https://doi.org/10.1016/j.renene.2021.05.023

Spatio-temporal Wind Speed Prediction of Multiple Wind Farms Using Capsule Network

Ling Zheng[a,b,c], Bin Zhou[a*], Siu Wing Or[b,c*], Yijia Cao[a], Huaizhi Wang[d], Yong Li[a], Ka Wing Chan[b]

[a]*College of Electrical and Information Engineering, Hunan University, Changsha 410082, China*
[b]*Department of Electrical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[c]*Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center, Hong Kong.*
[d]*College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China*

Abstract: Spatio-temporal wind speed prediction is of great significance to the grid-connected operation of multiple wind farms in smart grid. This paper proposes a spatio-temporal wind speed prediction method based on capsule network (CapsNet) for geographically dispersed wind farms over a region. In the proposed method, the historical wind speed data from the wind farms are originally converted into chronological images in a 3D space, and the spatial features implicit in the images are extracted by the convolutional operation. Then, the temporal information of wind speed spatial properties is encapsulated in multi-dimensional time-capsules and learned by the dynamic routing mechanism, thus capturing the nonlinear temporal dependencies based on the extracted spatial features. A regression layer activated by the leaky rectified linear unit (Leaky ReLU) function integrates the spatio-temporal features and generates the final prediction results. Furthermore, a two-layer iterative training approach is employed to well-tune the model parameters and accelerate the convergence speed. Finally, the real data of multiple wind farms from Ohio are collected in the case studies to demonstrate the superior performance of the proposed method compared with other forecasting methods.

Highlights

Multi-dimensional vectors are used to characterize spatio-temporal wind speed features.
Dynamic routing captures wind speed temporal dependencies among spatial properties.
A capsule network is proposed for spatio-temporal wind speed prediction problem.
A two-layer training approach is used to well-tune the model parameters of CapsNet.

Keywords: Capsule network, dynamic routing, renewable energy, spatio-temporal features, wind speed prediction

[*]E-mail addresses of corresponding authors: binzhou@hnu.edu.cn (B. Zhou), eeswor@polyu.edu.hk (S. W. Or)

1 Introduction

Due to the clean and renewable environmental benefits, the number and scale of wind farms have grown considerably in recent years [1]. Wind speed prediction of multiple wind farms over a local region contributes to providing rich and valuable renewable energy information for system operators and managers, thereby facilitating the operation, management and real-time control of smart grid [2]. Nevertheless, the majority of current investigations focus on the wind speed prediction problem for an individual wind farm, which merely consider the wind speed temporal correlations, i.e., the relationships between the wind speed at a certain site and its historical values [3]. In reality, wind farms are usually clustered within a certain area with abundant wind resources, and there are interactive influences of wind speed time series among different wind farms [4]. Specifically, with the geographical and meteorological factors such as terrain, air pressure and temperature, the wind speed at different sites over a region exhibits remarkable spatial correlations [5]. Therefore, it is crucial to investigate the spatio-temporal wind speed prediction problem of multiple wind farms by capturing wind speed correlations on both time and space scales.

The temporal correlations of wind speed at a certain wind site can be investigated from the historical data [4], while the spatial correlations, due to the influence of geographical factors, should be explored in connection with the wind speed information at the surrounding sites [6]. Consequently, the spatial and temporal correlations of wind speed from multiple wind farms need to be extracted in a targeted and differentiated manner to solve the wind speed prediction problem [3]. On the other hand, a large amount of spatio-temporal data from multiple wind farms increases the number of input variables and parameters to be trained in the prediction model, thereby increasing the complexity of data processing for the wind speed prediction problem [7]. Here, this paper is devoted to developing a capsule network (CapsNet) based model capable of capturing spatio-temporal features for the wind speed prediction of multiple dispersed wind farms. In the proposed model, a hybrid structure composed of convolution and capsule sub-networks is designed to hierarchically extract the underlying spatial features and intrinsic temporal dependencies of wind speed time series. Then, the wind speed prediction for multiple wind farms can be achieved using the extracted spatio-temporal features with a followed regression block. Moreover, the proposed prediction model is trained to encode spatio-temporal features with multi-dimensional vectors, offering a promising solution to handle the complex wind speed time series.

So far, the commonly-used wind speed prediction methods can be classified into three categories, including physical methods, statistical methods and machine learning methods [8]. Physical methods usually use meteorological and geographical information for wind speed prediction problems [5], but these methods are time-consuming and sensitive to initial conditions, and thereby not applicable for short-term prediction problems [9]. Statistical methods such as autoregressive integrated moving average model (ARIMA) [10] could obtain accurate short-term prediction results by capturing mathematical relationships between historical time series and future values. Nevertheless, the resulting residue of these methods will gradually accumulate when prediction steps increase [9]. Various machine learning techniques, with the strong capability of big-data training and nonlinear feature extraction, have been widely applied to the problems of renewable energy forecasting [8], [11], including support vector machine (SVM) [12], extreme learning machine (ELM) [13], stacked autoencoder (SAE) [14], deep belief networks (DBN) [15] and convolutional neural networks (CNN) [16], etc. However, most of these methods focused on the wind speed prediction problem for a single wind farm using wind speed temporal correlations. To further enhance the prediction performance, the spatial correlations of wind speed among the neighboring wind farms should also be taken into account [17], [18].

With the sufficient spatial and temporal information to be accessed from multiple wind farms, a few spatio-temporal prediction methodologies have been investigated in recent studies [19]-[24]. The wind farms were modeled as an undirected graph in [20] to extract the spatio-temporal characteristics of wind speed. In [21], a spatio-temporal forecasting model combining multi-output support vector machine and grey wolf optimizer was proposed to achieve the wind power prediction of multiple wind farms. The spatio-temporal correlations among wind farms was simulated by a joint distribution model in [22] based on the copula theory, and the Bayesian theory was used to deduce a conditional distribution of the aggregated wind power. The study in [23] adopted a computational numerical simulation approach to predict a spatial wind field under a complex terrain. A probabilistic wind speed prediction approach was presented in [24] based on a spatio-temporal neural network (STNN) and variational Bayesian inference. With feeding both the spatial and temporal information into the prediction model, these methods have achieved the better prediction performance. Nevertheless, most of these prediction models often collected all the wind speed information from different wind farms indiscriminately, and the implicit spatial correlations cannot be fully exploited in the original wind speed data to some extent. Also, the thorny multi-dimensional computing problem caused by the large amount of wind speed data at multiple wind farms needs to be solved in an effective way.

In this study, a CapsNet based wind speed prediction model is developed for capturing the intrinsic spatio-temporal features of wind speed time series to indicate the interactions among geographically dispersed wind farms over a region. CapsNet algorithm was firstly introduced by Sabour *et al.* in [25] to overcome the shortcomings of CNN which cannot identify objects with spatial relationships between properties. CapsNet uses multi-dimensional vector neurons as capsules to encode part-whole relationships between an object and its various properties, and a dynamic routing (DR) mechanism is utilized to allow CapsNet to learn these relationships through an iterative process of sending information from lower-level capsules to the appropriate higher-level capsules. [26]. Previous studies in [25]-[27] demonstrate that the discriminatively trained CapsNet model can offer the state-of-the-art performance in handling the high-dimensional data and extracting complex intrinsic features. In recent years, the CapsNet algorithm has been applied to various fields, such as image recognition [28]-[31], financial time series forecasting [32], human pose estimation [33], and traffic forecasting [34], etc.

In this paper, a spatio-temporal prediction method based on CapsNet is proposed to simultaneously predict the wind speed from multiple dispersed wind farms. The proposed method can hierarchically and sufficiently capture spatial and temporal wind speed features to solve the wind speed prediction problem. The contributions of this paper are presented as follows:

(1) The inherent spatio-temporal features of wind speed time series from dispersed wind sites are encoded with multi-dimensional vectors and learned by dynamic routing mechanism. The underlying spatial features of wind speed are extracted by convolutional implementation of sliding windows, while the temporal dependencies among the extracted spatial properties can further be captured by the capsules.

(2) A CapsNet based forecasting model is proposed to cope with the spatio-temporal wind speed prediction problem for multiple wind farms. This model stacks a convolution structure with capsule sub-networks, which combines both scalar and vector computation to handle the spatio-temporal wind speed data. Furthermore, the regression layer is used to integrate the extracted spatio-temporal features and generate the prediction results through the leaky rectified linear unit (Leaky ReLU) function.

(3) A two-layer model tuning approach is designed to optimize the model parameters of the proposed method. The internal training process is used for capsule sub-networks through DR iteration to adjust coupling coefficients between capsules, while the external layer training for the overall network adopts the back propagation rule embedding a trial-and-error technique so as to dynamically update the learning rate for optimizing the model parameters.

The rest of this paper is organized as follows: Section 2 formulates the problem of spatio-temporal wind speed prediction; Section 3 describes the principles of the proposed CapsNet based spatio-temporal prediction method, including the extraction of wind speed features, the CapsNet structure as well as its model tuning approach; Section 4 investigates and evaluates the comparative performance of the proposed method through experimental studies. Section 5 presents the conclusions.

