Energy Conversion and Management 192 (2019) 37–51

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Long Short-Term Memory Network based on Neighborhood Gates for processing complex causality in wind speed prediction

Zhendong Zhang[a,b], Hui Qin[a,b,*], Yongqi Liu[a,b], Yongqiang Wang[c], Liqiang Yao[c], Qingqing Li[c], Jie Li[a,b], Shaoqian Pei[a,b]

[a] School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China
[b] Hubei Key Laboratory of Digital Valley Science and Technology, Wuhan, Hubei, China
[c] Changjiang River Scientific Research Institute of Changjiang Water, Resources Commission, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Wind speed prediction
Long Short-Term Memory Network
Neighborhood Gates
Equivalent tree causality

ABSTRACT

Obtaining high-precision wind speed prediction results is very beneficial to the utilization of wind energy and the operation of the power system. The purpose of this study is to develop a novel model for wind speed causality processing and short-term wind speed forecasting. In this study, the hybrid model combining causality processing strategy called "decomposition- virtual nodes-pruning" and Long Short-Term Memory Network based on Neighborhood Gates is proposed to obtain high-precision wind speed predictions. First, Pearson Correlation Coefficient, Maximal Information Coefficient and Granger causality test are used to explore the correlation and causality between wind speed and meteorological factors. Then, the causality is divided into five categories: center, chained, ring, tree and network causality, according to the topological structure of causality. Next, all types of causality can be unified into an equivalent tree causality by the causality processing strategy. Afterward, Long Short-Term Memory Network based on Neighborhood Gates is proposed to dynamically adjust the network structure according to the specific equivalent tree causality. Finally, the performance of the proposed model is verified by eight models from three aspects with different features, different methods and different equivalent trees in the case in Fuyun meteorological station, Xinjiang province, China. The evaluation metrics of the prediction results obtained by the proposed model are optimal among the eight models. The experimental results show that the proposed model is very competitive and very suitable for processing complex causality in wind speed prediction.

1. Introduction

Wind energy is a promising renewable and clean energy that has received widespread attention around the world [1]. More and more wind power is integrated into the power grid, making the power system unreliable due to the fluctuation and randomness of wind speed [2]. Therefore, accurately predicting wind speed is critical to the utilization of wind energy and operation of the power grid.

Wind speed prediction methods can usually be divided into two categories that are physical methods and statistical methods [3]. The physical methods simulate wind formation process through a complex mathematical physics model to predict wind speed, such as numeric weather prediction (NWP) [4]. Soman et al. [5] made an overview of existing wind speed and wind power forecasting methods including NWP, statistical methods and hybrid methods. Al-Yahyai et al. [6] made an overview of the NWP method and discussed how to solve the shortcomings for classical wind energy measurement. NWP's prediction accuracy is high and model interpretability is strong. However, its data collection is difficult, modeling is complex and solution is time-consuming [7].

Statistical methods use historical data and related factors to predict wind speed. Chen and Yu used Support Vector Regression (SVR) [8] based on unscented Kalman filter and state-space to achieve short-term wind speed prediction. Zhang and Nielsen used Gaussian Process Regression (GPR) [9] and Quantile Regression (QR) [10] to predict short-term wind speed and wind power, respectively, and evaluated the uncertainty of the prediction. Li et al. compared the performance of three different Artificial Neural Networks (ANN) [11] on wind speed prediction. In the past decade, optimizing the parameters of these methods or proposing variants of these methods is one of the research directions of wind speed prediction statistical method. At the same time, new machine learning methods are being tried to predict wind speed, such

* Corresponding author at: School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China.
E-mail address: hqin@hust.edu.cn (H. Qin).

https://doi.org/10.1016/j.enconman.2019.04.006
Received 10 January 2019; Received in revised form 15 March 2019; Accepted 2 April 2019

as Extreme Learning Machine (ELM) [12], Wavelet Neural Network (WNN) [13] and so on. Wang et al. used ELM and proposed an adaptive model to predict short-term wind speed [12]. Santhosh proposed adaptive WNN for wind speed prediction [13]. In addition to the use of statistical methods alone, some hybrid methods are used to improve wind speed prediction accuracy by combing the advantages of multiple methods. The hybrid approach includes a combination of multiple statistical methods and a combination of statistical method and data-preprocessing method. For the former, Liu et al. combined Auto Regressive Integrated Moving Average (ARIMA) [14] and ANN so that it (ARIMA-ANN) had both the ability to process time information and non-linearity in wind speed prediction. For the latter, Peng et al. proposed a two-stage decomposition algorithm combining the Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and the Variational Mode Decomposition (VMD) to deal with the nonlinearity of wind speed time series [15]. Then, Two-stage Decomposition Technique was combined with AdaBoost-ELM to construct a hybrid model for predicting wind speed [15]. The introduction of these new statistical methods has injected fresh blood into the field of wind speed prediction, providing more options for solving the actual wind speed prediction problem.

Due to the rapid development of deep learning in recent years, the performance of many traditional machine learning methods is inferior to that of deep learning methods [16]. Among the deep learning methods, Recurrent Neural Networks (RNN) is good at dealing with sequence problems [17]. However, RNN has long-term dependency problem when predicting sequence problems [18], so Long Short-Term Memory Network (LSTM) is proposed to solve this problem [19]. Since wind speed is time series data, LSTM has been used to predict wind speed [20]. Qin et al. proposed a hybrid forecasting model based on LSTM and deep learning neural network for wind signal [20]. Li et al. used EWT decomposition and LSTM for multi-step wind speed forecasting [21]. Han et al. proposed a hybrid model based on copula function and LSTM and applied it to predict mid-to-long term wind and photovoltaic power generation [22]. Hu and Chen proposed a nonlinear wind speed forecasting model using LSTM and hysteretic ELM and used Differential Evolution algorithm to optimize the model [23]. Liu et al. proposed a smart multi-step deep learning model for wind speed forecasting based on VMD, singular spectrum analysis, LSTM and ELM [24]. Some variants of LSTM can also be used to predict wind speed, such as Gated Recurrent Unit (GRU) [25]. Yu et al. proposed a novel framework for wind speed prediction based on RNN and SVR, and they specifically combined LSTM and SVR, GRU and SVR [26].

Data preprocessing techniques have been a new tendency to improve wind speed prediction accuracy. In addition to the decomposition techniques mentioned above and their variants, feature selection and feature optimization are also an effective data preprocessing techniques for improving wind speed prediction accuracy. Wang et al. combined a variety of preprocessing techniques including VMD, Kullback-Leibler divergence, energy measure and sample entropy for feature selection and extraction, which greatly helped to improve the accuracy of LSTM [27]. Carta et al. performed feature selection by correlation analysis for meteorological data [28]. Salcedo-Sanz first used the coral reefs optimization algorithm for feature optimization and then used ELM to predict wind speed [29]. The results of these papers show that accuracy of wind speed prediction is significantly improved both in feature selection and feature optimization. In fact, feature selection and feature optimization are part of the feature engineering in the field of deep learning. Fan et al. improved energy prediction performance by deep learning-based feature engineering [30]. Their experimental results confirm the ability of feature engineering in aspects of reducing data dimensionality, decreasing prediction model complexity and tackling the problem of corrupted and noisy information. Feng et al. used principal component analysis, autocorrelation analysis and Granger causality test in feature engineering to deeply select features, which significantly improved the prediction accuracy of their models [31].

In feature engineering, correlation analysis and causality analysis are important means to improve model prediction performance [32]. It is well known that wind speed is affected by many meteorological factors, including air pressure, temperature, humidity and so on [33]. Analyzing the correlation and clarifying the causality between these factors are important to improve the accuracy of wind speed prediction and enhance the interpretability of the model. There are several methods are commonly used for correlation analysis in feature engineering, such as correlation coefficient method [34], covariance method [35] and maximum information coefficient (MIC) method [36]. The methods commonly used for causality analysis include theoretical analysis [37], transfer entropy [38] and Granger causality test [39]. In this study, Pearson Correlation Coefficient (PCC) is used to explore linear correlation since it is a classical method and is widely used [34]. MIC is used to explore nonlinear correlation since it can analyze correlations of many complex types [36]. Granger causality test is used to explore causality because it is also a classic method and is widely used [39].

However, due to the complexity of the causality between wind speed and meteorological factors, most of the prediction methods and techniques mentioned above will become nail-biting. Because these prediction methods can only consider meteorological factors from the perspective of correlation and not from the perspective of causality, which loses a lot of useful information. The problems faced by these prediction methods in dealing with complex causality are explained in detail in Section 2. The purpose of this study is to propose a causality processing strategy to convert all types of causality into a general equivalent causality, and propose a new Long Short-Term Memory Network based on Neighborhood Gates (NLSTM) to dynamically adjust the network structure to be consistent with the equivalent causality, thereby improving the accuracy of wind speed prediction.

