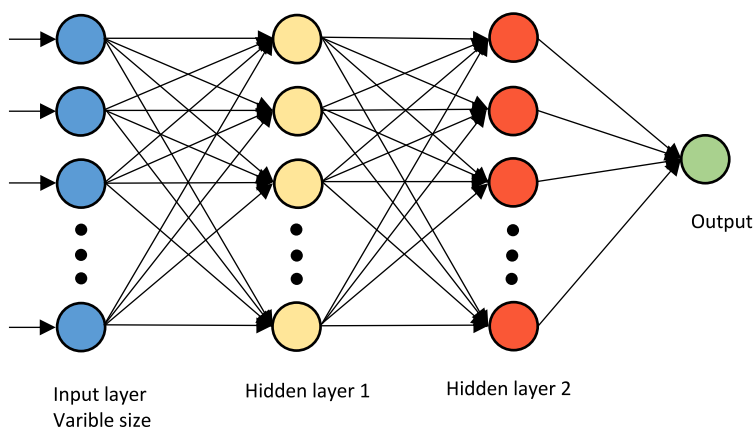


Fig. 4. Fully connected neural network architecture.

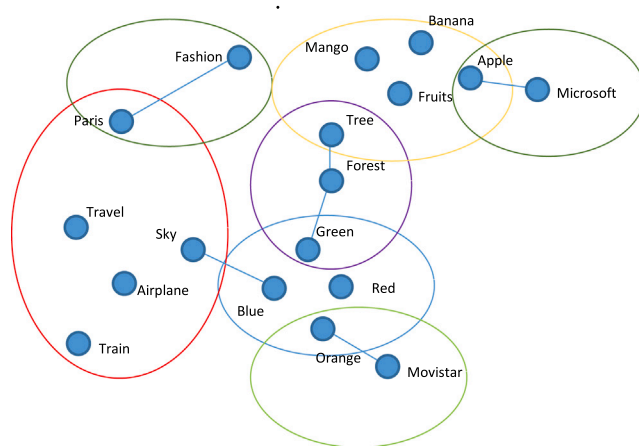


Fig. 5. Example of embedding vectors generated using Word2Vec, projected onto two dimensions using PCA. This illustrates the close proximity of related words. (Colored circles indicate semantically similar words.).

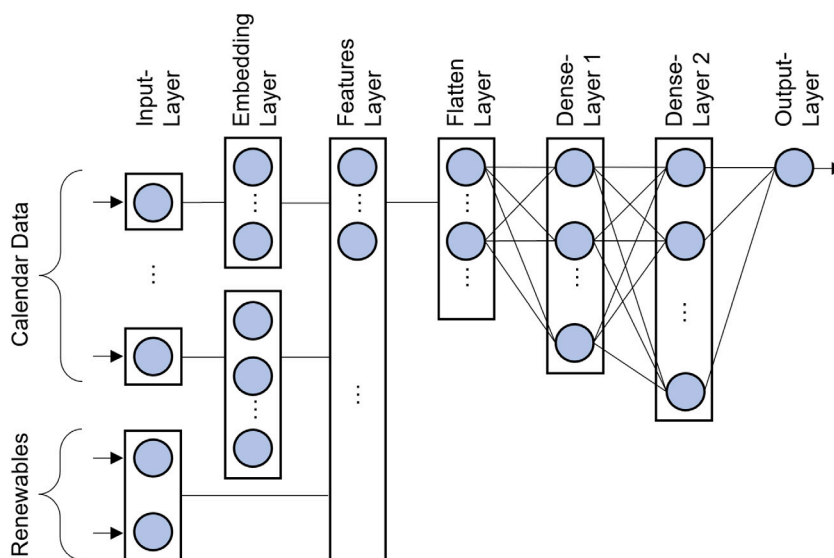


Fig. 6. Scheme of the proposed DNN with an embedding layer.

to gain additional insights, as shown in Fig. 5 for an NLP example. We observe that words with a similar semantic meaning are close to each other in the 2-dimensional projection of the embedding space. In this paper, we utilize the advantage that an embedding layer turns categorical variables into vectors. We use the concept of word embedding to encode the calendar features (month, weekday, hour) into a neural network. The resulting embedding vectors are used for two purposes:

1. As features in the neural network representing calendar information
2. To graphically understand how electricity prices behave depending on time variables and derive economical insights

Proposed neural network

In this paper we propose a dense neural network with an embedding layer to encode the calendar information. Fig. 6 illustrates the network. The input features are calendar data and, in some applications, infeed forecasts for renewables. The actual neural network consists of two hidden layers with Relu (Rectified Linear Unit) activation and an output layer with linear activation (as this is a regression task). In Section 4 we will describe the models and parameters in more detail.

Embedding layer for calendar features. The embedding layer is used to encode the calendar information. Embeddings are an alternative to one-hot-encoding of categorical features. As for every layer in a neural network including the embedding layer, the number of neurons has to be chosen carefully (*the number of dimensions is a hyperparameter you can tweak* (Géron, 2019) on embeddings). Scientific literature does not provide explicit guidelines and, to the best of our knowledge, embeddings have not been used in EPF so far. However, there are some studies on the dimensionality of word embeddings (Yin and Shen, 2018; Gu et al., 2021; Wendlandt et al., 2018). The authors point out the importance of a correct choice of the dimensionality of the embedding, as a high dimension can lead to overfitting and a very low dimension cannot capture all the meaning of the categorical variable. Most of these studies agree that the dimension must be selected empirically as it depends on the data. We optimized the dimension between 15 % and 35 % in 5% intervals and found that 25 % gives the best overall results, even though the differences were rather small.

We consider the following embedding variables:

- Hour: The categorical variable *hour* takes values in $\{0, 1, \dots, 23\}$, so in a one-hot-encoding it has dimension 24. For its embedding dimension we chose six.
- Weekday: The categorical variable *Weekday* takes values in $\{0, 1, \dots, N\}$, where N depends on the representation of holidays outlined below. Its embedding dimension is two. We consider three approaches to deal with holidays:
 - Approach 1: $\text{weekday} \in \{0, 1, \dots, 7\}$ considering each weekday (Sunday, Monday, ..., Saturday) separately and adding a category for holidays.
 - Approach 2: $\text{weekday} \in \{0, 1, \dots, 9\}$ considering seven weekdays and three types of holidays: Partial holiday, public holiday and bridge day.³
 - Approach 3: This approach uses two embedding variables, i.e. treating holiday and weekday separately. This is motivated by the fact, that every holiday also has an associated weekday (i.e. Easter Sunday is a holiday, but also a Sunday): $\text{weekday} \in \{0, 1, \dots, 6\}$ and $\text{type_holiday} \in \{0, 1, 2, 3\}$.

