

Data-Driven Missing Data Imputation for Wind Farms Using Context Encoder

Liao, Wenlong; Bak-Jensen, Birgitte; Pillai, Jayakrishnan Radhakrishna; Yang, Dechang ; Wang, Yusen

Published in:
Journal of Modern Power Systems and Clean Energy

DOI (link to publication from Publisher):
[10.35833/MPCE.2020.000894](https://doi.org/10.35833/MPCE.2020.000894)

Creative Commons License
CC BY 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Liao, W., Bak-Jensen, B., Pillai, J. R., Yang, D., & Wang, Y. (2022). Data-Driven Missing Data Imputation for Wind Farms Using Context Encoder. *Journal of Modern Power Systems and Clean Energy*, 10(4), 964-976. Article 13. <https://doi.org/10.35833/MPCE.2020.000894>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Data-driven Missing Data Imputation for Wind Farms Using Context Encoder

Wenlong Liao, Birgitte Bak-Jensen, Jayakrishnan Radhakrishna Pillai, Dechang Yang, and Yusen Wang

Abstract—High-quality datasets are of paramount importance for the operation and planning of wind farms. However, the datasets collected by the supervisory control and data acquisition (SCADA) system may contain missing data due to various factors such as sensor failure and communication congestion. In this paper, a data-driven approach is proposed to fill the missing data of wind farms based on a context encoder (CE), which consists of an encoder, a decoder, and a discriminator. Through deep convolutional neural networks, the proposed method is able to automatically explore the complex nonlinear characteristics of the datasets that are difficult to be modeled explicitly. The proposed method can not only fully use the surrounding context information by the reconstructed loss, but also make filling data look real by the adversarial loss. In addition, the correlation among multiple missing attributes is taken into account by adjusting the format of input data. The simulation results show that CE performs better than traditional methods for the attributes of wind farms with hallmark characteristics such as large peaks, large valleys, and fast ramps. Moreover, the CE shows stronger generalization ability than traditional methods such as auto-encoder, K -means, k -nearest neighbor, back propagation neural network, cubic interpolation, and conditional generative adversarial network for different missing data scales.

Index Terms—Data-driven, missing data imputation, wind farm, deep learning, context encoder.

NOMENCLATURE

$\sigma_{\text{con}}^i(\cdot)$	Activation function of the i^{th} convolutional layer
$\sigma_{\text{dense}}^i(\cdot)$	Activation function of the i^{th} dense layer
$\sigma_{\text{tran}}^i(\cdot)$	Activation function of the i^{th} transposed convolutional layer
λ	Weights between reconstructed loss and adversarial loss

B_{con}^i	Bias vector of the i^{th} convolutional layer
B_{dense}^i	Bias vector of the i^{th} dense layer
B_{tran}^i	Bias vector of the i^{th} transposed convolutional layer
$D(\cdot)$	Output of discriminator
E_X	Expectation of sample X
$F(\cdot)$	Generated sample
L_{adv}	Adversarial loss
L_{rec}	Reconstructed loss
M	A binary mask
MAE_1	Mean absolute error
MAE_2	The maximum absolute error
n	Number of missing sample points in test set
R	The maximum pooling area
$RMSE$	Root mean square error
W_{con}^i	Weight of the i^{th} convolutional layer
W_{dense}^i	Weight of the i^{th} dense layer
W_{tran}^i	Weight of the i^{th} transposed convolutional layer
X	Complete samples of wind farms
X_{con}^i	Input feature of the i^{th} convolutional layer
X_{dense}^i	Input feature of the i^{th} dense layer
$X_{j,k}^i$	Input feature of the i^{th} maximum pooling layers
X_{tran}^i	Input feature of the i^{th} transposed convolutional layer
Y_{con}^i	Output feature of the i^{th} convolutional layer
Y_{dense}^i	Output feature of the i^{th} dense layer
y_i	The i^{th} element of real data
y'_i	The i^{th} element of filled data through a model
Y_{pool}^i	Output feature of the i^{th} maximum pooling layers
Y_{tran}^i	Output feature of the i^{th} transposed convolutional layer

Manuscript received: December 23, 2020; revised: March 3, 2021; accepted: April 28, 2021. Date of CrossCheck: April 28, 2021. Date of online publication: June 7, 2021.

This work was supported by the China Scholarship Council.

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

W. Liao, B. Jensen, and J. R. Pillai are with the Department of Energy Technology, Aalborg University, Aalborg, Denmark (e-mail: weli@et.aau.dk; bbj@et.aau.dk; jrp@et.aau.dk).

D. Yang (corresponding author) is with the College of Information and Electrical Engineering, China Agricultural University, Beijing, China (e-mail: yang-dechang@cau.edu.cn).

Y. Wang is with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: yusenw@kth.se).

DOI: 10.35833/MPCE.2020.000894

I. INTRODUCTION

THE volatility and intermittence of the output power of wind farms pose challenges to the operation and planning of the power system [1]. In order to ensure the safety and stability of the power system, it is necessary to accurately predict the output power of the wind farm using historical data collected through the supervisory control and data acquisition system.

sition (SCADA) system. However, those collected data may be incomplete, since the SCADA system is often interfered with by various factors, such as sensor failure, cyber-attack, and communication congestion [2]. Therefore, the missing data imputation of wind farms is of great significance for wind power forecasting.

The traditional methods for missing data imputation of wind farms can be summarized into four categories: interpolation-based methods, regression-based methods, similarity-based methods, and parameter-estimation-based methods. The first category mainly includes the linear interpolation, spline interpolation, and Cubic interpolation [3]. These kinds of methods fill the missing data by constructing polynomials whose parameters are obtained by using the information around the missing position. High accuracy can be achieved when the scale of missing data is very small, but they ignore the correlation among multiple attributes in the filling process, which restricts the application scope of these methods [4]. The second category mainly includes recurrent neural networks (RNNs), random forest, and back propagation (BP) neural networks [5], [6]. For example, the long short-term memory network is utilized as a predictor to fill the missing wind power data, which shows better performance than traditional RNNs in [7]. To deal with the problem of filling the large-scale missing wind power data, the improved random forest is proposed to combine the matrix combination, linear interpolation, and matrix transposition [8]. The simulation results show that the improved random forest is applicable to fill wind power data in various missing forms. Compared with interpolation-based methods, regression-based methods make full use of the correlation among attributes, resulting in higher accuracy and wider application scope. The third category mainly includes the mean substitution method, k -nearest neighbor (KNN), and K -means [9]. Specifically, the mean substitution method uses the mean value of the existing incomplete attributes to fill all the missing data. This method is simple and widely used, but it limits the diversity and volatility of attributes, resulting in low accuracy [10]. KNN uses the mean value of the nearest k samples to fill the missing data, while K -means uses the cluster center to fill the missing attributes for samples with missing data in the cluster [11]. Similar to the interpolation-based methods, the similarity-based methods also ignore the correlation among multiple attributes, which limit the accuracy of these algorithms [12]. The fourth category mainly includes point estimation and interval estimation [13], [14]. This kind of method uses the existing data to fill in missing attributes through the maximum likelihood estimation. For example, the expectation-maximization algorithm uses missing data as variables to participate in the process of parameter estimation, alternately updating missing data and parameters to be estimated in an iterative manner, so as to achieve the goal of missing data imputation [15]. However, the accuracy is greatly affected by the form of the probability distribution function assumed artificially [16].

As a branch of artificial intelligence, deep learning has shown outstanding performance in many fields such as representation learning, image classification, and natural language processing [17], [18], which has brought new opportunities

for the development of missing data imputation of wind farms. The existing methods for missing data imputation of wind farms mainly include the de-noising automatic encoder (AE) and the conditional generative adversarial networks (CGANs). Specifically, AE corrupts the input data and requires the decoder to minimize the damage, so as to learn a semantically meaningful representation of samples [19]. However, its corruption process is typically low and localized, which does not need to undo much semantic information. The disadvantages of CGAN such as exploding gradients and vanishing gradients in the training process still exist in [20]-[22], which leads to low accuracy of filling data.

The context encoder (CE) is a new deep neural network for missing data imputation developed from AE. Compared with AE, CE shows a much deeper semantic understanding of the scenario, and a stronger ability to represent high-dimensional features over large spatial extents [23], leading to its wide applications in various fields. For example, a long short-term CE is proposed to fill missing air pollution data in [24]. While in [25], CE is designed to mine spatial features and capture high-level information for 2-dimensional image segmentation. The simulation results show that CE achieves the state-of-art performance with superior accuracy. To restore missing medium gaps of audio, CE including convolutional and fully connected layers is proposed to capture the context information of missing samples in [26]. The successful applications of CE in the image and audio fields prove that it can learn complex objective laws through unsupervised training. Theoretically, CE can not only use deep convolutional layers with strong learning ability to effectively mine the complex nonlinear correlation among the multiple attributes of wind farms, but also use the reconstructed loss and adversarial loss to represent the spatial-temporal relationship between missing parts and complete samples, so as to greatly improve the accuracy of missing data imputation for wind farms. However, the existing network structures of CE are designed for computer vision, which is not suitable for the 1-dimensional data of wind farms [27]. Therefore, it requires to design a structure of CE with strong feature extraction ability and high filling accuracy according to the characteristics of data from wind farms.

