Energy and AI 13 (2023) 100239

Contents lists available at ScienceDirect

Energy and AI

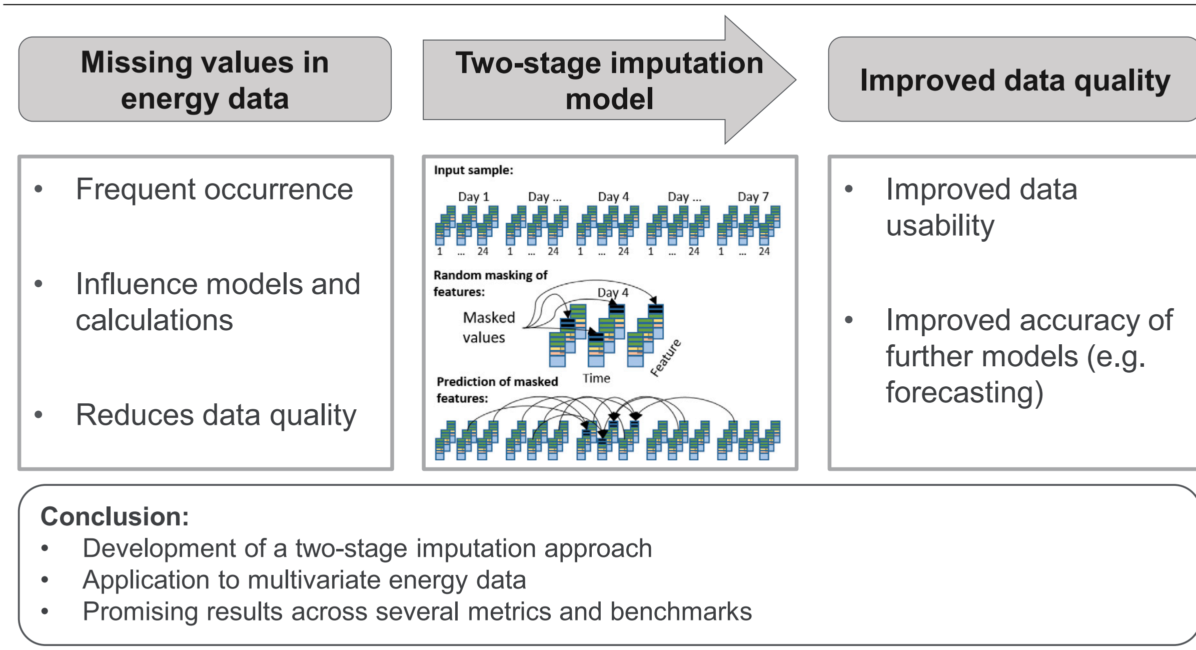
journal homepage: www.elsevier.com/locate/egyai

Multivariate time series imputation for energy data using neural networks

Christopher Bülte [a,*], Max Kleinebrahm [a], Hasan Ümitcan Yilmaz [a], Juan Gómez-Romero [b]

[a] Institute for Industrial Production (IIP), Chair of Energy Economics, Karlsruhe Institute of Technology (KIT), Germany
[b] Department of Computer Science and Artificial Intelligence, University of Granada, Spain

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Development of a two-stage imputation approach for multivariate energy time series.
- Estimation of missing value distribution using a recurrent neural network.
- Imputation of missing values based on the estimated distribution of missing values.
- Promising results for several evaluation metrics and benchmark methods.

ARTICLE INFO

Keywords:
Missing value estimation
Multivariate time series

ABSTRACT

Multivariate time series with missing values are common in a wide range of applications, including energy data. Existing imputation methods often fail to focus on the temporal dynamics and the cross-dimensional correlation simultaneously. In this paper we propose a two-step method based on an attention model to impute missing

* Corresponding author.
E-mail address: christopher.buelte@web.de (C. Bülte).

https://doi.org/10.1016/j.egyai.2023.100239
Received 16 October 2022; Received in revised form 6 February 2023; Accepted 8 February 2023
Available online 10 February 2023

Neural networks
Attention model
Energy data

values in multivariate energy time series. First, the underlying distribution of the missing values in the data is learned. This information is then further used to train an attention based imputation model. By learning the distribution prior to the imputation process, the model can respond flexibly to the specific characteristics of the underlying data. The developed model is applied to European energy data, obtained from the European Network of Transmission System Operators for Electricity. Using different evaluation metrics and benchmarks, the conducted experiments show that the proposed model is preferable to the benchmarks and is able to accurately impute missing values.

1. Introduction

Multivariate time series occur through a wide range of applications, such as energy, finance or meteorology. In the context of energy data, multivariate time series can be energy demand or supply, energy stock market prices, energy storage levels or many more. Missing values are very common in this type of data, due to communication or measurement errors and can have a large impact on further downstream usage of the time series in energy system analysis applications. A variety of methods have been developed for the purpose of imputing missing values in multivariate time series. However, classical imputation methods usually only focus either on the temporal dynamics of the time series or the cross-dimensional correlations, rarely both. On the other hand there has been an increase in the usage of deep learning methods for time series analysis and in particular time series imputation. While these methods are usually based on sequence-to-sequence networks [1–3], other approaches have been transferred to the task of imputing, including adversarial networks [4], autoencoders [5] and attention based models [6]. This paper proposes a two-step imputation method that is centered around an attention based neural network. The approach consists of a first stage model that learns the multivariate distribution of the missing values and a second stage model that estimates the actual missing data. Since the distribution of missing values is learned prior to the imputation, it can be used to generate artificial missing values for the training of the imputation model, making it flexible to learn the specific characteristics of the data. For readability, our approach is referred to as MVTSI (multivariate time series imputation) throughout the article. While the presented model is applicable to any type of (multivariate) time series, the focus is on an application to energy data obtained from the ENTSO-E transparency platform. This article includes the following contributions:

- We design a LSTM-based model that learns the distribution of the missing values in the multivariate time series data. After training, the model can be used to generate synthetic missing values that follow the distribution of the data. There are no restrictions or assumptions regarding the type of the data or the temporal dynamics.
- We implement an attention based estimation model that uses the predicted distribution from the LSTM model for its own training process. The interaction between both models allows for flexible adaptation to the input data. The imputation model can learn temporal and cross-dimensional correlations at the same time and estimates all missing values for a given timeframe simultaneously.
- We design and evaluate our model specifically for the purpose of imputing missing values in the ENTSO-E data. We carry out extensive analysis on the performance of our model in comparison to several benchmarks and on different subsets of the data. Our model shows favorable results in comparison to the benchmark methods.

The paper is structured as follows. Section 2 starts with a literature review, followed by an overview of the ENTSO-E data in Section 3. Section 4 introduces the setup and methodology of the proposed model, while Section 5 includes the experiments, as well as results. Section 6 concludes with a discussion.

2. Related work

This section contains a review of related work used in this article. Section 2.1 provides a brief overview of the literature on the importance and effect of missing values in energy data. Section 2.2 concludes with a review of different methods for time series analysis.