In this study, a novel method called NLSTM is proposed to handle complex causality in wind speed prediction. The main contributions are outlined as follows:

(1) The classification and unification of the causality between wind speed and meteorological factors are pioneered.
(2) NLSTM is proposed to dynamically adjust the network structure to accommodate all types of causality between wind speed and meteorological factors.
(3) NLSTM is applied to predict wind speed in Fuyun meteorological station in Xinjiang province, China and compared in these aspects: different feature, different methods and different equivalent tree causality. The experimental results show that NLSTM is very competitive and very suitable for processing complex causality in wind speed prediction.

The remainder of this paper is organized as follows. In Section 2, the problems faced by traditional wind speed prediction method in dealing with complex causality are explained. In Section 3, causality processing and the implementation details of NLSTM are introduced. In Section 4, the method evaluation metrics are explained. In Section 5, NLSTM is applied to predict wind speed in Fuyun meteorological station in Xinjiang province, China. In Section 6, the work of this paper is summarized and the conclusions are given.

2. Problem description

There are many factors that affect wind speed, such as historical wind speed, air pressure, temperature, humidity and other factors [33]. Clarifying the causality between these factors helps to improve the accuracy of wind speed prediction. However, the causality may be very complex and cannot be directly used in actual predictions. Suppose that Fig. 1 is a schematic diagram of the causality between wind speed, air pressure, temperature and humidity. In the Fig. 1, red circle represents the variable that need to be predicted (VNP), such as the wind speed of

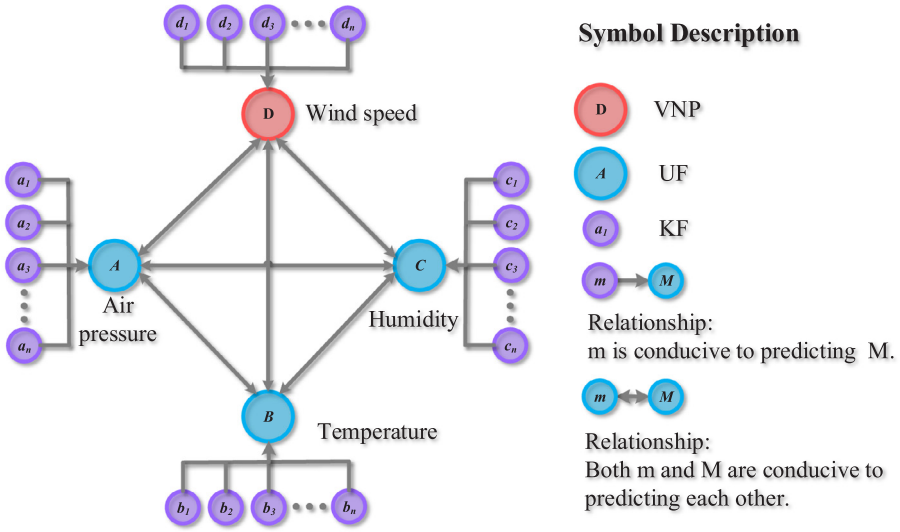

Fig. 1. Schematic diagram of the causality between VNP (wind speed) and factors (air pressure, temperature, humidity).

next period. Blue circle represents the unknown factor (UF) when predicting VNP of next period, such as the air pressure, humidity and temperature of next period. Purple circle represents the known factor (KF) when predicting VNP of next period, such as the historical wind speed, air pressure, temperature, humidity.

Now if the variable D is selected as the VNP, $[A, B, C]$ and $[d_1, d_2, \cdots, d_n]$ should be input as features. Unfortunately, $[A, B, C]$ are unknown and only $[a_1, a_2, \cdots, a_n]$, $[b_1, b_2, \cdots, b_n]$ and $[c_1, c_2, \cdots, c_n]$ are known. How to deal with these features more rationally and effectively is the focus of this study. There are five ideas to handle these features with ordinary wind speed prediction method, such as standard LSTM.

(1) Idea I: Only $[d_1, d_2, \cdots, d_n]$ are input as features. The obvious problem with this idea is that the factors considered are not comprehensive enough.
(2) Idea II: Combine all KFs

$[a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n, c_1, c_2, \cdots, c_n, d_1, d_2, \cdots, d_n]$ as feature inputs to predict D. Although the second idea considers all the factors, D and $[A, B, C]$ have a great correlation but it does not mean that D and $[A, B, C]$'s all KFs have a great correlation. These combined features are likely to add some noise, which will affect the accuracy of prediction.

(3) Idea III: Apply the feature selection again to the combined factors, removing features that have little correlation with D. This approach is a compromise in comparison with the first two approach. It not only considers the information obtained by the feature engineering but also eliminates unrelated noise, which are the advantages of this approach. However, this approach requires more than one feature screening. In addition, this approach does not fully consider the causality obtained by the feature engineering, which reduces the interpretability of model.
(4) Idea IV: First predict the value of A, B and C of the next period separately, and then use the predicted value and $[d_1, d_2, \cdots, d_n]$ as inputs to predict D. Although this method retains the causality of these factors, the problem faced in the prediction of A is the same as the prediction of D. This method is almost inoperable.
(5) Idea V: First ignore the causality between A, B, C and D, and use some other methods to predict A, B and C of the next period separately, and then use the predicted value and $[d_1, d_2, \cdots, d_n]$ as inputs to predict D. Idea V makes Idea IV seem feasible, but it brings new stack of errors. The predictions of A, B and C must have errors. Inputting the predicted value with errors into the model to predict D will cause a superposition of errors. Moreover, predicting A, B and C before the prediction of D makes the model cumbersome. And as the UF gets more, the error will become larger and the process for predicting UF will take more time.

These five ideas mainly discuss the feature input of the model, which is the first guarantee to improve the prediction accuracy of wind speed. The point prediction method such as standard LSTM is the second guarantee to improve the prediction accuracy of wind speed. This is the relation between five ideas and standard LSTM. In summary, when standard LSTM is dealing with complex causality, it is difficult to simultaneously take into account the model's prediction accuracy, interpretability and operability. In response to this difficulty, causality processing strategy and LSTM based on Neighborhood Gates (NLSTM) are proposed to process complex causality in time series regression.

3. Methods

In this section, causality processing method is first introduced, which is the basis for establishing NLSTM. After that, the specific implementation method of NLSTM is introduced. Finally, why the proposed method can improve the accuracy of wind speed prediction is qualitatively explained.

3.1. Causality processing

In this section, how to obtain the causality between wind speed and meteorological factors is first introduced. Then all types of causality are classified. On this basis, how to unify all types of causality into a general form of causality is explored.

3.1.1. Analysis of causality

In this study, Pearson Correlation Coefficient (PCC) [34] and Maximal Information Coefficient (MIC) [36] are used to explore the line and non-line correlation between wind speed and meteorological variables, respectively. The meteorological variable whose absolute value of PCC or MIC with wind speed is greater than 0.5 can be left as prediction factor. Of course, in order to avoid spurious correlation, the prediction factor must pass the significance test [40] when performing the correlation analysis. Based on the correlation analysis, Granger Causality Test [39] is used to explore the causality between factors and wind speed. Granger Causality is not the real causality but the statistical causality, no matter which causality is beneficial to improve the accuracy of wind speed prediction [41].

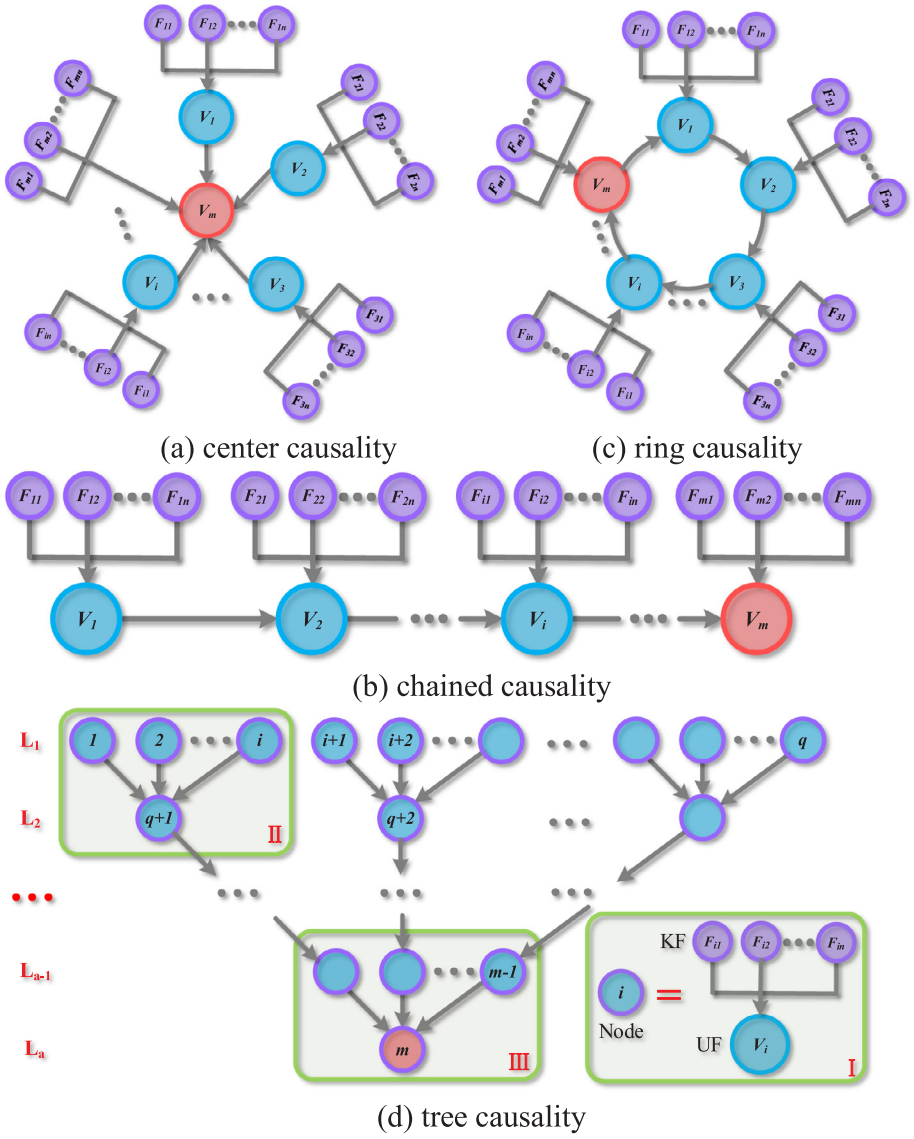

Fig. 2. Schematic diagram of causality.