2 Problem formulation

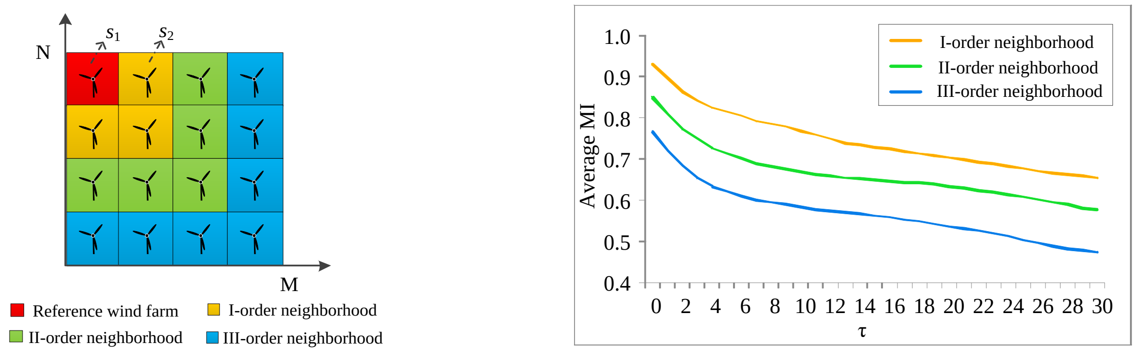
2.1 Spatio-temporal correlations in wind speed time series

The spatio-temporal correlations of wind speed reflect the physical nature of atmospheric motion and contain valuable information for wind speed prediction [5]. Therefore, it is vital to capture the spatio-temporal correlations for achieving accurate wind speed prediction. For preliminary analysis of the spatio-temporal correlations of wind speed at multiple wind farms, mutual information (MI) [35] is used to calculate the correlation values. Suppose there is a pair of wind farms s_1 and s_2, and the wind speed time series of these wind farms are denoted as $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively. The MI between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is computed as,

$$I(\boldsymbol{X},\boldsymbol{Y})=\sum_{x\in\boldsymbol{X}}\sum_{y\in\boldsymbol{Y}}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

where $p(x)$, $p(y)$ is the marginal distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ respectively, and $p(x, y)$ is the joint distribution. The mutual information $I(\boldsymbol{X}, \boldsymbol{Y})$ is the relative entropy of the joint distribution $p(x, y)$ and the marginal distribution $p(x)$ and $p(y)$. The geographical relationship of wind farms located in a certain region is illustrated in Fig. 1(a). Suppose s_1 is the reference wind farm, and the spatial region directly adjacent to s_1 is defined as the I-order neighborhood. Here, neighborhood refers to the area where the wind farms adjacent to the reference wind farm are located. Likewise, the region adjacent to the I-order neighborhood is defined as the II-order neighborhood, while the III-order neighborhood is the region adjacent to the II-order neighborhood. For each neighborhood, the average value of MI between the wind speed time series of s_1 and the lagged time series of each wind farm in the corresponding neighborhood is calculated and the results are shown in Fig. 1(b) with the time-lag τ ranging from 0 to 30.

The spatio-temporal correlations of wind speed time series are revealed by Fig. 1, which lie in two aspects: 1) Wind speed time series at the reference wind farm show different degrees of spatial correlations with those at different neighborhood. The closer the neighborhood is to the reference wind farm, the more relevant the wind speed data will be. 2) Wind speed time series exhibit strong temporal correlations, which will reduce with the increase of the time-lag τ. It can be concluded that the wind speed time series at multiple wind farms over a certain region have significant spatio-temporal correlations, which need to be further learned and utilized to enhance the prediction performance of wind speed.



(a) (b)

Fig. 1 (a) Geographical relationship of wind farms located in a certain region (b) Average value of MI between the wind speed time series of the reference wind farm and the lagged time series of wind farms in each neighborhood

2.2 Spatio-temporal wind speed prediction problem

The region where the wind farms are located can be represented by a $M \times N$ grid, as shown in Fig. 1(a), and the wind speed at a certain site can be expressed by $x(m,n)_t$, $(1 \le m \le M,\ 1 \le n \le N)$. In time dimension, the wind speed data of each site are 1D time series with the length of T. Therefore, the wind speed time series at multiple wind farms can be represented by a 3D tensor $\boldsymbol{X}_t \in R^{M \times N \times T}$. At time t, the wind speed at all sites in the grid can accordingly be denoted by a spatial matrix, as follows,

$$\boldsymbol{x}_t = \begin{bmatrix} x(1,1)_t & x(1,2)_t & \cdots & x(1,N)_t \\ x(2,1)_t & x(2,2)_t & \cdots & x(2,N)_t \\ \vdots & \vdots & & \vdots \\ x(M,1)_t & x(M,2)_t & \cdots & x(M,N)_t \end{bmatrix} \in R^{M \times N} \tag{2}$$

Thus, the wind speed spatial matrix at time $t+\lambda$ in the future can be predicted by previous spatial matrices, as follows,

$$\hat{\boldsymbol{x}}_{t+\lambda} = f(\boldsymbol{x}_{t-h+1}, \boldsymbol{x}_{t-h+2}, \cdots, \boldsymbol{x}_t \,|\, \boldsymbol{\theta}) \tag{3}$$

where f represents the mapping between the input and output of the prediction model, h is the number of the previous time points, and $\boldsymbol{\theta}$ is a set of model parameters to be learned through the training procedure. From the above analysis, the wind speed prediction of multiple wind farms can be achieved by the prediction of the wind speed spatial matrix, which is formulated as a spatio-temporal prediction problem.

3 CapsNet based spatio-temporal wind speed prediction

In this section, the extraction of spatial and temporal features of wind speed time series is described in detail, and the prediction architecture based on CapsNet is formulated. Moreover, a two-layer model tuning approach for optimizing the model parameters is proposed.

3.1 Spatial features captured by convolution operation

With superior performance on extracting low-level spatial features, convolutional neural network (CNN) has been extensively applied in image identification [36], [37]. Conventional CNN consists of three types of layers, i.e., the convolution layer for feature extraction, the pooling layer for compressing the feature maps, and the linear layer for integrating the local features. Among them, although the pooling layer can reduce the data dimension, it loses some important information during the sampling process. Here, in order to extract wind speed spatial features more accurately, only the convolution layer and the linear layer of the convolution network are retained.

The wind speed spatial matrix at a time point is originally displayed as an image, and the wind speed of the wind farms over a previous time period is correspondingly displayed as several frames of images. These input maps are first convolved with the filters, which are also termed shared weights. The local spatial features are learned by convolutional implementation of sliding windows and then passed through a non-linear activation function, as follows,

$$\boldsymbol{x}_{tp} = g(\boldsymbol{x}_t \otimes \boldsymbol{\beta}_{tp} + \boldsymbol{b}_p) \tag{4}$$

where $\boldsymbol{x}_t$ denotes the input map corresponding to the wind speed spatial matrix at time t, $\boldsymbol{x}_{tp}$ denotes the pth feature map, $\otimes$ denotes the convolution operation, $\boldsymbol{\beta}_{tp}$ represents the shared weights connecting the input map, $\boldsymbol{b}_p$ represents the bias, and $g(\cdot)$ represents the activation function. Note that the activation function is selected as the Leaky ReLU function which has a strong ability to solve the gradient disappearance problem and can accelerate the convergence speed [38]. In the convolutional network, the neurons in the upper layer only need to be connected to some neurons in the lower layer to perceive the local part of the image, known as the local receptive field, which is suitable for learning the local spatial features of wind speed.

Then through the linear layer, the feature maps are flattened to 1D form and passed through the activation function to form the spatial features, expressed as,

$$\boldsymbol{r}_t = g(\boldsymbol{w}\boldsymbol{x}_{tq} + \boldsymbol{b}) \tag{5}$$

where $\boldsymbol{r}_t = \{r_{t|1}, r_{t|2}, \cdots\}$ represents the extracted spatial features at time t, $\boldsymbol{w}$ and $\boldsymbol{b}$ denote the weight matrix vector and the bias respectively.

3.2 Temporal features extracted by dynamic routing mechanism

3.2.1 Capsule computing

A capsule is a multi-dimensional vector neuron encapsulating important information about the features of an object [25]. Specifically, the length of an output vector represents the detection probability of an object, while the direction characterizes the state of the features such as the size, location and orientation. CapsNet is a novel deep neural network characterized by using vector computation between the capsules instead of conventional scalar operation between the scalar neurons. The capsule computation can detect the presence of a particular object, which is illustrated in Fig. 2 by using the detection of wind speed spatio-temporal feature as an example.