A list of holidays used is in Appendix. In this paper we present the results for Approach 2, as it gave the mean absolute deviation. The interested reader can get detailed results for Approaches 1 and 3 by contacting the authors.

- Month: The categorical variable *Month* takes values in $\{1, 2, \dots, 12\}$ and its embedding dimension is three.
- Year: The categorical variable *Year* takes values in $\{2010, 2011, \dots, 2019\}$ and its embedding dimension is three.
- Cross-feature month-hour: Due to differences in daylight hours, there is a relationship in electricity demand between month and hour-of-the-day. The categorical variable *Month – hour* takes values in $\{1, 2, \dots, 12 * 24\}$ and its embedding dimension is ten.
- Cross-feature weekday-hour: Due to differences in human behavior (e.g. people get up later on the weekend) there is a relationship in electricity demand between weekday and hour. The categorical variable *weekday – hour* takes values in $\{0, 1, \dots, N * 24\}$ and its embedding dimension is fifteen.

4. Empirical performance on the German electricity market

In this section we carry out an experimental study on the German electricity market. We distinguish the two applications as outlined in detail in Section 2. In both applications we conduct a statistical analysis to show the significance of the results.

³ Bridge, partial, and public holiday describe days with influence through public holidays. Public is the actual public holiday, partial is a public holiday in only parts of Germany and bridge describes days between a public holiday and weekends.

Table 1
Data description.

	EPEX DE(/AT)	Expected production	
		Photovoltaic	Wind
Features	Date, price	Date, expected volume	Date, expected volume
Start date	1/1/2010	1/1/2010	1/1/2010
Final date	31/12/2019	31/12/2019	31/12/2019

Short-term forecasting: For the short-term forecasting we use a DNN with an embedding layer to encode the calendar information and six different benchmark models. Four of them are neural network models, one is a time series approach and one is a naive method. In addition to calendar information we also use forecasts on renewable infeed (wind and photovoltaic) as features (see below for details).

We follow a training framework with daily recalibration, so we retrain the model every day with historical data available. As we forecast the day-ahead market, which is traded at 12 o'clock for the next day, we can rely our model on prices up to the current day (which was traded the day-before) and forecasts on renewable infeed for the next day. In our experimental study we use a history of five years to train the model. In other words, using a five-year history of the data available up to the time the prediction model is run, the model predicts the prices for each of the next 24 h.

Long-term profile forecasting: For the long-term forecasting we use a DNN with an embedding layer and, as a benchmark, popular methods from the literature (dummy variables and sinusoidal). We forecast four years ahead.

Data. For our study we use data from the German⁴ Day-Ahead electricity market (EPEX DE) from 2010 through 2019,⁵ as traded on EPEX Spot.⁶ We also use data on the expected generation from renewables in Germany, which we collect from the EEX transparency platform.⁷ We compiled our data sets from FTP-files which we licensed from the German Energy Exchange EEX,⁸ but the data can be viewed without a license on the corresponding websites. More details on the data are in Table 1. We use an Ex-Ante timestamp (i.e. hour 1 describes the price or renewable infeed for the time between 1:00 AM and 2:00 AM).

It is a known fact that neural networks work much better when variables are normalized. That is why in our experimental study the renewable variables are scaled using a technique known as standard scaler:

$$X_{normalized} = \frac{X - X_{mean}}{X_{stddev}} \quad (3)$$

The values X_{mean} and X_{stddev} are computed over the training set only. The prices were not normalized in our models because they are the output feature (makes no difference in training), except for the LSTM models, which use prices also as an input feature.

The evaluation metric used in this study is the mean absolute error (MAE):

$$MAE = \frac{1}{H} \sum_{i=1}^H |P_r - P_p|, \quad (4)$$

where H is the number of hours, P_r is the realized price on the exchange and P_p is the predicted price.

Short-term forecasting

This section summarizes the results of the day-ahead forecasting of hourly prices. We compare our proposed neural networks using embeddings and benchmarks from the literature.

Setup. For the experimental study we design different configurations for DNN and LSTM models, see Tables 2 and 3.

In general there are alternative approaches to include calendar information in neural networks. We compare the embedding approach to two alternatives:

- Using ordinal variables:
 - weekday: numeric variable {0..9}
 - month: numeric variable {1..12}
 - hour: numeric variable {0..23}
- Using a periodicity function for month and hour. To take into account the periodicity of months and hours we project these values onto a circle and use the two-dimensional projections as features as in Schnürch and Wagner (2020).

⁴ Note that the German spot market had been a larger market including Austria (EPEX DE/AT) until October 1st 2018.

⁵ Due to missing data from EEX Transparency, we excluded 11/01/2010 and 10/02/2010 from our analysis.

⁶ <https://www.epexspot.com/en/market-data>

⁷ <https://www.eex-transparency.com/power/de/production/usage/>.

⁸ <https://www.eex.com/de/marktdaten/strom>

Table 2
Dense neural network parameters.

Parameters	c1	c2	c3	c4	c5
Hidden layers	1	2	2	2	2
Neurons per layer	2085	128/128	2285/1024	484/381	234/203
Activation layers	Relu	Relu	Relu	Sigmoid	Relu
Epochs	10	10	10	Auto	Auto
Optimizer	RMSprop	RMSprop	RMSprop	Adam	Adam

Table 3
LSTM parameters.

Parameters	
Hidden layers	3
Neurons per layer	10/10/24
Type of layer	LSTM/LSTM/dense
Activation layers	Relu
Epochs	10
Optimizer	Adam

- weekday: numeric variable {0..9}
- month: see Eq. (5)
- hour: see Eq. (6)

$$month_x(t) := \sin\left(\frac{2\pi m}{12}\right), month_y(t) := \cos\left(\frac{2\pi m}{12}\right) \quad (5)$$

$$hour_x(t) := \sin\left(\frac{2\pi m}{24}\right), hour_y(t) := \cos\left(\frac{2\pi m}{24}\right) \quad (6)$$

To select the configurations for the DNN we follow recommendations from Kapoor et al. (2019) and use the formula in Eq. (7) (as also referenced in a blog-post Eckhardt (2018) on the *towardsdatascience.com*-website, a platform very popular among practitioners):

$$N_h = \frac{N_s}{(\alpha * (N_i + N_o))} \quad (7)$$

N_i is the number of input neurons, N_o the number of output neurons, N_s the number of samples in the training data, and α represents a scaling factor that is usually between 2 and 10. We calculate and compare the following models to forecast the 24 hourly prices of the next day.

We compare the following model architectures. For most of them we do calculations with two sets of features, namely only calendar information as well as calendar information and renewables (forecasts on the infeed of wind and photovoltaic/solar energy).

Naive A naive model using past prices. The output of the naive method for hour h of date d is the price at hour h of the last observed day of the same type (e.g. working day, Saturday).