This paper aims to design a CE to improve the accuracy of missing data imputation for wind farms. The performance of the proposed method is tested by a real-world dataset. The key contributions of this paper are as follows.

- 1) A new data-driven, model-free, and scalable method is proposed for missing data imputation of wind farms. By employing the encoder-decoder-discriminator pipeline that consists of deep convolutional networks, it can fully explore the dynamic nonlinear correlations among the multiple attributes of wind farms that are difficult to be modeled explicitly, such as the temporal correlation and dynamic changes of wind power sequences.

- 2) This paper innovatively applies CE to missing data imputation of wind farms. To generate a plausible hypothesis for the missing data of wind farms, both a reconstructed loss and an adversarial loss are designed as the loss function for CE. Specifically, the reconstructed loss is responsible for exploring the overall information of the missing data and co-

herence with regard to its surrounding context. In addition, the adversarial loss makes filling parts look real, and reduces the maximum absolute error of the model.

3) Extensive experiments on a real-world dataset of wind farms are performed to validate the effectiveness of missing data imputation. The influence of key parameters of CE (e.g., the number of iterations, the choices of optimizer and batch size, the number of middle layers, and the weights of adversarial loss and reconstructed loss) on the performance is analyzed by simulation, and some constructive suggestions for the selection of key parameters are given.

The remainder of this paper is organized as follows. Section II formulates the missing data imputation problem based on a CE. Section III presents the process of the missing data imputation using CE. In Section IV, the effectiveness of the proposed approach is verified by simulation. Section V discusses the limitation and generalization of the proposed approach. Section VI summarizes the paper.

II. FORMULATION OF MISSING DATA IMPUTATION PROBLEM BASED ON CE

A. Encoder-decoder-discriminator Pipeline

As shown in Fig. 1, the overall framework of CE is an encoder-decoder-discriminator pipeline. First of all, the reshaping function from Python 3.6 is used to reconstruct the incomplete sample from wind farms into a feature matrix to facilitate connection to the encoder with convolutional (Conv) layers and maximum pooling (Maxpool) layers. Then, the feature matrix is passed through the encoder to obtain the low-dimensional feature representation of the incomplete sample, which is connected to the decoder with transposed convolutional (ConvTran) layers. The decoder takes this low-dimensional feature representation and generates the complete sample. The discriminator with convolutional layers, flatten layers, and dense layers is regarded as a detector, whose purpose is to identify the samples generated by the encoder and decoder as much as possible.

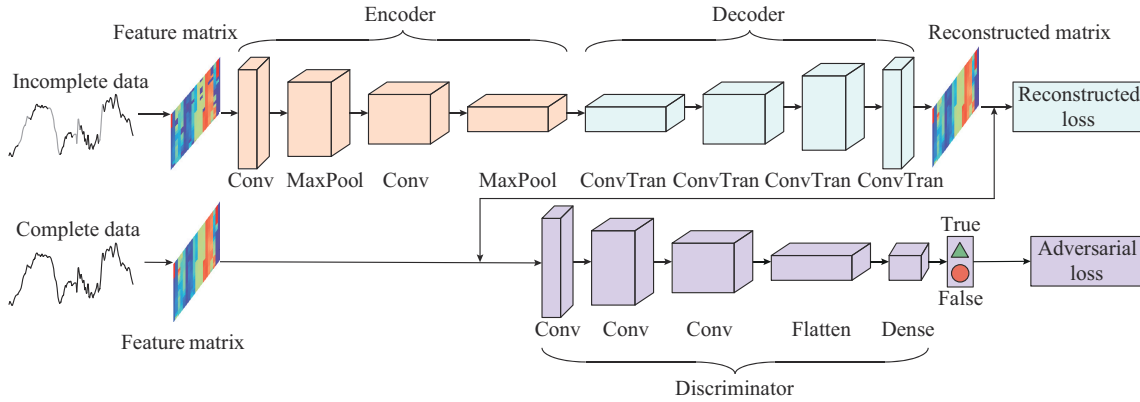


Fig. 1. Framework of CE for missing data imputation.

The convolutional neural network (CNN) is a feed-forward neural network with a convolutional operation. Its emergence has greatly promoted the development of artificial intelligence. Because of its powerful feature extraction capabilities, CNN has been widely used in various fields such as fault diagnosis, object detection, speech recognition, and semantic segmentation [28]. Therefore, CNN is chosen to build the encoder, decoder, and discriminator.

The encoder consists of several Conv layers and MaxPool layers as seen in Fig. 1. Specifically, the key operation of the Conv layer is to perform a convolutional operation on the output features of the previous layer, and then add a bias vector as the input features to the next layer. Its mathematical formula is expressed as:

$$Y_{\text{con}}^i = \sigma_{\text{con}}^i (X_{\text{con}}^i * W_{\text{con}}^i + B_{\text{con}}^i) \quad (1)$$

where $*$ is the operation of convolution. Note that the output feature of a Conv layer is used as the input feature to the following MaxPool layer in the encoder.

As shown in Fig. 2, the MaxPool layer reduces the dimensionality of the data output of the Conv layers to obtain a low-dimensional feature representation of the incomplete sample. Its mathematical formula is expressed as:

$$Y_{\text{pool}}^i = \max_{j,k \in R} (X_{j,k}^i) \quad (2)$$

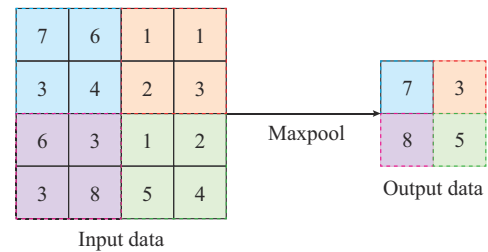


Fig. 2. Visualization of MaxPool layer.

Note that the output feature of a MaxPool layer is used as the input feature to the next Conv layer in the encoder.

The decoder consists of the fully transposed Conv layers as shown in Fig. 1 to obtain the same dimensionality as the original input data. Specifically, the key operation of the transposed Conv layer is to perform a transposed Conv operation on the output features of the previous layer and then add a bias vector as the input features of the next layer. Note that the first input to the transposed convolutional layer is the one from the last MaxPool layer. Its mathematical formula is expressed as:

$$Y_{\text{tran}}^i = \sigma_{\text{tran}}^i(X_{\text{tran}}^i \cdot W_{\text{tran}}^i + B_{\text{tran}}^i) \quad (3)$$

where \cdot is the operation of the transposed convolution.

The discriminator consists of several convolutional layers, a flatten layer, and a dense layer as shown in Fig. 1. Specifically, the flatten layer is considered as a bridge between the last convolutional layer and the dense layer, and it reshapes the multi-dimensional features into a 1-dimensional vector without changing the amplitude of the data. In other words, the function of the flatten layer is to change the shape of the data. The output of the dense layer is 0 or 1, which is used to determine whether the input data is a reconstructed sample or a real sample. Note that the input feature of the dense layer is the output data of the flatten layer in the discriminator. The mathematical formula of the dense layer is expressed as:

$$Y_{\text{dense}}^i = \sigma_{\text{dense}}^i(X_{\text{dense}}^i W_{\text{dense}}^i + B_{\text{dense}}^i) \quad (4)$$

B. Loss Function

To generate a plausible hypothesis for the missing data of wind farms, both the reconstructed loss and the adversarial loss are designed as the loss function for CE. Specifically, the reconstructed loss is responsible for exploring the overall information of the missing data and coherence with regard to its surrounding context. The mathematical formula of the reconstructed loss L_{rec} is:

$$L_{\text{rec}} = \|M \odot (X - F((1 - M) \odot X))\|_2 \quad (5)$$

where the binary mask M corresponds to the missing position with a value of 1 wherever the data is dropped and 0 for input data; and \odot is the Hadamard product operation.

In addition, the adversarial loss makes filling parts look real, and has the effect of choosing a specific mode from the probability distribution [29]. The mathematical formula of the adversarial loss L_{adv} is:

$$L_{\text{adv}} = \max E_X \left(\lg(D(X)) + \lg(1 - D(F((1 - M) \odot X))) \right) \quad (6)$$

This equation is improved from the loss function of the CGAN.