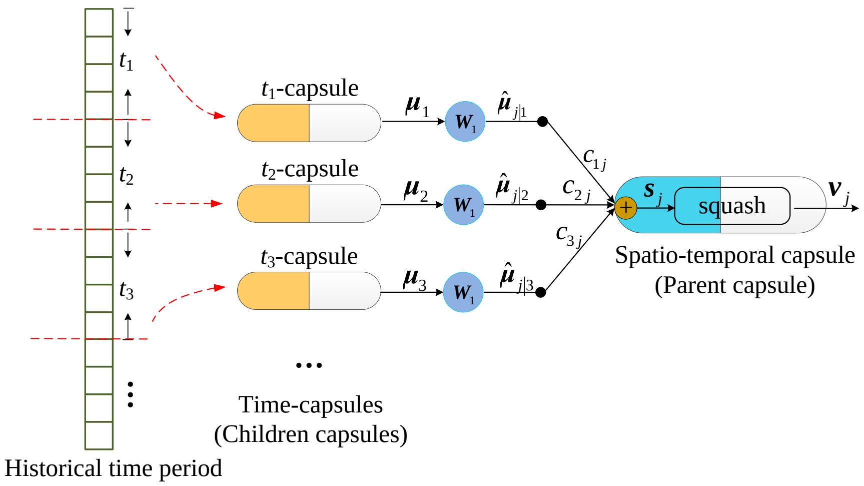


Fig. 2 Capsule computing process in the detection of wind speed spatio-temporal features

Considering that the wind speeds at adjacent time points exhibit strong correlations, the historical time period is divided into several parts with an hourly basis, and children time-capsules are correspondingly formulated, i.e., t_1-capsule, t_2-capsule, etc. Each child time-capsule is a multi-dimensional vector represented by $\boldsymbol{\mu}_i$. The child time-capsule encodes the temporal features of the extracted wind speed spatial properties over a period of time, and each dimension of the vector represents an abstract temporal feature. Then, they predict the parent capsule $\boldsymbol{v}_j$, i.e., the spatio-temporal capsule, which characterizes a wind speed spatio-temporal feature. The part-whole relationship between the ith child capsule and the jth parent capsule is encoded by a weight matrix $\boldsymbol{W}_{ij}$, as follows,

$$\hat{\boldsymbol{\mu}}_{j|i} = \boldsymbol{W}_{ij}\boldsymbol{\mu}_i \tag{6}$$

where $\hat{\boldsymbol{\mu}}_{j|i}$ is the prediction for the spatio-temporal capsule from a time-capsule $\boldsymbol{\mu}_i$, $\boldsymbol{W}_{ij}$ is the weight matrix, which can be learned by backward training the network.

3.2.2 Dynamic routing mechanism and temporal feature extraction

As can be seen from Fig.2, the lower-level time-capsules are connected to the higher-level spatio-temporal capsules through the coupling coefficient c_{ij} which determines the output of a spatio-temporal capsule, namely the wind speed spatio-temporal feature to be extracted. The key of wind speed spatio-temporal feature extraction is to transfer the temporal information of the extracted wind speed spatial properties implied in the time-capsules to the spatio-temporal capsules through the calculation of c_{ij}. Dynamic routing is an iterative routing-by-agreement mechanism for information selection, and ensures that the temporal information of the extracted wind speed spatial properties from the time-capsules can be sent to the appropriate spatio-temporal capsules which agree the most with predictions of the time-capsules. The schematic diagram of dynamic routing mechanism for temporal feature extraction is shown in Fig. 3.

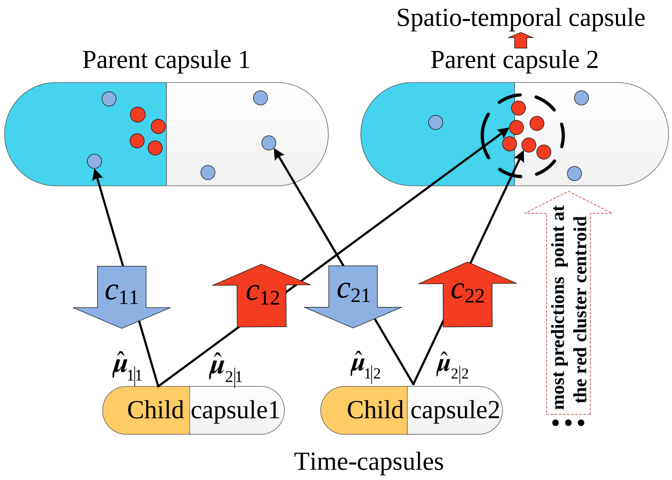


Fig. 3 Schematic diagram of dynamic routing mechanism

The red and blue dots represent the predicted vectors from time-capsules, i.e., the wind speed spatio-temporal features. The red dots are clustered together, which represents that the predictions are similar with each other, while the blue dots are scattered indicating that the predictions are quite different. If most predictions of the time-capsules point at the red cluster centroid of the same parent capsule, it must be the spatio-temporal capsule. As shown in Fig. 3, the child time-capsule routes its prediction $\hat{\boldsymbol{\mu}}_i$ to the parent capsules by adjusting the coupling coefficient c_{ij}, which is calculated by the Softmax function, i.e.,

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{7}$$

where $\sum_j c_{ij} = 1$ and $c_{ij} \geq 0$, b_{ij} is a temporary variable and can be originally set to 0.Then the input vector of the jth spatio-temporal capsule $\boldsymbol{s}_j$ can be calculated by the weight sum of all predictions from the time-capsules, as follows,

$$\boldsymbol{s}_j = \sum_i c_{ij} \hat{\boldsymbol{\mu}}_{j|i} \tag{8}$$

The Squash function is adopted to make the length of the output vector $\boldsymbol{v}_j$ of the spatio-temporal capsule no more than 1, thus representing the detection probability of a spatio-temporal feature, as follows,

$$\boldsymbol{v}_j = \frac{\|\boldsymbol{s}_j\|^2}{1+\|\boldsymbol{s}_j\|^2} \frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|} \tag{9}$$

"Routing-by-agreement" determines the similarity of $\hat{\boldsymbol{\mu}}_{j|i}$ and $\boldsymbol{v}_j$ by calculating the agreement factor a_{ij}, and then the temporary variable b_{ij} can be updated, as follows,

$$a_{ij} = \boldsymbol{v}_j \cdot \hat{\boldsymbol{\mu}}_{j|i} \tag{10}$$

$$b_{ij} = b_{ij} + a_{ij} \tag{11}$$

where "·" represents the dot product of $\hat{\boldsymbol{\mu}}_{j|i}$ and $\boldsymbol{v}_j$. If the predictions of time-capsules are similar with the wind speed spatio-temporal feature, namely if $\hat{\boldsymbol{\mu}}_{j|i}$ and $\boldsymbol{v}_j$ agree, the agreement factor a_{ij} will possess a big internal product, thereby increasing the coupling coefficient c_{ij} using Eq. (11) and Eq. (7). The more these vectors are similar, the more wind speed information from the time-capsules will be sent to the spatio-temporal capsule. Through iteratively adjusting the coupling coefficient c_{ij}, the temporal features among the wind speed spatial properties, namely the wind speed spatio-temporal features, are captured in the spatio-temporal capsules.

3.3 CapsNet based wind speed prediction for multiple wind farms

Wind speed time series from multiple wind farms are represented by different spatial matrices arranged in chronological order, whereas a single spatial matrix with wind speed spatial information can be learned by convolution operation. It is necessary to further capture temporal dependencies among the extracted spatial properties with dynamic routing mechanism. Therefore, the proposed CapsNet stacks a convolution structure with capsule sub-networks to hierarchically capture spatial and temporal features of wind speed, and further integrate the extracted spatio-temporal features through a regression layer for achieving wind speed prediction of multiple wind farms. The detailed CapsNet based prediction architecture is depicted in Fig. 4.

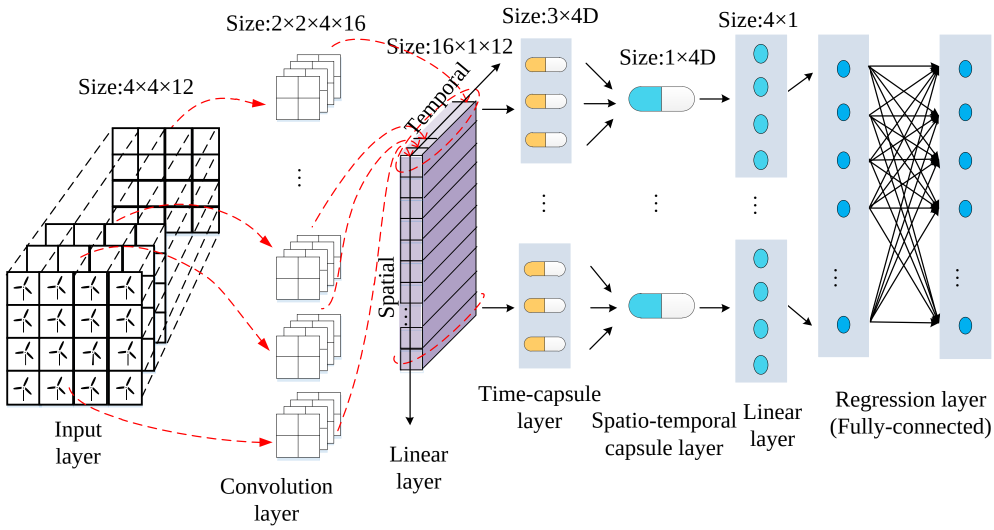


Fig. 4 CapsNet based prediction architecture

In the input layer of CapsNet, the wind speed spatial matrices over a historical time period are displayed as images. The lower-level wind speed spatial features of multiple wind farms are first extracted by convolution operation. As revealed in the analysis of spatio-temporal correlations of wind speed time series in Section 2.1, the closer the wind farms are located, the stronger the spatial correlations of wind speed will be. Therefore, the filters only need to be connected with local regions, i.e., local receptive fields, of the input images and perform convolution operation through sliding windows. In order to extract the intrinsic spatial features such as the wind direction, the distance, etc., implied in the wind speed information, the filters are designed differently by the weight matrices. In this way, the spatial features are extracted and abstracted into network parameters. Then through the linear layer, the local features are integrated to form global wind speed spatial features of multiple wind farms. Based on the above steps, the wind speed spatial matrix $\boldsymbol{x}_t$ can be represented by the extracted spatial features $\boldsymbol{r}_t = \{r_{t|1}, r_{t|2}, \cdots\}$, and thus the prediction problem can be further expressed by,

$$\hat{\boldsymbol{x}}_{t+\lambda} = f'(\boldsymbol{r}_{t-h+1}, \boldsymbol{r}_{t-h+2}, \cdots, \boldsymbol{r}_t | \boldsymbol{\theta}') \tag{12}$$

where f' denotes the implicit function for capturing the temporal features, $\boldsymbol{\theta}'$ denotes the network parameters.