3.1.2. Classification of causality

The causality between wind speed and meteorological factors is very complex and may be of various shapes. According to the shape of the causality, causality is divided into five categories: center causality, chained causality, ring causality, tree causality and network causality, as shown in the Fig. 2. Red circle V_m is the VNP, blue circle $[V_1, V_2, \cdots, V_{m-1}]$ are UFs, and purple circle F_{ij} is the j-th KF of V_i.

(1) Center causality

As shown in the Fig. 2(a), all UFs affect the VNP but the VNP does not affect any UFs.

(2) Chained causality

As shown in the Fig. 2(b), the causality between VNP and UFs is progressive. In other words, V_1 affects V_2 but V_2 does not affect V_1; V_2 affects V_3 but V_3 does not affect V_2 and so on.

(3) Ring causality

As shown in the Fig. 2(c), if V_m further affects V_1 but V_1 does not affect V_m, the chained causality becomes ring causality.

(4) Tree causality

A more complex causality structure is discussed, as shown in the Fig. 2(d), whose overall shape is tree-like. In part I, the symbol KF and UF are combined into a new symbol called node. In part II, node $q + 1$ is called the parent node of nodes $[1, 2, \cdots, i]$ and nodes $[1, 2, \cdots, i]$ are the child node of node $q + 1$. A node can have many child nodes or no child nodes but at most one parent node. If a node has no child nodes, then this node is called leaf node, such as nodes $[1, 2, \cdots, i]$. If a node has no parent node, then this node is called root node. In the tree causality, there is only one root node that can only be the last node. The root node is VNP, as the node m. $[L_1, L_2, \cdots, L_a]$ are layer numbers. The order of the node numbers needs to be increased from the layer with a small number to the layer with a large number. And in the same layer, the node number needs to be increased from left to right.

(5) Network causality

Network causality is the general form of causality, which can be seen as a combination of center causality, chained causality, ring

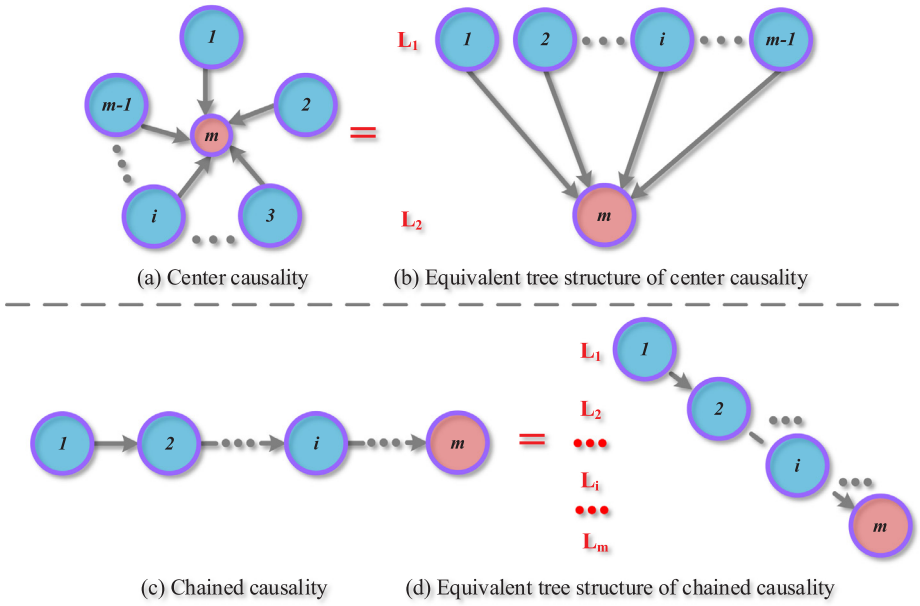

Fig. 3. Equivalent tree structure of center and chained causality.

causality and tree causality.

3.1.3. Unity of causality

Looking closely at center causality, chained causality and tree causality, it is found that both center and chained causality are two special cases of tree causality. Center causality is a special case of horizontal tree causality and chained causality is a special case of the vertical tree causality, as shown in the Fig. 3. Ring causality can be first decomposed into a set of chained causality and then equivalent to the vertical tree causality. It can be seen that center, chained and ring causality can all be converted to tree causality.

Network causality is the most common form of causality, as shown in the Fig. 4(a). This causality structure is complex and cannot be directly used for model prediction. How to transform it into equivalent tree causality structure reasonably and efficiently is crucial for designing one model to predict all types of causality. Decomposition, virtual node and pruning can effectively convert any network causality to tree causality.

(1) Decomposition

Since the network causality usually has loop lines and the causality in the nodes may exist many-to-many situation, it is necessary to separate the causality lines by decomposition to form a series of chained causality. For example, nodes [4,5,2,1,4] is one loop line and node 4 is one many-to-many situation, as shown in the Fig. 4(a). The method of decomposition is to connect each node from the root node to the leaf node in the direction of the reverse arrow, which complete one decomposition line. Repeat this step until all causality are listed. Since there is ring in the network causality, it is stipulated that each node appears only once in a decomposition line. All decomposition lines in the example are shown in the Fig. 4(b).

(2) Virtual node

The same node will appear multiple times in different decomposition lines, such as node 4 in the Fig. 4(b). When these decomposition lines are reorganized into tree causality, one node is not enough to clearly show causality. Therefore, virtual node is proposed to distinguish the same nodes in the equivalent tree structure. The virtual node is treated as a new node, but all attributes of a virtual node are the same as the real node, except the number. The number of virtual node is in the form of A.B, where A is real node number and B is the order of the virtual nodes in all nodes with the same real node, such as nodes 4.1 and 4.2 in the Fig. 4 (c). In the diagram, the circle of the real node is a solid line while the circle of the virtual node is a dotted line.

(3) Pruning

The intricate network causality often creates a number of decomposition lines, which in turn makes the reorganized tree causality very wide in the lateral direction and deep in the longitudinal direction. The complex large tree causality structure will increase the complexity of the model and may also lead to over-fitting. Therefore, it is necessary to pruning complex large tree. The purpose of pruning is to simplify the equivalent tree causality without significantly reducing the prediction accuracy. Its principles have the following three points:

① Child nodes that are not particularly relevant to the parent node and whose causality is complex can be clipped entirely. In this study, the specific situation is that the node whose absolute value of PCC or MIC with parent node is between 0.5 and 0.7 and causality is complex can be clipped entirely.
② Child nodes that are particularly relevant to the parent node and whose causality is complex, its virtual nodes can be partially cut off, but at least one real node needs to be retained.
③ The result of pruning is not unique and has certain subjectivity, which is related to the computing resources and time consuming allowed by the case. If the computing resources are rich or allowed time is long, the tree can be slightly larger; otherwise the tree needs to be trimmed smaller. The equivalent tree causality after pruning of the example can be the one shown in the Fig. 4(d).