Ultimately, the overall loss function L for missing data imputation of wind farms can be defined as:

$$L = \lambda L_{\text{rec}} + (1 - \lambda) L_{\text{adv}} \quad (7)$$

where the weight λ is in the range of 0 to 1.

C. Missing Types

Different missing formats may have different effects on the validity of research conclusions. According to the factors leading to missing data, the types of missing forms can be summarized into the following three categories: complete random missing forms, non-random missing forms, and random missing forms [30]. Normally, the formats of missing data caused by human error or sensor failure are considered as the first category. For example, the operator may inadvertently omit certain values when inputting data. Due to cyber-attack or communication congestion, the SCADA system may generate continuous missing data in a period, which is a special case of non-random missing forms. Random missing forms mean that the probability of missing data is only

related to non-missing variables, not related to missing variables. For example, men are more willing to announce their weight than women, so the lack of weight attributes is often related to gender. In general, the main factors leading to missing data of wind farms include sensor failures, cyber-attacks, and communication congestion. Therefore, the missing data forms of wind farms mainly belong to the two categories shown in Fig. 3.

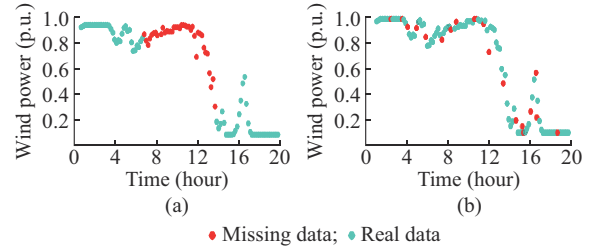


Fig. 3. Missing data forms of wind farms. (a) Continuous missing forms. (b) Complete random missing forms.

III. PROCESS OF MISSING DATA IMPUTATION USING CE

The process of missing data imputation for wind farms using CE is shown in Fig. 4, and the detailed steps are as follows.

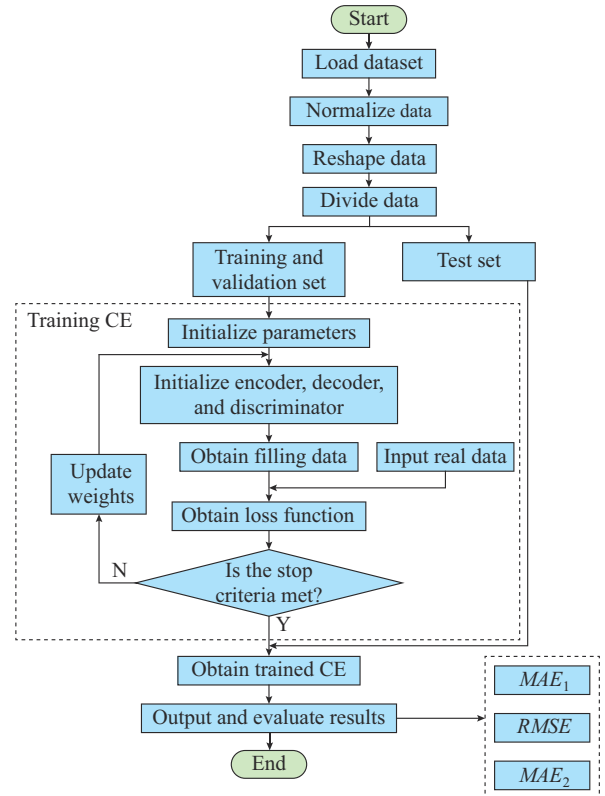


Fig. 4. Process of missing data imputation using CE.

1) Load and normalize data: in addition to wind power, the data of wind farms also include environmental attributes such as wind direction, wind speed, air temperature, and density. There is a strong correlation among these attributes, and using them as input data helps improve the accuracy of missing data imputation. Before the data are fed to CE, it is nec-

essary to normalize the samples of the wind farm; otherwise, the loss function may not converge. This paper uses the minimum-maximum normalization method to convert the input data into the range of 0 to 1.

2) Reshape and divide data: to meet the format requirements of convolutional layers and account for the correlation among multiple attributes, the time series of multiple attributes from wind farms are transformed into a feature matrix with the same number of columns and rows by the method in [31]. Moreover, 80% of the samples are randomly selected to train the CE, and 10% of the samples are randomly selected as the validation set. The remaining samples are used as the test set to evaluate the performance of the trained CE.

3) Initialize parameters and train CE: before starting to train CE, it is necessary to initialize the network structure and parameters such as the number of iterations, the number of middle layers, and the weights of loss functions. Then, the BP algorithm consisting of the forward incentive propagation and backward weight update is utilized to train CE. Specifically, the input matrices are processed by the encoder and decoder pipeline. The filling data output by the decoder and real data are used to calculate the reconstructed loss and adversarial loss. Next, the chain rule is utilized to transfer errors from the output layer to the middle layer. The weights of each layer are updated by the gradient descent algorithms. If the number of iterations is reached, the iteration will be stopped and the result will be output.

4) Evaluate performance of CE: after training CE, the test set will be used to evaluate the performance of the model. To fully measure the variation in the errors in a set of imputation, the mean absolute error MAE_1 , root mean square error $RMSE$, and the maximum absolute error MAE_2 are selected to evaluate the performance:

$$MAE_1 = \text{mean} |y'_i - y_i| \quad i = 1, 2, \dots, n \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y'_i - y_i)^2}{n}} \quad (9)$$

$$MAE_2 = \max |y'_i - y_i| \quad i = 1, 2, \dots, n \quad (10)$$

IV. CASE STUDY

A. Dataset and Model Details

To fully test the performance of various models for missing data imputation of wind farms, a real-world dataset collected from [32] is used for simulation and analysis. In this dataset, the statistical attributes include wind power, wind direction at 100 m, wind speed at 100 m, air temperature at 2 m, surface air pressure, and density at hub height. These attributes are recorded every half an hour from January 1, 2011 to December 31, 2012. 80% of the data is randomly selected as the training set, 10% as the validation set, and the rest as the test set.

The programs of CE for missing data imputation of wind farms are implemented in Spyder 3.2.8 with Keras 2.2.4 and Tensorflow 1.12.0 library. The programming language is Py-

thon. The parameters of the computer are: Intel^(R) Core^(TM) i5-10210U, the processor is @1.60 GHz and 2.11 GHz with 8 GB of memory.

In order to make CE have high performance for missing data imputation of wind farms, the control variable method in [33] is employed to find the suitable structures and parameters of CE, as shown in Fig. 5.

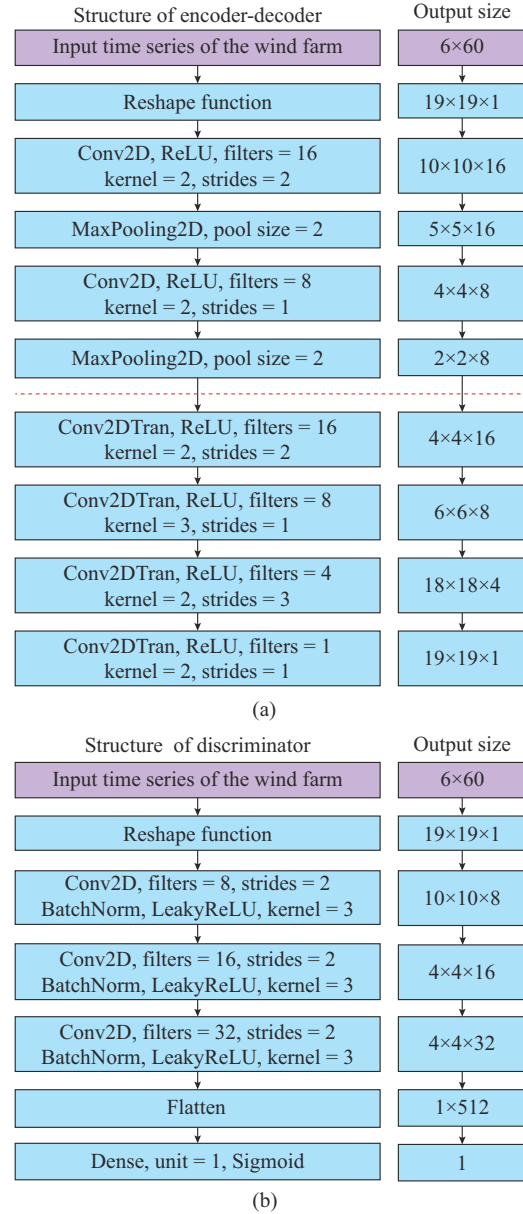


Fig. 5. Structure and parameters of CE. (a) Encoder-decoder. (b) Discriminator.

