Applied Energy 238 (2019) 249–257

Contents lists available at ScienceDirect

Applied Energy

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

# Scene learning: Deep convolutional networks for wind power prediction by embedding turbines into grid space

Ruiguo Yu[a,b,c], Zhiqiang Liu[a,b,c], Xuewei Li[a,b,c], Wenhuan Lu[a,b,c], Degang Ma[d], Mei Yu[a,b,c,*], Jianrong Wang[a,b,c], Bin Li[e]

[a] College of Intelligence and Computing, Tianjin University, China
[b] Tianjin Key Laboratory of Cognitive Computing and Application, China
[c] Tianjin Key Laboratory of Advanced Networking, China
[d] School of Environmental Science and Engineering, Tianjin University, China
[e] School of Electrical and Information Engineering, Tianjin University, China

## HIGHLIGHTS

- The spatio-temporal feature is proposed, instead of traditional feature, time series.
- Convolutional network is used to predict wind power based on spatio-temporal feature.
- Much higher accuracy is achieved within much less training time than existing works.

## ARTICLE INFO

*Keywords:*
Wind
Embedding
Spatio-temporal feature
Prediction
Convolutional networks

## ABSTRACT

Wind power prediction is of vital importance in wind power utilization. There have been a lot of researches based on the time series of the wind power or speed. But in fact, these time series cannot express the temporal and spatial changes of wind, which fundamentally hinders the advance of wind power prediction. In this paper, a new kind of feature that can describe the process of temporal and spatial variation is proposed, namely, spatio-temporal feature. We first map the data collected at each moment from the wind turbines to the plane to form the state map, namely, the scene, according to the relative positions. The scene time series over a period of time is a multi-channel image, i.e. the spatio-temporal feature. Based on the spatio-temporal features, the deep convolutional network is applied to predict the wind power, achieving a far better accuracy than the existing methods. Compared with the state-of-the-art methods, the mean-square error in our method is reduced by 49.83%, and the average time cost for training models can be shortened by a factor of more than 150.

## 1. Introduction

Wind power has become a significant renewable resource that can be developed and utilized on a large scale [1]. Thanks to the mass production of equipment, wind power has turned to be the fastest growing renewable energy in the world. By 2017, the worldwide wind power installed capacity has reached 539 GW, and 52 GW was added in 2017 [2], thus making wind power expected to be one of the major power sources in the 21st century. However, due to the influence of wind speed and direction, randomness and volatility of wind turbines can not be avoided, bringing severe challenges to the safety and stability of the operation of power systems [3]. Accurate wind power prediction can enhance the controllability of wind power, ensure the stable operation of the power grid, and promote the ability of the grid to accept wind power.

Smart grid [4] is a topic of great concern in recent years [5], and wind power forecasting technology is conducive to smart grid. At present, scholars have done a lot of related researches, including physical methods [6], statistical methods [7] and machine learning methods. Among them, machine learning methods, including support vector machine regression (SVR) [8], k-nearest neighbor regression (kNN) [9] or multi-layer perceptron neural network (MLP) [10] are used to model wind speed time series or power time series to achieve prediction. Machine learning methods simplify the wind power forecasting

* Corresponding author.
*E-mail addresses:* rgyu@tju.edu.cn (R. Yu), tjubeisong@tju.edu.cn (Z. Liu), lixuewei@tju.edu.cn (X. Li), wenhuan@tju.edu.cn (W. Lu), dgma@tju.edu.cn (D. Ma), yumei@tju.edu.cn (M. Yu), wjr@tju.edu.cn (J. Wang), binli@tju.edu.cn (B. Li).

https://doi.org/10.1016/j.apenergy.2019.01.010
Received 16 October 2018; Received in revised form 6 December 2018; Accepted 1 January 2019

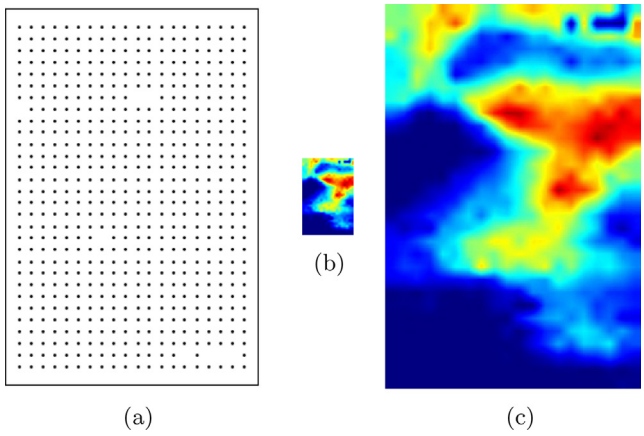


**Fig. 1.** (a) The image produced by scaling down the real coordinates. White pixels indicate blanks and black pixels indicate wind motors. Black pixels are extremely sparse, that is, the ratio of effective pixels in the picture is very low. (b) The *scene* produced by the proposed algorithm. (c) Made by (b) through bidirectional, and is used to show details.

problem, but the accuracy rate has failed to be improved in the past several years.

We think that the wind is temporal and spatial correlation process, however, the time series can only express the information at the time level, but say nothing at the space level, let alone the spatio-temporal process of air flow, thus fundamentally standing in the way of the progress of wind power prediction. Therefore, finding the features that can better express the state of the wind farm is the key to breaking through the bottleneck of accuracy.

Such being the case, this paper put forward a new feature that can express the spatio-temporal process of air flow, called spatio-temporal feature (STF). The scene time series over a period of time is a multi-channel image, in which each scene is a sample of the true distribution of physical data in space, expressing spatial-related information, as shown in Fig. 1. The scene sequence represents the change of wind farm state over time, expressing time information, so the multi-channel image is called spatio-temporal feature. Compared with wind speed or power series, STF implies factors such as wind speed, wind direction and air density, which greatly expands the ability to express wind-related information and lays a foundation for breaking through the bottleneck of wind power prediction accuracy.

Based on the STF, the spatio-temporal process of the wind farm is simulated and predicted by using the deep convolutional network, which has achieved good effects. The experimental results on two wind farms with 592 wind turbines (farm1) and 454 wind turbines (farm2) respectively show that, the proposed methods are better than the existing state-of-the-art series modeling methods, for the reason that the *MSE* of the proposed method decreases by an average of 26.69% and 49.83% at most in wind farm1, and by an average of 24.37% and 46.94% at most in wind farm2, and the time for training models is both shortened by more than 150 times.

The innovations of this paper are as follows:

- The spatio-temporal feature in the form of the multichannel image is proposed for the first time by embedding wind turbines into the grid space, which fully expresses the spatio-temporal variation process of the air flow and can perfectly combine with the most advanced theory of deep learning at present.
- The convolutional neural network is reasonably used to predict wind power for the first time based on the spatio-temporal feature, it can predict the wind power of a large number of turbines in parallel. And both the accuracy and time cost of the prediction have been greatly optimized.