3.2. Long Short-Term Memory network based on Neighborhood Gates

After processing the causality, how to improve the LSTM so that it can preserve the structure of tree causality and increase the prediction accuracy is the top priority of this study. The idea of this study is that each node in the causality uses a standard LSTM [19] corresponding to it, and nodes are connected by the neighborhood gate, so that the

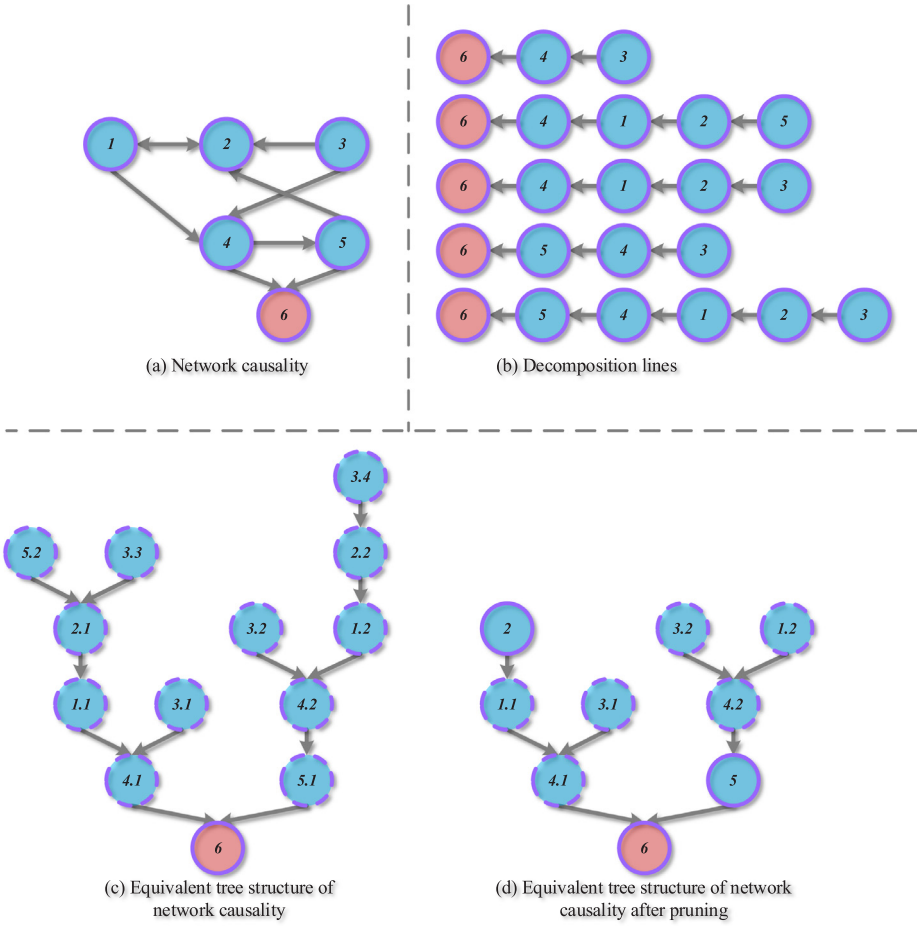

Fig. 4. The process of converting network causality to tree causality.

causality structure can be completely preserved.

In the Fig. 5, part IV in the lower right corner is the part of a general form of tree causality, whose corresponding NLSTM structure is shown as part IV on the left. Part III of the Fig. 2(d) is the last part of predicting VNP, which corresponds to the NLSTM structure as shown in the part III of the Fig. 5. x_{it} is the feature input of i-th node in the t-th period and $x_{it} = [F_{i1,t}, F_{i2,t}, \cdots, F_{in,t}]$. The standard LSTM can be found in the paper [19]. h_{it} is the output of standard LSTM. R_{it} is the central hub. n_{1mt} and n_{2it} represent Neighborhood Gates. N_{it} is the neighborhood. r_{i_t} is intermediate variable. y_t is the predicted value.

Forward propagation of NLSTM in the t-th period are as follows:

(1) All nodes complete the forward propagation of the standard LSTM. Take node i as an example:

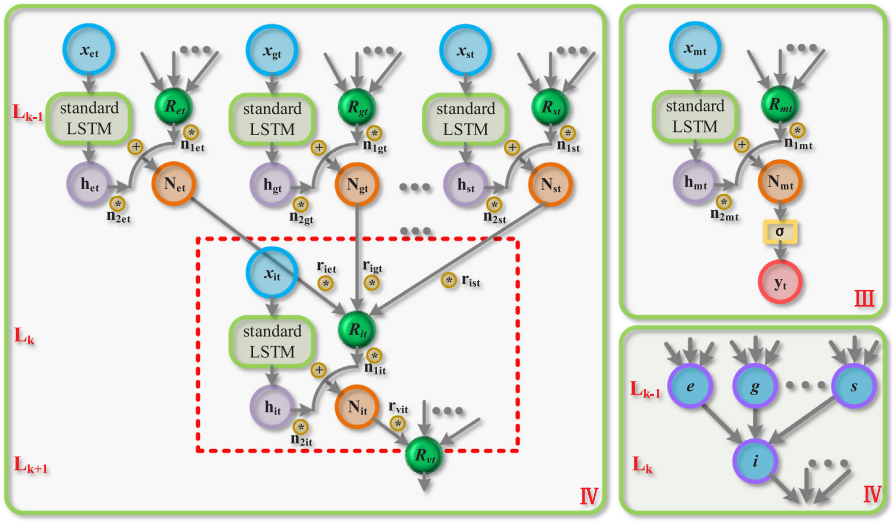

Fig. 5. Schematic diagram of NLSTM in the t-th period.

$$f_{it} = \sigma(net_{f,i,t}) = \sigma(w_{fh,i}\cdot h_{i,t-1} + w_{fx,i}\cdot x_{it} + b_{f,i}) \quad (1)$$

$$i_{it} = \sigma(net_{i,i,t}) = \sigma(w_{ih,i}\cdot h_{i,t-1} + w_{ix,i}\cdot x_{it} + b_{i,i}) \quad (2)$$

$$a_{it} = \tanh(net_{a,i,t}) = \tanh(w_{ah,i}\cdot h_{i,t-1} + w_{ax,i}\cdot x_{it} + b_{a,i}) \quad (3)$$

$$C_{it} = f_{it}*C_{i,t-1} + i_{it}*a_{it} \quad (4)$$

$$o_{it} = \sigma(net_{o,i,t}) = \sigma(w_{oh,i}\cdot h_{i,t-1} + w_{ox,i}\cdot x_{it} + b_{o,i}) \quad (5)$$

$$h_{it} = o_{it}*\tanh(C_{it}) \quad (6)$$

where $[w_{fh,i}, w_{fx,i}, b_{f,i}]$, $[w_{ih,i}, w_{ix,i}, b_{i,i}]$, $[w_{ah,i}, w_{ax,i}, b_{a,i}]$ and $[w_{oh,i}, w_{ox,i}, b_{o,i}]$ are weights and bias. f_{it}, i_{it}, o_{it} and a_{it} denote for the forget gates, the input gates, the output gates and the current information state. $C_{i,t-1}$ and $C_{i,t}$ represent for the cell state of the previous period and current period. $h_{i,t-1}$ and h_{it} stand for the standard LSTM outputs of the previous period and current period. $net\_{,i,t}$ is intermediate variable. The symbol $\cdot$ indicates matrix multiplication and the symbol $*$ indicates multiplication between matrix elements. $\sigma(x)$ and $\tanh(x)$ are activation function of *Sigmoid* and *Tanh* [19].

(2) Forward propagation along the tree. Take node i as an example:

$$r_{ijt} = \sigma(net_{r,i,j,t}) = \sigma(w_{rh,i,j}\cdot h_{i,t-1} + w_{rx,i,j}\cdot x_{it} + b_{r,i,j}),\ (P_i \neq \varnothing) \quad (7)$$

$$R_{it} = \begin{cases} \sum_{j=1}^{len(P_i)} r_{ijt}*N_{P_{ij},t} & (P_i \neq \varnothing) \\ 0 & else \end{cases} \quad (8)$$

$$n_{1it} = \sigma(net_{n1,i,t}) = \sigma(w_{n1h,i}\cdot h_{i,t-1} + w_{n1x,i}\cdot x_{it} + b_{n1,i}) \quad (9)$$

$$n_{2it} = \sigma(net_{n2,i,t}) = \sigma(w_{n2h,i}\cdot h_{i,t-1} + w_{n2x,i}\cdot x_{it} + b_{n2,i}) \quad (10)$$

$$N_{it} = n_{1it}*R_{it} + n_{2it}*h_{it} \quad (11)$$

where $[w_{n1h,i}, w_{n1x,i}, b_{n1,i}]$, $[w_{n2h,i}, w_{n2x,i}, b_{n2,i}]$ and $[w_{rh,i,j}, w_{rx,i,j}, b_{r,i}]$ are weights and bias. P_i is the set of child nodes of node i. P_{ij} is the j-th child node number of node i. Function *len()* is used to calculate the length of the set. Other variables have the same meaning as mentioned before.

(3) Prediction:

$$y_t = \sigma(z_t) = \sigma(w_y\cdot N_{mt} + b_y) \quad (12)$$

where $[w_y, b_y]$ are weights and bias. z_t is intermediate variable.

The framework of the whole method is presented in the Fig. 6. The framework consists of four major parts: Data collection, Data processing, Modeling and Evaluation. The innovations of this research are mainly concentrated in the second part and the third part, which are the causality processing and the NLSTM with the same structure as the equivalent tree causality.

3.3. Qualitative analysis of prediction accuracy

Qualitatively speaking, there are six reasons to explain why the method proposed in this paper is conducive to improve the accuracy of wind speed prediction.