2.1. Importance of missing values in energy data

While missing data occurs in nearly every field of study, it is especially important in the context of energy data, since most energy system models require complete data. These models cover a huge variety of tasks, for example proposing a future energy system, obtaining measures to reduce CO_2 or to estimate investment expenditures. Ruggles et al. [7] show that after filling the gaps only in electricity demand time series, the results of a power system model can vary by 5% between using two sophisticated data imputation approaches, even for a very simple analysis considering only one region. Since energy system models cannot operate with missing data, some choice of imputing has to be made. Besides, the paper also shows that the results from simple data imputation methods are generally not of sufficient quality. These insufficiency is highlighted for models that are planning power generation capacity and production for the long term. Since security of supply must be ensured and electricity demand must be met at all times, differences in estimated missing values can lead to different planning strategies. However, even for short-term forecasting models this can pose a problem. In the case of wind farms, short-term planning is important to balance electricity supply and demand [8]. Since electricity demand and supply must be in balance at all times, incorrect assumptions and estimates can lead to instabilities in the electricity grid and require additional grid stabilization measures. In such cases, operators of renewable power generation plants pay penalties due to their incorrect estimates. For such estimates, the statistical data forecasting methods are usually preferred [9]. These methods however, require recent observations of power and wind speed from potentially several wind farms. These observations are prone to data loss, due to reasons like maintenance, communication errors or delays. For example, Akçay and Filik [10] use a dataset providing wind speed measures for five different stations in Turkey and report an average of 2.17% values missing. Two frequently used data sources for energy system analysis are the PJM data [11] in the US and the transparency platform ENTSO-E [12] in Europe. These datasets contain time series with information about the national electricity demand, prices, as well as the energy production of all available technologies. In this article we will focus on the latter. Compiling these measurements for a very large number of electricity production units, including small generation such as rooftop solar plants, is difficult and produces time series with many missing values. For instance, in the ENTSO-E transparency platform an average of 1000 values per week are missing in the production mix data for 2015 and 2016 [13]. As mentioned before, this can have a huge impact on energy models that require (sufficiently) complete data. In order to deal with this problem, it is necessary to impute the data in some way.

2.2. Models for time series prediction and imputing

A vast amount of literature focuses on the problem of imputing missing data in time series. Simple "naive" methods include mean imputing, interpolation or last observation carried forward. These methods however, fail to incorporate more complex time dependencies. Autoregressive methods, for example ARIMA or GARCH models remove the trend from the time series and estimate the relationships between the individual timesteps. State space models further extend the use of ARIMA with Kalman filters [14]. Approaches centered around k-nearest-neighbor estimation [15] aim to replace the missing value with a value of the most similar neighborhood. These methods have proven to perform quite well in time series imputing and can be applied in different areas [16,17].

However, the advances in machine and deep learning have led to an increasing number of such methods used for imputation. A frequent approach is using sequence models like recurrent neural networks (RNNs or LSTMs), to impute time series data [1–3]. The approach described by Zhang et al. [1] proposes a deep neural network based on bi-directional LSTMs to recover missing values in sensor data. Applying the model to water quality data of the Great Barrier Reef led to an improvement of up to 70% on evaluation metrics against chosen benchmark models. Cao et al. [18] propose BRITS, a self-supervised algorithm based on a network of bidirectional LSTMs. Their approach can be used for univariate time series imputing, as well as for multivariate time series with correlated features. The study compares the method against various baselines, including RNN-based methods, and achieve an improvement in accuracy over three different independent datasets. Further approaches exist, making use of different state-of-the-art neural network architectures. Generative Adversarial Networks (GANs) can be used efficiently to learn the distribution of a multivariate time series [4]. This can then be further used to estimate the missing values in the data. The method GRUI, as proposed by Luo et al. [4], shows an improved accuracy against chosen baselines on different time series datasets. Fortuin et al. [5] use a combination of variational autoencoders and Gaussian processes to estimate the distribution of missing values with a focus on the application in medical data. The study compares the model on different datasets, including medical time series data. The performance on different tasks (imputing, classification after imputing) is significantly higher than most baselines methods and similar to state-of-the-art methods like BRITS [18].

Some recent approaches focus on using an attention based architecture to learn dependencies in multivariate time series. The self-attention mechanism introduced by Vaswani et al. [19] has become the state-of-the-art in several different fields of machine learning, mainly in natural language processing. However, the approach itself can also be applied to computer vision or sequential modeling. Zerveas et al. [20] use an attention framework for time series regression and classification. Their model consists of an attention based neural network encoder that is adapted to dealing with multivariate time series. The study shows that the model is able to obtain a higher performance than the respective baselines, both in regression and classification tasks. However, it is also mentioned that the approach is transferable to forecasting or imputation. Ma et al. [6] propose CDSA, a cross-dimensional self-attention architecture with application to multivariate geo-tagged time series. They propose to use a decomposition of the attention map of the model architecture in order to incorporate cross-dimensional dependencies while keeping the computational costs low. The model shows a lower imputation error than chosen baseline metrics over several geo-tagged time series datasets. Table 1 shows a summary of the review literature on time series imputing. While many approaches are available, we are not aware of any deep learning methods focusing specifically on the area of energy time series.

Table 1
An overview of relevant models and literature.

Method	Study	Approach	Application
KNN	[15]	K-nearest-neighbor	Power data
SSIM	[1]	LSTM	Sensor data
BRITS	[18]	LSTM	Medical data, sensor data
GRUI	[4]	GAN	Medical data, sensor data
GP-VAE	[5]	Autoencoder	Medical data
CDSA	[6]	Attention based	Sensor data
MVTSI	This study	Attention based	Energy data

3. Data

This section describes the data used in the experiments in more detail. The data originates from the ENTSO-E transparency platform.[1] ENTSO-E, the European Network of Transmission System Operators for Electricity, is the association for the cooperation of different European transmission system operators (TSOs). The association includes 39 TSOs from 35 different countries. Details about ENTSO-E, for example the mission statement and the individual members are accessible on the project's website.[2] The ENTSO-E transparency platform offers a way to easily access data on the European electricity system, via an API. A systematic review of the platform would extend the scope of this article and can be found in Hirth et al. [13]. However a short analysis of the data quality is given below with a special focus on the distribution of missing values. For that purpose this section will briefly cover the study by Hirth et al. [13], which provides an in-depth analysis of the data quality of the ENTSO-E data. However, since it only refers to the years 2015–2016, an additional statistical analysis is included.

Hirth et al. [13] provide an extensive study about the ENTSO-E transparency platform, considering its ambitions, characteristics, methodology as well as data quality and usability. Since for the presented model the main focus is on data quality, further information about the platform is not discussed here. Hirth et al. [13] use data retrieved in April 2017. This means, the more recent data that is used in this study might differ, due to adjustments. It is shown that the amount of missing values heavily depends on the country, year and the covered variable. For example in Denmark there is a 100% coverage of all variables, except for "Hydro Run-of-river and poundage", where 32.4% of values are missing. Similarly, Italy only has a reported coverage of around 50%, since one year is missing in most features. The Completeness of "Aggregated Generation per Type" table from Hirth et al. [13] is replicated using all available data from 2015–2020 and can be found in Fig. A.8 in the Appendix. However, for further data analysis the focus is not on the completeness of the raw ENTSO-E data, but on the data as it is processed for the model presented in this article.

The used data includes 29 countries provided by ENTSO-E. These countries contain the EU27 excluding Cyprus and Malta, but including Switzerland, Serbia, Norway, and the United Kingdom. However, the countries Luxembourg and Germany are added together, as they form a bidding zone on the electricity market. The data covers all years from 2015 to 2020 (inclusive) and was retrieved using the public API on 09.06.2021. Furthermore, several features were aggregated. The reason being that the model is evaluated only on a reduced subset of the data, but also because a lot of the features do not carry important information. For example the feature "Fossil Peat" does only exist for three out of 29 countries and is therefore removed. The remaining features were aggregated in the following way: The features "Hydro Run-of-river and poundage" and "Hydro Water Reservoir" were summed up into "Hydro". The features "Fossil Peat", "Other renewable", "Waste", "Fossil Oil", "Fossil Coal-derived gas", "Geothermal" were added on top of "Other". Missing values were carried through the aggregation