## 2. Related work

### 2.1. Machine learning methods in WPP

Machine learning methods perform well in short term prediction. By means of the regression model or neural network, researchers map the time series to the wind power of the future moment, so as to make the prediction. The commonly used methods are SVR [8], kNN [9], Multilayer Perceptron Network (MLP) [10] and its variant [11] and Long and Short Term Memory Neural Network (LSTM) [12], etc., among which SVR and kNN are the representatives [13].

SVR has a perfect mathematical foundation in theory and performs best among numerous regressions. It realizes regression by finding a hyperplane to make all the data closest to the plane. This process can be abstracted as that, when Eq. (2) is satisfied, the parameter should be found to make the value of Eq. (1) minimized. In these two equations, $C$ and $\varepsilon$ are empirical parameters, $\xi_i$, and $\xi_i^*$ are called relaxation factors, and $w$ and $b$ represent hyperplanes [14].

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{1}$$

$$\begin{cases} y_i - \langle w, X_i\rangle - b & \leqslant \varepsilon + \xi_i \\ \langle w, X_i\rangle + b - y_i & \leqslant \varepsilon + \xi_i^* \\ \xi_i, \xi_i* & \geqslant 0 \end{cases} \tag{2}$$

It has been proved in many works that SVR is one of the best methods in the field of wind power prediction currently, such as stated in literature [8] and literature [15].

KNN is the most simply equipped machine learning model based on similarity metric and it still has a good performance in practice. As the similarity between the two vectors is negative correlated with the distance between them (such as Euclidean distance, Manhattan distance, etc.), the similarity can be represented by distance. The $k$ vectors in a set with the minimum distance from the target vector $x$ can be called $k$ nearest neighbors of $x$. If $\aleph_k(x)$ denotes the subscript of $k$ nearest neighbors of $x$ in the training set, then the prediction result $p(x)$ of kNN model for $x$ is generated by Eq. (3), and the function $f$ can use arithmetic average method, weighted average method or other methods that are more complex.

$$p(x) = f_{i\in\aleph_k(x)}(yi) \tag{3}$$

Based on the algorithms such as k-d tree, a kNN model can be trained very quickly.

In recent years, there have been some new ideas in this research field. Sequence decomposition is a popular idea to simplify complex problems [16], and is therefore widely used in wind data processing. For example, wavelet transform was used in literature [17] and literature [18] to decompose the power series to form multiple new subseries that would be predicted in turn. And then the results were combined [19]. Li et al. decompose the original wind speed time series into a set of modes and into one bias series. and then subsequently, use Gram-Schmidt orthogonal to select the important features in their wind speed prediction work [20]. Wang Jianzhou et al. proposed an effective decomposing technique to eliminate redundant noise [21]. Liu et al. used twice decomposing processes to process the wind data [22]. This type of methods need to build a model for each sub-series, thereby leading to much higher costs. Feature selection is also a research hotspots for there is noise in the time series. Wang Jianzhou et al. proposed to eliminate redundant noise through decomposing technique and extract the primary characteristics of wind speed data [21]. The researchers modeled the prediction error to improve the prediction effect by error analysis [23]. But the error is produced by the specific prediction model, which has limits, difficult to be applied to the production [24]. In addition, the process of error analysis increases the computational costs. Although using ensemble learning [25] to predict could

improve the accuracy, such as in literature [26] and literature [27], many models working at the same time also consume computational resources substantially [28]. Moreover, in the literature [29], the series data with the length of $p \times q$ were filled in the grid of $p \times q$ in order to achieve a two-dimensional image, following which the convolutional neural network was utilized for the prediction. But the constructed image had no explicit physical meaning. And the required time series were far too long to add the computational costs.

In sum, all of the above approaches used the series data for modeling in essence, and achieved a higher accuracy via the complicated models. However, their computational cost was largely increased and their models could not reflect the spatio-temporal variation of air.

### 2.2. Convolutional neural network

This chapter introduces the convolutional neural network (CNN) [30], which lays the foundation for the third chapter to introduce the method proposed in this paper. At present, CNN is the most successful method in deep learning that has been widely used in auxiliary medical treatment, speech recognition, intelligent city and automatic driving system [31]. Besides, CNN can speed up computing by GPU [32]. With the rapid development of hardware in recent years, the computing ability of computers has been greatly improved, thus leading the CNN model to a significant progress in many fields, such as image segmentation [33] and image recognition [34].

The central operation of CNN is the convolution. As both the input and output of convolutions are multichannel images, these images are usually called as feature maps. There are abundant types of CNN models, but as a whole they can be divided into two basic types. The first is the coding machine-decoder model, whose core operations are the convolution, pooling and deconvolution. The convolution process is to extract deep features, pooling is to narrow the size of images, and the deconvolution aims at enlarging the image size by up-sampling. FCN network [35] is typical in this method. The second type is a convolutional network with a fully connected layer, the core operations of which include the convolution, pooling and full connection. In this type of model, the convolution and pooling process produce deep features, while full connection maps deep features to predictive values. On account of the excellent expression of full connection, the model, VGGNet [36] in particular, can always fit a very complex nonlinear relation.

## 3. Proposed method

The information related to wind such as wind speed or turbine's output power can be strongly combined with convolutional networks. On one hand, convolutional networks are quite suitable to deal with data in tensor form, which can automatically extract features at different layers and realize the end-to-end learning. On the other hand, the wind turbines distributed on the plane is easy to be embedded into the grid to construct a two-dimensional tensor, namely, a matrix or grid. But the current researches have little experience in combining the both.

In this chapter, the ***scene*** and ***spatio-temporal feature (STF)*** are introduced, and then two kinds of convolutional networks models based on STF are put forward. These two models represent ideas of two main convolutional network structures respectively, we use both of them to show that STF can be combined with various convolutional networks in practical utilization.

The rest of this chapter will further elaborated on the above contents.

### 3.1. Scene and STF

Feature extraction has always been a hot topic in wind power prediction. In this paper, the feature extracted only from the data of the target turbine itself is called "single-feature (SF)", and the feature extracted from the data of the target turbine and several adjacent turbines is called "local-feature (LF)". Basically, the local-feature is an extended form of the single-feature. When the local-feature selects a distance threshold of 0 for adjacent turbines, it degenerates into the single-feature.

Most of the features used in existing works are single-features, and some researchers have also studied local-features. For example, in the literature [13], the local-feature is generated by connecting the single-feature of each turbine. The feature extracted in this way contains more information, but it is not enough, covering only the information of temporal but in devoid of the spatial information.