(1) Correlation analysis and causality analysis in feature engineering are beneficial to improve wind speed prediction accuracy [32].
(2) Classification of causality facilitates the handling of all types of causality.
(3) Decomposition, virtual node and reorganized equivalent tree causality preserves the causality information of feature engineering, making complex causality structures available.
(4) Pruning avoids over-fitting.
(5) The structure of NLSTM is dynamically adjusted to be consistent according to the equivalent tree causality structure, so that the causality information in the feature engineering can be accurately utilized.
(6) Wind speed is time series data. LSTM in NLSTM considers time information and is very good at processing time series prediction problems [20].

In summary, NLSTM takes full advantage of the causality in feature engineering and has the ability to process time information, so it can improve wind speed prediction accuracy.

4. Method evaluation metric

4.1. Root mean square error

Root mean square error (RMSE) [42] is defined as the square root of the mean of squared error, whose formula is as follows. y_i and Y_i are prediction and observation, respectively. Tv is the size of validation set sample. The smaller the RMSE, the higher the prediction accuracy.

$$RMSE = \sqrt{\frac{1}{Tv}\sum_{i=1}^{Tv}(y_i - Y_i)^2} \quad (13)$$

4.2. Mean absolute percentage error

Mean absolute percentage error (MAPE) [43] is a measure of prediction accuracy. It usually expresses accuracy as a percentage and is defined as follows. The smaller the MAPE, the higher the prediction accuracy.

$$MAPE = \frac{1}{Tv}\sum_{i=1}^{Tv}\left|\frac{y_i - Y_i}{Y_i}\right| \quad (14)$$

5. Case study

In this section, the case description is first introduced. Then, the correlation and causality between wind speed and meteorological factors are analyzed. Next, the network causality of wind speed is converted into equivalent tree causality. After that, the experimental design and parameter settings are introduced. Finally, the performance of the proposed method is verified from three aspects: different features, different methods and different equivalent trees.

5.1. Case introduction

The data in this case is the meteorological data of Fuyun meteorological station (N 46.59°, E 89.31°) in Xinjiang province, China. The collected meteorological data has a total of 12 factors. The time of the meteorological data is from 0:00 on July 15, 2018 to 23:00 on August 14, 2018. The time step is one hour. The total length of the sample is 744. The basic information of meteorological factors are shown in the Table 1. The abbreviations num, abb, min, max, std, hpa represent number, abbreviation, minimum, maximum, standard deviation and hectopascal, respectively. The 12 factors in the Table 1 can be roughly divided into five categories: pressure (factor 1–2), temperature (factor 3–4), wind speed (factor 5–7), humidity (factor 8–9) and others (factor 10–12). In this case, there are five tasks that need to be completed.

①Explore the correlation and causality between these 12 meteorological factors;
②Select average wind speed (AWS) as VNP and convert its network causality to equivalent tree causality;
③Compare different features: the feature considered all relevant historical meteorological factors and the feature only considered historical wind speed;
④Compare different methods: NLSTM (consider causality) and the state-of-the-art wind speed prediction methods (not consider causality), such as LSTM [19], Gated Recurrent Unit (GRU, one of the most important variants of LSTM) [25], SVR [8] and AdaBoost-ELM

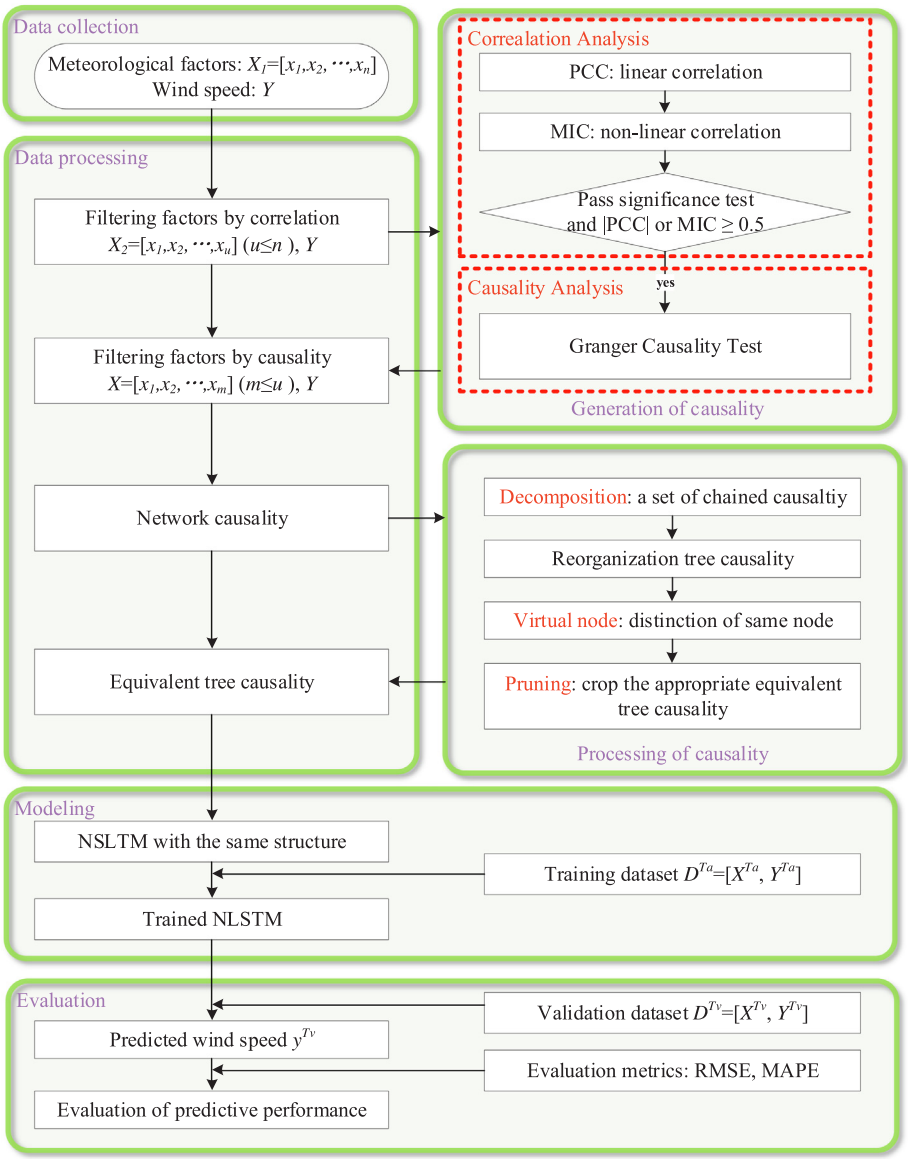

Fig. 6. The framework of the whole method.

Table 1
Basic information of meteorological factors.

Num	Factor	Abb	Unit	Min	Mean	Max	Std
1	air pressure	AP	hpa	902	909.99	919.6	3.46
2	sea level pressure	SLP	hpa	993.9	1004.54	1017.4	4.43
3	air temperature	AT	℃	9.8	22.56	35.3	5.17
4	apparent temperature	APT	℃	8.82	21.68	32.36	4.94
5	extreme wind speed	EWS	m/s	1.2	6.37	15.65	3.66
6	**average wind speed**	**AWS**	m/s	0.15	3.37	9.52	2.33
7	wind level	WL	\	0	2.40	6	1.31
8	relative humidity	RH	%	9	40.76	91	17.20
9	minimum relative humidity	MIRH	%	8	38.18	90	16.46
10	vapor pressure	VP	hpa	4.3	10.36	18.4	2.41
11	precipitation	P	mm	0	0.00	0.7	0.04
12	horizontal visibility	HV	m	7500	34464.84	35,000	

with Two-stage Decomposition Technique (AELM) [15].
⑤Compare different equivalent trees: equivalent tree causality after shallow pruning and depth pruning.

5.2. Task I: Explore the correlation and causality

5.2.1. Correlation analysis

The PCC and MIC between these meteorological factors are calculated and significance test is performed, as shown in the Table 2. Since both the PCC and the MIC matrix are symmetric matrices, the upper triangle represents the PCC matrix and the lower triangle represents the MIC matrix in the Table 2. The absolute values of PCC above 0.5 are highlighted in red bold fonts and MIC values above 0.5 are highlighted in green bold fonts. If two factors do not pass the significance test, they are filled in gray. All factors with red or green fonts and without gray fill may be selected as feature inputs. For example, the relevant factors of AWS (6) are EWS (5) and WL (7); the relevant factors of EWS (5) are AT (3), APT (4), AWS (6) and WL (7); the relevant factors of AT (3) are APT (4), RH (8), MIRH (9) and EWS (5). The correlation of these factors can be strung together like the example. Whether these factors can ultimately be selected as features also requires Granger causality test.

5.2.2. Causality analysis

The Granger causality test is used to explore the causality of these meteorological factors. In this study, the maximum lag periods of two

Table 2
Correlation between meteorological factors.