Then in time dimension, the extracted spatial features from different input images are chronologically encoded with multi-dimensional time-capsules, which embody multiple nonlinear temporal features among the extracted spatial features. Through an iterative process of routing the temporal information of the extracted wind speed spatial properties from the lower-layer time-capsules to the appropriate higher-layer spatio-temporal capsules, the spatio-temporal features of wind speed are captured by the spatio-temporal capsules. Subsequently, the followed linear layer reshapes the spatio-temporal data to 1D form. The last full-connected regression layer is employed to obtain wind speed prediction results for multiple wind farms.

3.4 Implementation of proposed prediction method

3.4.1 Two-layer model tuning approach

Three types of model parameters need to be trained in CapsNet, namely coupling coefficients, weights and bias. A two-layer model tuning approach is correspondingly proposed to adjust the parameters of CapsNet. The inner-layer training is used to tune coupling coefficients between the time-capsules and the spatio-temporal capsules, while the outer-layer training is responsible for optimizing the weights and bias.

For the internal training of the capsule layers, the lower-level time-capsules predict the spatio-temporal capsules by iteratively adjusting the coupling coefficient c_{ij}. The detailed internal iterative training process between the capsule layers is illustrated in Fig.5. The temporary variable b_{ij} is initially set to be 0, and the coupling coefficient c_i^1 equals 1/n according to Eq. (7). In the following iterations, c_{ij} is updated using Eq. (7) and Eq. (10)-(11). The number of iterations is determined as 3, which optimizes the model fast and leads to a low loss according to the previous research experience [28].

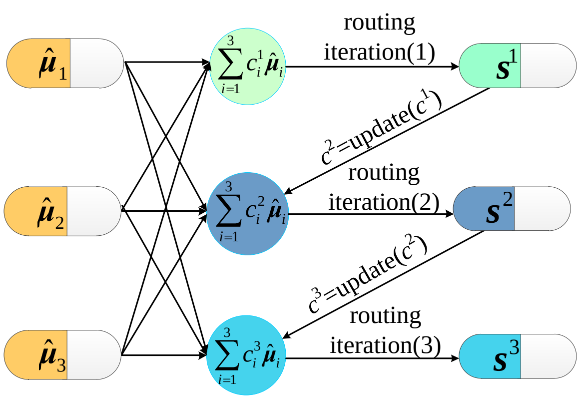


Fig. 5 Internal iterative training process between the capsule layers

The outer-layer training process is used to optimize the weights and bias of CapsNet. These parameters are trained by the back propagation rule (BP) employing adaptive moment estimation optimization algorithm (Adam) [39]. The BP training aims at minimizing the loss function E, defined as,

$$E = \frac{1}{M \times N} \frac{1}{|\boldsymbol{T}|} \sum_{t \in \boldsymbol{T}} \left\| \boldsymbol{x}_{t+\lambda} - \hat{\boldsymbol{x}}_{t+\lambda} \right\|_F \tag{13}$$

where $\boldsymbol{T}$ denotes a set of historical time points corresponding to the training sample, $|\boldsymbol{T}|$ denotes the size of the training samples, $\|\cdot\|_F$ represents the Frobenius norm, $\boldsymbol{x}_{t+\lambda}$ and $\hat{\boldsymbol{x}}_{t+\lambda}$ represent the actual and the predicted wind speed of the wind farms respectively. The error differentials are propagated in a top-down manner to adjust the weights and bias towards the optimal states. The parameters are updated based on the following rules:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} - \alpha_k \boldsymbol{m}_k / (\sqrt{\boldsymbol{e}_k} + \eta) \tag{14}$$

$$\boldsymbol{b}_k = \boldsymbol{b}_{k-1} - \alpha_k \boldsymbol{m}_k / (\sqrt{\boldsymbol{e}_k} + \eta) \tag{15}$$

where $\boldsymbol{w}_k$ and $\boldsymbol{b}_k$ denote the weight vector and bias vector respectively; k represents the number of iterations; $\boldsymbol{m}_k$ and $\boldsymbol{e}_k$ indicate the first-order and second-order moment estimation vectors; α_k represents the learning rate, and the parameter η is utilized to avoid zero denominator. In each iteration, $\boldsymbol{m}_k$ and $\boldsymbol{e}_k$ can be updated as:

$$\boldsymbol{m}_k = \varsigma_1 \boldsymbol{m}_{k-1} + (1-\varsigma_1) \cdot \partial E_k / \partial \boldsymbol{w}_k \tag{16}$$

$$\boldsymbol{e}_k = \varsigma_2 \boldsymbol{e}_{k-1} + (1-\varsigma_2) \cdot (\partial E_k / \partial \boldsymbol{w}_k)^2 \tag{17}$$

where ς_1 and ς_2 represent two exponential decay rate parameters, and their values can be set as 0.9 and 0.999 respectively. The periodic training process is completed when the training epoch reaches the preset value N_T. In order to accelerate the convergence speed and reduce the model errors, a trial-and-error method is embedded to the training process. Concretely, every few training periods, all alternative learning rates are tested once to select the one which results in the smallest prediction error. Then the updated learning rate is adopted in the next training.

3.4.2 Implementation steps for CapsNet based prediction method

The proposed spatio-temporal wind speed prediction method combines various techniques such as feature extraction, dynamic routing mechanism, back propagation rule and trial-and-error technique. In summary, the detailed implementation steps for the proposed CapsNet based spatio-temporal wind speed prediction method are graphically presented in Fig. 6, where the two-layer iterative training process is also included.

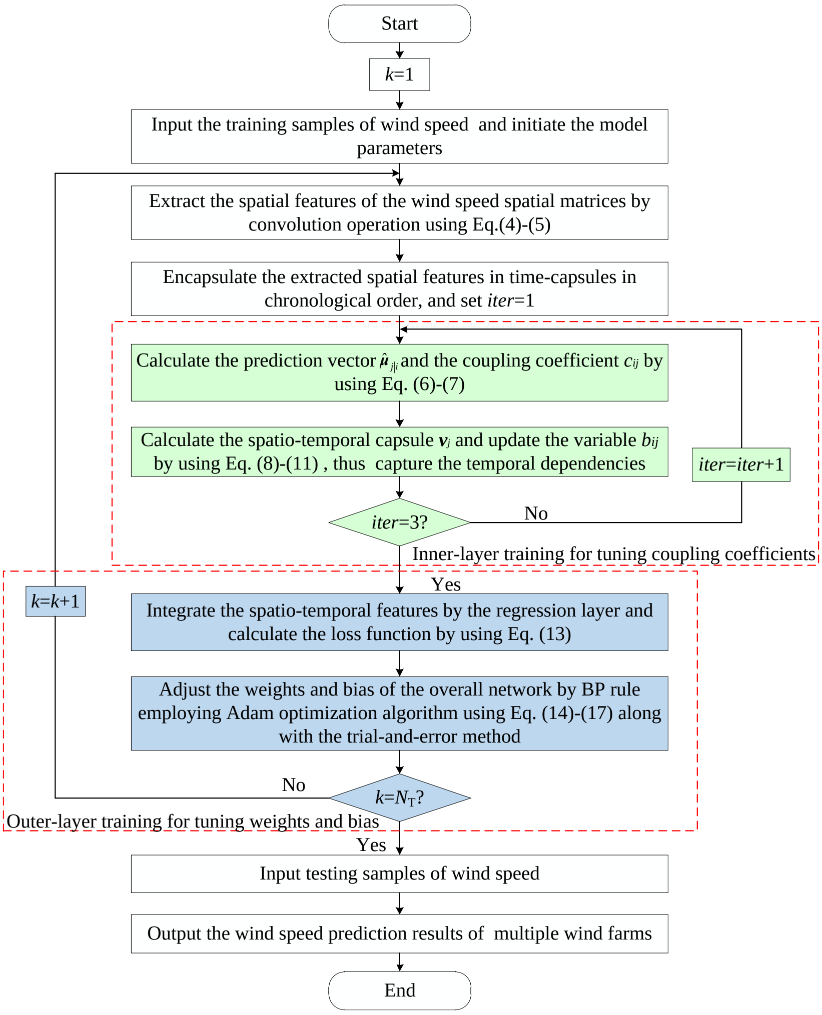


Fig. 6 Implementation framework of the proposed method

4 Case study

In this section, the proposed spatio-temporal wind speed prediction method for multiple wind farms has been comprehensively evaluated and benchmarked using real-world data.

4.1 Experimental settings

In this case study, 16 wind farms, represented by a 4×4 grid, from Ohio in the middle east of the United States are selected for wind speed prediction. The data is collected from the Wind Integration National Dataset (WIND) which provides wind speed data in the United States from 2007 to 2013 with an observation interval of 5 minutes. The dataset used in this study covers the annual wind speed data of 2012 and the time resolution is transferred to 15 minutes. The proportion of the training samples and the testing samples of the prediction model is set to 4:1.