LEAR LASSO Estimated Auto-Regressive model is one of two benchmark models presented in Lago et al. (2021). The model is based on a parameter-rich ARX (Auto-Regressive with exogenous features) structure which is estimated by LASSO (least-absolute-shrinkage and selection operator). They show that long calibration windows (three and four years) lead to the best results. For our LEAR model we chose five years as calibration window, as we used the same time frame for the recalibration of our DNN models.

LSTM Neural-network with LSTM-architecture and configuration as in Table 3.

DNN-Lago Dense neural network, which is the second benchmark model presented in Lago et al. (2021). This model is based on a multivariate framework. The input features and hyperparameters of every model configuration are calculated in a separate pre-processing using Bayesian optimization. We use the first five years of our dataset for the hyperparameter optimization. The resulting configurations are shown as configurations c4 and c5 in Table 2.

DNN-ordinal Dense neural network with three different configurations c1, c2 and c3 as shown in Table 2. This model is also evaluated without calendar information using only renewables as features.

DNN-sin-cos Dense neural network with three different configurations c1, c2 and c3 as shown in Table 2. Calendar information is encoded as a circle function as in Schnürch and Wagner (2020).

Table 4
Mean absolute error for all benchmark methods and the proposed models.

Method	Features	Config.	2015	2016	2017	2018	2019	all
LSTM	Renewables	–	7.12	6.52	7.94	9.38	8.77	7.94
Naive	–	–	7.34	6.19	9.89	10.43	9.77	8.72
LEAR	Renewables	–	4.22	4.26	4.70	5.92	4.84	4.79
DNN-Lago	Renewables	c4	3.68	3.70	5.32	5.01	4.52	4.45
		c5	3.76	3.52	4.56	4.77	4.53	4.23
DNN-ordinal	Renewables	c1	8.30	6.98	9.76	9.77	9.49	8.85
		c2	8.23	6.69	9.66	9.52	9.11	8.64
		c3	8.51	6.88	9.56	9.83	8.72	8.70
	Calendar	c1	7.76	6.76	9.87	10.06	9.10	8.71
		c2	7.78	6.71	9.77	10.15	9.11	8.70
		c3	7.21	6.47	9.38	9.87	8.65	8.31
	+ renewables	c1	5.44	4.98	7.05	6.73	6.97	6.23
		c2	5.57	4.92	7.03	6.83	7.14	6.30
		c3	5.16	4.66	6.52	6.63	6.15	5.82
DNN-sin-cos	Calendar	c1	7.06	5.93	8.93	10.05	8.18	8.03
		c2	6.71	5.67	8.75	9.33	8.01	7.69
		c3	6.26	5.47	8.48	8.70	7.59	7.30
	+ renewables	c1	4.29	3.93	5.58	5.89	5.83	5.10
		c2	4.13	3.72	5.43	5.51	5.39	4.84
		c3	3.88	3.49	4.85	4.90	4.77	4.38
DNN-embedding	Calendar	c1	5.92	5.06	8.11	8.44	8.41	6.98
		c2	6.04	5.25	8.16	8.38	7.37	7.04
		c3	5.86	5.00	7.87	8.19	7.24	6.83
	+ renewables	c1	3.78	3.42	5.12	4.93	5.00	4.45
		c2	3.82	3.38	5.09	4.98	4.78	4.41
		c3	3.50	3.21	4.69	4.65	4.46	4.10

DNN-embedding Our approach as presented in the previous section. We use three different configurations c1, c2 and c3 as shown in Table 2.

The models are trained on the last five years preceding the day we forecast and are retrained daily. Training the LSTM models is computationally extensive and could take more than one day in practice. For this reason we use a simple configuration for our experimental study. All DNN configurations are summarized in Table 2. We use MSE as loss function. Note that we do show only the most important hyperparameters for the sake of clarity.

Results. We predict the next 24 h using the past five years of data. We repeat these experiments for every day starting on the 1st of January 2015. Note that, therefore, all results are out-of-sample and provide a valid benchmark for use in practice. In the following, we present the mean hourly absolute error per year, for every different configuration and training sample. Any interested reader can get detailed results (every hour) by contacting the authors. Table 4 shows the results.

Our proposed approach based on embeddings performs well compared to the benchmarks. The DNN-embeddings + renewables approach has an overall MAE of 4.10 EUR/MWh for its best configuration (DNN-emb-renew-c3). Other competitive approaches include DNN-Lago (4.23) and DNN-sin-cos (4.38). Fig. 7 shows an example of hourly prediction for September of 2016 using renewables and embeddings for calendar information. It can be observed that our model is able to nicely capture the seasonal structure.

We can conclude that our proposed method is very competitive with the existing state-of-the-art machine-learning based forecast of electricity prices. However, we think that it provides those good results with a fairly simple model architecture. In order to statistically support our findings we conduct an extensive analysis using non-parametric tests in the following.

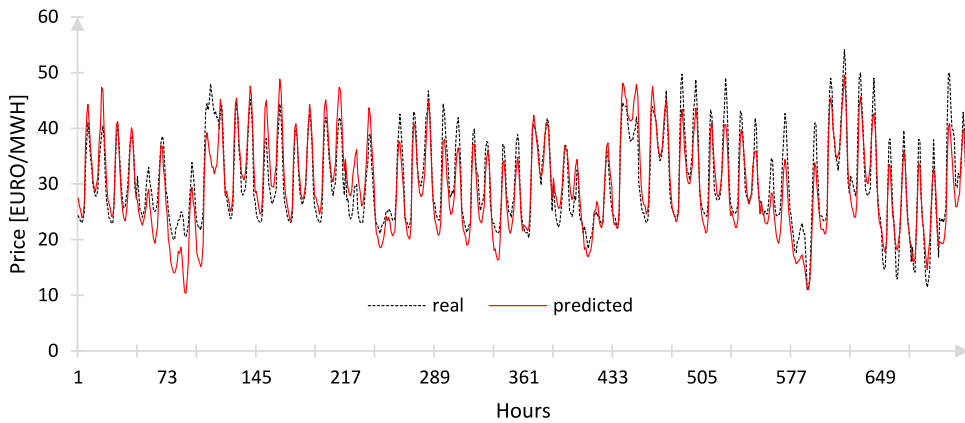


Fig. 7. Hourly and predicted prices for September of 2016 using model DNN-emb-renew-c3.