MIC \ PCC		1	2	3	4	5	**6**	7	8	9	10	11	12
		AP	SLP	AT	APT	EWS	**AWS**	WL	RH	MIRH	VP	P	HV
1	AP	1.00	**0.98**	-0.39	-0.41	-0.10	-0.07	-0.07	0.17	0.18	-0.18	-0.06	0.09
2	SLP	**0.85**	1.00	-0.47	-0.48	-0.07	-0.03	-0.03	0.25	0.26	-0.14	-0.04	0.06
3	AT	0.19	0.23	1.00	**0.99**	**0.57**	0.49	0.46	**-0.84**	**-0.82**	-0.30	-0.04	-0.01
4	APT	0.19	0.24	**0.99**	1.00	**0.57**	0.49	0.46	**-0.82**	**-0.80**	-0.24	-0.03	-0.02
5	EWS	0.19	0.17	0.38	0.40	1.00	**0.95**	**0.86**	-0.46	-0.46	-0.12	0.03	-0.05
6	**AWS**	0.17	0.15	0.32	0.34	**0.81**	1.00	**0.85**	-0.40	-0.40	-0.10	0.02	-0.08
7	WL	0.18	0.15	0.24	0.26	**0.66**	**0.61**	1.00	-0.39	-0.41	-0.14	0.02	-0.07
8	RH	0.14	0.21	**0.67**	**0.69**	0.31	0.25	0.21	1.00	**0.98**	**0.73**	0.11	-0.07
9	MIRH	0.15	0.20	**0.63**	**0.66**	0.31	0.25	0.20	**0.89**	1.00	**0.72**	0.07	-0.06
10	VP	0.19	0.18	0.16	0.21	0.15	0.14	0.14	0.34	0.35	1.00	0.13	-0.09
11	P	0.04	0.04	0.05	0.06	0.06	0.05	0.01	0.04	0.03	0.04	1.00	-0.12
12	HV	0.09	0.10	0.09	0.12	0.12	0.10	0.06	0.10	0.10	0.12	0.02	1.00

Table 3
F values of Granger causality test for meteorological factors.

F		1	2	3	4	5	**6**	7	8	9	10	11	12
		AP	SLP	AT	APT	EWS	**AWS**	WL	RH	MIRH	VP	P	HV
1	AP	\	**21.5**	0.7	0.9	10.4	12.4	12.2	1.5	8.8	12.7	2.9	1.8
2	SLP	**5.3**	\	13.5	12.2	9.4	12.4	6.2	0.7	12.9	13.3	4.3	0.9
3	AT	12.8	10.1	\	0.9	**29.5**	31.1	34.7	2.9	2.8	18.6	1.9	0.4
4	APT	11.0	9.0	0.4	\	**26.6**	28.7	35.0	1.1	2.5	18.0	2.1	0.1
5	EWS	8.1	5.4	**16.6**	**17.5**	\	**26.6**	**469.0**	13.3	22.4	7.3	1.0	4.6
6	**AWS**	6.6	3.9	9.5	10.4	**27.3**	\	**890.9**	5.7	12.0	4.8	0.7	4.5
7	WL	4.1	3.8	8.4	8.8	**10.2**	**19.0**	\	0.7	3.2	4.0	2.0	1.0
8	RH	7.1	8.0	**4.8**	**4.4**	16.7	17.2	15.0	\	**321.5**	**17.2**	2.2	0.6
9	MIRH	7.5	7.8	**4.2**	**5.0**	16.9	17.8	21.7	**6.3**	\	**8.1**	1.3	0.5
10	VP	1.1	2.5	2.6	1.7	4.9	4.5	2.9	2.4	2.8	\	5.0	2.5
11	P	1.8	3.8	9.1	8.7	1.6	1.0	2.6	12.4	2.9	3.0	\	2.8
12	HV	6.9	6.5	4.8	4.8	1.7	2.0	3.4	4.8	5.3	2.5	0.1	\

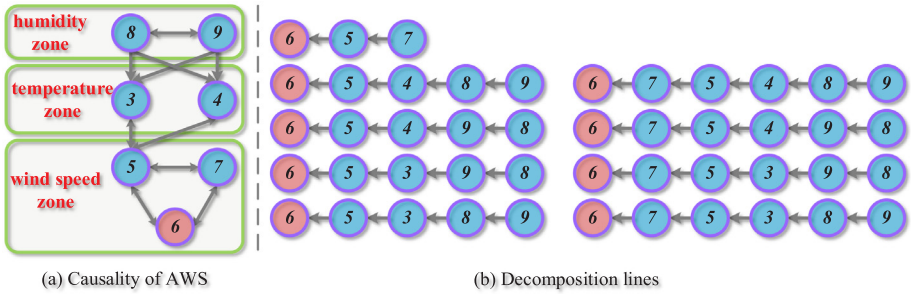

Fig. 7. Causality structure of AWS (6).

factors are all set as 2 and the number of samples is 744. So the critical value $F_{95\%}$ of F test under 95% confidence level is 3.0078 [39]. F values of Granger causality test for meteorological factors are calculated in the Table 3. Cells with F values greater than $F_{95\%}$ and highlighted in the Table 2 are highlighted with purple bold font in the Table 3. If the cell is highlighted with purple bold font, the factor of the horizontal axis is the Granger cause of the factor of the vertical axis. For example, Granger causes of AWS (6) are EWS (5) and WL (7); Granger causes of EWS (5) are AT (3), APT (4), AWS (6) and WL (7); Granger causes of AT (3) are EWS (5), RH (8), MIRH (9); RH (8) and MIRH (9) are each other's Granger causes. The causality of these factors can be strung together like this example. The graphical results of their causality are shown in

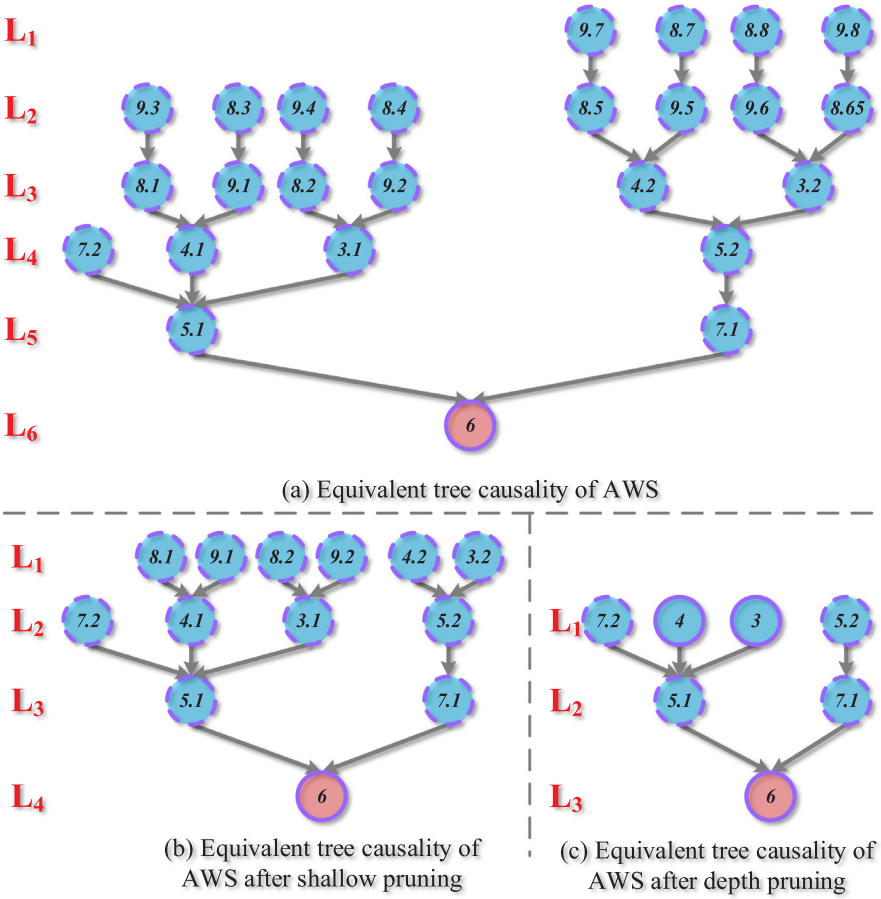

Fig. 8. Equivalent tree causality of AWS (6).

the Fig. 7(a).

5.3. Task II: Causality processing

According to the Table 3, AWS' network causality can be summarized as shown in the Fig. 7(a). According to the decomposition method proposed in this paper, the network causality can be decomposed into the nine chained causality shown in the Fig. 7(b). The AWS' equivalent tree causality can be reorganized according to these decomposition lines, where the same nodes are distinguished by virtual nodes, as shown in the Fig. 8(a). Nodes 5, 7, 3, 4, 8 and 9 all appear on multiple decomposition lines. These nodes are distinguished by virtual nodes in equivalent tree causality. Causality in the Fig. 8(a) is a complete equivalent tree causality, which has s total of six layers. From layer L1 to L2 and L2 to L3, there are eight arrows that describe the causality between node 8 and node 9, which seems a bit complicated. Therefore, it is necessary to crop the virtual nodes 8 and 9 of the L1–L3 layer in the Fig. 8(a). The MIC and PCC between node 8 and 9 are both greater than 0.7, indicating that they are significantly correlated. This situation is the second point of the pruning principle: Node 8 and 9 need to be partially cropped, leaving at least one real node. The equivalent tree causality after shallow pruning is shown in the Fig. 8(b). If the computational resources are limited, the equivalent tree in the Fig. 8(b) needs further pruning. The MICs between nodes 8, 9 and nodes 3, 4 are less than 0.7, which is the first case of the pruning principle, so nodes 8, 9 can be completely clipped. The situation between nodes 3, 4 is the same as nodes 8, 9 in the Fig. 8(b). The equivalent tree causality after depth pruning is shown in the Fig. 8(c).