[1] https://transparency.entsoe.eu/

[2] https://www.entsoe.eu/about/

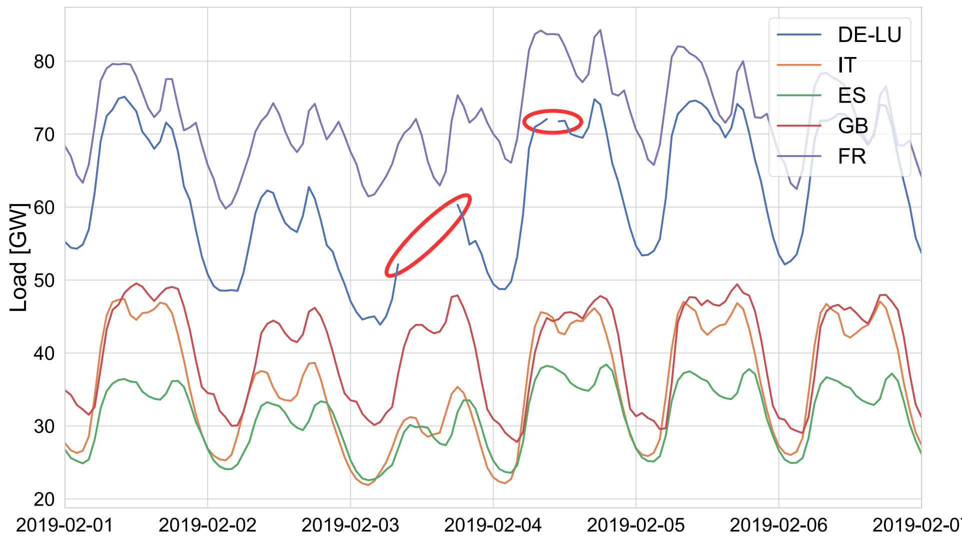


Fig. 1. Weekly load profile for five different countries. During the weeks there are two gaps of different length with missing values for the bidding zone DE-LU (Germany & Luxembourg), indicated by the red ellipses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

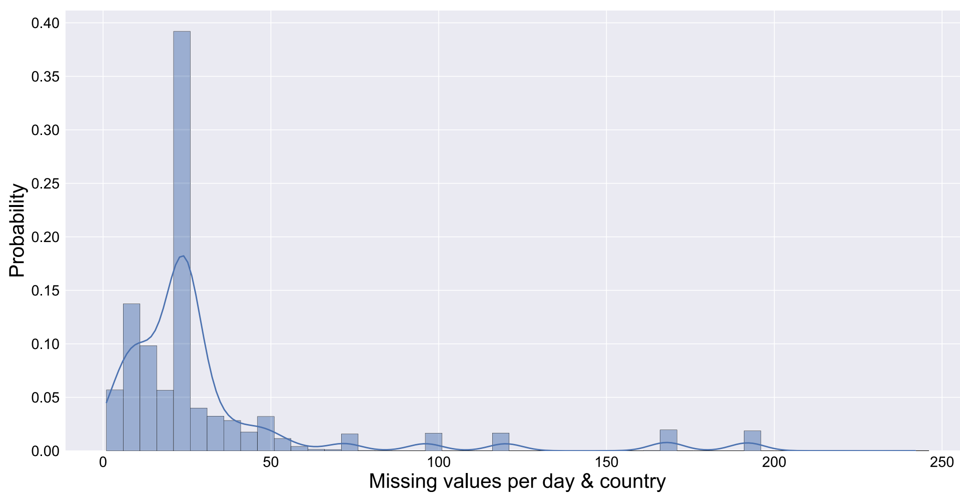


Fig. 2. Histogram of the amount of missing values per day and country, including only incomplete days.

Table 2
Summary statistics of different aggregation schemes for the missing values.

Aggregation method	Mean	Min	25%	50%	75%	Max
Per day & country[a]	10.33	0	0	0	13	242
Per day & country[b]	31.51	1	13	24	27	242
Per day & feature[a]	21.58	0	0	16	45	150
Per day & feature[b]	34.23	1	24	28	49	150
Per timestamp[a]	11.69	3	7	10	15	72
Per timestamp[b]	11.69	3	7	10	15	72
Gap length[b]	21.36	1	1	3	6	46222

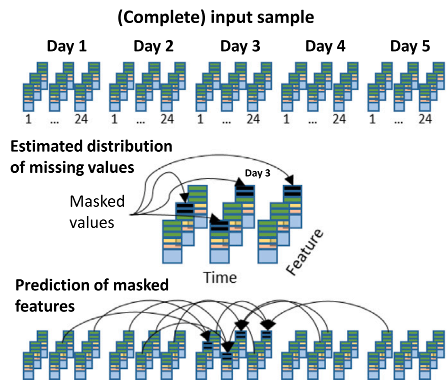
[a]Refers to complete data.
[b]Only includes data with missing values in the specific aggregation.

process, meaning that if one feature was missing, the summed feature is also declared as missing, since the actual value cannot be determined. This process resulted in a total of $n = 13$ features. More information on the processing step can be found in the Appendix. Fig. 1 shows an example of the used data.

A closer look on the distribution and characteristics of the missing values in the data is given in the following. Table 2 shows summary statistics for different aggregation levels of the missing values. For example, the amount of missing values during one day per country is analyzed, which shows that on average 10.33 values are missing. However, for more than 50% of the days the data is complete, as can be seen in Table 2. Since the goal is to impute the missing values later on, it is more interesting to focus on the amount of daily missing values, when data is actually missing. Because in any day that has no missing value nothing needs to be imputed. Summary statistics for this data can also be found in Table 2. Fig. 2 shows a histogram for the amount of missing values per day and country, only including incomplete days. It is noticeable that if data is missing during one day, it is most likely that the amount of missing values is 24. This is probably due to a whole feature missing for the day. Furthermore, we analyze the length of consecutive missing values (gap length). While on average about 21 consecutive timesteps are missing, this value is highly distorted, since for some features almost all of the data from one year is missing. Table 2 shows that the 75%-quantile of the gap length is 6, implying that 25% of the gaps are made up of 6 consecutive missing values or more. Our analysis of the missing values shows that the general amount of missing data over a specific time period (day, timestamp) is quite high, if we exclude complete data. Due to consecutive missing values across different dimensions, univariate and simple statistical methods are not suitable to solve this kind of imputation problem, hence the proposed method of this paper.

Year	Country	Timestamp	S_1	S_2	...	S_n
		T_1	.	.	...	.
	C_1	T_2	.	.	...	.
		...	...	...	...	...
		T_{8760}	.	.	...	.
Y_1		T_1	.	.	...	.
	C_2	T_2	.	.	...	.
		...	...	...	...	...
		T_{8760}	.	.	...	.
	...	...	...	...	...	...

(a) Demonstration of the input data.



(b) Example of the masking setup. Missing values are imputed for the middle day of five consecutive days.

Fig. 3. The left figure shows the raw multivariate time series, as obtained by the ENTSO-E data. Additional features, like the masking vector, are added before the data is fed into the model. The right figure shows an imputing example for a complete input sample. However, dealing with additional missing values across the time domain is not a problem for the imputation model.

It is worth noting at this point that we assume the data to be missing completely at random or missing at random. The first would suggest that missing data exists just because of random distortions. In the context of this study this seems to be a strong assumption, since a lot of missing values are time dependent, e.g. if exactly 24 hours are missing. Missing at random would imply that the fact that specific data is missing might be due to measured characteristics. For example, the probability that data is missing for one country might be higher than for another country. While missing completely at random might be an extreme assumption, missing at random should be justifiable for the proposed model, as the information that can influence the probability of missing values is incorporated in the model. However, the data could also be missing not at random. This would suggest that there is additional information, not reflected in the data, that determines the missingness of data. In that case the model estimations can potentially be biased.

4. Methodology

This section explains the proposed approach in detail and includes the methodology behind the two-stage model. Fig. 3(a) shows the structure of the processed data used as input to the model. Let $C = \{C_1, \dots, C_m\}$ be a set of m countries (e.g., "France", "Germany") and $S = \{S_1, \dots, S_n\}$ a set of n time series types (e.g., "electricity production from solar", "electricity prices"). $Y = \{Y_1, \dots, Y_l\}$ represents the years and $T = \{T_1, \dots, T_{8760}\}$ an hourly timestamp of the year, including information about the day, week, month and hour.