Table 5

Average Friedman Ranking and adjusted p -values using Holm's Posthoc procedure for the 5 years of daily predictions, using DNN-emb-renew_c3 as the control algorithm.

i	Method	Friedman ranking	Adjusted p-value
1	DNN-emb-renew-c3	4.0975	–
2	DNN-Lago-renew-c5	4.262	0.669005
3	DNN-Lago-renew-c4	4.592	0.000093
4	DNN-emb-renew-c2	4.825	0
5	DNN-emb-renew-c1	4.9181	0
6	LEAR-renew	5.0452	0
7	DNN-emb-cal-c3	7.1679	0
8	DNN-emb-cal-c1	7.3182	0
9	DNN-emb-cal-c2	7.6054	0
10	Naive	8.043	0
11	LSTM	8.1257	0

Statistical analysis

In this section we carry out a statistical analysis using non-parametric tests as recommended in Demšar (2006) in order to compare the different algorithms and configurations appropriately. We use the open-source software tool KEEL (Knowledge Extraction based on Evolutionary Learning) from Alcalá-Fdez et al. (2011).⁹ Our goal is to find out if there are significant differences between our models using embeddings and the benchmark algorithms.

First, we use Friedman's aligned-ranks test to detect statistical differences among a set of algorithms (Friedman, 1937). The Friedman test computes the average aligned-ranks of each algorithm, obtained by computing the difference between the performance of the algorithm and the mean performance of all algorithms for each dataset. In our setting, every daily MAE can be considered a new result for a different dataset. It holds: the lower the average rank, the better the algorithm. If the Friedman test finds significant differences between the compared algorithms, we check if the control algorithm (the one with the smallest rank) is significantly better than the others using Holm's Posthoc test (Holm, 1979). We use a significance level of $\alpha = 0.05$. We are aware that the Diebold–Mariano test is a common method when comparing different forecasting approaches. However we refer to Diebold (2015), where the originator of the Diebold–Mariano test evaluates the usage of his test in research. He argues that it is misused in a setting where different models are compared in a pseudo out of sample analysis. We are in this setting and therefore follow the advice by using another test, i.e. the Friedman test. In the comparison we have 1826 samples, i.e. one per day for the five years (2015–2019) we are forecasting. We compare a total of 11 models. Table 5 shows the average ranks obtained by each method in the Friedman test and the adjusted p -values obtained by Holm's Posthoc using the DNN-embddings + renewables configuration 3 (DNN-emb-renew_c3) as control algorithm. The p -value computed by the Friedman test is 0 or close to 0, which means that there exist significant differences between the compared algorithms and the hypothesis of equivalence can be rejected. We can observe that the best rank corresponds to the DNN-emb-renew_c3 followed by DNN-Lago-renew-c5 and DNN-Lago-renew-c4. From the p -values computed by the Holm's Posthoc test we can conclude that our approach based on embedding variables (configuration 3) is statistically superior to all the compared algorithms in Table 5 with the exception of DNN-Lago-renew-c5, in which case the null hypothesis cannot be rejected.

To compare all models to each other, we perform a $N \times N$ Friedman's aligned-ranks test and compute the p -values using the Holm's Posthoc test. Fig. 8 shows a heat map with all the p -values obtained with the Holm's test. The algorithms have been ordered

⁹ <http://www.keel.es>.

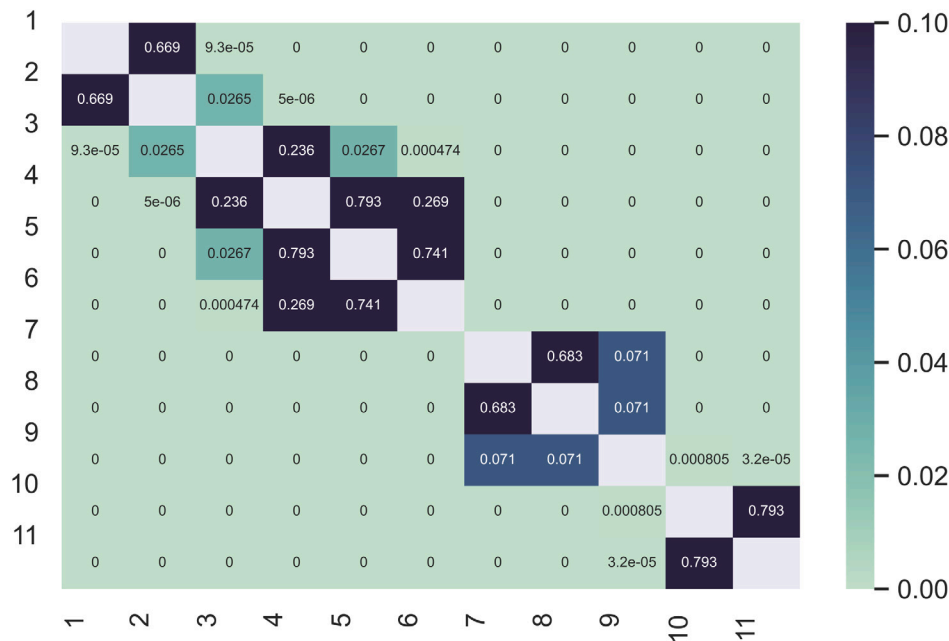


Fig. 8. p -values obtained by Holm's test. The algorithms have been ordered according to their Friedman rank. The numbers on the axes correspond to the number (i) in Table 5.

according to the Friedman ranking. In the x -axis the algorithms are ordered from left to right (best to worst). In the y -axis the algorithms are ordered from top to bottom (best to worst).

From the statistical study performed, we conclude that the use of embeddings to turn the categorical calendar variables into vectors significantly improves the predictions of electricity prices and allows the use of simple neural network architectures.

Analysis of errors

The claim that embeddings can be used to gain insight into the *black box* neural network only holds true if there is no seasonality structure left in the errors. Therefore, we present an analysis of the errors and show that there is no significant pattern in the errors caused by calendar information (hour, type of day, etc.). We use the best model configuration, i.e. DNN-emb-renew-c3.

Boxplots of the errors for different embedding variables in Fig. 9 show that there is no particular pattern. There are differences in the scatter range from time to time. In particular it is noticeable that the mean values (green dashed lines) are close or even to zero. Therefore it can be assumed that the model is unbiased or at least without significant bias and it accounts – at least in the mean – for seasonal effects. A comparison of the autocorrelation of absolute errors and day-ahead prices further justifies this conclusion. Fig. 10 shows autocorrelations for one-week ($7 * 24$ h). A seasonality is obvious in day-ahead prices (as expected), whereas for the absolute errors almost all autocorrelations are below 0.05 and thus negligible. We also analyzed autocorrelation for multiple months with the same conclusions. In particular, the errors do not reproduce the autocorrelation structure of day-ahead prices, proving that the model indeed captures the seasonality of prices. For these reasons, we conclude that seasonal effects are largely captured by the neural network.

Long-term profile forecasting

This section shows the use of a simple dense neural network with (embedded) calendar features in order to generate a long-term profile of expected electricity prices. We mimic the practical application: we train the model on historical data and use it to forecast an hourly profile. Although the models can be used to forecast multiple years (in practice the next four years may be relevant), for the presentation of results we only forecast one year ahead in order to be able to calculate out-of-sample error measures on our data. For the evaluation, we define suitable measures on the quality of the generated hourly profile. We use two benchmark models from the literature for comparison.