In order to describe the spatial distribution of wind in a certain area at a certain time, the concept of scene is put forward in this paper. We map the output power of the wind turbines at a certain time to the plane according to the geographical coordinates of the turbines to form a two-dimensional image, namely, the scene. Mapping the real coordinates to the plane is the main problem while constructing a scene. The most direct solution is to scale down the real geographic coordinates and then to draw them onto the plane, as shown in Fig. 1a. This method can successfully represent the spatial position, but the size of the constructed image is relatively large while containing only sparse effective pixels, which is not conducive to calculation. To solve this problem, this paper proposes a method to embed turbines into grids as small an area as possible, which is called *grid space embedding method*, as shown in Fig. 1b, c. In this algorithm, the longitude and latitude coordinates are firstly processed by sorting and discretization, in order to determine the shape of the scene, then we construct a grid (matrix). Each turbine is mapped to the corresponding cell of grid in the order sorted by its horizontal and vertical coordinates. More details are shown in Algorithm 1. The output is of the algorithm is the mapping matrix $G$ of turbines to cells, each position serving as the serial number of the corresponding turbine. Ineluctable, there are some cells not related to any turbine. The vacant position is filled with $-1$, and the output power of turbines at a certain time is filled into the matrix according to the position specified by $G$, then the scene corresponding to the time can be obtained, as shown in Algorithm 2.

The proposed embedding algorithm constructs the grids as small as possible to avoid invalid pixels, and the constructed scene is suitable for convolutional computation.

The scene represents the spatial distribution of wind power at a certain time. And connecting several continuous scenes in series can convey the process of spatial state changing with time. Although the air motion is complicated, it still shows certain regularity on the whole, and the scene series can reflect this regularity to some extent. In this paper, the multichannel image got by the scenes arranged in time series is named as the spatio-temporal feature (STF).

**Algorithm 1.** Algorithm for embedding wind turbines into a grid

**Input:** *ids*: Id list of turbines, *coordinates*: Coordinate list of turbines
**Output:** $G$: *ids* embeding result
1: $latitudes = coordinates[:,0]$
2: $longitudes = coordinates[:,1]$
3: $latitudes = unique(latitudes)$, $longitudes = unique(longitudes)$
4: $sorted\_latitudes = sort(latitudes)$//increasing order
5: $sorted\_longitudes = sort(longitudes)$//increasing order
6: $array\_shape = [len(latitudes), len(longitudes]$
7: $G = array(-1, shape = array\_shape)$
8: **for** $id, (x, y)$ in enumerate(*coordinates*) **do**
9: $index_x = get\_index(sorted\_latitudes, x)$
10: $index_y = get\_index(sorted\_longitudes, y)$
11: $G[index_x][index_y] = id$
12: **end for**
13: **return** $G$

**Algorithm 2.** Algorithm for constructing scene

**Input:** $G$: Output of embeding algorithm, *ids*: Id list of turbines, $I$: Information (e.g., output power) of the wind turbine at a certain time
**Output:** $S$: Scene

```
S = array(0, shape = get_shape_of(G))
for x in range(len(G)) do
  for y in range(len(G[0])) do
    if G[x][y] = =−1 then
      continue
    else
      id = G[x][y]
      S[x][y] = I[id]
    end if
  end for
end for
return S
```

Each channel of STF independently represents spatial information, and the combination of the multichannel sorting represents temporal information. It is a kind of global-feature for it can synthetically deliver the information in a large geographical area and a long time range. In fact, each channel of the STF can also be used to represent different types of information, such as wind power output, wind speed, temperature and so on. The STF, which combines many kinds of data, is called multi-spatio-temporal feature (MSTF). The STF can be processed by deep convolutional neural network. Convolution neural network is the most complete theory of deep learning at present, which, given perfect tools and frameworks, can give full play to the advantages of new technologies such as GPU acceleration.

### 3.2. E2E model

The first kind of convolutional neural network model for wind power prediction based on STF is introduced in this section, which is called E2E model, using the idea of autoEncoder [37].

After received, the input image will be handled in two stages. The first stage is down-sampling, that is, the coding stage, in which the deep features are extracted step by step and the image size is shrunk by means of multiple nested convolution layers and a pooling layer. The second is up-sampling, that is, the decoding stage, which mainly includes deconvolutional layers. By deconvolution, the size of the feature map is initially increased, and finally the output of the same size as the input image is obtained. As a result, the pixels of the input image and the pixels of the output image can be corresponded one-to-one to realize the end-to-end mapping.

In the down-sampling stage, under the guidance of the idea of "short circuit" in DenseNet, the outputs of multiple prepositive convolutional layers are connected in series, and then input to the next convolutional layer to preserve the spatial information of the original input image. Since the major task of this stage is to fully extract features, the number of channels in the feature image increases rapidly. The main task of the upper sampling stage is the fusion of features in order to produce the output. In this stage, the outputs of each convolutional layer are no longer connected in series, and the output of each deconvolution reduces the channels. In this way the single channel image is finally output. The structure of the E2E model is shown in Fig. 2.

### 3.3. FC-CNN model

The second model is a convolutional neural network containing a fully connected layer, called FC-CNN. After receiving the input image, the model also performs the operations of two stages.The first stage is similar to the down-sampling stage of E2E model, but the deeper layers are in demand in FC-CNN and the size of feature map is smaller. The second stage is the fully connected network. The deep features are mapped to the output of each turbine by fitting the complex function relationship with the fully connected layer. The output vector length of the last full connected layer, equal to the number of pixels in the input image, is reshaped to be two dimensional, and mapped to the pixels of the input image one by one. The down-sampling process of the model also incorporates the idea of DenseNet, and the model structure is shown in Fig. 3.

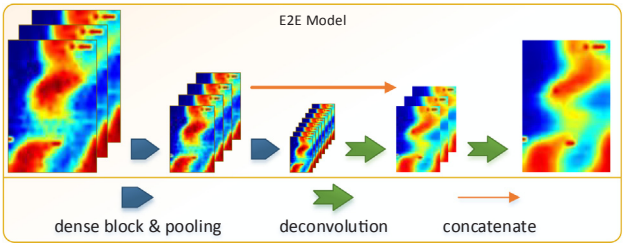


**Fig. 2.** Structure of the E2E model.

## 4. Experiment and analysis

### 4.1. Data sets and evaluation criteria

The data set used in this paper is the wind data set from the NREL,[1] which contains the output values of every 10 min of wind turbines in the United States from 2004 to 2006. To validate our method, two wind farms are selected. The longitude of wind farm1 is range from $105.00W$ to $105.34W$ and latitude is range from $41.40N$ to $41.90N$ (17.3 km × 38.7 km), it is located in the central United States. Wind turbines are densely distributed reaching a number of 592. Wind farm2 contains 454 turbines and with the longitude range from $105.358W$ to $105.675W$ and latitude range from $34.908N$ to $35.392N$ (28.9 km × 53.8 km), which is located in the southern United States, as shown in Fig. 4. We make the prediction about the wind power output of the turbines after 10, 20, 30, and 60 min respectively based on the data set. For a clearer description, the following experiments, if not specified, are the results of wind farm1.