According to Section 3, the CapsNet based spatio-temporal prediction model is established. The time length of the look-back period is 3 hours, and thus the previous 12 wind speed spatial matrices represented by 12 frames of images are used for prediction. Accordingly, the size of the input layer is 4×4×12. The first layer is a convolution layer with 4 filters (kernel size: 3 × 3, stride: 1). Then the linear layer with the size of 16×1×12 is followed to integrate the local spatial features. There are 16 channels in the time-capsule layer, and each channel consists of 3 4D capsules. The spatio-temporal capsule layer is composed of 16 4D capsules. The next linear layer reshapes the data to 1D form. Finally, the regression layer generates the 1D prediction results for the wind farms. The model parameters including the coupling coefficient c_{ij}, the shared weights $\boldsymbol{\beta}_{tp}$, the weight matrix $\boldsymbol{W}_{ij}$ and the bias $\boldsymbol{b}_p$ are randomly initiated, and optimized based on the two-layer model tuning approach. To prevent over-fitting problem, early-stopping can be adopted, that is, the outer-layer iterative training is stopped when the training epoch N_T reaches the pre-determined value 100. The alternative learning rates are selected from {0.5 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001, 0.00005}. The forecasting horizon λ ranges from 15 minutes to 3 hours.

To verify the superiority of the proposed method, various algorisms, including statistical method-ARIMA [9], machine learning methods-SVM [12], MLP [40], DBN [15], RNN [41], CNN [16], ComPonentNet (CPNet) [42], and hybrid methods-ST-GWO-MSVM [43] and CNN+MLP [44], are used as benchmarks. The prediction algorithms are implemented in Matlab R2018b and conducted on a 64-bit personal computer with Intel(R) core i7-7700 CPU/16.00 GB RAM.

4.2 Performance metrics

In this paper, three metrics, including mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) are used to assess the prediction performance. For a single site (m, n), the MAE, MAPE, and RMSE are denoted as,

$$\varepsilon_m(m,n) = \frac{1}{|\boldsymbol{T}|}\sum_{t\in\boldsymbol{T}} \left| y(m,n)_{t+\tau} - \hat{y}(m,n)_{t+\tau} \right| \tag{18}$$

$$\varepsilon_p(m,n) = \frac{1}{|\boldsymbol{T}|}\sum_{t\in\boldsymbol{T}} \frac{\left| y(m,n)_{t+\tau} - \hat{y}(m,n)_{t+\tau} \right|}{y(m,n)_{t+\tau}} \times 100\% \tag{19}$$

$$\varepsilon_r(m,n) = \sqrt{\frac{1}{|\boldsymbol{T}|}\sum_{t\in\boldsymbol{T}} \left(y(m,n)_{t+\tau} - \hat{y}(m,n)_{t+\tau} \right)^2} \tag{20}$$

where $y(m,n)_{t+\tau}$ and $\hat{y}(m,n)_{t+\tau}$ represent the observed and predicted value respectively, $\boldsymbol{T}$ denotes the set of time points corresponding to the testing samples, $|\boldsymbol{T}|$ denotes the size of the testing samples. For multiple wind farms, the evaluation indices are modified as,

$$MAE = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} \varepsilon_m(m,n) \tag{21}$$

$$MAPE = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} \varepsilon_p(m,n) \tag{22}$$

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} \varepsilon_r(m,n)^2} \tag{23}$$

4.3 Results and analysis

(1) Case 1

To validate the effectiveness of the CapsNet based spatio-temporal wind speed prediction method, the individual prediction methods including ARIMA, MLP and CNN are employed as comparison algorithms. Note that, the individual prediction methods merely capture the temporal denpendencies of wind speed time series from a single wind site, and achieve wind speed prediction of multiple wind farms one by one. The dataset covers wind speed from August 3, 2012 to August 13, 2012, and the 15-min ahead prediction results are shown in Table 1.

Table 1 15-min ahead prediction results of CapsNet and individual prediction models

Performance metrics	Models			
	ARIMA	MLP	CNN	CapsNet
MAE (m/s)	0.5743	0.5629	0.5462	**0.4035**
MAPE (%)	12.4619	11.2867	11.205	**6.1834**
RMSE (m/s)	0.8369	0.7848	0.7063	**0.6074**

From table 1, it can be clearly seen that all evaluation metrics of CapsNet are superior to those of individual prediction models. Specifically, compared to CNN, the improvements CapsNet shows in MAE, MAPE and RMSE are 26.13%, 5.02% and 14.01% respectively. Those improvements reach 28.32%, 5.10% and 22.64% when compared with MLP. By contrast, the statistical model ARIMA exhibits the worst forecasting results with the MAE, MAPE and RMSE of 0.5743 m/s, 12.4619% and 0.8369 m/s, respectively. The reasons can be explained as follows: 1) Susceptible to the over-fitting problem, ARIMA has limited capability to learn the strong randomness and nonlinearity in wind speed series. 2) Although CNN and MLP perform better results than the statistical model ARIMA due to their deep learning architectures, the lack of considering spatial correlations of wind speed leads to suboptimal predictions. Case 1 demonstrates that the proposed prediction method based on CapsNet using both spatial and temporal correlations of wind speed can obtain more accurate prediction results than individual prediction methods which merely capture wind speed features on time scales.

(2) Case 2

To verify the superiority of the CapsNet based prediction method among numerous spatio-temporal prediction methods, state-of-the-art deep learning methods, including DBN, RNN, CNN, CPNet, as well as hybrid methods, i.e., ST-GWO-MSVM and CNN+MLP, are used as benchmarking algorithms. CPNet is a family of architectures, which consists of the core CPNet, the fully-fused CPNet and the bottom-fused CPNet, based on fully convolutional neural networks. Spatio-temporal wind speed prediction is achieved by processing u- and v-components of wind speed spatial information and predicting both u- and v-components of wind speed. Here, the standard CPNet, namely the core CPNet is used for comparison. Regarding to ST-GWO-MSVM, wind speed spatio-temporal (ST) correlations are analyzed, and multi-output support vector machine (MSVM) with grey wolf optimizer (GWO) is adopted to obtain wind speed for multiple wind farms. The GWO algorithm is utilized for parameters optimization of the MSVM model. In terms of the hybrid architecture CNN+MLP, CNN is used to extract wind speed spatial features while MLP is used to capture the temporal dependencies. The spatio-temporal wind speed data is flattened to 1D vectors before being fed to ST-GWO-MSVM, RNN and DBN, and reshaped as 2D images for CNN. The u- and v-components of wind speed during the historical time period are reshaped as two-branch 2D images for CPNet. The 15-min ahead wind speed prediction results are shown in Table 2, and the histogram of statistical performance indices is illustrated in Fig.7.

The results demonstrate that CapsNet outperforms the other six competitive models in terms of all error indices. Specifically, compared with the hybrid prediction methods, the MAE, MAPE and RMSE of CapsNet are 6.38%, 0.87% and 5.15% respectively lower than those of CNN+MLP, and 8.90%, 1.02% and 7.69% respectively lower than those of ST-GWO-MSVM. The presented results indicate that the proposed model with convolution and capsule sub-networks can more effectively capture the spatio-temporal features of wind speed time series to predict the wind speed for multiple wind farms. Also, the prediction errors of CapsNet are lower than those of CNN+MLP model, which implies that although the spatial features of wind speed can be extracted by CNN of the hybrid model CNN+MLP, MLP cannot sufficiently capture the temporal correlations due to the lack of modeling mechanism for time series. As for ST-GWO-MSVM, a multi-output wind speed prediction model is built to predict the wind speed of strong correlation, moderate correlation and weak correlation. However, the wind speed data of multiple wind farms is input as flattened 1D vectors, which weakens the spatio-temporal information implicit in the original data to a certain extent.

Compared with single-model prediction methods, CapsNet improves the MAE, MAPE and RMSE by 10.65%, 1.34% and 10.35% respectively compared to CPNet, and 17.62%, 2.13% and 12.5% respectively compared to CNN. Such improvements reach 21.51%, 3.47% and 13.24% compared to RNN, and 24.9%, 3.85% and 13.64% compared to DBN. The reason can be explained that despite taken into account both spatial and temporal information, the single models cannot directly process 3D spatio-temporal wind speed data. Either the reshaped 1D vectors of RNN and DBN, or the input 2D images of CNN lose considerably valuable spatial information originally contained in the wind speed spatial matrices to varying degrees. Although the horizontal and vertical spatial characteristics of wind speed are considered for CPNet, there is still a lack of effective mechanism for temporal features extraction. Among the single-model prediction methods, CPNet and CNN perform higher accuracy than RNN and DBN. It implies that the 1D form of wind speed data makes it difficult for RNN and DBN to handle the wind speed spatio-temporal correlations in an appropriate way. In addition, the prediction errors of RNN are lower than those of DBN because the recurrent signals of RNN can help it to capture the wind speed temporal dependencies.