For a specific country the features can be expressed as a multivariate time series $\mathbf{S} = \{\mathbf{s}_1 \dots \mathbf{s}_T\}$ with T observations. Any observation from a timestamp t consists of N features, $\mathbf{s}_t = \{s_t^1, \dots, s_t^N\} \in \mathbb{R}^N$. In order to model the missing data, a masking vector $\mathbf{x}_t$, that includes the information, whether a feature is missing or not, is used. It is defined in the following way:

$$\mathbf{x}_t^n = \begin{cases} 1 & \text{if } s_t^n \text{ is missing} \\ 0 & \text{otherwise} \end{cases}$$

Similar to before, this can be interpreted as a multivariate time series $\mathbf{X} = \{\mathbf{x}_1 \dots \mathbf{x}_T\}$. In order to take the interdependencies of the time series into account, the input data is transformed into a specific format. For each country, a multivariate time series is extracted using a sliding-window approach, with a 24 hour shift. Therefore, each sample includes a window of five days over all features. The MVTSI approach aims at imputing the missing values in the middle day, as can be seen in Fig. 3(b). This approach has the advantage that the model is able to directly incorporate the data temporarily close to the missing values. The further relations are learned implicitly from the data. Since the model receives the masking vector as an input it is flexible in dealing with additional missing values across all days and features. In principle, different window sizes can be chosen for the approach. However, since an attention architecture scales quadratically in memory, large window sizes can lead to computational infeasibility, depending on the available memory. In addition, the window size can be treated as one of many parameters of the model that could potentially be optimized. Since the approach already shows promising results without extensive parameter tuning, the window size is kept fixed as five days.

The proposed approach to imputing missing values consists of the steps that are depicted in Fig. 4. After the data is prepared and transformed using the sliding window, in the first stage, a LSTM based neural network is implemented to learn the distribution of missing values in the data. This model is then used to create artificial missing values in the data, that are similar to the original distribution. Because the imputation cannot be evaluated on actual missing data, since the true value is unknown, this step is necessary to calculate the loss and actually train the imputation model in the second step. This model is then trained using the generated missing values and therefore learns to predict missing values in the way they actually occur in the underlying data. For the inference, in order to predict the missing values, only the attention model is needed. Although this approach incorporates two specific model architectures, they could in principle be exchanged. One could for example use the first stage of the model in the first step and train any suitable neural network as described above, or vice versa. In the following the two mentioned models are described in more detail.

4.1. Modeling of missing data distribution

The first step of the modeling approach is to estimate the distribution of missing values in the data that is to be used. In order to estimate the distribution, a LSTM-based sequence model is used. Fig. 5(a) shows how data in the form of a time series is processed in this model. First a start-of-sequence (SOS) token is added to tell the model to predict the first value of the time series. The model then sequentially predicts the next value in the time series. In the training process teacher-forcing is used, which means that after each prediction, the model uses the true value for the next prediction. This approach can significantly reduce the training time of a neural network [21]. In the inference process, the whole time series is estimated using the step-wise predictions of

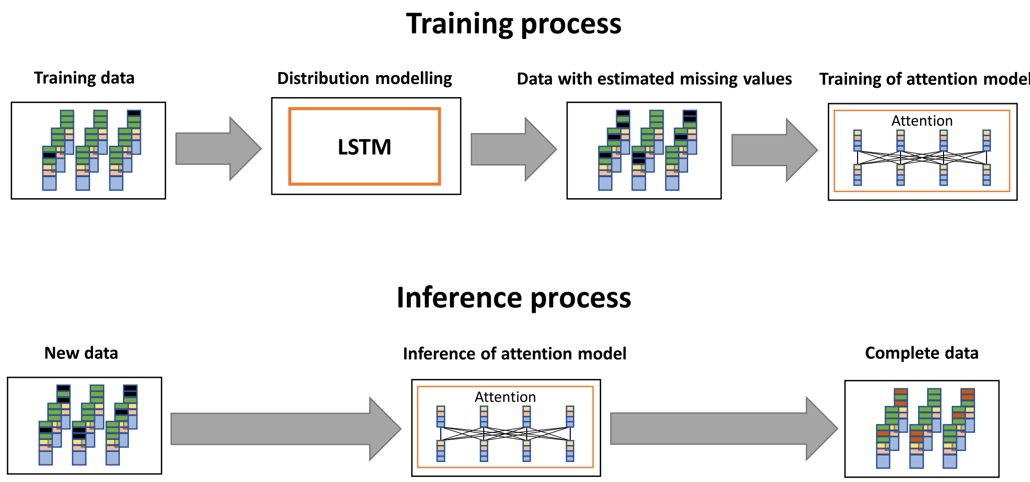
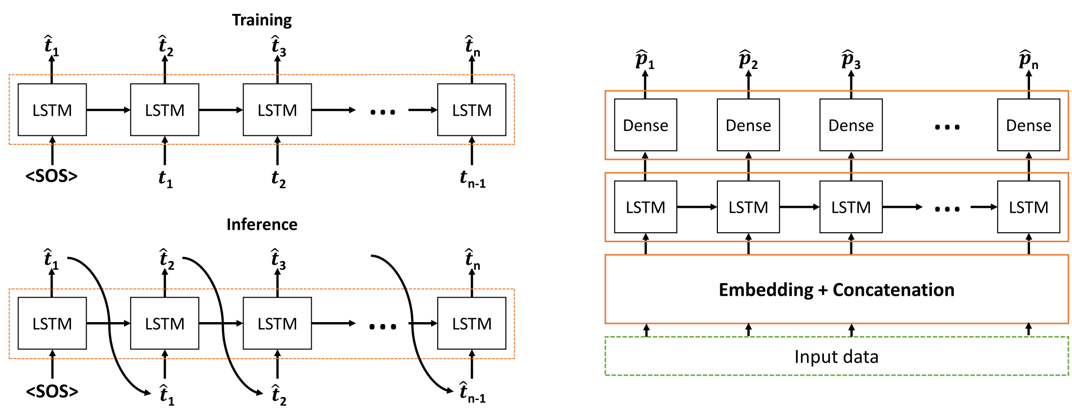


Fig. 4. An abstract scheme for the model workflow. In the training process, missing data are generated artificially by the first stage model and then further used to train the attention model. For performing inference on unseen data, only the second model is used to impute the actual missing values.



(a) An abstract scheme for the sequence processing. (b) The LSTM model structure.

Fig. 5. The LSTM model for estimating the distribution of missing data. In the training process the true data is used as an input for the next step, while in the inference process the predictions are fed back into the model. The model consists of several embedding and concatenation layers, followed by the bi-directional LSTM units. The final output is the probability of each individual value being missing.

the model. Fig. 5(b) shows the basic structure of the LSTM sequence model. Instead of using the whole time series, subsets of 24 h are used as input to the model, since the aim is to estimate missing values in this specific time horizon. The input time series is concatenated with feature embeddings, like positional encoding or country embeddings. For the MVTSI approach the time series is a binary vector, containing the information whether a specific datapoint is missing or not. As a start token the value −1 is used. In the model, the data is first fed through one (or several) LSTM layers and afterwards a final Dense layer with a sigmoid activation function:

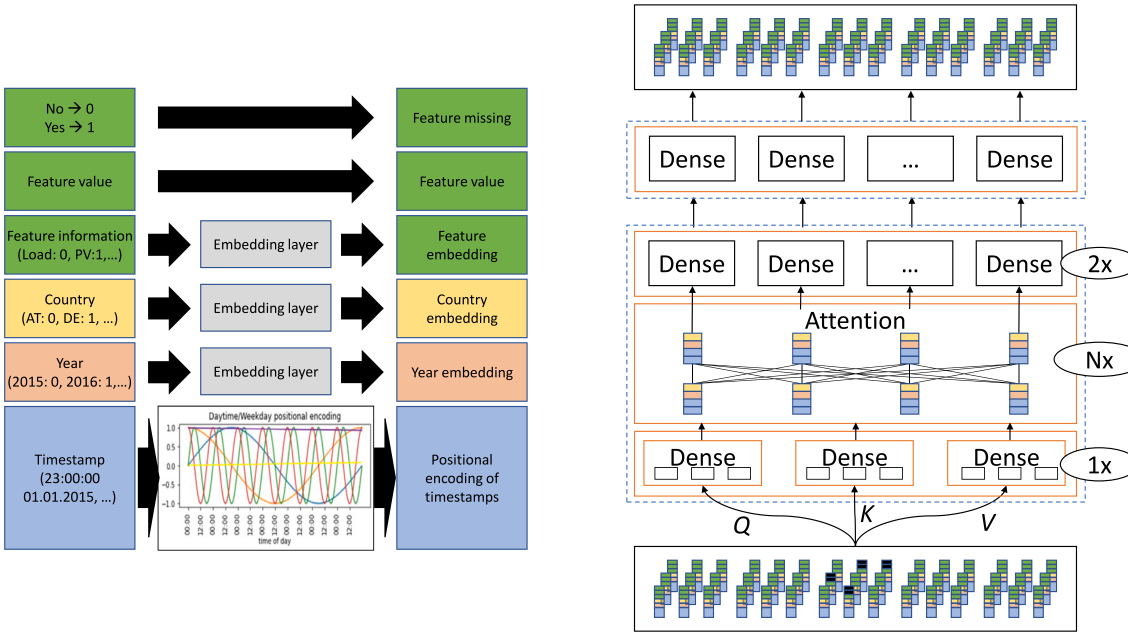
$$S(x) = \frac{1}{1 + e^{-x}}.$$

Since $S(x) \in [0, 1]$ the output of the model can be interpreted as the probability that the next value is missing. In the training process the binary cross entropy loss is minimized using teacher-forcing, as mentioned before. For the inference process, the model uses its own step-wise predictions, to predict the whole time series. However, the probability predictions of the model are converted into a binary vector using a Bernoulli random variable with the predicted probability.

4.2. Attention model for imputation

This section describes the second stage of the proposed approach, the attention based imputation model. Attention based models do not process time series sequentially, but learn the dependencies between all time steps across all dimensions at the same time. To simplify the learning process for the model, positional encodings are added to the data, carrying information about the relative position in the time series. The model also incorporates categorical embeddings for the year, the country and the feature, as can be seen in Fig. 6(a). Embedding layers can learn multidimensional representations of the discrete input data.

The actual attention mechanism mainly consists of the matrices $Q, K, V \in \mathbb{R}^{T,d}$ (query, key, value) and the attention mask $M \in \mathbb{R}^{T,T}$. In our case, T is the dimension of the stacked inputs, as described in Fig. 6(a) and d is a hyperparameter called the embedding dimension. Since our proposed approach incorporates self-attention, the matrices Q, K, V are equal. Furthermore, the causal attention mask is not needed, because an imputation approach is used and the model is "allowed" to see all data. In forecasting for example, the mask M would ensure, that no future values are incorporated in the prediction.



(a) Visualization of the first input processing layers. (b) The structure of the attention model.

Fig. 6. The proposed attention model. Several embedding layers are used to implement categorical features, for example the country, into the model. The data is fed through the attention layers and the model learns to impute missing values in the middle day of the time window. In the setting of self-attention, the three matrices are equal and of dimension $Q, K, V \in \mathbb{R}^{T,d}$, where T is the input dimension and d is the embedding dimension.

The attention is then calculated as $\mathbf{A}(Q,K,V) = \text{softmax}(\frac{QK^\top}{\sqrt{d}})V$. This is referred to as scaled dot-product attention [19]. One can interpret this mechanism in the following way. Using the softmax activation, the model can learn to focus its "attention" on specific parts of the data, in dependence of the input. In the energy context for example, the model could learn not to put attention on the feature Nuclear in Denmark, since the values are always zero (Denmark does not produce nuclear energy). Fig. 6(b) shows the structure of the proposed attention model. As mentioned before, the input data is created using a sliding-window approach of 5 days, shifted by one day, including all features and countries. In each sample the previously described first stage model is used to estimate the distribution of missing values in the middle day. Using this artificial data, the attention model is trained to impute these missing values, using the mean squared error.

It is worth noting, that the presented model is used for deterministic forecasting of missing values and does not consider the uncertainty in the predictions. However, to change into a more probabilistic forecasting setting can be very helpful in order to quantify the uncertainty that goes along with the predictions. Several approaches are available specifically for turning deterministic into probabilistic forecasts. A simple way is the Single Gaussian technique [22] that uses the model output to create a predictive density based on the normal distribution. Lakshminarayanan et al. [23] propose deep ensembles, an algorithm that trains several ensembles of a neural network by using a proper scoring rule [24] as a training criterion. Gal and Ghahramani [25] suggest to use dropout in the training process of neural networks to represent the model uncertainty. These are just examples of several approaches available, tailored specifically to deep neural networks. However, since most approaches are independent of the specific network architecture, this article focuses on deterministic forecasting and the uncertainty prediction is left open as possible future research.

4.3. Implementation details

Each sample $\mathbf{s}_t$ is normalized by subtracting the mean and dividing by the standard deviation of the training data. Every missing datapoint is set to zero, since the neural network itself cannot handle non encoded missing values. Both model stages are trained using the Adam optimizer [26], minimizing the mean squared error (MSE). A batch size of 32 and a custom learning rate scheduler, that includes early stopping is used. For that purpose the training data is split into a training and validation set. The model will stop training, if the validation loss has not improved for a set number of steps. The model is implemented in python and tensorflow and was trained on different GPUs using Google Colab. The aggregated runtime of all experiments did not exceed ~ 3 hours. The accompanying code can be found on github.[3]

5. Results

The following section describes the experimental setup, as well as the model evaluation process. The evaluation process is split into two parts, since the two model stages are evaluated separately. First, the evaluation focuses on the first stage model and its ability to estimate the distribution of missing values. Afterwards a detailed analysis of the performance of the attention model is provided across different metrics and in comparison to several benchmarks.

5.1. Evaluation of distribution modeling

In order to evaluate the first stage model, it is helpful to take a look at what the model is actually supposed to estimate. The aim is not to overfit on the missing data, since this would lead to an estimator that learns the exact location of missing values. Instead, the goal is to obtain a model that creates a distribution that is in general similar to the distribution of missing values. Due to the stochastic nature of the model, the created missing values can be different for each iteration, but should generally reflect the properties of the original data. Since the model estimates a different distribution for every time series, it is not practical to compare every single estimated distribution to its original