Setup. We train the models on January 1st 2015, 2016, 2017, 2018, 2019 and use in each case the whole past from 1/1/2010 for training. We compare two neural networks (one with cross-features and one without) and compare them to standard approaches from the literature:

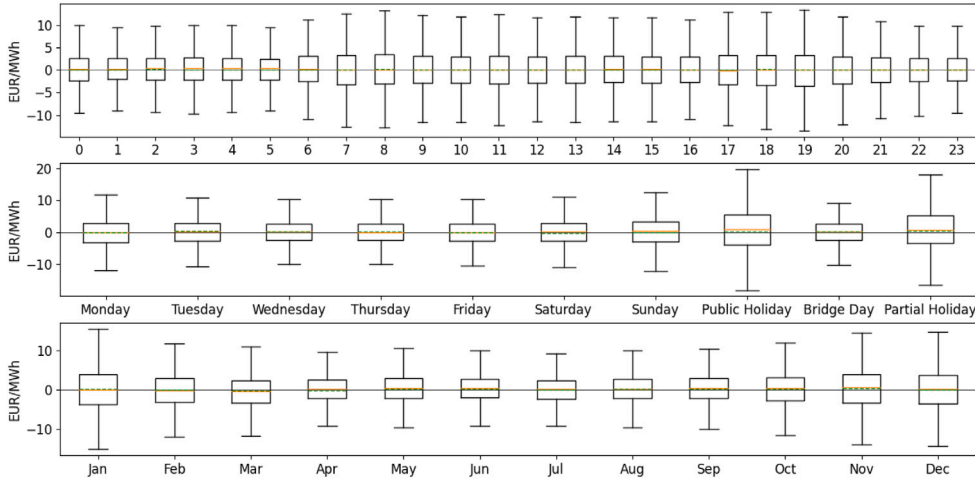


Fig. 9. Boxplots of the errors for different embedding variables for DNN-emb-renew-c3.

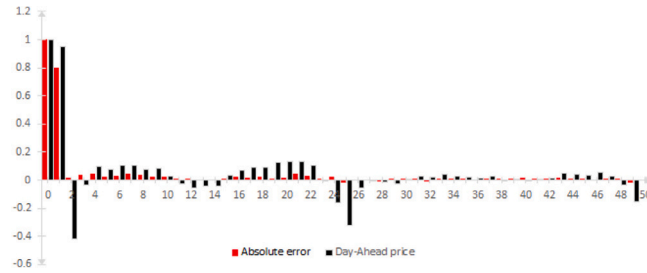


Fig. 10. Partial autocorrelations of the absolute error of the forecast and the Day-Ahead prices.

LTF(1) Dense neural network, using the calendar features month, weekday, hour, and year as described in Section 3. The different configurations are detailed in Table 2.

LTF(2) Same as LTF(1), but additionally using cross-features for hour and type of day as well as hour and month.

LTF(3) Benchmark method using dummy variables as outlined in Section 2 and Eq. (1) using the calendar features month, day type, hour, and year.

LTF(4) Benchmark method using sinusoidals as outlined in Section 2 and Eq. (2) using the calendar features month, day type, hour, and year.

In a nutshell, we evaluate four models, use three different configurations for the neural-networks and five test periods, resulting in a total of 40 experiments. In the following we only give summaries, important results, and make conclusions. Additional results can be obtained by contacting the authors.

As detailed above, the generated profiles are used to create hourly price forward curves (HPFC). In order to generate the HPFC, the hourly profiles are shifted according to observed market prices for futures.¹⁰ For this reason, the quality or goodness-of-fit of the profiles cannot be measured as a classical forecast error. We have to define a suitable measure. The measure must rate the quality of the structure of the daily and hourly profile. Therefore we do not compare the forecasted prices with the realized prices directly, but use the following modified time-series in order to calculate error measures like an MAE or L2-error.

We use the deviation of the daily average price to the corresponding monthly average (dDev) and the deviation of hourly price to the corresponding daily average (hDev) for the hourly seasonality. This is

$$hDev_d^h = p_d^h - \frac{1}{24} \sum_{h=1}^{24} p_d^h$$

¹⁰ Traded futures have delivery periods of years, quarters, months or weeks. In particular no hourly prices can be observed on the market until a day before delivery.

Table 6

MAE of dDev (validating **daily** seasonality). Yearly recalibration, training with data from 2010 to the year previous to the one simulated.

Year	LTF(1)			LTF(2)			LTF(3)	LTF(4)
	c1	c2	c3	c1	c2	c3		
2015	4.7264	4.8842	4.8637	4.6813	4.7146	5.0050	4.8461	4.7952
2016	4.5485	4.7153	4.7812	4.2544	4.3026	4.2401	4.6787	4.3661
2017	7.1219	7.1328	7.3774	6.9513	7.0069	6.9197	7.0708	6.9616
2018	6.4918	6.5174	6.5266	6.4808	6.3927	6.5746	6.4396	6.3725
2019	6.2172	6.2296	6.3075	6.1461	6.1218	6.2829	6.1734	6.0571
Mean	5.8212	5.8959	5.9713	5.7028	5.7077	5.8045	5.8417	5.7105

Table 7

MAE of hDev (validating **hourly** seasonality). Yearly recalibration, training with data from 2010 to the year previous to the one simulated.

Year	LTF(1)			LTF(2)			LTF(3)	LTF(4)
	c1	c2	c3	c1	c2	c3		
2015	4.1228	4.5492	4.4911	4.0262	3.9669	4.0012	4.6098	4.3846
2016	3.5656	4.2070	3.8957	3.5573	3.7369	3.3727	4.7161	3.9993
2017	5.2106	4.6969	5.0940	4.6332	4.7876	4.7680	5.6534	5.071
2018	4.7313	4.8363	4.5289	4.6598	4.6360	4.6511	5.3895	4.8848
2019	4.7318	4.8349	5.0043	4.5092	4.6469	4.5611	5.6929	4.8285
Mean	4.4724	4.6248	4.6028	4.2771	4.3549	4.2708	5.2123	4.6336

and

$$dDev_d = \frac{1}{24} \sum_{h=1}^{24} \left[p_d^h - \frac{1}{|m(d)|} \sum_{d \in m(d)} p_d^h \right]$$

where p_d^h is the price at day d in hour h and $m(d)$ is the month corresponding to day d . In the following, we further use the MAE (Eq. (4)) to analyze the forecast quality.