Accuracy is the most important factor to measure the effect of wind power prediction, and the main indexes of evaluating accuracy are mean square error (*MSE*) and root mean square error (*RMSE*), *RMSE* being the square root of *MSE*. So in this paper, *MSE* is chosen as the standard of evaluation, whose calculation process is shown in Eq. (4), in which *real* represent the series of true values, *predictions* represent the series of the predicted values, and *n* represents the length of the series.

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(real_i - predictions_i)^2 \tag{4}$$

### 4.2. Scene display

As shown in Fig. 5, there are 8 scenes sequenced in time series, with the warm tone regions in each scene representing the larger value. This figure is used to show the spatial information expressed by scene and the spatio-temporal information expressed by STF.

It can be seen from the figure that the air flow in this region obviously shows regularity during this period (70 min). Firstly, the output power of the wind turbine is strongly correlated with the spatial position at any certain time. Secondly, as time goes by, a visible displacement is shown among the scenes. So it can be inferred that the west wind has crossed the border during this period, thus expanding the affected areas. These laws are the basis of prediction using machine learning methods. These scene series show the advantage of STF, that is, STF is able to express the spatio-temporal change process of wind. The traditional single-feature can be visualized into a curve, but it is difficult to find obvious regulation no matter for human eyes or computer algorithms, thus having no access to a better prediction accuracy.

[1] https://www.nrel.gov.

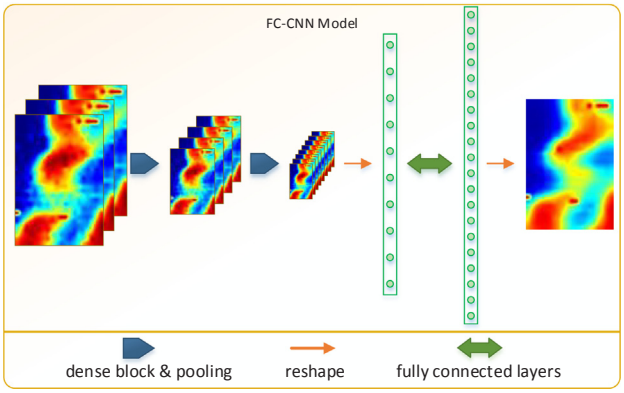


**Fig. 3.** Structure of the FC-CNN model.

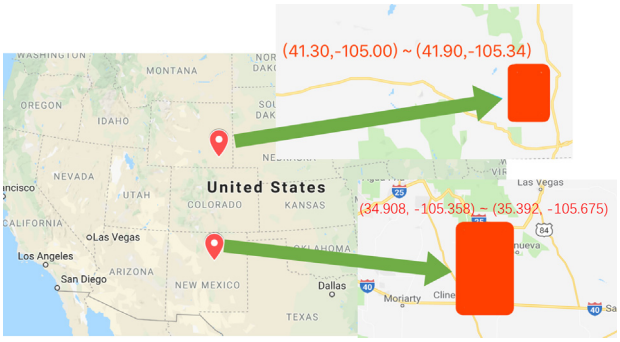


**Fig. 4.** Locations of two study wind farms.

corresponding to each of these methods. The columnar section in each subgraph corresponds to the distribution of *MSE*, in which the curve illustrates the probability density. The horizontal scale represents the value of *MSE*, and the ordinate represents the corresponding probability density (*PDF*) in each subgraph. The first five images show the results of each method, and the last image compares all the results.

It is clear that the *MSE* of FC-CNN and E2E model are distributed in the region with the smaller values. So on the whole, for most wind turbines, the method proposed in this paper can get a lower error rate during prediction. In fact, our proposed method performs best on all the selected 592 wind turbines. Therefore, the proposed methods outperform the SVR and kNN.

The above results have proved the advantages of the proposed method. In Figs. 7 and 8, wind turbines are analyzed in turn, to quantitatively compare the results of optimization. In the figures, *M* denotes the models, LF + SVR and LF + kNN, used for comparison. The effect of the method using SF is inferior to that of the method using LC, so it is no longer comparison. The values got from Eq. (5) reflect the reduced ratio of *MSE* of FC-CNN compared with *M*. And Figs. 7 and 8 are the probability density curves obtained by fitting these values.

$$p_i = \frac{MSE(M)_i - MSE(FC-CNN)_i}{MSE(M)_i} \tag{5}$$

In Figs. 7 and 8, the area of the region whose horizontal coordinate is less than 0 is almost none, which means that the prediction effect of FC-CNN on almost all wind turbines is optimized compared with the above two methods. According to the statistics, compared with LC + SVR, its *MSE* had an average reduction of 24.10%, and maximally decreased by 45.55%. And compared with LC + kNN, its *MSE* decreased by 30.10%, and highest by 45.55%.

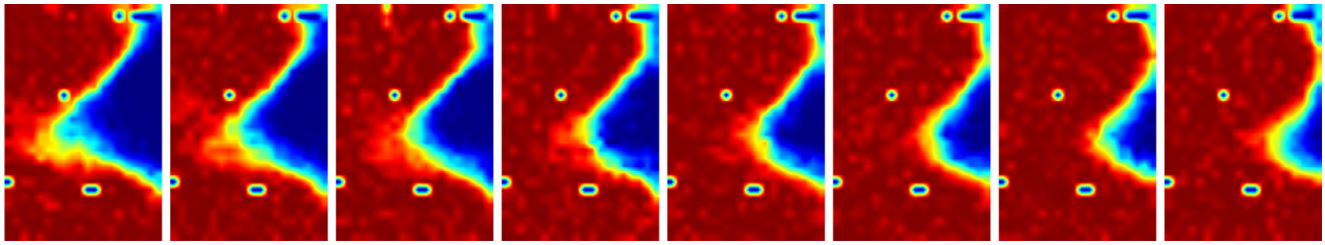

**Fig. 5.** Scene time series related wind power. The sequence involves 8 moments with a total duration of 70 min, wherein the warmer the pixels in the picture, the bigger the corresponding value of the wind power. The series clearly shows the state changes of the wind farm.