[3] https://github.com/cbuelt/transformer-imputing

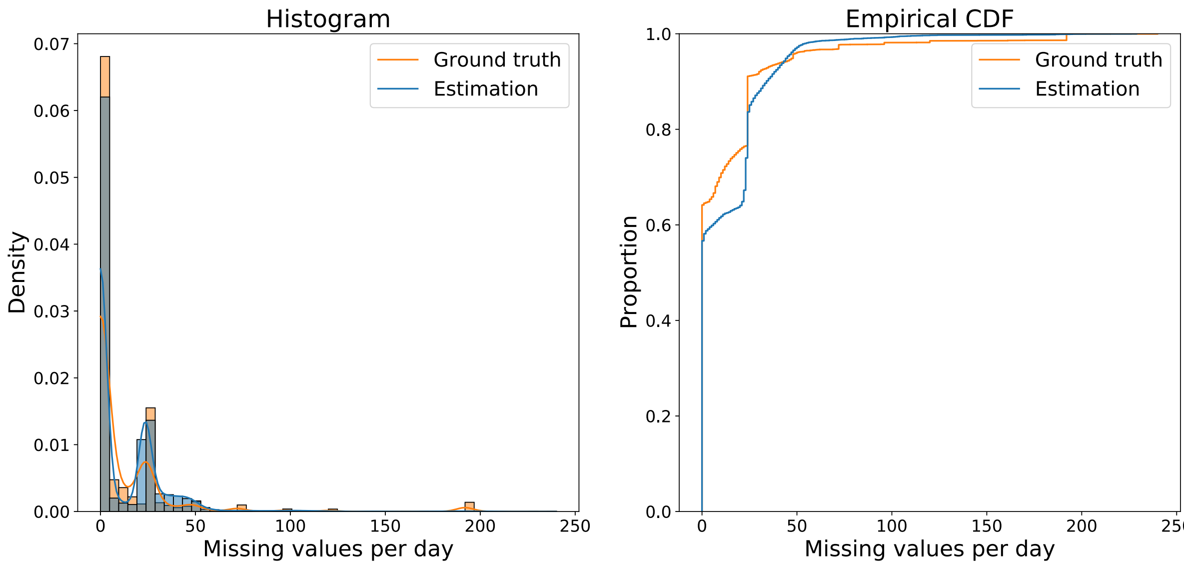


Fig. 7. Both figures show the estimated and true (empirical) distribution of the amount of missing values per day in the test data. The left plot shows the values in a histogram, while the right plot shows the corresponding empirical cumulative distribution function.

Table 3
Distance measures between the model predictions and the test data over different variants of distribution characteristics. A smaller value indicates a higher similarity between the two distributions.

Missing values	Hellinger	Wasserstein
Per day	0.2978	0.0014
Per feature (Mean over all features)	0.1343	0.0048
Per timestep (Mean over all timesteps)	0.1231	0.0105
Sequence length (per feature and day)	0.3391	0.0154

Table 4
Result metrics and their respective standard deviations. The lowest result is highlighted in bold. For all models the standard deviation is quite high, since some time series are hard to replicate, which leads to high errors, affecting the aggregated MSE.

Model	MSE	MAE
MVTSI	**0.0610** ± 0.5822	**0.0845** ± 0.2321
LSTM	0.0687 ± 0.5810	0.1217 ± 0.2320
LOCF	0.1030 ± 0.9196	0.0959 ± 0.3062
KNN	0.0668 ± 0.5645	0.0927 ± 0.2412

one. Instead we evaluate the performance of our model by comparing summary statistics over all test samples, in order to determine whether the approach can generally reflect the properties of the true distribution of missing values. For this purpose we chose several specific summary statistics. First, we compare the distribution of the amount of missing values per day over all features. Furthermore, we compare the amount of missing values per feature and per timestep. Finally, we also compare the sequence length of missing values per feature and day. All of these summary statistics can be transformed into a histogram or an estimated distribution function. Therefore we need tools or measures to compare the similarity of two (empirical) distributions. For that purpose, we use two distance measures, the earth mover's distance (also referred to as the first Wasserstein metric) and the Hellinger distance. Both measures have a value of zero, if the (empirical) distributions are identical. The Hellinger distance for discrete distributions is defined as

$$H(P,Q) = \frac{1}{\sqrt{2}} \| \sqrt{P} - \sqrt{Q} \|_2,$$

where P, Q are probability vectors. In our case they refer to the probability of a specific value being missing. $H(P,Q)$ fulfills the property $0 \leq H(P,Q) \leq 1$. Let U, V be distributions with cumulative distribution functions F_U, F_V, respectively. The first Wasserstein distance is then defined as:

$$W_1(U,V) = \int_{-\infty}^{\infty} |F_U(x) - F_V(x)| dx.$$

See Nguyen [27] for further information on characteristics of the Wasserstein metric and Meshgi and Ishii [28] for further ways to measure the similarity of histograms.

The next step is to analyze the performance of the LSTM model that is used to predict the distribution of missing values in the data. Fig. 7 shows the histogram and the empirical cumulative distribution function (ECDF) of the true and estimated distribution. It can be seen that the empirical distributions are quite similar, meaning that the model is able to match the general distribution of missing values in the data. As explained earlier, different metrics are used to quantify the similarity of the two distributions. Table 3 shows the distance measures for the summary distributions. As already mentioned, a value of zero would determine a perfect fit. Therefore, the results show that the proposed model is able to generate a distribution of missing values which accurately reflects the characteristics of the underlying distribution. In the next step this model is used to predict missing values in the test data and train the attention model on the imputation task.

5.2. Evaluation of imputation model

To evaluate the quality of the data imputing different metrics are used to examine the performance of each method. The mean squared error (MSE) and the mean absolute error (MAE) are very common methods to evaluate the deviance of the original value from the prediction. In order to evaluate the model properly, the data needs to be split into a training and a test set. Since the data is ordered, in the sense of a time series, it is not possible to only take random samples from the data. However, if only one time slice is taken as a test set, for example one year, the evaluation might be biased. In our approach a test set with a monthly resolution is used. From all unique countries, years and months, which is a total of 1958 different monthly data samples, 10% of the data is chosen as the test set. The data is then transformed using the rolling window approach, as described earlier. In this way the timely nature of the data is considered, while still keeping it random

Table 5
Mean squared error calculated per individual feature, as well as share of total missing values. The best model for each feature is highlighted in bold.

Feature	Missing	MVTSI	LSTM	LOCF	KNN
Hard coal	23%	**0.0222**	0.0291	0.0369	0.0380
Other	21.46%	**0.0108**	0.0209	0.0209	0.0169
Pumped Storage	16.93%	**0.1662**	0.1785	0.3110	0.1929
Fossil Gas	16.11%	0.0636	**0.0607**	0.1150	0.0719
PV	10.45%	0.0778	0.0990	**0.0660**	0.0662
Hydro	9.49%	**0.0194**	0.0527	0.0237	0.0268
Biomass	5.97%	**0.0110**	0.0343	0.0144	0.0160
Day ahead price	5%	0.1012	**0.0942**	0.1865	0.0957
Wind Offshore	4.14%	0.2103	0.1429	0.3651	**0.1311**
Load	3.15%	**0.0060**	0.0205	0.0107	0.0074
Wind Onshore	2.52%	0.1107	**0.0871**	0.1845	0.1129
Nuclear	1.71%	0.0102	0.0338	0.0077	**0.0054**
Lignite	1.32%	**0.0040**	0.0251	0.0199	0.0209

Table 6
Different technologies for electricity production, as well as their respective greenhouse gas emissions [29], calculated for 2021 in Germany [30]. The right column shows the Model with the lowest MSE per feature.

Feature	CO_2 eq [kg/MWh]	CO_2 eq [Mt]	Best model
Lignite	1229	121.55	MVTSI
Hard Coal	1227.3	56.95	MVTSI
Fossil gas	488.4	24.96	LSTM
Biomass	201.4	8.68	MVTSI
PV	85.8	4.15	LOCF
Pumped storage	56.8	0.22	MVTSI
Wind Onshore	27.1	2.43	LSTM
Nuclear	11.7	0.76	KNN
Wind Offshore	27.1	0.65	LOCF
Hydro	4.6	0.07	MVTSI
Other	–	–	MVTSI

enough to create a stable evaluation process. The first stage model, is used to estimate the underlying distribution of the missing values. It is trained only on the training data and predicts a missing value mask for the test data. This mask is put on top of the data and determines where the missing values are located. Since some of these values might be completely missing in the ENTSO-E data, the evaluation metrics are only calculated over the data that the LSTM model predicts to be missing, but is not actually missing. Since the true value is not known in that case, an evaluation would not be possible. For the evaluation, the LSTM model is only trained on data which includes missing values. The reason is that the focus is on imputing missing values and a model that predicts a lot of complete days is not useful in that case. By training only on incomplete data, the model tends to predict more days with missing values than can be seen in the original data. However, since complete days are not relevant for our imputation models, this method can be justified. The attention model is tested against several other methods for time series imputing, including classical statistical models, as well as deep learning approaches. The first "naive" method is LOCF (Last Observation Carried Forward). This methods replaces missing values by using the observation from the previous day. The second method we use as a comparison is the k-nearest-neighbor-imputing (KNN-imputing). Furthermore, we are comparing our method to another deep learning method, namely a bidirectional LSTM, which is similar to the approach proposed by Cao et al. [18].