Specifically, each attribute includes 48 sampling points per day. In addition, the 12 sampling points at the end of the previous day are also used as input data for each attribute to capture the surrounding context information of missing parts. Therefore, the size of each attribute is 1×60. In other words, the original input data are a vector of 1×360 scales. A zero element is added to the end of the input data, which causes it to become a vector of 1×361 scales. In this case, the input data can be converted into a matrix of 19×19

scales by the method proposed in [31], so as to be fed to the encoder-decoder pipeline and the discriminator. For the encoder, it includes two convolutional layers and two maximum pooling layers, and their activation functions are rectified linear unit (ReLU) functions. For the decoder, it includes four transposed convolutional layers whose activation functions are ReLU functions. For the discriminator, it includes three convolutional layers, a flatten layer, and a dense layer. The activation functions of convolutional layers are leaky rectified linear unit (LeakyReLU) functions and the activation function of the dense layer is a sigmoid function. The batch normalizations are applied to the discriminator so as to alleviate over-fitting. The number of iterations is 1000, and the optimizer is the Adam algorithm. λ is set to be 0.999.

Previous works show that the types of missing data in wind farms include complete random missing forms and continuous missing forms [8]. In order to verify the effectiveness of the proposed method, Section IV-B discusses the impact of key parameters, and the Section IV-C discusses the performance of different methods for the data in complete random missing forms. Furthermore, the correlation between the filling accuracy and the missing data scale in complete random missing forms is analyzed in Section IV-D, and the performance of CE for continuous missing forms is presented in Section IV-E.

B. Discussions on Impact of Key Parameters

In order to observe the training stability and convergence of CE, Fig. 6 shows how the loss function decays as the number of iterations increases.

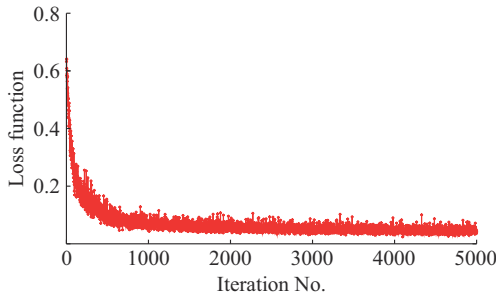


Fig. 6. Training process of CE.

The loss function of CE decreases rapidly with the increase of iteration times. When the number of iterations is more than 1000, its loss function tends to be a constant, indicating that CE has converged. Compared with existing methods such as CGAN, the training process of CE is relatively stable, and there is no gradient vanishing problem that makes the loss function difficult to converge.

To explore the number of middle layers in encoder, decoder, and discriminator, the number of middle layers is gradually increased, and MAE_1 and $RMSE$ of the test set in different middle layers are counted, as shown in Fig. 7.

Obviously, MAE_1 and $RMSE$ first become smaller and then larger with the increase of the number of middle layers in the encoder, decoder, and discriminator, which shows that the number of middle layers cannot be too small or too large. The appropriate number of middle layers for the en-

coder and decoder is between 2 and 4, and the appropriate number of intermediate layers for the discriminator is between 1 and 4.

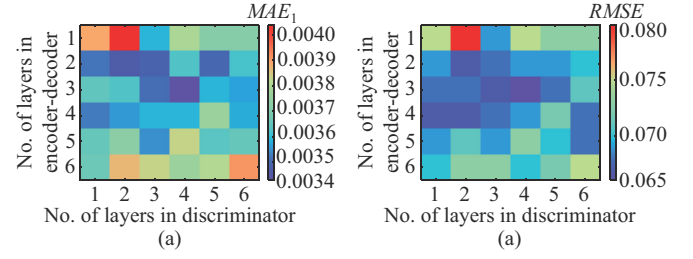


Fig. 7. MAE_1 and $RMSE$ of test set under different middle layers. (a) MAE_1 . (b) $RMSE$.

In order to analyze the influence of the weight λ on the performance of CE, the size of λ is gradually increased, and the MAE_1 and $RMSE$ of the test set are calculated, as shown in Fig. 8.

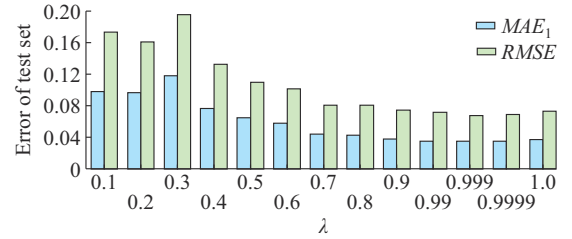


Fig. 8. MAE_1 and $RMSE$ of test set.

When λ is less than 0.9, the performance of CE improves with the increase of λ . Furthermore, if λ is larger than 0.9, MAE_1 and $RMSE$ decrease first and then increase with the increase of λ . The optimal size of λ is 0.999 for this dataset.

In order to find the appropriate batch size of the CE, the batch size is gradually increased, and MAE_1 and $RMSE$ of the test set in different batch sizes are counted, as shown in Table I.

TABLE I
 MAE_1 AND $RMSE$ OF TEST SET WITH DIFFERENT BATCH SIZES

Batch size	MAE_1 (p.u.)	$RMSE$ (p.u.)	Batch size	MAE_1 (p.u.)	$RMSE$ (p.u.)
8	0.036	0.070	64	0.037	0.075
16	0.035	0.068	128	0.045	0.089
32	0.039	0.080	257	0.047	0.093

As the batch size increases, MAE_1 and $RMSE$ of the CE first decrease and then increase. When the batch size is 16, the MAE_1 and $RMSE$ are the smallest, which indicates that the performance of missing data imputation is the best. In general, a too large batch size leads to poor generalization, and the model with a too small batch size has difficulties in converging to the global optimal solution. Sixteen is a good starting point for batch size, and larger or smaller values may be fine for some datasets.

After initializing the above parameters, a gradient descent

method is employed to optimize the loss function of the CE. Normally, the popular gradient descent methods include Adamax, Adam, Nadam, RMSprop, Adadelta, SGD, and Adagrad, which are used as black boxes in deep learning libraries (e.g., Keras and Tensorflow). To show how to choose an appropriate optimizer for the CE in missing data imputation of wind farms, the above-mentioned optimizers are set up and simulated, and then MAE_1 and $RMSE$ of the test set are calculated, as shown in Table II.

TABLE II
 MAE_1 AND $RMSE$ OF TEST SET IN DIFFERENT OPTIMIZERS

Optimizer	MAE_1 (p.u.)	$RMSE$ (p.u.)	Optimizer	MAE_1 (p.u.)	$RMSE$ (p.u.)
Adadelta	0.191	0.223	Nadam	0.036	0.069
Adagrad	0.152	0.181	RMSprop	0.039	0.072
Adam	0.035	0.068	SGD	0.154	0.180
Adamax	0.037	0.071			

Obviously, the CE has good performance when Adamax, Nadam, RMSprop, and Adam are used as optimizers. Specifically, the MAE_1 and $RMSE$ of the Adam algorithm are slightly smaller than those of the first three algorithms, which indicates that the Adam algorithm is the most suitable optimizer for the CE in missing data imputation of wind farms. Furthermore, the $RMSE$ values of the SGD, Adadelta, and Adagrad algorithms are all larger than 0.18, which shows that they are not suitable for missing data imputation based on the CE.

C. Comparison of Different Methods for Data in Complete Random Missing Forms

To illustrate the effectiveness of the CE, the existing methods such as the Cubic interpolation, BP, KNN, K -means, CGAN, and AE are used as benchmarks. The proportion of missing data in each sample is 5% and the missing data belong to the complete random missing forms. The controlled

variable method in [33] is used to find the best parameters and structure of these methods as follows.

1) For Cubic interpolation, the function `interp1` from MATLAB 2018a is used to obtain the missing data of the wind farm.

2) For BP, the middle layers consist of three dense layers, and the numbers of neurons are 10, 15, and 5, respectively. The maximum number of epochs is 200. The learning rate is 0.1, and the performance goal is 0.00004. The neural fitting toolbox from MATLAB 2018a is used to obtain the missing data of the wind farm.

3) For KNN and K -means, the size of K is adjusted adaptively according to the error of the training set.

4) For AE, its structure is consistent with the encoder-decoder pipeline of the CE, and its loss function is a reconstructed loss. The maximum number of epochs is 500. Other parameters are the same as those of the CE.

5) For CGAN, its framework and loss function can be found in [34]. Besides, the generator is similar to the encoder-decoder pipeline of the CE, and the discriminator is as that of the CE.