### 4.3. Experimental results and comparison

The methods based on LF such as SVR have reached the level of state-of-the-art in wind power prediction. In order to prove the validity of the method proposed in this paper, SVR, the most accurate method for prediction, and kNN, the fastest method for training are compared with. In the experiment, SF and LF are both used for training each SVR model or KNN model. The experimental results are shown in Table 1 and Fig. 6. For fairness, a prediction horizon 30 mins is chosen, which is widely used in existing works.

In the experiment, the *MSE*s of each method on 592 wind turbines are calculated respectively firstly. Table 1 compares the results of these methods according to the maximum, minimum and average values of these *MSE* values. The average *MSE* of the two methods proposed in this paper are 7.91 and 7.78 respectively, and for the ensemble of the two models, *MSE* reaches the number of 7.61. However, the optimal value of other methods is 10.05, compared with the value in our methods which is reduced by 24. 28%. Therefore, according to the above numerical results, the two methods proposed in this paper are superior to other methods in prediction accuracy.

Table 1 provides a quantitative comparison of the overall performance of the methods. And Fig. 6 further shows the distribution of *MSE*

Fig. 9 shows the predicted value curves of each method on a randomly selected turbine. It can be seen from the figure that the predicted results of the model using STF are more stable, whose stability is even better than that of the true value. As a matter of fact, wind is a natural phenomenon, but the conversion process from wind to wind power output is complex, with many interference factors related to the characteristics of the wind turbine itself. In order to further analyze the experimental results, the wind power prediction is divided into two stages. The first stage is to predict the information such as wind speed, and another stage is to convert the wind state information from the prediction to wind power output. And it has been believed in this paper that the prediction errors mainly come into existence during the second stage. To verify this idea, this paper uses the proposed methods to separately predict wind speed and wind power. The typical results are shown in Fig. 10, in which the true and predicted value of wind power and wind speed at 8 moments are visualized. Obviously, the predicted value of wind power, the true value and the predicted value of wind speed are relatively smooth, but the true value of wind power is far from smooth. This shows that the output power rates of two wind turbines with similar wind speed are different even if the turbines are quite close to each other, which fully indicates that the conversion from wind speed to wind power is related to the characteristics of the wind

**Table 1**
Comparison with existing methods on two wind farms. Wind Farm1 is located in the central United States and contains 592 wind turbines and covers an area of 17.3 km × 38.7 km. Farm2 is located in the southern United States and contains 454 wind turbines and covers an area of 28.9 km × 53.8 km. The STFs of these two wind farms have the same shape, both 30 × 20.

| | Methods | PERSISTENCE | LF + LSTM | SF + kNN | LF + kNN | SF + SVR | LF + SVR | STF + E2E | STF + FC-CNN | STF ensemble |
|---|---|---|---|---|---|---|---|---|---|---|
| MSE ($MW^2$) Farm1, 30 * 20 592 turbines | MAX | 19.10 | 15.70 | 25.44 | 16.79 | 18.70 | 15.84 | 11.94 | 12.23 | 11.39 |
| | MIN | 8.61 | 7.37 | 8.83 | 7.30 | 8.37 | 6.64 | 5.25 | 5.00 | 5.00 |
| | AVE | 12.78 | 10.75 | 13.28 | 10.90 | 12.50 | 10.05 | 7.91 | 7.78 | 7.61 |
| | Train time (s) | – | 92755 | 509 | 10592 | 91191 | 207081 | 1200 | 1059 | – |
| MSE ($MW^2$) Farm2, 30 * 20 454 turbines | MAX | 17.56 | 15.09 | 18.02 | 15.03 | 17.33 | 14.91 | 13.08 | 12.62 | 12.18 |
| | MIN | 10.23 | 8.94 | 10.61 | 8.61 | 9.98 | 8.33 | 7.42 | 6.30 | 6.52 |
| | AVE | 12.45 | 10.77 | 12.91 | 10.47 | 12.17 | 10.25 | 8.90 | 7.80 | 7.94 |
| | Train time (s) | – | 70913 | 386 | 8217 | 72380 | 156287 | 1147 | 986 | – |

turbine itself. In addition, at the same moment and in the same region, the *MSE* of wind speed and wind power predicted with the same method are 0.92 and 7.17, respectively. We can see that the *MSE* of wind speed is much lower than that of wind power. It further shows the wind speed is easier to predict, when the wind power is difficult to predict due to the wind turbine' specific features. The STF presented in this paper can express wind-related information in a large geographical area and a long time span. Convolutional network can be used to predict the overall variation of the wind in the region and can reduce the effect of "noise" caused by the specific features of the wind turbine. So in the result shown in Fig. 8, the predicted value is more stable than the true value.

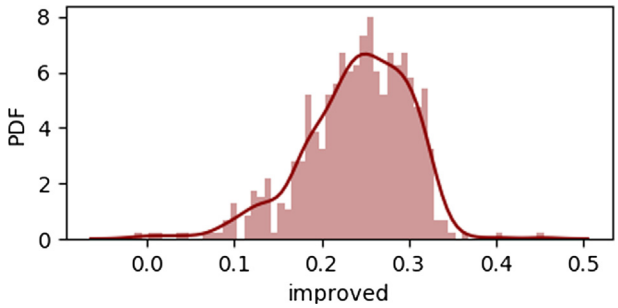


**Fig. 7.** The ratio of *MSE* reduction when FC-CNN compared to SVR using LF. The figure shows the results of experiments on 592 wind turbines. The abscissa represents the value calculated according to the Eq. (5), the vertical axis is the probability density, and the curve is the fitted probabilit.y density curve.

### 4.4. Changing feature window and prediction horizon

In our experiments, 30 mins is a typical prediction horizon. However, the proposed method can represent the changes of wind both in time and space, which innovated the expression of the wind features. And it has the potential for a better result on different prediction horizon when compared with existing methods. In the following