As already mentioned, the models are evaluated using the mean squared- and the mean absolute error. Table 4 shows the two error metrics for each model, including their respective standard deviations. Both, in terms of MSE and MAE, the attention model obtains the lowest error and a standard deviation similar to the benchmark models. It is noticeable that for all methods the standard deviation is quite high compared to the mean value. This is due to some missing values being very hard to predict, thus leading to an abnormally large error and affecting the standard deviation. A possible explanation could be that some missing data occurs completely at random and is therefore difficult to predict by a model. However, since this is the case for all models equally, the results are still comparable. Table 5 shows the MSE of the different models grouped by the individual features of the data. It can be seen that the attention model has the lowest error for the most features. More importantly, the model shows a good performance for features with a lot of missing values. This means that the proposed approach is suitable to estimate the features, responsible for the most gaps in the data. The dependency between the amount of missing values in the data and the model accuracy is highly interesting and could be analyzed further. The first stage model can easily be modified in order to simulate a distribution with more or less missing values and the imputation approaches could be analyzed accordingly. However, since the focus of this study is tailored to the ENTSO-E dataset and its characteristics, these experiments are left as a direction for future research.

This last section aims to give a brief outlook on further processing of the imputed data. For that purpose, the imputed features are compared in terms of their greenhouse gas emissions. Table 6 shows the kilogram CO_2 equivalent per MWh for each specific feature, as well as the model with the lowest MSE. The respective values are taken from Xu et al. [29]. It is noticeable, that the MVTSI model obtains the best performance for technologies with the highest carbon emissions, especially Lignite and Hard coal. However, the performance is worse for highly volatile technologies, such as wind power or photovoltaic. The results suggest that the chosen approach could be suitable for imputing data that is to be used in an energy system model, which aims at estimating future technology dispatch and impact of greenhouse gas emissions. However, a more detailed study of this topic would be needed, extending the scope of this paper.

6. Conclusion

Imputing missing data in multivariate time series is an important task in many areas of application, including energy time series. A lot of research has been put into developing such methods, more recently using machine learning approaches. This paper proposes a two-step attention based data model for imputing multivariate time series. In the first part of the model the distribution of missing values in the original data is learned. This information is further used to train an attention based neural network that imputes the missing values. While we focus on the use-case of the ENTSO-E data, our model is flexible for any kind of multivariate time series and any desired time frame. Our experimental results show that the model can estimate missing data more accurately than the chosen benchmark methods. Apart from the evaluation with a focus on time series metrics, we also give more detail on the model performance in dependence of a specific technology for electricity generation. These results show that our approach shows the lowest error (in terms of the MSE) for most features. However, this is not the case for all features and also heavily depends on the distribution of the missing values. Further analysis of the feature and model dynamics in dependence on the amount of missing values can offer interesting insights. Further improvement and tuning of the model could potentially lead to an improvement for these features as well, while keeping the ability to generalize over all features. As previously mentioned, several approaches are available to shift into a probabilistic imputing setting. Further use of different approaches and an analysis of the uncertainty in the predictions of our model is a promising area of future research. As for the model structure, implementing the cross-dimensional self-attention [6] for the proposed architecture could be an interesting approach to further improve performance and reduce computational complexity. In addition to the mentioned results, we provide an outlook on the potential of using the proposed approach for imputing missing values for further processing of the data, for example in energy system models. Further work could go into examining how the choice of an imputation methods affects the output of a model that further processes the data, for example an energy system model.

Country	Country availability	Day ahead price	Biomass	Fossil Gas	Fossil Hard coal	Fossil Oil	Hydro Pumped Storage	Hydro Run-of-river and poundage	Hydro Water Reservoir	Other	Solar	Wind Onshore	Nuclear
AT	0.00%	12.54%	0.00%	0.00%	0.00%	n/e	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	n/e
BE	0.00%	0.18%	0.00%	0.00%	32.78%	0.00%	47.53%	0.00%	n/e	0.00%	0.00%	0.00%	0.00%
BG	16.65%	17.57%	0.11%	34.31%	34.31%	n/e	0.16%	34.31%	0.16%	n/e	0.00%	0.00%	0.16%
CH	0.00%	0.09%	n/e	88.19%	n/e	n/e	8.25%	13.45%	8.25%	n/e	0.04%	0.04%	8.12%
CZ	0.00%	0.18%	0.04%	0.04%	0.04%	0.07%	0.02%	0.04%	0.04%	0.06%	0.07%	0.04%	0.04%
DE-AT-LU	37.50%	0.36%	0.22%	0.22%	0.22%	0.00%	0.22%	0.22%	0.00%	0.22%	0.37%	0.28%	0.22%
DE-LU	62.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
DK	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	n/e	0.00%	n/e	n/e	0.00%	0.00%	n/e
EE	0.00%	0.00%	0.54%	0.54%	n/e	0.54%	n/e	0.54%	n/e	n/e	0.54%	0.54%	n/e
ES	0.00%	0.00%	0.04%	0.04%	0.04%	0.04%	n/e	0.04%	0.04%	0.04%	0.04%	0.04%	0.04%
FI	0.00%	0.05%	0.12%	0.12%	0.12%	0.13%	n/e	0.12%	n/e	0.12%	n/e	0.12%	0.11%
FR	0.00%	0.18%	0.08%	0.08%	7.22%	0.08%	43.90%	0.09%	0.16%	n/e	0.06%	0.08%	0.08%
GB	0.00%	0.14%	44.03%	0.64%	0.64%	0.64%	0.64%	0.64%	n/e	0.46%	0.17%	0.17%	0.64%
GR	0.00%	0.00%	n/e	0.09%	n/e	0.09%	0.15%	n/e	0.09%	n/e	0.12%	0.12%	n/e
HR	50.00%	0.36%	33.30%	33.30%	33.30%	33.30%	33.30%	33.30%	33.30%	n/e	33.30%	33.30%	n/e
HU	0.00%	0.00%	0.10%	0.10%	0.00%	48.32%	n/e	0.10%	0.52%	0.10%	78.74%	5.36%	0.01%
IE	0.00%	1.61%	n/e	0.72%	0.71%	0.71%	6.61%	6.61%	n/e	2.89%	n/e	0.64%	n/e
IT	0.00%	0.23%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	n/e
LT	0.00%	0.05%	1.81%	1.81%	n/e	51.81%	1.81%	1.81%	n/e	1.81%	1.81%	1.81%	n/e
LV	0.00%	0.00%	0.16%	0.16%	n/e	n/e	n/e	3.96%	96.20%	0.16%	n/e	0.15%	n/e
NL	0.00%	0.18%	0.49%	0.01%	0.85%	n/e	n/e	n/e	n/e	8.92%	0.23%	0.34%	1.43%
NO	0.00%	0.00%	n/e	0.00%	n/e	n/e	0.00%	0.00%	0.00%	0.00%	n/e	0.00%	n/e
PL	0.00%	8.72%	0.04%	0.04%	0.04%	8.53%	0.04%	0.04%	0.04%	n/e	87.86%	0.00%	n/e
PT	0.00%	0.09%	0.00%	0.00%	0.00%	n/e	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	n/e
RO	0.00%	0.23%	0.31%	0.31%	0.25%	n/e	n/e	0.37%	0.45%	n/e	0.31%	0.31%	0.31%
RS	16.65%	19.59%	19.67%	19.61%	n/e	n/e	19.61%	19.56%	19.61%	19.56%	n/e	n/e	n/e
SE	0.00%	0.00%	n/e	n/e	n/e	n/e	n/e	n/e	0.00%	0.00%	n/e	0.00%	0.00%
SI	0.00%	0.00%	0.08%	0.07%	n/e	0.07%	0.07%	0.07%	n/e	n/e	0.01%	0.08%	0.07%
SK	0.00%	0.23%	1.00%	1.00%	1.00%	1.00%	37.37%	1.00%	14.79%	1.00%	1.00%	1.00%	1.00%

Fig. A.8. Amount of missing values of chosen features in the used ENTSO-E data. Country availability refers to the country submitting any data at all in the period 2015–2020. The figure is restricted to a subset of available technologies.