Table 2 15-min ahead prediction results of spatio-temporal prediction models

Performance metrics	Models						
	DBN	RNN	CNN	CPNet	ST-GWO-MSVM	CNN+MLP	CapsNet
MAE (m/s)	0.5373	0.5137	0.4898	0.4516	0.4429	0.4310	**0.4035**
MAPE (%)	10.0309	9.6556	8.3131	7.5281	7.2068	7.0509	**6.1834**
RMSE (m/s)	0.7033	0.7001	0.6942	0.6775	0.6580	0.6404	**0.6074**

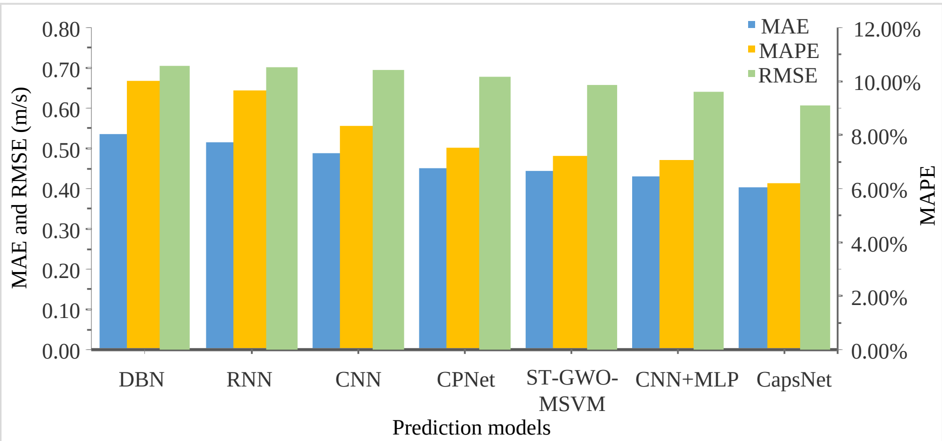
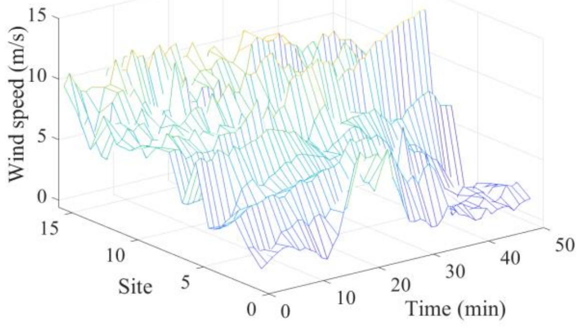
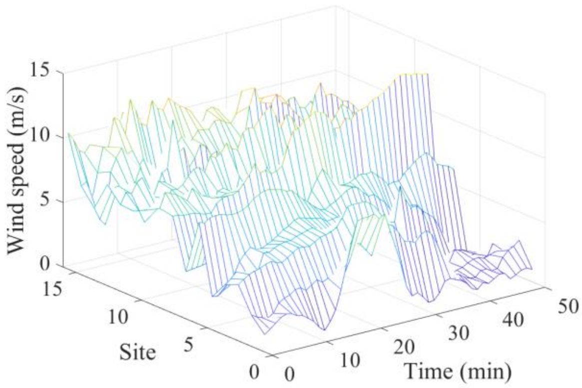


Fig. 7 15 min ahead wind speed prediction errors in August

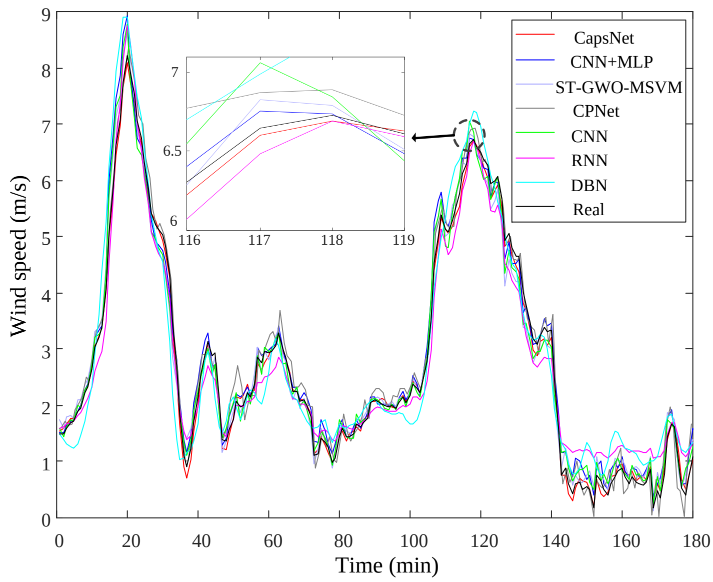


(a)

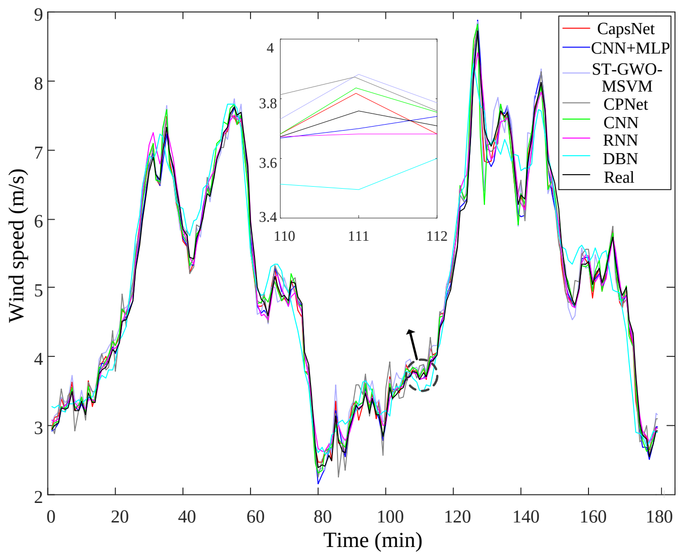


(b)

Fig. 8 (a) Wind speed prediction results of 16 wind farms in August by CapsNet (b) Actual wind speed values of 16 wind farms in August



(a)



(b)

Fig. 9 (a) Wind speed prediction results of the site (1, 2) in August (b) Wind speed prediction results of the site (2, 3) in August

To intuitively verify the prediction performance of CapsNet, the wind speed prediction results and the actual wind speed values of the 16 wind farms are respectively depicted in Fig. 8 (a) and (b). The figures describe that the prediction results of the proposed method have very similar contours and variation trends to the real wind speed data. Moreover, the prediction curves of the site (1, 2) and the site (2, 3) by the benchmarking algorithms are graphically presented in Fig. 9 (a) and (b) respectively. From Fig. 9, it can be seen that the prediction curves of CapsNet are concentrated near the real wind speed curves. In contrast, there are large gaps between the prediction curve of DBN and the actual wind speed curve. This is because the one-dimensional input form and the sample size limit the capability of DBN to learn the intrinsic spatio-temporal features of wind speed time series.

For a more comprehensive comparison between the prediction models, the monthly prediction results for the year of 2012 are evaluated by the three indices, as shown in Table 3. It can be seen that the MAE of CapsNet varies from 0.2596 to 0.7593 with an average of 0.4458, while the average MAE of CNN+MLP, ST-GWO-MSVM, CPNet, CNN, RNN and DBN is 0.4931, 0.5081, 0.5230, 0.5448, 0.6086, and 0.6937, respectively. Also, compared to the other algorithms, the average MAPE of CapsNet is improved from 0.93% to 4.11% with an average of 2.10%, and the improvements of the average RMSE range from 9.15% to 28.70% with an average of 16.74%. In addition, it can be found that the proposed method has a slightly worse fitting degree in July and September than that in other months, which can be explained that wind speed in these months are more irregular and unpredictable. It can be concluded from Table 3 that the CapsNet based prediction method outperforms the other four benchmarking algorithms throughout the year, proving that the proposed method can maintain strong wind speed prediction stability with different degrees of fluctuation.

Table 3 Monthly MAE (m/s), MAPE (%) and RMSE (m/s) of the prediction models

Month	Evaluation indices	CapsNet	CNN+MLP	ST-GWO-MSVM	CPNet	CNN	RNN	DBN
Jan.	MAE	**0.4755**	0.5276	0.5292	0.5314	0.5328	0.5727	0.9216
	MAPE	**7.6050**	8.4850	8.5189	8.6858	8.7450	9.1452	14.6705
	RMSE	**0.8097**	0.7340	0.7965	0.8118	0.8211	0.8710	1.2166
Feb.	MAE	**0.3748**	0.4113	0.4398	0.4604	0.4951	0.5355	0.6191
	MAPE	**6.9574**	7.2150	7.6855	8.4874	9.7226	9.8350	11.6118
	RMSE	**0.5493**	0.6688	0.6817	0.6832	0.7019	0.6964	0.7816
Mar.	MAE	**0.3989**	0.4439	0.4594	0.4932	0.5192	0.5239	0.5560
	MAPE	**4.2733**	4.7414	4.9342	5.2494	5.5359	5.5950	5.9253
	RMSE	**0.5736**	0.6638	0.6678	0.6783	0.6805	0.6935	0.7191
Apr.	MAE	**0.3015**	0.3627	0.3969	0.4128	0.4213	0.4577	0.4913
	MAPE	**6.3550**	7.5850	7.9715	8.3181	8.6650	9.2500	9.9450