Results. The forecast quality is defined as the MAE between $hDev$ ($dDev$) of the forecasted prices and $hDev$ ($hDev$) of the realized prices. Note that for analysis we only compare the one-year-ahead forecasts, since this is the most relevant time-horizon for practitioners. Regarding the daily deviations, we cannot find a significant difference between any of the methods, see Table 6. The range of the overall MAEs is just 0.2685 or 4.71% of the smallest overall MAE. Moreover, when comparing the best results of the benchmarks and the embedding forecasts, the difference in MAE is just 0.0077 or 0.14% of the smallest overall MAE. We conclude that the embedding approach is as well suited to reproduce the structure of daily prices as established methods but yields no significant improvement to the benchmark. Concerning the hourly deviations (Table 7) we see a different picture. In this case, the best embedding configuration – LTF(2)_c1 – improves the overall MAE of the best benchmark method – the sinusoidal approach LTF(4) – by 0.3565 or 8.35%. In particular, the results are strictly better for every year and do not interchange. For a graphical overview, we reduce the number of considered methods to the best neural network configuration and the best benchmark method, i.e. LTF(2)_c1 and LTF(4). Fig. 11 (Fig. 12) shows differences between the daily (hourly) deviations of the two methods and the real spot prices, i.e. the error of the daily (hourly) deviations. In general, the errors follow a similar curve, which can be expected as both methods make use only of calendar information. However, as the numbers suggest, the hourly deviations for the embedding-based neural network are smaller.

We can conclude that our neural network is applicable to the task of creating long-term profiles for forwarding curve generation. It is even slightly better than the benchmark methods used for comparison.

5. Economic insights from embedding layer

In 2018, the new General Data Protection Regulation (GDPR) approved by the European Union entered into force, requiring citizens to have the right to an explanation regarding any algorithmic decision-making (Goodman and Flaxman, 2016). The new GDPR states that if an algorithm makes an automatic decision regarding a user, this user will have the right to obtain an explanation of how the decision was made (Qureshi and Greene, 2017). In this section, we briefly show the use of the embedding vector obtained

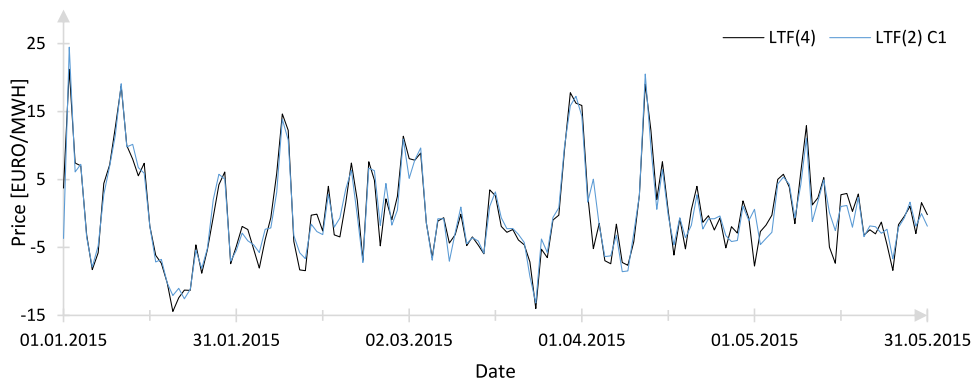


Fig. 11. Errors of daily deviations (LTF(2)_c1 and LTF(4)) for training with 5 years of real spot prices.

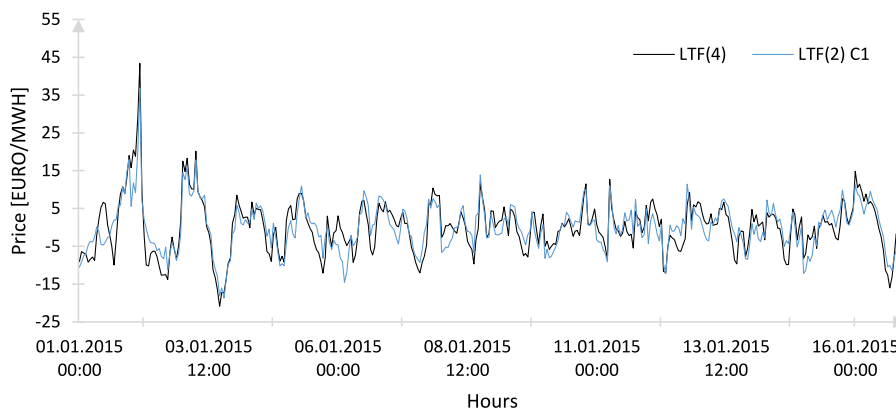


Fig. 12. Errors of hourly deviations (LTF(2)_c1 and LTF(4)) for training with 5 years of real spot prices.

during the training of the DNN to graphically understand how the models use the calendar information in the forecast. To visualize the resulting embedding vectors we use Tensorflow Projector.¹¹ We carry out the embedding vectors analysis using the vectors trained with data from 2010 to 2018 and testing with data from 2019 using LTF(2).

Fig. 13 shows a visualization of the variable month in the embedding space. It has three dimensions, as explained in subsection 3. It can be seen that the embedding separates the winter months of January and December as well as August, which is the holiday season in Germany. Also, the spring months March, April, and May are very close, similar to November and February, which are winter months without long periods of public holiday (in contrast to December and January). Overall, the embedding shows a very reasonable picture from an energy economist's point-of-view. Fig. 14 shows the resulting visualization of the variable weekday in the embedding space. We additionally show the use of the cosine-distance as a tool for the analysis of embeddings. We calculate the distance of "Sunday" and find that it is very close to "Saturday", "partial holiday" and "public holiday", indicating that the network learned that the hourly profile is similar for those types of days. This is again reasonable.

Although these findings seem to be well-known conclusions for electricity price experts, finding relationships between the days of the week, the months, and price variations is a very valuable tool in model analysis. Neural networks typically are black-box models that hardly allow any insight into the model's logic. This reduces the acceptance in practice, as there remains a significant risk, that the models "learned the wrong thing". From our experience, one can reduce these risks by choosing simple network architectures and extensive out-of-sample performance testing. Both are fulfilled in this study, and additionally, we are able to provide graphical insight into the model's logic, which can be used to gain the trust of decision-makers.

6. Conclusion

This paper provides a novel method to forecast electricity prices using machine learning, which competes with existing approaches, but uses easier-to-understand model architectures, and provides insight into the model's logic. In an extensive study on the German electricity market, we compared our approach to benchmarks from the literature. We showed that the embedding

¹¹ <https://projector.tensorflow.org/>.