Data availability

Data is publicly available via an API, but the author has no permission to share it.

Acknowledgments

Juan Gómez-Romero is partially supported by the projects IA4TES (MIA.2021.M04.0008, funded by the European Union – NextGenerationEU) and DeepSim (A.TIC.244.UGR20, funded by FEDER, Spain – Junta de Andalucía).

Declaration of competing interest

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

Appendix. Processing of ENTSO-E data

This section gives a more detailed overview of the data processing of the ENTSO-E data, since this has an influence on the characteristics of the missing data. First, if a whole year (or part of a year) is missing for a country, the data is treated as not existent and not as missing. This is for example the case in Serbia, where the whole year of 2015 is not available. The same applies if a country does not incorporate a specific production feature. In this case, the production is just set to 0 MWh, since removing the feature does not make sense, due to the structure of our model. This is the case for example with Denmark and the feature "Nuclear". Furthermore, some countries incorporated features after the transparency platform was already running, leading to "n/e" for some years and actual data for others. Here the "n/e" is again replaced with a production of 0 MWh. For instance, Poland only started submitting data for the feature "Solar" from 10.04.2020 onwards. These processing steps are not heavily important for the data or the model, but they do change the characteristics of the missing values. Therefore the processed data in this article does not necessarily show the same characteristics as the data used in Hirth et al. [13].

References

[1] Zhang Y-F, Thorburn PJ, Xiang W, Fitch P. SSIM—A deep learning approach for recovering missing time series sensor data. IEEE Internet Things J 2019;6(4):6618–28. http://dx.doi.org/10.1109/JIOT.2019.2909038.

[2] Yuan H, Xu G, Yao Z, Jia J, Zhang Y. Imputation of missing data in time series for air pollutants using long short-term memory recurrent neural networks. In: Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers. UbiComp '18, New York, NY, USA: Association for Computing Machinery; 2018, p. 1293–300. http://dx.doi.org/10.1145/3267305.3274648.

[3] Verma H, Kumar S. An accurate missing data prediction method using LSTM based deep learning for health care. In: Proceedings of the 20th international conference on distributed computing and networking. ICDCN '19, New York, NY, USA: Association for Computing Machinery; 2019, p. 371–6. http://dx.doi.org/10.1145/3288599.3295580.

[4] Luo Y, Cai X, Zhang Y, Xu J, Xiaojie Y. Multivariate time series imputation with generative adversarial networks. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. Advances in neural information processing systems. vol. 31, Curran Associates, Inc.; 2018, URL https://proceedings.neurips.cc/paper/2018/file/96b9bff013acedfb1d140579e2fbeb63-Paper.pdf.

[5] Fortuin V, Baranchuk D, Rätsch G, Mandt S. GP-VAE: Deep probabilistic time series imputation. 2020, arXiv:1907.04155.

[6] Ma J, Shou Z, Zareian A, Mansour H, Vetro A, Chang S-F. CDSA: cross-dimensional self-attention for multivariate, geo-tagged time series imputation. 2019, arXiv preprint arXiv:1905.09904.

[7] Ruggles TH, Farnham DJ, Tong D, Caldeira K. Developing reliable hourly electricity demand data through screening and imputation. Sci Data 2020;7(1):1–14.

[8] Pinson P. Wind energy: Forecasting challenges for its operational management. Statist Sci 2013;28(4):564–85. http://dx.doi.org/10.1214/13-STS445.

[9] Giebel G, Brownsword R, Kariniotakis G, Denhard M, Draxl C. The state of the art in short-term prediction of wind power a literature overview. 2nd ed. 2011, http://dx.doi.org/10.13140/RG.2.1.2581.4485.

[10] Akçay H, Filik T. Short-term wind speed forecasting by spectral analysis from long-term observations with missing values. Appl Energy 2017;191:653–62. http://dx.doi.org/10.1016/j.apenergy.2017.01.063, URL https://www.sciencedirect.com/science/article/pii/S0306261917300739.

[11] PJM data miner. Data miner 2. 2021, https://dataminer2.pjm.com/list. [Accessed on 05/27/2021].

[12] ENTSO-E. ENTSO-e transparency platform. 2021, https://transparency.entsoe.eu/. [Accessed on 05/27/2021].

[13] Hirth L, Mühlenpfordt J, Bulkeley M. The ENTSO-E Transparency Platform – A review of Europe's most ambitious electricity data platform. Appl Energy 2018;225:1054–67. http://dx.doi.org/10.1016/j.apenergy.2018.04.048.

[14] Fung DS. Methods for the estimation of missing values in time series. 2006.

[15] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Springer series in statistics, New York, NY, USA: Springer New York Inc.; 2001.

[16] Kim M, Park S, Lee J, Joo Y, Choi JK. Learning-based adaptive imputation methodwith kNN algorithm for missing power data. Energies 2017;10(10). http://dx.doi.org/10.3390/en10101668, URL https://www.mdpi.com/1996-1073/10/10/1668.

[17] Hsu H-H, Yang AC, Lu M-D. KNN-dtw based missing value imputation for microarray time series data.. J Comput 2011;6(3):418–25.

[18] Cao W, Wang D, Li J, Zhou H, Li L, Li Y. BRITS: bidirectional recurrent imputation for time series. 2018, arXiv:1805.10572. URL http://arxiv.org/abs/1805.10572.

[19] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2017, arXiv preprint arXiv:1706.03762.

[20] Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, Eickhoff C. A transformer-based framework for multivariate time series representation learning. 2020, arXiv:2010.02803. URL https://arxiv.org/abs/2010.02803.

[21] Williams RJ, Zipser D. A learning algorithm for continually running fully recurrent neural networks. Neural Comput 1989;1(2):270–80. http://dx.doi.org/10.1162/neco.1989.1.2.270.

[22] Doubleday K, Van Scyoc Hernandez V, Hodge B-M. Benchmark probabilistic solar forecasts: Characteristics and recommendations. Sol Energy 2020;206:52–67. http://dx.doi.org/10.1016/j.solener.2020.05.051, URL https://www.sciencedirect.com/science/article/pii/S0038092X20305429.

[23] Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Proceedings of the 31st international conference on neural information processing systems. NIPS '17, Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 6405–16.

[24] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 2007;102(477):359–78. http://dx.doi.org/10.1198/016214506000001437.

[25] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd international conference on international conference on machine learning. ICML '16, vol. 48, JMLR.org; 2016, p. 1050–9.

[26] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2017, arXiv:1412.6980.

[27] Nguyen X. Wasserstein distances for discrete measures and convergence in nonparametric mixture models. 2011, arXiv:1109.3250.

[28] Meshgi K, Ishii S. Expanding histogram of colors with gridding to improve tracking accuracy. In: Proceedings of the 14th IAPR international conference on machine vision applications, MVA 2015. 2015, p. 475–9. http://dx.doi.org/10.1109/MVA.2015.7153234.

[29] Xu L, Wang Z, Yilmaz HU, Poganietz W-R, Ren H, Guo Y. Considering the impacts of metal depletion on the European electricity system. Energies 2021;14(6). http://dx.doi.org/10.3390/en14061560, URL https://www.mdpi.com/1996-1073/14/6/1560.

[30] Fraunhofer. Public net electricity generation in Germany in 2021: Renewables weaker due to weather. 2022, https://www.ise.fraunhofer.de/en/press-media/news/2022/public-net-electricity-in-germany-in-2021-renewables-weaker-due-to-weather.html.