The above-mentioned various algorithms are repeatedly tested and the average filling errors of the test set are presented in Table III and Fig. 9, which visualizes the median, interquartile ranges with the box plot, and analyzes the full probability distribution of filling errors with the violin plot.

TABLE III
ERRORS OF DIFFERENT METHODS

Method	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)
AE	0.030	0.082	0.801
CE	0.035	0.068	0.567
K -means	0.204	0.275	0.975
BP	0.048	0.100	0.879
KNN	0.075	0.135	1.011
Cubic interpolation	0.018	0.077	0.995
CGAN	0.053	0.097	0.590

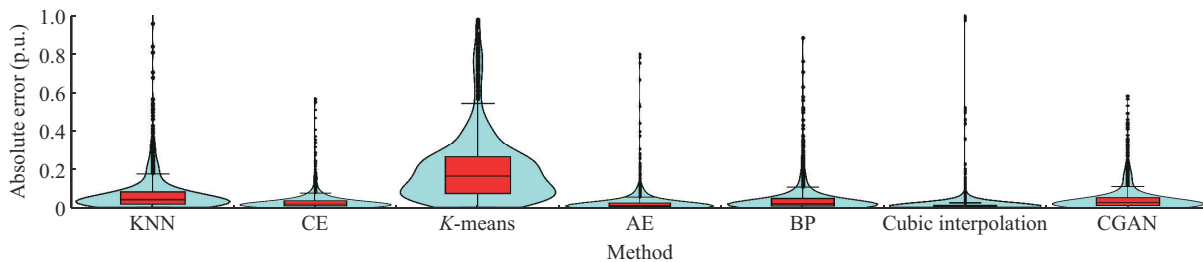


Fig. 9. Absolute errors of different methods.

The following conclusions can be drawn from Fig. 9 and Table III.

1) Although the principles of KNN and K -means are simple and easy to be applied, their absolute error is very large. For example, the upper quartile of KNN is greater than 0.18, and the upper quartile of K -means is larger than 0.5, while the upper quartile of CE is less than 0.1. In addition, the maximum absolute errors of KNN and K -means are also far larger than that of CE.

2) The CE has the similar upper quartile, median, and lower quartile with AE, BP, and CGAN, which are slightly inferior to those of Cubic interpolation. Moreover, the maximum absolute error of CE is much smaller than AE, BP, and Cubic interpolation. Specifically, the maximum absolute errors of CE, AE, BP, and Cubic interpolation are 0.567, 0.801, 0.879, and 0.995, respectively.

3) Most of the structures of CE and AE are the same, but CE has one more discriminator and adversarial loss function

than AE. Comparing the absolute errors of CGAN, AE, and CE, it is found that the discriminator and adversarial loss are helpful to reduce the maximum absolute error of the model. In addition, the MAE_1 , $RMSE$, and MAE_2 of CE are smaller than those of CGAN, which shows that the framework and loss function of CE are more suitable for missing data imputation of wind farms than CGAN.

The analysis of variance is a popular statistical method, which is often used to test the significance of mean differences between two groups. Moreover, the analysis of variance is employed to analyze differences between the results obtained by CE and other methods. The probability values (p -values) between CE and other methods are shown in Table IV.

TABLE IV
PROBABILITY VALUES BETWEEN CE AND OTHER METHODS

Method	p -value	Method	p -value
KNN	3×10^{-24}	BP	4×10^{-5}
K -means	1×10^{-139}	Cubic interpolation	4×10^{-7}
AE	0.162	CGAN	2×10^{-9}

The following conclusions can be drawn from Table IV.

1) The small p -values of KNN, K -means, BP, CGAN, and Cubic interpolation indicate that the mean differences between CE and other methods are significant, which is consistent with the simulation results in Table III.

2) Although the mean differences between CE and AE are not significant, the maximum absolute error of CE is much smaller than that of AE, as can be seen from Fig. 9.

In order to visually compare the differences of various algorithms, a sample from the test set is randomly selected and visualized as shown in Fig. 10.

The following conclusions can be drawn from Fig. 10.

1) For the attributes with small changes (e.g., wind speed, wind direction, air temperature, surface air pressure, and density), the performances of KNN and K -means are significantly inferior to those of other algorithms, since the similarity-based methods ignore the temporal correlation of attributes, while the AE, CE, BP, CGAN, and Cubic interpolation make good use of the surrounding context of the missing data, resulting in high accuracy.

2) For the attributes with large changes (e.g., wind power), Cubic interpolation no longer shows slightly better performance than CE. For example, there is a large peak in wind power at 10 p.m. as shown in Fig. 10(f). In this case, the accuracy of Cubic interpolation that only fills in missing data based on surrounding context information is very limited, since there may be a great difference in the information between the previous time and later time due to the strong fluctuation of wind power curves. Relatively, CE not only takes into account the overall information of the wind power curve, but also considers the correlation among multiple factors, which lead to higher accuracy of missing data imputation than those of other algorithms for fast ramps.

In addition to filling accuracy, the cost-effectiveness of the proposed method should be further discussed. Therefore, Table V shows the running time of each method.

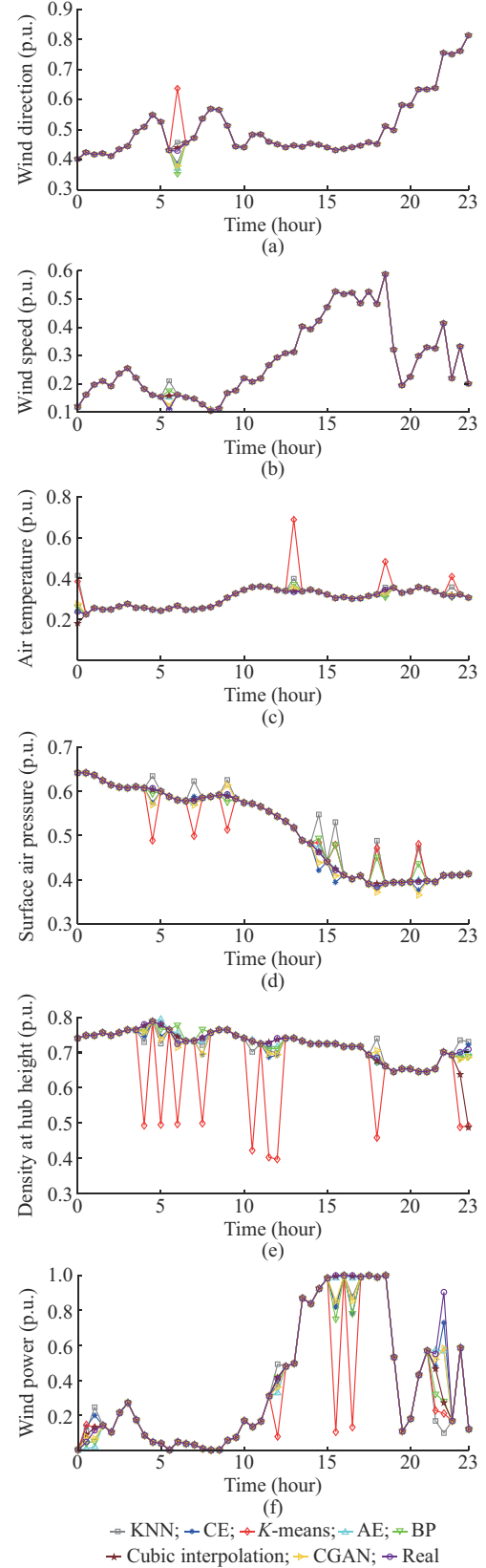


Fig. 10. Visualization of each attribute in one day. (a) Wind direction. (b) Wind speed. (c) Air temperature. (d) Surface air pressure. (e) Density at hub height. (f) Wind power.

The following conclusions can be drawn from Table V.

TABLE V
RUNNING TIME OF EACH METHOD

Method	Offline time (s)	Real time (s)	Method	Offline time (s)	Real time (s)
BP	19.45	134.47	KNN	0.00	0.27
AE	190.26	0.23	K-means	0.00	0.65
CE	71.53	0.15	Cubic interpolation	0.00	1.96
CGAN	85.60	0.15			

1) Different from the interpolation-based methods and similarity-based methods, the deep neural network-based methods need to train the model in advance. Hence, the pre-training time of the BP, AE, CE, and CGAN is larger than 0, while the pre-training time of the KNN, K-means, and Cubic interpolation is equal to 0. Although these deep neural net-

works need to be pre-trained, the pre-training time is less than 4 min, which is acceptable.

2) Furthermore, the real-time calculation of the BP is longer than those of other methods, since it needs to employ the model more frequently to predict missing values in the sample. Except for the BP, the real-time calculations of other deep neural networks are slightly less than those of KNN, K-means, and Cubic interpolation.