	RMSE	**0.4120**	0.4729	0.5062	0.5362	0.5587	0.6581	0.6656
May.	MAE	**0.4658**	0.4819	0.5015	0.5189	0.5386	0.6044	0.7402
	MAPE	**8.8879**	9.3350	9.8122	10.4670	10.7052	11.8591	14.2704
	RMSE	**0.6950**	0.7090	0.7103	0.7548	0.7019	0.8266	0.9365
Jun.	MAE	**0.4368**	0.5139	0.5242	0.5327	0.5487	0.6725	0.7247
	MAPE	**5.9850**	6.9307	7.1704	7.3165	7.3415	8.7423	9.4650
	RMSE	**0.6337**	0.7143	0.7259	0.7384	0.7465	0.8672	0.9391
Jul.	MAE	**0.7593**	0.7727	0.7708	0.7937	0.8087	0.9742	1.0224
	MAPE	**18.5514**	18.9650	19.1002	19.4637	19.9213	23.8050	24.9514
	RMSE	**1.0972**	1.2402	1.2337	1.3629	1.5109	1.5585	1.6370
Aug.	MAE	**0.4035**	0.4310	0.4429	0.4516	0.4898	0.5137	0.5373
	MAPE	**6.1834**	7.0509	7.2068	7.5281	8.3131	9.6556	10.0309
	RMSE	**0.6074**	0.6404	0.6580	0.6775	0.6942	0.7001	0.7033
Sept.	MAE	**0.7030**	0.8380	0.8445	0.8562	0.8575	0.8657	0.9978
	MAPE	**10.0350**	12.3197	11.8109	12.0203	11.9250	12.6450	13.2716
	RMSE	**1.0438**	1.0986	1.1076	1.1884	1.1159	1.2536	1.2663
Oct.	MAE	**0.2596**	0.2951	0.3127	0.3228	0.3355	0.3752	0.3980
	MAPE	**7.7850**	9.2350	9.4852	10.2265	10.6673	11.9047	10.5548
	RMSE	**0.3611**	0.4138	0.4147	0.4295	0.4296	0.4877	0.4848
Nov.	MAE	**0.3382**	0.3420	0.3537	0.3789	0.3908	0.5822	0.6968
	MAPE	**6.4525**	7.3150	6.7536	7.0485	6.6100	10.4173	10.6450
	RMSE	**0.4391**	0.4518	0.4675	0.4977	0.4989	0.7115	0.8055
Dec.	MAE	**0.4327**	0.4973	0.5220	0.5229	0.5991	0.6260	0.6192
	MAPE	**6.2852**	7.3329	8.6252	8.7814	9.0567	9.6227	9.3450
	RMSE	**0.7389**	0.8310	0.8443	0.8423	0.8614	0.8308	0.9041
AVG	MAE	**0.4458**	0.4931	0.5081	0.5230	0.5448	0.6086	0.6937
	MAPE	**7.9463**	8.8759	9.0896	9.4661	9.7674	11.0397	12.0572
	RMSE	**0.6571**	0.7233	0.7345	0.7668	0.7768	0.8463	0.9216

(3) Case 3

Case 3 is studied on the wind speed dataset of August in 2012 and the prediction horizon is extended from 15 minutes to 3 hours. The results are presented in Table 4 to Table 6. As shown in the tables, the proposed model based on CapsNet holds the dominant position over other benchmarking algorithms in terms of MAE, MAPE, as well as RMSE regarding different prediction horizons. Besides, with the extension of the prediction horizon, the prediction accuracy of all prediction models decreases, and the prediction errors of other benchmarking algorithms grow faster than CapsNet. Specifically, the RMSE of CapsNet is 5.15%, 7.69%, 10.35%, 12.50%, 13.24%, and 13.64% lower than that of CNN+MLP, ST-GWO-MSVM, CPNet, CNN, RNN and DBN respectively in 15-min ahead wind speed prediction. In 3-hour ahead prediction, these improvements reach 48.19%, 50.50%, 53.26%, 54.49%, 61.02%, and 62.84%, respectively. It can also be observed that the hybrid CNN+MLP has generally better performance compared to the other prediction methods over all prediction tasks. Moreover, among the single-model prediction methods, DBN generates fairly poor results when the prediction horizons are extended longer. For instance, regarding the 15-min ahead prediction, the MAE of DBN is 24.90% higher than that of CapsNet, while this value increases to 56.14% when performing 3-hour ahead prediction.

The reasons for the results of Case 3 lie in four aspects: 1) The hierarchical sub-networks of CapsNet and CNN+MLP enables these models to learn the spatial and temporal correlations of wind speed in a targeted manner, therefore achieving more accurate prediction results based on richer knowledge. 2) In CapsNet, the wind speed dependencies among the spatial properties are encoded with multi-dimensional capsules and learned through an iterative information transfer process between the capsules, and thus the intrinsic spatio-temporal features are comprehensively extracted by the discriminatively trained CapsNet. 3) CapsNet adopts novel activation functions such as the Leaky ReLU function and the Squash function, which have better activation performance compared with conventional non-linear activation functions such as Sigmoid and help the proposed model to cope with complex high-dimensional wind data. 4) As the prediction horizon is extended, the temporal correlations of wind speed time series decrease, and thus it is more significant to capture the spatial correlations of wind speed for wind speed prediction. Case 3 proves the state-of-the-art performance of CapsNet in handling spatio-temporal wind speed prediction problem under various prediction horizons.

Table 4 MAE of the prediction models (m/s)

Models	Prediction horizons (min)				
	15	30	60	120	180
DBN	0.5373	0.8618	0.9381	1.4860	1.6857
RNN	0.5137	0.7936	0.8151	1.2422	1.4210
CNN	0.4898	0.6909	0.7211	0.8436	1.1405
CPNet	0.4516	0.6238	0.7094	0.8159	1.1197
ST-GWO-MSVM	0.4429	0.5452	0.6523	0.7922	0.9138
CNN+MLP	0.4310	0.5257	0.6235	0.7585	0.8233
CapsNet	**0.4035**	**0.4970**	**0.5304**	**0.6290**	**0.7393**

Table 5 MAPE of the prediction models (%)

Models	Prediction horizons (min)				
	15	30	60	120	180
DBN	10.0309	11.2117	15.8245	21.2534	29.8739
RNN	9.6556	10.5749	15.2782	20.1404	28.0984
CNN	8.3131	9.8601	13.1329	19.2867	25.8377
CPNet	7.5281	9.3034	12.7128	19.1085	24.6385

ST-GWO-MSVM	7.2068	8.9105	12.1335	18.8154	24.3931
CNN+MLP	7.0509	8.4450	11.3574	18.6062	23.8926
CapsNet	**6.1834**	**7.3916**	**10.1939**	**17.2335**	**21.9634**

Table 6 RMSE of the prediction models (m/s)

Models	Prediction horizons (min)				
	15	30	60	120	180
DBN	0.7033	0.9208	1.702	2.151	2.4292
RNN	0.7001	0.8799	1.6964	2.0798	2.3144
CNN	0.6942	0.8296	1.3096	1.7428	1.9834
CPNet	0.6775	0.8163	1.2529	1.5681	1.9313
ST-GWO-MSVM	0.6580	0.7912	1.1553	1.3119	1.8235
CNN+MLP	0.6404	0.7761	0.9784	1.2513	1.7422
CapsNet	**0.6074**	**0.7040**	**0.8299**	**0.8822**	**0.9026**

4.4 Discussions

The model parameters of CapsNet are optimized with the two-layer iterative training process, and as suggested in previous research experience [25], [28], the number of iterations of inner-layer training is determined as 3, while the outer-layer layer training for the entire network employs BP rule with Adam optimization algorithm. Note that a trial-and-error technique is embedded in the training process for convergence acceleration. To verify the convergence of the CapsNet based prediction method, the training process for the 15-min ahead wind speed prediction is inspected by recording the loss function E. Fig. 10 illustrates the values of E for CapsNet trained by the two-layer model tuning approach with different optimization algorithms including Adam, RMSprop, Adagrad, and Adam without the trial-and-error technique. It can be found that CapsNet trained by the two-layer model tuning approach with Adam optimization algorithm converges faster than the others. In addition, the trial-and-error technique considerably accelerates the convergence speed, thereby enhancing the convergence performance of the proposed prediction method.

Furthermore, a comparison of training time for 15-min ahead spatio-temporal wind speed prediction with all the prediction models is shown in Table 7. It can be seen that the proposed CapsNet takes more time for training than the other models. This is because CapsNet is trained with the two-layer model tuning approach, which is relatively computational expensive due to the inner and outer iterative training process. Owing to the complex layer-wise training and fine-tuning process of DBN, and the recurrent connections of RNN, these two methods are more time-consuming than the other comparison methods. In addition, the hybrid prediction methods including ST-GWO-MSVM and CNN+MLP, as well as the CPNet with u- and v-components processing need more training time than CNN with relatively simpler structure. Although requiring a comparatively longer computational time, CapsNet is still attractive since the training time is within an acceptable range.