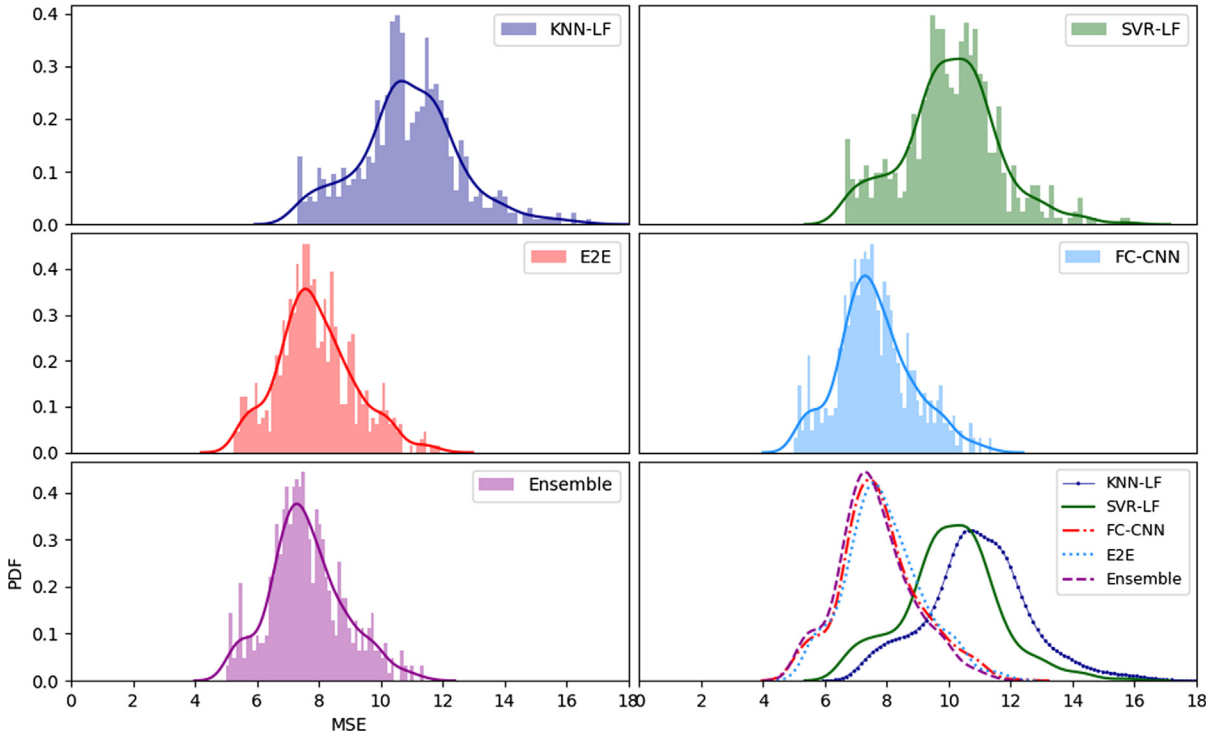


**Fig. 6.** Predictive error distribution for each method.

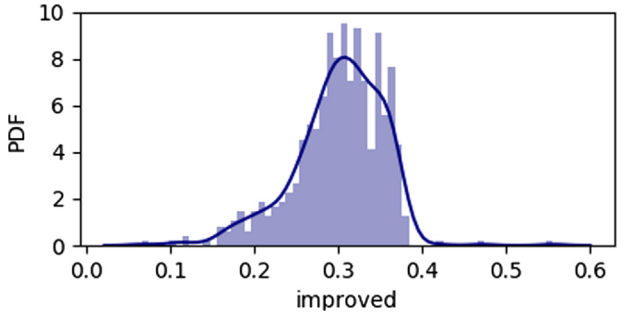


**Fig. 8.** The ratio of *MSE* reduction when FC-CNN compared to kNN using LF. The figure shows the results of experiments on 592 wind turbines. The abscissa represents the value calculated according to the Eq. (5), the vertical axis is the probability density, and the curve is the fitted probabilit.y density curve.

experiments, we change the values of feature window and prediction horizon, and the results are show in Table 2. What's more, it must be stated that the time cost of building SVR models is too high, and we only built 20 models with the randomly selected turbines. Nevertheless, it takes about 480 min to build the SVR models under each prediction horizon with feature window equals 3 and 5. When the feature window equals at 7, it costs too much time and we cant't get the result within 10 h. The average *MSE* of the proposed method are calculated by all 592 turbines, and the time cost of the proposed method are almost equal under different parameters, which take about 1100 s to train and predict all the 592 turbines.

As shown in 2, *MSE* corresponding to all the methods are quite different when choosing different prediction horizons, that is, the bigger the prediction horizon is, the more difficult it is to predict. However, there is no big difference whether we change the feature window. especially for our proposed methods. It is clear that our methods benefit more from the introduction of the increasing feature window than from others while KNN even suffer from this. The results also demonstrate the potential of applying the proposed methods to a long-term forecasting, for the reason that they have great advantages when prediction horizon is 60 mins and compared with LF-SVR model, which is the state-of-the-art method.

### 4.5. Efficiency analysis

The two convolutional networks proposed in this paper can achieve the end-to-end prediction. And since each pixel point at the output end corresponds to a turbine, the prediction of a scene is actually the prediction of all turbines in parallel. Meanwhile, the convolutional network can make full use of GPU acceleration, so the training time has been greatly shortened. The comparative effect of the time for training the model is shown in the last line of Table 1. It can be seen that, overall, the training time is qualitatively optimized, which has been shortened by a factor of more than 150, in contrast with that of SVR.

### 4.6. MSTF experiment

As described in Chapter 3, the STF carrying multiple types of information is called MSTF. Using MSTF can further improve the effect of wind power prediction. This paper uses simple experiments to prove this view but will not discuss it in detail. As shown in Table 3, the *MSE* of MSTF + FC-CNN compared with that of LF + SVR was reduced by 26.69% on average and 49.83% at most. Compared with the *MSE* of LF + kNN, it decreased by 32.49% on average and 56.63% at most. The effect is also better than that of using STF. The average *MSE* of E2E model and FC-CNN model, both of which use MSTF, in comparison with the model using STF are respectively reduced by 7.08% and 6.81%.

## 5. Conclusion

This paper proposes spatio-temporal feature for wind power prediction, and uses convolutional network to predict wind power. Compared with the existing methods, the proposed method greatly optimizes the prediction accuracy and the time cost for training models. In addition, this paper also proposes an approach to fuse various types of data by multi-spatio-temporal feature, which is then proved to be effective in the experiment.

In fact, spatio-temporal feature is modeling the spatio-temporal state of wind farm, in which wind turbines play the role of information collectors. The denser the wind turbines are, the more completed the information collected is, so spatio-temporal feature is quite suitable to describe the state of a large wind farm. It is also worth noting that the spatio-temporal feature uses plane to represent the spatial state, which will lose the terrain information. So spatio-temporal feature is more suitable for the flat area. In the past several years, offshore wind power has grown rapidly. Thanks to the large scale and flat area of offshore wind farms, spatio-temporal feature is naturally ideal for modeling and forecasting offshore wind farms. In future work, this paper will focus on offshore wind farms as the main area of application and further develop the following researches.