5.4. Experimental design

In this study, three comparisons are designed: different features (Task III), different methods (Task IV) and different equivalent trees (Task V). The contrasting model details and parameter details are shown in the Tables 4 and 5, respectively. In order to ensure the fairness of the comparison, the same parameters are set to be the same in different models. Some special parameters are either optimized using Grid Search (GS) [44] algorithm or reference to the original text of the method. All neural network models are optimized by Adam optimization algorithm [45]. For each factor, its previous two periods of historical data are inputted as known features. 80% of the dataset are used as training set and the rest are used as verification set. Run 10 times per model, using the average as the final result.

5.5. Task III: Compare different features

In Task III, the feature considered all relevant historical meteorological factors and the feature only considered historical wind speed are compared. Two features are verified on LSTM and SVR. The predictive evaluation metrics of Task III are shown in the Table 6. For Model 1, the average values of RMSE and MAPE are 0.73 m/s and 14.01%, respectively. For Model 2, the average values of RMSE and MAPE are 0.85 m/s and 17.73%, respectively. Comparing Model 1 and Model 2, it can be found that RMSE and MAPE of Model 1 are lower than those of Model 2, indicating that Model 1 has higher prediction accuracy than Model 2, further indicating that the feature considered all relevant historical meteorological factors is superior to the feature only considered historical wind speed. For Model 3, the average values of RMSE and MAPE are 0.84 m/s and 16.19%, respectively. For Model

Table 4
Contrasting model details.

Task	Model	Method	Feature	Consider causality	Equivalent tree causality
III	1	LSTM	All relevant historical meteorological factors [6,5,7,3,4,8,9]	FALSE	Null
	2	LSTM	Only historical wind speed [6]	FALSE	Null
	3	SVR	All relevant historical meteorological factors [6,5,7,3,4,8,9]	FALSE	Null
	4	SVR	Only historical wind speed [6]	FALSE	Null
IV	5	NLSTM	All relevant historical meteorological factors [6,5,7,3,4,8,9]	TRUE	Shallow pruning, Fig. 7(b)
	1	LSTM	All relevant historical meteorological factors [6,5,7,3,4,8,9]	FALSE	Null
	6	GRU	All relevant historical meteorological factors [6,5,7,3,4,8,9]	FALSE	Null
	3	SVR	All relevant historical meteorological factors [6,5,7,3,4,8,9]	FALSE	Null
	7	AELM	All relevant historical meteorological factors [6,5,7,3,4,8,9]	FALSE	Null
V	5	NLSTM	All relevant historical meteorological factors [6,5,7,3,4,8,9]	TRUE	Shallow pruning, Fig. 7(b)
	8	NLSTM	All relevant historical meteorological factors [6,5,7,3,4,8,9]	TRUE	Depth pruning, Fig. 7(c)

4, the average values of RMSE and MAPE are 0.96 m/s and 18.83%, respectively. The comparison between Model 3 and Model 4 has the same conclusion.

In order to more intuitively compare the differences between the two features, the prediction results of Task III are plotted as shown in the Fig. 9. The histogram in the figure is the ordering of these models on the evaluation metrics RMSE and MAPE. The comparison of the prediction accuracy of Model 1, 2, 3 and 4 can be seen in the Fig. 12(a), (b), (c) and (d). The same conclusion can be obtained more intuitively from these figures. The reason why the feature considered all relevant historical meteorological factors is superior to the feature only considered historical wind speed is that the former utilizes more factors related to wind speed than the latter and extracts more information that is conducive to improving the accuracy of wind speed prediction.

5.6. Task IV: Compare different methods

In Task IV, the proposed method (NLSTM) is compared with the state-of-the-art wind speed prediction methods, such as LSTM, GRU, SVR and AELM. All methods use the feature considered all relevant historical meteorological factors. The biggest difference between these comparison methods is that NLSTM can accurately consider the causality of features, while other methods can only consider these features from the perspective of correlation. The predictive evaluation metrics of Task IV are shown in the Table 6. For Model 5, the average values of

Table 6
Evaluation metrics of Task III, IV, V.

Task	Model	Metric	RMSE (m/s)			MAPE (%)		
		Method	Min	Mean	Max	Min	Mean	Max
III	1	LSTM	0.65	0.73	0.83	12.18%	14.01%	15.08%
	2	LSTM	0.74	0.85	0.95	16.35%	17.73%	19.48%
	3	SVR	0.74	0.84	0.97	14.70%	16.19%	17.87%
	4	SVR	0.82	0.96	1.04	16.66%	18.83%	21.44%
IV	5	NLSTM	0.39	0.45	0.52	7.45%	8.16%	9.43%
	1	LSTM	0.65	0.73	0.83	12.18%	14.01%	15.08%
	6	GRU	0.60	0.66	0.74	11.61%	12.90%	13.60%
	3	SVR	0.74	0.84	0.97	14.70%	16.19%	17.87%
	7	AELM	0.58	0.64	0.74	10.87%	11.89%	13.38%
V	5	NLSTM	0.39	0.45	0.52	7.45%	8.16%	9.43%
	8	NLSTM	0.61	0.68	0.76	10.98%	12.56%	13.59%

Table 5
Contrasting model parameter details.

Model	Symbol	Meaning	Value	Reason
1(LSTM)	n_i	Number of input layer nodes	7 × 2	Equal to the number of feature inputs
6(GRU)	n_h	Number of hidden layer nodes	16	Common value [4,8,16,32,64,…]
8(NLSTM)	n_o	Number of output layer nodes	1	Equal to the number of prediction output
	η	Learning rate	0.01	Common value [0.001,0.005,0.01,0.05,0.1,…]
	T	Size of batch	32	Common value [8,16,32,50,100,…]
	Ep	Epochs of training	1000	converged
2 LSTM	n_i	Number of input layer nodes	1 × 2	Equal to the number of feature inputs
	n_h	Number of hidden layer nodes	16	Common value [4,8,16,32,64,…]
	n_o	Number of output layer nodes	1	Equal to the number of prediction output
	η	Learning rate	0.01	Common value [0.001,0.005,0.01,0.05,0.1,…]
	T	Size of batch	32	Common value [8,16,32,50,100,…]
	Ep	Epochs of training	1000	Converged
3,4 SVR	*kernel*	Kernel function	Gaussian	A competitive kernel functions
	σ	Kernel parameter	0.5	Obtained by GS with (0:0.5:2]
	C	Penalty parameter	1	Obtained by GS with (0:0.5:5]
5 NLSTM	n_i	Number of input layer nodes	13 × 2	Equal to the number of feature inputs
	n_h	Number of hidden layer nodes	16	Common value [4,8,16,32,64,…]
	n_o	Number of output layer nodes	1	Equal to the number of prediction output
	η	Learning rate	0.01	Common value [0.001,0.005,0.01,0.05,0.1,…]
	T	Size of batch	32	Common value [8,16,32,50,100,…]
	Ep	Epochs of training	1000	Converged
7 AELM	n_i	Number of input layer nodes	7 × 2	Equal to the number of feature inputs
	n_h	Number of hidden layer nodes	16	Common value [4,8,16,32,64,…]
	n_o	Number of output layer nodes	1	Equal to the number of prediction output
	ϕ	Initial threshold	0.2	Refer to the original paper [15]
	K	Maximum iteration number	20	Refer to the original paper [15]

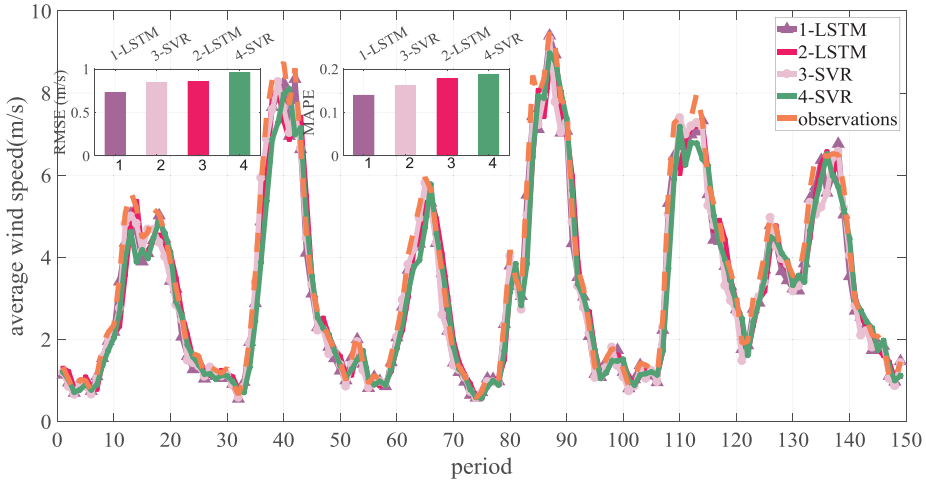

Fig. 9. Prediction results of Task III.