D. Correlation Between Filling Accuracy and Missing Data Scale

To explore the correlation between the missing data scale and the filling accuracy, the proportion of complete random missing data scale in each sample is set to be 10%, 20%, 30%, 40%, and 50%, respectively. The starting position of missing data is random. Then, the various algorithms are repeatedly tested and the filling errors of the test set are shown in Table VI and Fig. 11.

TABLE VI
FILLING ERRORS OF DIFFERENT METHODS FOR COMPLETE RANDOM MISSING DATA

Method	Missing rate is 10%			Missing rate is 20%			Missing rate is 30%			Missing rate is 40%			Missing rate is 50%		
	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)
AE	0.031	0.083	0.904	0.030	0.078	0.900	0.033	0.085	1.019	0.038	0.093	0.980	0.040	0.098	0.962
CE	0.037	0.071	0.619	0.038	0.075	0.606	0.041	0.078	0.590	0.044	0.083	0.615	0.045	0.085	0.652
K-means	0.191	0.265	0.939	0.189	0.260	0.956	0.198	0.269	0.968	0.198	0.268	0.960	0.194	0.266	0.963
BP	0.047	0.104	0.954	0.046	0.098	0.941	0.046	0.096	0.860	0.047	0.096	0.851	0.047	0.097	0.974
KNN	0.080	0.130	1.068	0.085	0.143	1.337	0.094	0.147	1.257	0.100	0.153	1.388	0.107	0.166	1.448
Cubic interpolation	0.018	0.074	0.991	0.021	0.144	8.053	0.021	0.126	8.053	0.029	0.226	9.965	0.033	0.253	13.031
CGAN	0.042	0.093	0.692	0.043	0.089	0.694	0.047	0.083	0.727	0.049	0.091	0.713	0.050	0.092	0.758

The following conclusions can be drawn from Table VI and Fig. 11.

1) For different missing rates, MAE_1 , $RMSE$, and MAE_2 of the CE are always smaller than those of the K-means, KNN, CGAN, and BP, which shows that CE has higher filling accuracy. Specifically, MAE_1 of CE is always smaller than 0.045, $RMSE$ of CE is always smaller than 0.085, and MAE_2 of CE is always smaller than 0.652.

2) MAE_1 evaluates the mean magnitude of the filling errors without considering their direction. In addition, it is a linear index where all the individual errors are weighted equally in the average. Relatively, $RMSE$ makes larger errors with higher weights, which means that $RMSE$ is more useful if larger errors are particularly undesirable. Specifically, MAE_1 of AE and Cubic interpolation is slightly smaller than that of CE, but their $RMSE$ is greater than that of CE, which indicates that the maximum filling errors of AE and Cubic interpolation are much larger than that of CE. In other words, CE controls the maximum filling error better than AE and Cubic interpolation.

3) As the missing rate increases, MAE_2 of the Cubic interpolation increases rapidly, indicating that it is very sensitive to the scales of missing data and is only suitable for the dataset of wind farms with a small amount of missing data. In

contrast, MAE_2 of other methods slowly increases with the enlargement of the missing data scale, which demonstrates that they are also suitable for datasets with high missing rates.

E. Performance of CE for Continuous Missing Forms

Due to cyber-attack or communication congestion, the SCADA system may have continuous missing data in a period. To test the performance of the proposed method for continuous missing data imputation of the wind farm, the proportion of missing data scale in each sample is set to be 10%, 20%, 30%, 40%, and 50%, respectively. Then, the various algorithms are repeatedly tested and the filling errors of the test set are shown in Table VII.

The following conclusions can be drawn from Table VII.

1) When the missing rate is 10%, MAE_1 of CE is better than those of AE, K-means, CGAN, and KNN, but slightly inferior to those of BP and Cubic interpolation. In addition, $RMSE$ and MAE_2 of CE are smaller than other algorithms. This phenomenon shows that if one wants to choose an algorithm with small MAE_1 , $RMSE$, and MAE_2 to fill the continuous missing data of wind farms, CE, AE, CGAN, and BP are very suitable. In particular, CE makes a trade-off between MAE_1 and MAE_2 .

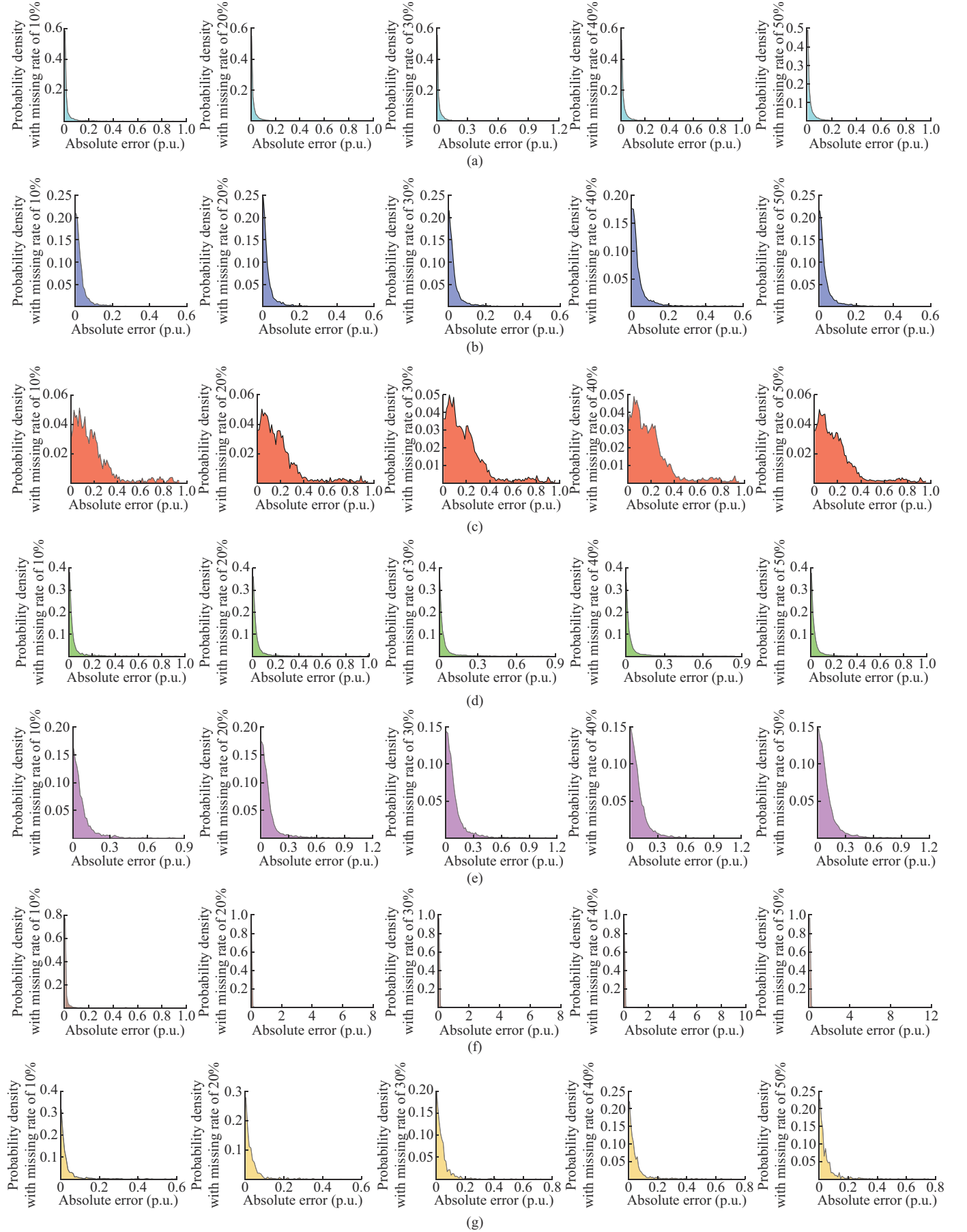


Fig. 11. Visualization of absolute error for each attribute. (a) AE. (b) CE. (c) *K*-means. (d) BP. (e) KNN. (f) Cubic interpolation. (g) CGAN.

2) Similar to the simulation results of data in complete random missing form, the filling accuracy of various algorithms decreases with the increase of the missing rate. Cubic interpolation is also very sensitive to the missing rate of continuous missing data. MAE_2 of Cubic interpolation is far larger than the existing maximum value of various attributes, which makes Cubic interpolation not suitable for filling continuous missing data for wind farms. For example, when the missing rate is 50%, MAE_2 of Cubic interpolation is 259.131

p.u., which is unacceptable.