- The way of multi-spatio-temporal feature's fusion of multiple types of data will be studied in order to continuously improve the accuracy of prediction.
- In this paper, two kinds of simple models of convolutional network are constructed to make a prediction, and good prediction results

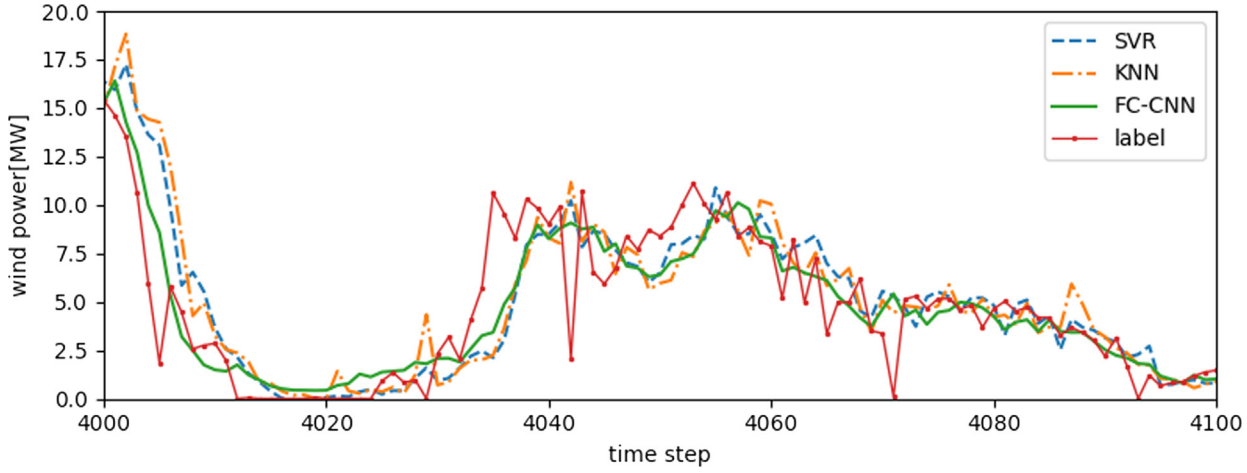


**Fig. 9.** Comparison of the true wind power time series of a turbine and the predicted series corresponding to different methods.

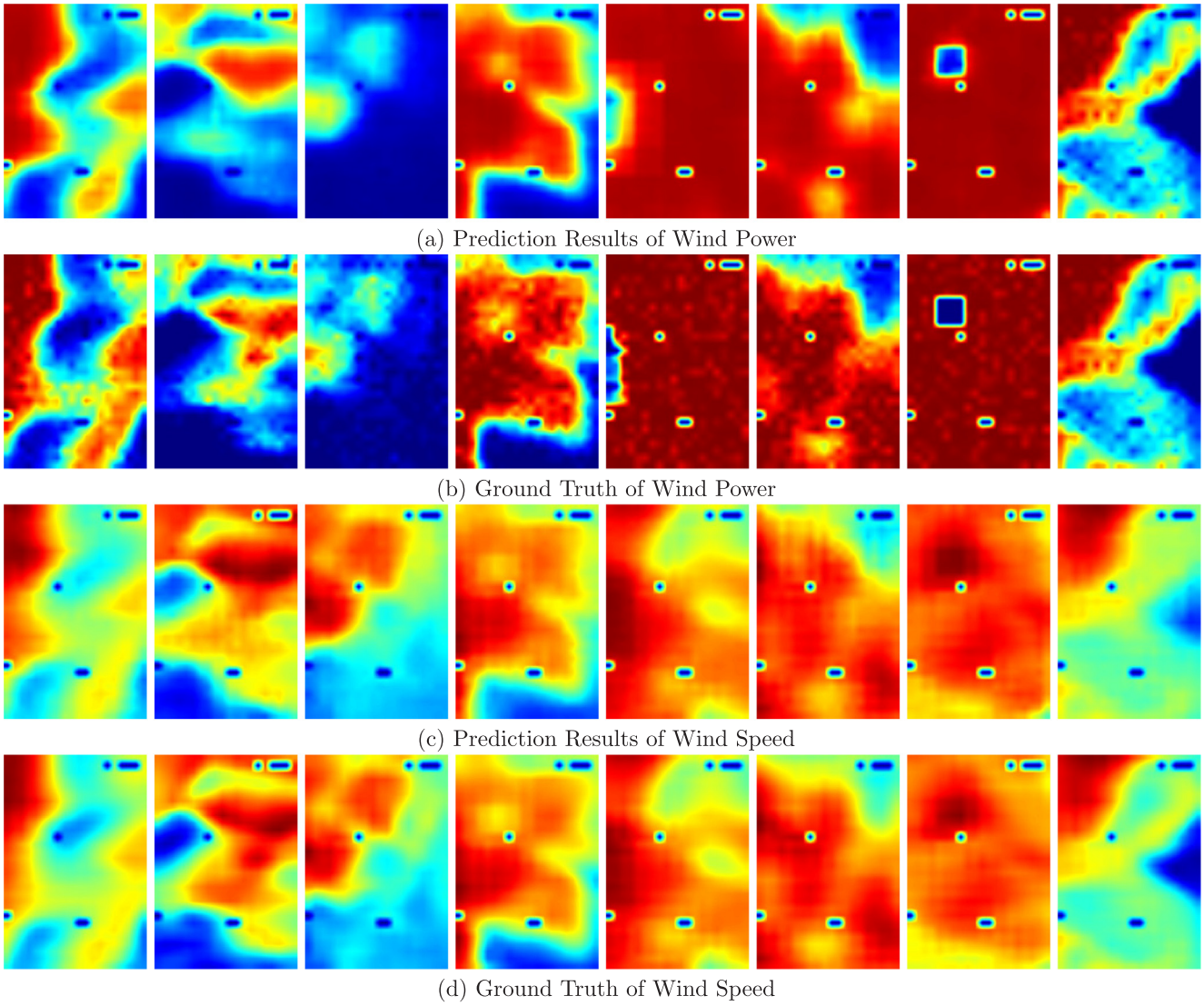


**Fig. 10.** Scene predictions results corresponding to wind speed and wind power.

**Table 2**
Average MSE under different feature windows and prediction horizons. "PH" means "Prediction Horizon", $PH = 30$ min means the result to predict half an hour latter. "FW" means "Feature Window", the time interval between which to abstract features from.

| PH (mins) | 10 | | | 20 | | | 30 | | | 60 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FW (× 10 mins) | 3 | 5 | 7 | 3 | 5 | 7 | 3 | 5 | 7 | 3 | 5 | 7 |
| LF-SVR | 3.840 | 3.883 | – | 7.602 | 7.583 | – | 10.054 | 11.611 | – | 23.393 | 22.764 | – |
| LF-KNN | 4.678 | 5.058 | 5.476 | 8.807 | 9.134 | 9.617 | 10.901 | 13.365 | 13.857 | 24.314 | 25.580 | 25.941 |
| FC-CNN | 3.232 | 3.177 | 3.158 | 5.389 | 5.368 | 5.368 | 7.782 | 7.854 | 7.843 | 15.458 | 15.462 | 15.680 |
| E2E | 3.240 | 3.205 | 3.28 | 5.289 | 5.290 | 5.297 | 7.910 | 7.741 | 7.696 | 15.162 | 15.464 | 15.220 |

**Table 3**
Comparison of effects between STF-based models and MSTF-based models.

| Feature | FC-CNN | | E2E | |
|---|---|---|---|---|
| Method | STF | MSTF | STF | MSTF |
| MAX | 12.23 | 11.33 | 11.94 | 11.05 |
| MIN | 5.00 | 4.72 | 5.25 | 4.96 |
| AVE | 7.78 | 7.25 | 7.91 | 7.35 |

have been obtained. In fact, convolutional networks have developed rapidly in recent years, and the next step is to introduce more advanced models.

## Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant No. 61877043), Major Scientific and Technological Projects for A New Generation of Artificial Intelligence of Tianjin (Grant No. 18ZXZNSY00300), and Key Project for Science and Technology Support from Key R&D Program of Tianjin (Grant No. 18YFZCGX00960).

## References

[1] Zhao Y, Ye L, Li Z, Song X, Lang Y, Su J. A novel bidirectional mechanism based on time series model for wind power forecasting. Appl Energy 2016;177:793–803.

[2] R. REN21. Global status report, ren21 secretariat, Paris, France. Tech Rep; 2017.
[3] Xydas E, Qadrdan M, Marmaras C, Cipcigan L, Jenkins N, Ameli H. Probabilistic wind power forecasting and its application in the scheduling of gas-fired generators. Appl Energy 2017;192:382–94.
[4] Maharjan S, Zhu Q, Zhang Y, Gjessing S, Basar T. Dependable demand response management in the smart grid: a stackelberg game approach. IEEE Trans Smart Grid 2013;4(1):120–32.
[5] Zhang Y, Yu R, Nekovee M, Liu Y, Xie S, Gjessing S. Cognitive machine-to-machine communications: visions and potentials for the smart grid. IEEE Network 2012;26(3):6–13.
[6] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-arima models. Renew Energy 2009;34(5):1388–93.
[7] Ziel F, Croonenbroeck C, Ambach D. Forecasting wind power-modeling periodic and non-linear effects under conditional heteroscedasticity. Appl Energy 2016;177:285–97.
[8] Chen K, Yu J. Short-term wind speed prediction using an unscented kalman filter based state-space support vector regression approach. Appl Energy 2014;113:690–705.
[9] Becker R, Thrän D. Completion of wind turbine data sets for wind integration studies applying random forests and k-nearest neighbors. Appl Energy 2017;208:252–62.
[10] Deo RC, Ghorbani MA, Samadianfard S, Maraseni T, Bilgili M, Biazar M. Multi-layer perceptron hybrid model integrated with the firefly optimizer algorithm for wind-speed prediction of target site using a limited set of neighboring reference station data. Renew Energy 2018;116:309–23.
[11] Marvuglia A, Messineo A. Monitoring of wind farms' power curves using machine learning techniques. Appl Energy 2012;98:574–83.
[12] Xiaoyun Q, Xiaoning K, Chao Z, Shuai J, Xiuda M. Short-term prediction of wind power based on deep long short-term memory. Power and energy engineering conference (APPEEC), 2016 IEEE PES Asia-Pacific. IEEE; 2016. p. 1148–52.
[13] Treiber NA, Heinermann J, Kramer O. Wind power prediction with machine learning. Computational sustainability. Springer; 2016. p. 13–29.
[14] Khosravi A, Machado L, Nunes R. Time-series prediction of wind speed using machine learning algorithms: a case study osorio wind farm, brazil. Appl Energy 2018;224:550–66.
[15] Díaz S, Carta JA, Matías JM. Performance assessment of five mcp models proposed for the estimation of long-term wind turbine power outputs at a target site using three machine learning techniques. Appl Energy 2018;209:455–77.
[16] Zhang Y, Yu R, Xie S, Yao W. home m2m networks: architectures, standards, and qos improvement. Commun Mag IEEE 2011;49(4):44–52.
[17] Tascikaraoglu A, Sanandaji BM, Poolla K, Varaiya P. Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using wavelet transform. Appl Energy 2016;165:735–47.
[18] Lixin M, Lei W. Prediction model of wind power based on lifting wavelet transform. Electric Power Sci Eng 2018;34:20–5.
[19] Saroha S, Aggarwal S. Wind power forecasting using wavelet transforms and neural networks with tapped delay. CSEE J Power Energy Syst 2018;4(2):197–209.
[20] Li C, Xiao Z, Xia X, Zou W, Zhang C. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. Appl Energy 2018;215:131–44.
[21] Wang J, Du P, Niu T, Yang W. A novel hybrid system based on a new proposed algorithm–multi-objective whale optimization algorithm for wind speed forecasting. Appl Energy 2017;208:344–60.
[22] Liu H, Tian H-Q, Liang X-F, Li Y-F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157:183–94.
[23] Wang Z, Shen C, Liu F. A conditional model of wind power forecast errors and its application in scenario generation. Appl Energy 2018;212:771–85.
[24] De Giorgi MG, Ficarella A, Tarantino M. Error analysis of short term wind power prediction models. Appl Energy 2011;88(4):1298–311.
[25] Zhao J, Guo Z-H, Su Z-Y, Zhao Z-Y, Xiao X, Liu F. An improved multi-step forecasting model based on wrf ensembles and creative fuzzy systems for wind speed. Appl Energy 2016;162:808–26.
[26] Pinson P, Nielsen HA, Madsen H, Kariniotakis G. Skill forecasting from ensemble predictions of wind power. Appl Energy 2009;86(7–8):1326–34.
[27] Feng C, Cui M, Hodge B-M, Zhang J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 2017;190:1245–57.
[28] Heinermann J, Kramer O. Machine learning ensembles for wind power prediction. Renew Energy 2016;89:671–9.
[29] Wang H-Z, Li G-Q, Wang G-B, Peng J-C, Jiang H, Liu Y-T. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[30] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision. Springer; 2014. p. 818–33.
[31] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436.
[32] Li Y, Liu Z, Xu K, Yu H, Ren F. A gpu-outperforming fpga accelerator architecture for binary convolutional neural networks. ACM J Emerg Technol Comput Syst 2018;14(2):18.
[33] LeCun Y, Kavukcuoglu K, Farabet C, et al. Convolutional networks and applications in vision. In: ISCAS, vol. 2010; 2010. p. 253–6.
[34] Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification. Advances in neural information processing systems. 2015. p. 649–57.
[35] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431–40.
[36] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[37] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th international conference on machine learning. ACM; 2008. p. 1096–103.