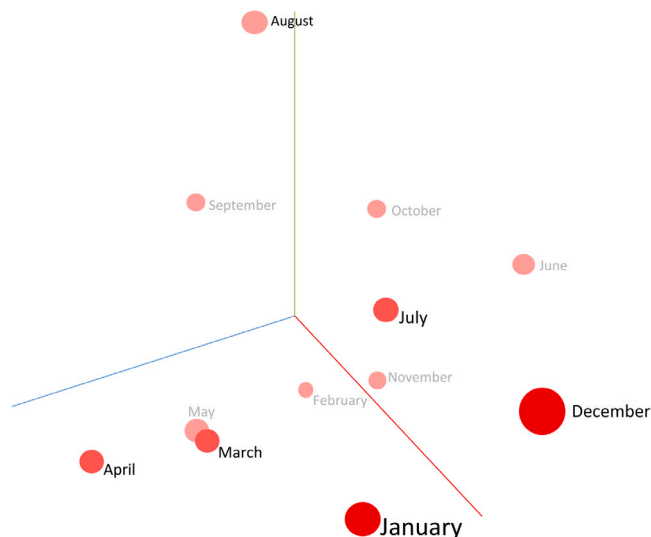


Fig. 13. Embedding vector for the months variable using Tensorflow projector.

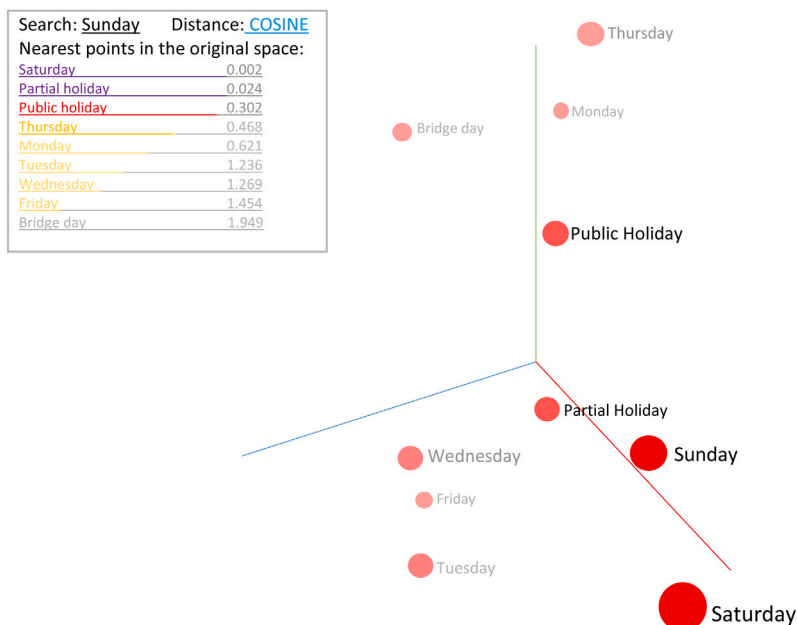


Fig. 14. Embedding vector for the weekday variable using Tensorflow projector.

approach can be used for the generation of long-term price profiles, as needed for the construction of hourly price forward curves. Therefore utilities can base both applications on the same logic, which highly reduces the operational effort. Additionally, we present tools to analyze the “black-box” neural network in order to reduce the model risk and to increase acceptance of our approach in practice.

The research questions have been answered. Calendar information is well included in neural networks using embeddings. From the embeddings we can gain economic insights and analyze which calendar variables lead to similar behavior of electricity prices. We believe that the presented approach of embedding calendar information can be used in many regression or classification tasks with a strong calendar dependency. This includes potentially all electricity markets worldwide. Many other commodities show a seasonal behavior in prices, however, such a dependency is usually not a feature in equity markets. In a further research project we also successfully applied the proposed approach to the forecasting of air quality near roads (traffic also has a strong calendar dependency due to commuting patterns). Therefore, future research could explore new applications of the calendar embedding, not only in price forecasting, but also in other domains. Our technique, however, is not well suited for applications, which require

not only solid forecasts but also deep model insights. This is a natural limitation of all models using neural networks. Further improvements of the forecasting quality of the model could be gained by adding additional fundamental factors (like system load or demand) or explicitly including a history of prices as features to account for its auto-regressive behavior.¹² There is also no evidence from our perspective that LSTM is a competitive approach in EPF, which is also stated in [Lago et al. \(2021\)](#): “Considering the most complete benchmark study in terms of forecasting models [57], it seems that a simple DNN with two layers is one of the best ML models. In particular, while more complex models, e.g. LSTMs, could potentially be more accurate, there is at the moment no sound evidence to validate this claim”. Further research may show if the proposed embedding layer and the benchmark models from [Lago et al. \(2021\)](#) can be combined.

CRediT authorship contribution statement

Andreas Wagner: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Enislay Ramentol:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Florian Schirra:** Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Hendrik Michaeli:** Data curation, Formal analysis, Methodology, Software, Writing – review & editing.

Acknowledgments

This work has partially been supported by the German Federal Ministry for Economic Affairs and Energy in grant 01186724/1 (FlexEuro: Wirtschaftliche Optimierung flexibler stromintensiver Industrieprozesse) and the German Federal Ministry of Education and Research in grant 05M18AMC (ENets: Modellierung und Steuerung zukünftiger Energienetze). The research of Enislay Ramentol has been funded by the European Research Consortium for Informatics and Mathematics (ERCIM) Alain Bensoussan Fellowship Programme and the Fraunhofer Institute for Industrial Mathematics ITWM. The authors would like to thank Tania Jacob for the valuable review of the manuscript. We also thank the three anonymous referees for their valuable comments, which significantly improved the paper.

Appendix. Detailed classification of public holidays

The embedding for calendar features builds upon the following classification of holidays:

- *public holidays*: Christmas, Day After Christmas, New Years Day, First of May (International Workers Day), Day of German Unity, Good Friday, Easter Sunday, Easter Monday, Ascension Day, Pentecost Monday
- *partial holidays*: assumption of Mary, Reformation Day, All Hallows Day, Day of Prayer and Repentance, Pentecost Sunday, the Christmas week
- *bridge days*: all days between public holidays, Fridays if Thursdays are public holidays and Mondays if Tuesdays are public holidays