3) In most cases, MAE_1 , $RMSE$, and MAE_2 of CE are smaller than those of other algorithms for high missing rates, which indicate that CE is more suitable for filling continuous missing data with a high missing rate than other algorithms. In addition to CE, the AE, CGAN, and BP perform slightly worse than CE, because they can also consider multiple factors, temporal correlation, and surrounding context information as well as CE.

TABLE VII
FILLING ERRORS OF DIFFERENT METHODS FOR CONTINUOUS MISSING DATA

Methods	Missing rate is 10%			Missing rate is 20%			Missing rate is 30%			Missing rate is 40%			Missing rate is 50%		
	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)	MAE_1 (p.u.)	$RMSE$ (p.u.)	MAE_2 (p.u.)
AE	0.082	0.120	0.766	0.096	0.141	0.867	0.097	0.142	0.971	0.103	0.147	0.847	0.105	0.149	0.987
CE	0.053	0.099	0.593	0.063	0.124	0.611	0.065	0.133	0.633	0.098	0.164	0.659	0.075	0.144	0.651
<i>K</i> -means	0.187	0.265	0.951	0.186	0.262	0.962	0.193	0.264	0.948	0.220	0.288	0.978	0.195	0.265	0.968
BP	0.048	0.109	0.911	0.074	0.142	0.957	0.097	0.169	0.898	0.072	0.135	0.921	0.094	0.161	0.978
KNN	0.084	0.136	0.976	0.083	0.145	1.137	0.102	0.173	1.313	0.095	0.162	1.366	0.103	0.174	1.182
Cubic interpolation	0.038	0.235	8.765	0.087	0.641	21.247	0.166	1.329	39.582	0.180	2.339	112.330	1.132	9.257	259.131
CGAN	0.064	0.106	0.646	0.070	0.130	0.692	0.071	0.113	0.715	0.103	0.143	0.738	0.089	0.148	0.727

In order to visually compare the performance of various algorithms for continuous missing data of the wind farm, Fig. 12 visualizes a sample selected from the test set where the missing rate is 10%.

The following conclusions can be drawn from Fig. 12.

1) Since KNN and *K*-means do not consider the context information around the missing data, their filling accuracies are also very poor for the continuous missing data of wind farms. When the position of missing data is at the head or tail of the time series, the filling accuracy of Cubic interpolation will be unacceptable. In addition, it is also difficult to capture the hallmark characteristics (e.g., fast ramps) of the time series, while CE, AE, CGAN, and BP have certain adaptability to these rapid changes. For example, the air temperature in Fig. 12(c) has a large valley between 5 a.m. and 10 a.m., which is captured by CE, AE, CGAN, and BP.

2) As shown in Fig. 12(b) and (f), the filling error of Cubic interpolation will become very large if the missing values are at the beginning and end of the sample. In contrast, CE can adapt to the situation of missing values in different positions, and always keep a low filling error.

V. DISCUSSION

The objective of this paper is to propose a data-driven method to fill missing data of wind farms via the CE. Moreover, the effectiveness of the proposed CE has been tested on a real-world dataset from the renewable energy lab of the United States. The simulation results show that the CE achieves state-of-art performance with superior accuracy for miss data imputation of attributes with large changes (e.g., large peaks, large valleys, and fast ramps). However, the CE has similar upper quartile, median, and lower quartile with AE, CGAN, and BP for the attributes with small changes

(e.g., wind speed, wind direction, air temperature, surface air pressure, and density), which are slightly inferior to those of Cubic interpolation. Furthermore, the CE and Cubic interpolation may be integrated to a hybrid model to achieve the highest filling accuracy for both attributes with large changes and attributes with small changes in wind farms.

Besides, the application of the CE is not limited to missing data imputation of wind farms. For example, GCN may also be suitable for scenario generations of renewable energy sources and power loads of distribution networks by fine-tuning the structures and parameters of the model.

VI. CONCLUSION

Missing data imputation of wind farms is of great significance for wind power forecasting. In order to improve the accuracy of missing data imputation for wind farms, a new data-driven, model-free, and scalable method is proposed in this paper. Through the simulation and analysis on a real-world dataset, the following conclusions are obtained.

1) The number of iterations, the choices of optimizer and batch size, the number of middle layers, and the weights of adversarial loss and reconstructed loss have a great influence on the performance of missing data imputation. Specifically, the training process of CE is relatively stable, and there is no gradient vanishing problem in CE. The appropriate numbers of middle layers for the encoder, decoder, and discriminator range from 2 to 4. The Adam algorithm is more suitable to be the optimizer of CE than other algorithms. As the batch size increases, the filling error of the CE first decreases and then increases. When the batch size is 16, the accuracy is the largest. The larger λ is, the smaller the filling accuracy will be. When λ is equal to 0.999, CE has an outstanding performance.

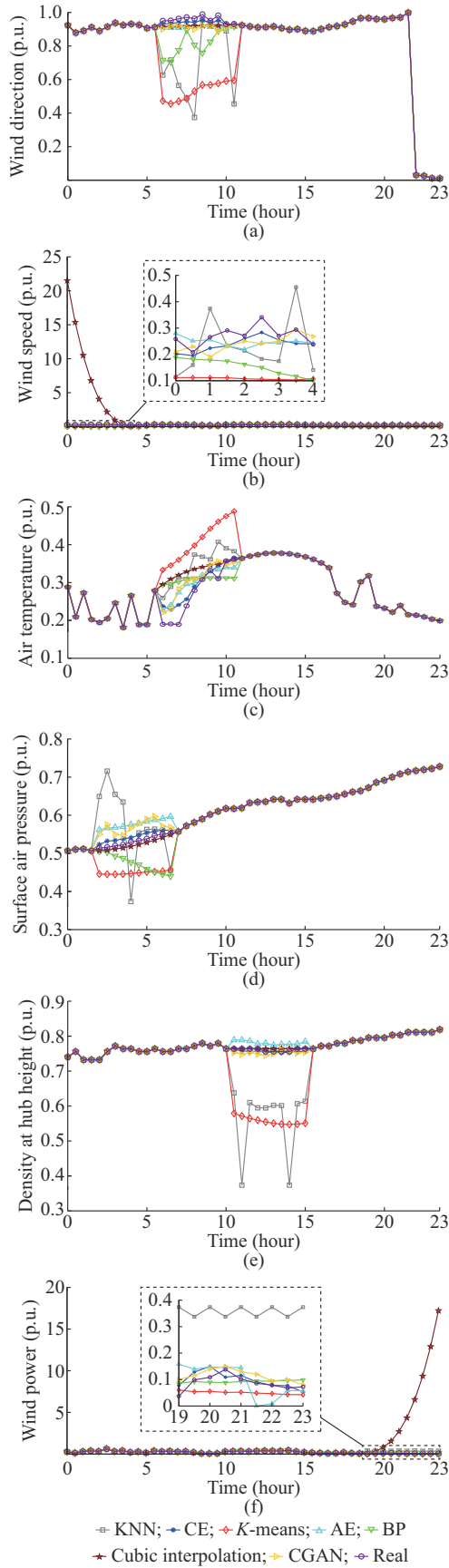


Fig. 12. Visualization of each attribute with missing rate of 10%. (a) Wind direction. (b) Wind speed. (c) Air temperature. (d) Surface air pressure. (e) Density at hub height. (f) Wind power.

2) The performances of KNN and K-means are significantly inferior to those of other algorithms, since the similarity-based methods ignore the temporal correlation of attributes, while the AE, CE, BP, CGAN, and Cubic interpolation make good use of the surrounding context of the missing data, resulting in high accuracy. In addition, Cubic interpolation is very sensitive to the missing rate, which means that its maximum absolute error will be very large for datasets with large-scale missing values. Relatively, MAE_1 , $RMSE$, and MAE_2 of CE are smaller than those of other algorithms for high missing rate, which indicates that CE is significantly more suitable for filling continuous data with a high missing rate than other algorithms.

3) Interpolation-based methods and similarity-based methods have difficulties in capturing the hallmark characteristics (e.g., large peaks, large valleys, and fast ramps) of the time series, while CE has certain adaptability to these rapid changes. Most of the structures of CE and AE are the same, but CE shows a better performance than that of AE, which indicates that the discriminator and adversarial loss are significantly helpful to reduce the maximum absolute error of the model.

4) Although the CE needs to be pre-trained, the pre-training time is about 71.25 s, which is acceptable. Furthermore, the real-time calculations of CE are slightly less than those of KNN, K-means, and Cubic interpolation.