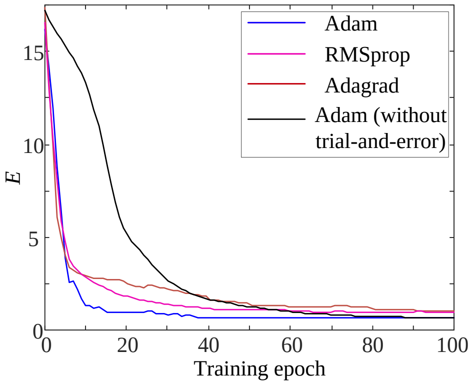


Fig. 10 Values of *E* for CapsNet during the training process

Table 7 Training time of the prediction models

	DBN	RNN	CNN	CPNet	ST-GWO-MSVM	CNN+MLP	CapsNet
Runtime/s	261.4	275.3	192.6	224.1	207.7	248.1	289.5

5 Conclusion

In this paper, a hybrid prediction method based on CapsNet is proposed to solve the spatio-temporal wind speed prediction problem for multiple wind farms. In the proposed method, the spatio-temporal features of wind speed are encoded with multi-dimensional capsules and learned by DR mechanism. The proposed method has been comprehensively tested and compared with individual prediction methods as well as other spatio-temporal prediction methods. A real-world dataset from a spatial region consisting of 16 wind farms is used to verify the feasibility and effectiveness of the proposed method. Statistically, for 15-min ahead wind speed prediction, compared to individual prediction methods, the highest improvements of MAE, MAPE and RMSE are 29.74%, 50.38%, and 27.42%, respectively. Also, compared with other spatio-temporal prediction methods, the highest improvements of MAE, MAPE and RMSE are 24.90%, 38.36%, and 13.64%, respectively. Moreover, the comparative results also show that the proposed prediction method outperforms other benchmarking algorithms in the whole year and under various prediction horizons. These experiment results prove that the proposed method can effectively deal with the spatio-temporal wind speed prediction problem and has great potentials for practical applications in smart grid.

Accurate wind power ramp events prediction helps system operators to make sensible scheduling decisions to mitigate impacts of drastic fluctuations in wind power, thereby facilitating secure and reliable operation of smart grid. Due to the low frequency of wind power ramp events, it is challenging to learn the mapping relationship between the historical wind power samples and wind power ramp events. Further ongoing research would focus on the application of CapsNet to wind power ramp events prediction for multiple wind farms.

Acknowledgment

This work was jointly supported by the Research Grants Council of the HKSAR Government (Grant No. R5020-18), the Innovation and Technology Commission of the HKSAR Government to the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center (Grant No. K-BBY1), the National Natural Science Foundation of China (51877072), and the Innovative Team Projects of Zhuhai City (ZH01110405180049PWC).

References

[1] Tao S, Xu Q, Feijóo A, et al. Wind farm layout optimization with a three-dimensional Gaussian wake model. Renew Energy 2020; 159: 553-569.

[2] Yang M, Lin Y, Zhu S, et al. Multi-dimensional scenario forecast for generation of multiple wind farms. J Mod Power Syst Clean Energy 2015; 3(3): 361-370.

[3] Zhu Q, Shi D, Zhu L, et al. Learning temporal and spatial correlations jointly: a unified framework for wind speed prediction. IEEE Trans Sustain Energy 2020; 11(1): 509-523.

[4] He M, Yang L, Zhang J, et al. A spatio-temporal analysis approach for short-term forecast of wind farm generation. IEEE Trans Power Syst 2014; 29(4): 1611-1622.

[5] Cai H, Jia X, Feng J, et al. Gaussian process regression for numerical wind speed prediction enhancement. Renew Energy 2020; 146: 2112-2123.

[6] Ye L, Zhao Y, Zeng C, et al. Short-term wind power prediction based on spatial model. Renew Energy 2017; 101: 1067-1074.

[7] Chen Y, Zhang S, Zhang W, et al. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. Energy Convers Manag 2019; 185: 783-799.

[8] Wang H, Li G, Wang G, et al. Deep learning based ensemble approach for probabilistic wind power

forecasting. Appl Energy 2017; 188: 56-70.

[9] Wang H, Lei Z, Zhang X, et al. A review of deep learning for renewable energy forecasting. Energy Convers Manag, DOI: 10.1016/j.enconman.2019.111799, in press, 2019.

[10] Yuan X, Tan Q, Lei X, et al. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 2017; 129: 122-137.

[11] Cui M, Wang J, Chen B. Flexible machine learning-based cyberattack detection using spatiotemporal patterns for distribution systems. IEEE Trans Smart Grid 2020; 11(2): 1805-1808.

[12] Shrivastava N A, Lohia K, Panigrahi B K, et al. A multiobjective framework for wind speed prediction interval forecasts. Renew Energy 2016; 87: 903-910.

[13] Li Z, Ye L, Zhao Y, et al. Short-term wind power prediction based on extreme learning machine with error correction. Protection and Control of Modern Power Systems 2016; 1(1): 9-16.

[14] Cai H, Jia X, Feng J, et al. A combined filtering strategy for short term and long term wind speed prediction with improved accuracy. Renew Energy 2019; 136: 1082-1090.

[15] Wang K, Qi X, Liu H, et al. Deep belief network based k-means cluster approach for short-term wind power forecasting. Energy 2018; 165: 840-852.

[16] Harbola S, Coors V. One dimensional convolutional neural network architectures for wind prediction. Energy Convers Manag 2019; 195: 70-75.

[17] Zhao Y, Ye L, Pinson P, et al. Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting. IEEE Trans Power Syst 2018; 33(5): 5029-5040.

[18] Dowell J, Weiss S, Hill D, et al. Short-term spatio-temporal prediction of wind speed and direction. Wind Energy 2013; 17(12): 1945-1955.

[19] Ezzat A A, Jun M, Ding Y, et al. Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting. IEEE Trans Sustain Energy 2018; 9(3): 1437-1447.

[20] Khodayar M, Wang J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans Sustain Energy 2019; 10(2): 670-681.

[21] Lu P, Ye L, Zhong W, et al. A novel spatio-temporal wind power forecasting framework based on multi-output support vector machine and optimization strategy. Journal of Cleaner Production, DOI: 10.1016/j.jclepro.2020.119993, in press, 2020.

[22] Sun M, Feng C, Zhang J. Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation. Appl Energy, DOI: 10.1016/j.apenergy.2019.113842, in press, 2019.

[23] Ren H, Laima S, Chen W, et al. Numerical simulation and prediction of spatial wind field under complex terrain. J Wind Eng Ind Aerod 2018; 180: 49-65.

[24] Liu Y, Qin H, Zhang Z, et al. Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. Appl Energy, DOI: 10.1016/j.apenergy.2019.114259, in press, 2020.

[25] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules. Proc 31st Int Conf Neural Inf Process Syst 2017.

[26] Peng D, Zhang D, Liu C, et al. BG-SAC: entity relationship classification model based on self-attention supported capsule networks. Appl Soft Comput, DOI: 10.1016/j.asoc.2020.106186, in press, 2020.

[27] Paoletti M E, Haut J M, Fernandez R, et al. Capsule networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens 2019; 27(4): 2145-2160.

[28] Yang M, Zhao W, Chen L, et al. Investigating the transferring capability of capsule networks for text classification. Neural Netw 2019; 118: 247-261.

[29] Vesperini F, Gabrielli L, Principi E, et al. Polyphonic sound event detection by using capsule neural networks. IEEE J Sel Top Signal Process 2019; 13(2):310-322.

[30] Yu Y, Guan H, Li D, et al. A hybrid capsule network for land cover classification using multispectral LiDAR data. IEEE Geosci Remote Sens Lett 2020; 17(7): 1263-1267.

[31] Yao H, Gao P, Wang J, et al. Capsule network assisted IoT traffic classification mechanism for smart cities. IEEE Internet Things J 2019; 6(5): 7515-7525.

[32] Sezer O B, Gudelek M U, Ozbayoglu A M. Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. Appl Soft Comput, DOI: 10.1016/j.asoc.2020.106181, in press, 2020.

[33] Ramírez I, Cuesta-Infante A, Schiavi E, et al. Bayesian capsule networks for 3D human pose estimation from single 2D images. Neurocomputing 2020; 379(28): 64-73.

[34] Ma X, Zhong H, Li Y, et al. Forecasting transportation network speed using deep capsule networks with nested LSTM models. IEEE Trans Intell Transp Syst, DOI: 10.1109/TITS.2020.2984813, in press, 2020.

[35] Khodayar M, Kaynak O, Khodayar M E. Rough deep neural architecture for short-term wind speed forecasting. IEEE Trans Ind Informat 2017; 13(6): 2770-2779.

[36] Kim T Y, Cho S B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019; 182: 72-81.

[37] Zhao X, Wei H, Wang H, et al. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. Solar Energy 2019; 181: 510-518.

[38] Lin B, Fu S, Zhang C, et al. Optical fringe patterns filtering based on multi-stage convolution neural network. Opt Lasers Eng, DOI: 10.1016/j.optlaseng.2019.105853, in press, 2020.

[39] Zhang Z, Ye L, Liu Y, et al. Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression. Appl Energy 2019; 247: 270-284.

[40] Islam F, Al-Durra A, Muyeen S M. Smoothing of wind farm output by prediction and supervisory-control-unit-based FESS. IEEE Trans Sustain Energy 2013; 4(4): 925-933.

[41] Yu C, Li Y, Bao Y, et al. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Convers Manag 2018; 178(15): 137-145.

[42] Bruno Q B, Fernando L. C O, Ruy L M. Componentnet: Processing U- and V-components for spatio-temporal wind speed forecasting. Electr Power Syst Res, DOI: 10.1016/j.epsr.2020.106922, in press, 2021.

[43] Lu P, Ye L, Zhong W, et al. A novel spatio-temporal wind power forecasting framework based on multi-output support vector machine and optimization strategy, J Clean Prod, DOI:

10.1016/j.jclepro.2020.119993.

[44] Zhu Q, Chen J, Zhu L, et al. Wind speed prediction with spatio-temporal correlation: a deep learning approach. Energies 2018; 11(4): 1-18.