References

- Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets: A review and evaluation. *Int. J. Electr. Power Energy Syst.* 31 (1), 13–22.
- Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., 2011. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J. Mult. Valued Log. Soft Comput.* 17 (2–3), 255–287.
- Amjadi, N., 2006. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Trans. Power Syst.* 21, 887–896.
- Bano, H., Tahir, A., Ali, I., Haseeb, A., Javaid, N., et al., 2019. Electricity load and price forecasting using enhanced machine learning techniques. In: *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. Springer, pp. 255–267.
- Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C., 2003. A neural probabilistic language model. *J. Mach. Learn. Res.* 3, 1137–1155.
- Benth, F.E., Koekebakker, S., Ollmar, F., 2007. Extracting and applying smooth forward curves from average-based commodity contracts with seasonal variation. *J. Deriv.* 15, 52–66.
- Bian, J., Gao, B., Liu, T.-Y., 2014. Knowledge-powered deep learning for word embedding. In: *Proceedings of the 2014th European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part I*. In: ECMLPKDD’14, Springer-Verlag, Berlin, Heidelberg, pp. 132–148.
- Blöchliger, L., 2008. Power Prices – A Regime-Switching Spot/Forward Price Model with Kim Filter Estimation (Ph.D. thesis). University of St. Gallen.
- Brusaferri, A., Matteucci, M., Portolani, P., Vitali, A., 2019. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Appl. Energy* 250, 1158–1175.
- Caldana, R., Fusai, G., Roncoroni, A., 2017. Electricity forward curves with thin granularity: Theory and empirical evidence in the hourly exepspot market. *European J. Oper. Res.* 261, 715–734.
- Chen, X., Dong, Z.Y., Meng, K., Xu, Y., Wong, K.P., Ngan, H., 2012. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Trans. Power Syst.* 27, 2055–2062.
- Chen, C., Song, M., 2017. *Representing Scientific Knowledge: The Role of Uncertainty*. Springer.
- Cludius, J., Hermann, H., Matthes, F.C., Graichen, V., 2014. The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications. *Energy Econ.* 44, 302–313.

¹² We thank the anonymous referee for her comment.

- Conejo, A.J., Plazas, M.A., Espinola, R., Molina, A.B., 2005. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Trans. Power Syst.* 20 (2), 1035–1042.
- Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* 7, 1–30.
- Deng, L., Liu, Y., 2018. *Deep Learning in Natural Language Processing*. Springer.
- Diebold, F.X., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *J. Bus. Econom. Statist.* 33 (1), 1–9.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *J. Bus. Econom. Statist.* 13 (3), 253–263.
- Eckhardt, K., 2018. Choosing the right hyperparameters for a simple LSTM using keras. URL: <https://towardsdatascience.com/choosing-the-right-hyperparameters-for-a-simple-lstm-using-keras-f8e9ed76f046>.
- Fleten, S.-E., Lemming, J., 2003. Constructing forward price curves in electricity markets. *Energy Econ.* 25, 409–424.
- Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Assoc.* 32 (200), 675–701.
- Géron, A., 2019. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media.
- Goodman, B., Flaxman, S., 2016. Eu regulations on algorithmic decision-making and a “right to explanation”. In: *ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*. New York.
- Gu, W., Tandon, A., Ahn, Y.-Y., Radicchi, F., 2021. Principled approach to the selection of the embedding dimension of networks. *Nature Commun.* 12, 3772.
- Hinderks, W., Wagner, A., 2020. Factor models in the german electricity market: Stylized facts, seasonality, and calibration. *Energy Econ.* 85, 104351.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Holm, S., 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- Hopfield, J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79, 2554–2558.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 2261–2269.
- Jiang, L., Hu, G., 2018. Day-ahead price forecasting for electricity market using long-short term memory recurrent neural network. In: *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, pp. 949–954.
- Kapoor, A., Guli, A., Pal, S., 2019. *Deep Learning with TensorFlow 2 and Keras: Regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API*, second ed. Packt Publishing.
- Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity prices applying artificial neural networks. *Appl. Energy* 162, 218–230.
- Khabiri, E., Gifford, W.M., Vinzamuri, B., Patel, D., Mazzoleni, P., 2019. Industry specific word embedding and its application in log classification. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. In: *CIKM '19*, Association for Computing Machinery, New York, NY, USA, pp. 2713–2721.
- Kiesel, R., Paraschiv, F., Sætherø, A., 2019. On the construction of hourly price forward curves for electricity prices. *Comput. Manag. Sci.* 16 (1), 345–369.
- Kuo, P.-H., Huang, C.-J., 2018. An electricity price forecasting model by hybrid structured deep neural networks. *Sustainability* 10 (4), 1280.
- Lago, J., De Ridder, F., De Schutter, B., 2018. Erratum to “forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms”. *Appl. Energy* 229, 386–405.
- Lago, J., Marcjasz, G., De Schutter, B., Weron, R., 2021. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Appl. Energy* 293, 116983.
- Marcjasz, G., Uniejewski, B., Weron, R., 2019. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. *Int. J. Forecast.* 35 (4), 1520–1532.
- McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–133.
- Mosbah, H., El-Hawary, M.E., 2016. Hourly electricity price forecasting for the next month using multilayer neural network. *Canad. J. Electr. Comput. Eng.* 39, 283–291.
- Narajewski, M., Ziel, F., 2020. Econometric modelling and forecasting of intraday electricity prices. *J. Commod. Mark.* 19, 100107.
- Nogales, F., Contreras, J., J. Conejo, A., Espinola, R., 2002. Forecasting next-day electricity prices by time series models. *Power Eng. Rev. IEEE* 22, 58.
- Qureshi, M.A., Greene, D., 2017. Eve: Explainable vector based embedding technique using wikipedia. *J. Intell. Inf. Syst.* 53, 137–165.
- Sætherø, A.S., 2018. *Hourly Price Forward Curves for Electricity Markets* (Ph.D. thesis). Universität Duisburg Essen.
- Schnürch, S., Wagner, A., 2020. Electricity price forecasting with neural networks on EPEX order books. *Appl. Math. Finance* 27 (3), 189–206.
- Troncoso, A., Santos, J., Gomez-Exposito, A., Martinez-Ramos, J., Riquelme, J., 2007. Electricity market price forecasting based on weighted nearest neighbors techniques. *IEEE Trans. Power Syst.* 22, 1294–1301.
- Ugurlu, U., Oksuz, I., Ta, O., 2018. Electricity price forecasting using recurrent neural networks. *Energies* 11, 1255.
- Wagner, A., 2014. Residual demand modeling and application to electricity pricing. *Energy J.* 35 (2), 45–73.
- Wendlandt, L., Kummerfeld, J.K., Mihalcea, R., 2018. Factors influencing the surprising instability of word embeddings. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics.
- Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* 30 (4), 1030–1081.
- Weron, R., Misiorek, A., et al., 2005. Forecasting spot electricity prices with time series models. In: *Proceedings of the European Electricity Market EEM-05 Conference*, pp. 133–141.
- Würzburg, K., Labandeira, X., Linares, P., 2013. Renewable generation and electricity prices: Taking stock and new evidence for Germany and Austria. *Energy Econ.* 40, 159–171.
- Yin, Z., Shen, Y., 2018. On the dimensionality of word embedding. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. In: *NIPS'18*, Curran Associates Inc., Red Hook, NY, USA, pp. 895–906.
- Zhu, Y., Dai, R., Liu, G., Wang, Z., Lu, S., 2018. Power market price forecasting via deep learning. In: *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*. pp. 4935–4939.
- Ziel, F., Steinert, R., 2016. Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Econ.* 59, 435–454.
- Ziel, F., Weron, R., 2018. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Econ.* 70, 396–420.