REFERENCES

- [1] Q. Zhao, W. Liao, S. Wang et al., "Robust voltage control considering uncertainties of renewable energies and loads via improved generative adversarial network," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1104-1114, Nov. 2020.
- [2] Q. Li, L. Cheng, W. Gao et al., "Fully distributed state estimation for power system with information propagation algorithm," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 627-635, Jul. 2020.
- [3] Q. Lin and J. Wang, "Vertically correlated echelon model for the interpolation of missing wind speed data," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 804-812, Mar. 2014.
- [4] Y. Hu, Y. Qiao, J. Liu et al., "Adaptive confidence boundary modeling of wind turbine power curve using SCADA data and its application," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 804-812, Aug. 2019.
- [5] J. Yoon, W. Zame, and M. Schaar, "Estimating missing data in temporal data streams using multi-directional recurrent neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1477-1490, May 2019.
- [6] Y. Mao and M. Jian, "Data completing of missing wind power data based on adaptive BP neural network," in *Proceedings of 2016 International Conference on Probabilistic Methods Applied to Power Systems*, Beijing, China, Oct. 2016, pp. 1-6.
- [7] T. Li, J. Tang, F. Jiang et al., "Fill missing data for wind farms using long short-term memory based recurrent neural network," in *Proceedings of 2019 IEEE 3rd International Electrical and Energy Conference*, Beijing, China, Sept. 2019, pp. 705-709.
- [8] W. Deng, Y. Guo, J. Liu et al., "A missing power data filling method based on improved random forest algorithm," *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, pp. 33-39, Dec. 2019.
- [9] P. Jonsson and C. Wohlin, "An evaluation of k-nearest neighbour imputation using Likert data," in *Proceedings of 10th International Symposium on Software Metrics*, Chicago, USA, Nov. 2004, pp. 108-118.
- [10] J. Du, M. Hu, and W. Zhang, "Missing data problem in the monitoring system: a review," *IEEE Sensors Journal*, vol. 20, no. 23, pp. 13984-13998, Dec. 2020.
- [11] P. Shi and L. Zhang, "A missing data complement method based on K-means clustering analysis," in *Proceedings of 2017 IEEE Conference on Energy Internet and Energy System Integration*, Beijing, China, Nov. 2017, pp. 1-5.

- [12] S. Tak, S. Woo, and H. Yeo, "Data-driven imputation method for traffic data in sectional units of road links," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1762-1771, Jun. 2016.
- [13] Y. Wang, Y. Sun, Z. Wei *et al.*, "Parameters estimation of electromechanical oscillation with incomplete measurement information," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5016-5028, Sept. 2018.
- [14] N. Eklund, "Using genetic algorithms to estimate confidence intervals for missing spatial data," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 36, no. 4, pp. 519-523, Jul. 2016.
- [15] J. Liu, S. Kumar, and D. Palomar, "Parameter estimation of heavy-tailed ar model with missing data via stochastic EM," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 1762-1771, Apr. 2019.
- [16] S. Zargar, M. Farsangi, and M. Zare, "Probabilistic multi-objective state estimation-based PMU placement in the presence of bad data and missing measurements," *IET Generation, Transmission & Distribution*, vol. 14, no. 15, pp. 3042-3051, Aug. 2020.
- [17] D. Cao, W. Hu, J. Zhao *et al.*, "Reinforcement learning and its applications in modern power and energy systems: a review," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1029-1042, Nov. 2020.
- [18] M. Lippi, M. Montemurro, M. Esposti *et al.*, "Natural language statistical features of LSTM-generated texts," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3326-3337, Nov. 2019.
- [19] S. Ryu, M. Kim, and H. Kim, "Denoising autoencoder-based missing value imputation for smart meters," *IEEE Access*, vol. 8, pp. 40656-40666, Feb. 2020.
- [20] S. Wang, H. Chen, Z. Pan *et al.*, "A reconstruction method for missing data in power system measurement using an improved generative adversarial network," *Proceedings of the CSEE*, vol. 33, no. 1, pp. 56-64, Jan. 2019.
- [21] F. Qu, J. Liu, Y. Ma *et al.*, "A novel wind turbine data imputation method with multiple optimizations based on GANs," *Mechanical Systems and Signal Processing*, vol. 139, pp. 1-16, May 2020.
- [22] F. Qu, J. Liu, Y. Ma *et al.*, "Data imputation of wind turbine using generative adversarial nets with deep learning models," in *Proceedings of International Conference on Neural Information Processing*, Siem Reap, Cambodia, Dec. 2018, pp. 152-161.
- [23] T. Do, Q. Nguyen, and V. Nguyen, "A multi-scale context encoder for high quality restoration of facial images," in *Proceedings of 2020 International Conference on Multimedia Analysis and Pattern Recognition*, Ha Noi, Vietnam, Oct. 2020, pp. 1-6.
- [24] Y. Yu, V. Li, J. Lam *et al.* (2020, Mar.). Missing air pollution data recovery based on long-short term context encoder. [Online]. Available: <https://ieeexplore.ieee.org/document/9028215>
- [25] Z. Gu, J. Cheng, H. Fu *et al.*, "CE-Net: context encoder network for 2D medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281-2292, Oct. 2019.
- [26] A. Marafioti, N. Perraudin, N. Holighaus *et al.*, "A context encoder for audio inpainting," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2362-2372, Dec. 2019.
- [27] L. Liao, R. Hu, and J. Xiao, "Edge-aware context encoder for image inpainting," in *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Canada, Apr. 2018, pp. 3156-3160.
- [28] J. Choung, S. Lim, S. H. Lim *et al.*, "Automatic discontinuity classification of wind-turbine blades using a-scan-based convolutional neural network," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 1, pp. 210-218, Jan. 2021.
- [29] D. Pathak, P. Krähenbühl, J. Donahue, *et al.*, "Context encoders: feature learning by inpainting," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, Dec. 2016, pp. 2536-2544.
- [30] S. Hosseini, C. Tang, and J. Jiang, "Calibration of a wind farm wind speed model with incomplete wind data," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 1, pp. 343-350, Jan. 2014.
- [31] L. Ge, W. Liao, S. Wang *et al.*, "Modeling daily load profiles of distribution network for scenario generation using flow-based generative network," *IEEE Access*, vol. 8, pp. 77587-77597, Apr. 2020.
- [32] C. Draxl, A. Clifton, B. Hodge *et al.*, "The wind integration national dataset (WIND) toolkit," *Applied Energy*, vol. 151, pp. 355-366, Aug. 2015.
- [33] X. Kong, F. Zheng, and Z. E., "Short-term load forecasting based on deep belief network," *Automation of Electric Power Systems*, vol. 42, no. 5, pp. 133-139, Jan. 2018.
- [34] Z. Wang, D. Wang, Q. Duan *et al.*, "Missing load situation reconstruction based on generative adversarial networks," in *Proceedings of IEEE/IAS Industrial and Commercial Power System Asia*, Weihai, China, Sept. 2020, pp. 1528-1534.

Wenlong Liao received the B.S. degree from China Agricultural University, Beijing, China, in 2017. He received the M.S. degree from Tianjin University, Tianjin, China, in 2020. He is currently pursuing the Ph.D. degree in Aalborg University, Aalborg, Denmark. His current research interests include smart grids, machine learning, and renewable energy.

Birgitte Bak-Jensen received the M.Sc. degree in electrical engineering and the Ph.D. degree in modeling of high voltage components from the Institute of Energy Technology, Aalborg University, Aalborg, Denmark, in 1986 and 1992, respectively. She is currently a Professor with the Department of Energy Technology, Aalborg University. Her current research interests include analysis of distribution system, smart grids, and intelligent energy systems.

Jayakrishnan Radhakrishna Pillai received the M.Tech. degree in power systems from the National Institute of Technology, Calicut, India, in 2005, the M.Sc. degree in sustainable energy systems from the University of Edinburgh, Edinburgh, U.K., in 2007, and the Ph.D. degree in power systems from Aalborg University, Aalborg, Denmark, in 2011. He is currently an Associate Professor with the Department of Energy Technology, Aalborg University. His current research interests include analysis of distribution system, grid integration of electric vehicles and distributed energy resources, smart grids, and intelligent energy systems.

Dechang Yang received the M.S. degree in electric engineering from China Agricultural University, Beijing, China, in 2008, and the Ph.D. degree in TU Dortmund Graz, Dortmund, Germany, in 2012. He is currently an Associate Professor in the China Agricultural University. His research interests are state estimation of active distribution network and planning analysis of regional energy systems.

Yusen Wang received the B.S. degree from China Agricultural University, Beijing, China, in 2017, and the M.S. degree from KTH Royal Institute of Technology, Stockholm, Sweden, in 2019. He is currently pursuing the Ph.D. degree in KTH Royal Institute of Technology. His current research interests include machine learning and smart grids.