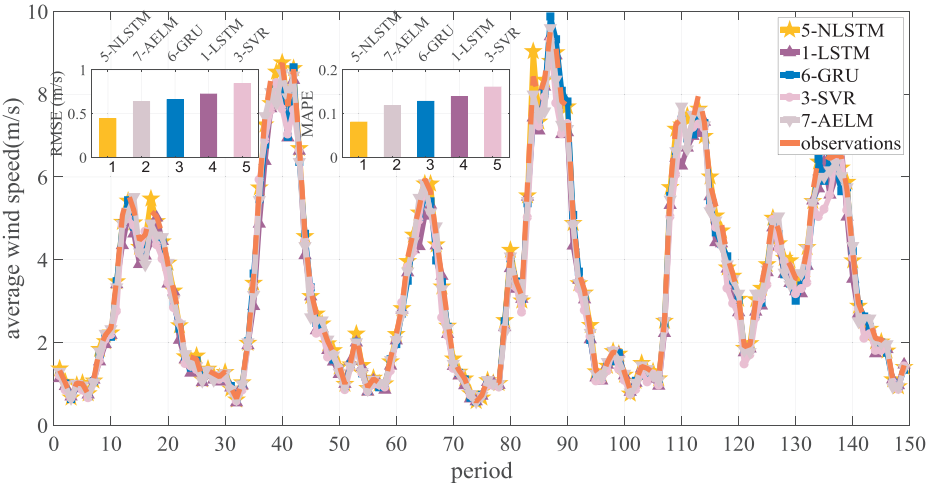

Fig. 10. Prediction results of Task IV.

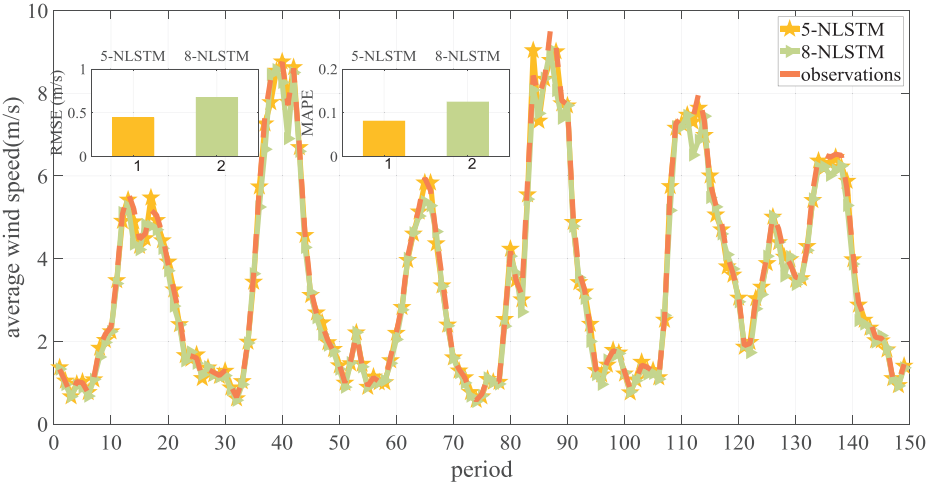

Fig. 11. Prediction results of Task V.

RMSE and MAPE are 0.45 m/s and 8.16%, respectively. For Model 1, the average values of RMSE and MAPE are 0.73 m/s and 14.01%, respectively. For Model 6, the average values of RMSE and MAPE are 0.66 m/s and 12.90%, respectively. For Model 3, the average values of RMSE and MAPE are 0.84 m/s and 16.19%, respectively. For Model 7, the average values of RMSE and MAPE are 0.64 m/s and 11.89%, respectively. These results show that RMSE and MAPE of NLSTM are the smallest of the five models, indicating that NLSTM has the highest prediction accuracy. This proves that NLSTM proposed in this paper is very competitive and very suitable for processing complex causality in wind speed prediction. The prediction results of Task IV are shown in the Fig. 10. The histogram in the figure is the ordering of these models

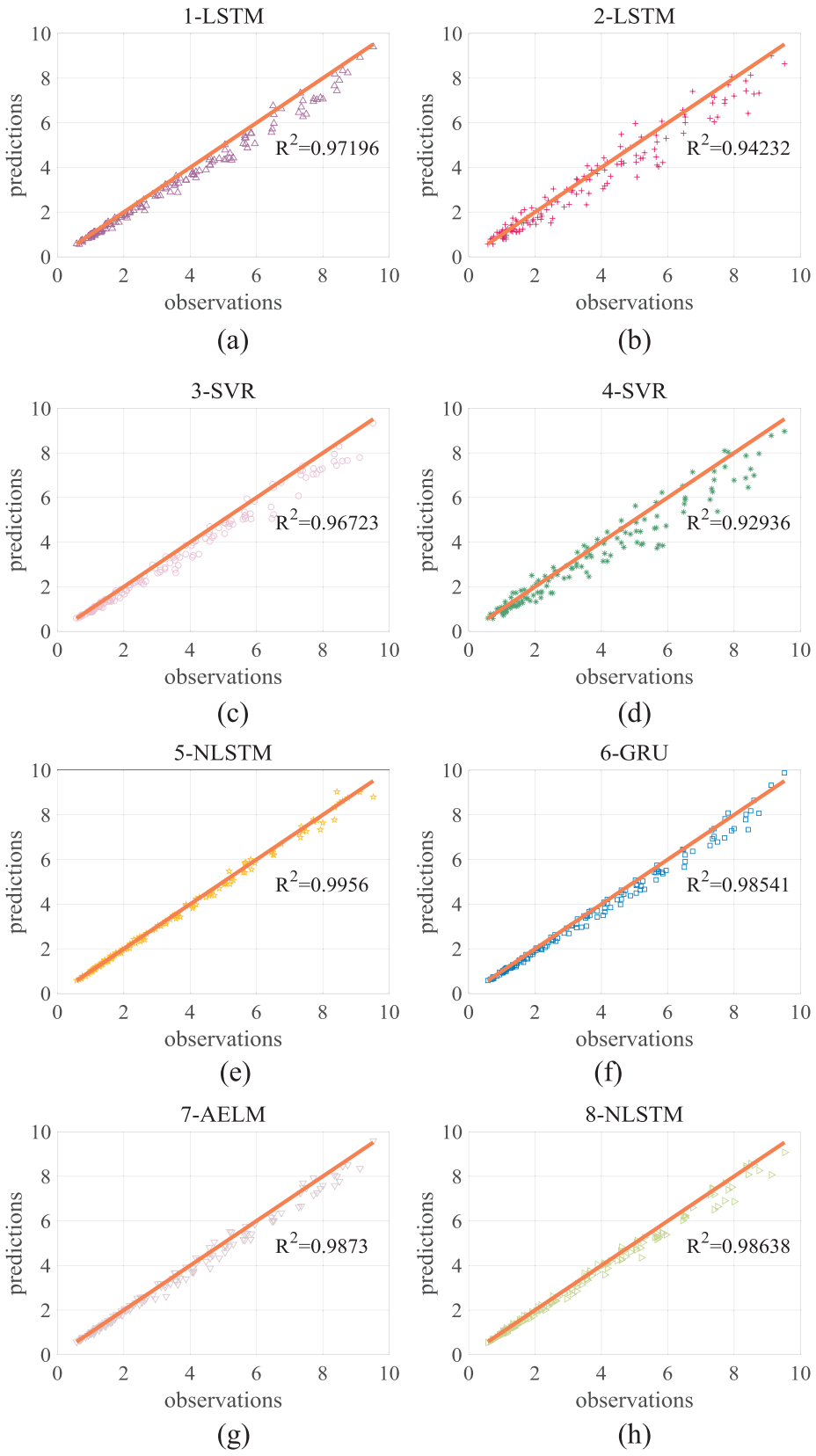

Fig. 12. Comparison of prediction accuracy of eight models.

on the evaluation metrics RMSE and MAPE. The comparison of the prediction accuracy of Model 5, 1, 6, 3 and 7 can be seen in the Fig. 12(e), (a), (f), (c) and (g), and the same conclusions can be seen more intuitively from these figures. The reason why NLSTM has the highest prediction accuracy among the five models is that it accurately considers the causality between wind speed and meteorological factors, which is the information that other models cannot obtain only through correlation.

5.7. Task V: Compare different equivalent trees

In Task V, different equivalent trees are compared. The features and method of Model 5 and Model 8 are the same. The only difference is the size of equivalent tree. The equivalent tree causality of Model 5 is shallow pruned and the Model 8 is deep pruned. The predictive evaluation metrics of Task V are shown in the Table 6. For Model 5, the average values of RMSE and MAPE are 0.45 m/s and 8.16%, respectively. For Model 5, the average values of RMSE and MAPE are 0.68 m/s and 12.56%, respectively. Obviously, the prediction accuracy of Model 5 is higher than that of Model 8, which indicates that the complex equivalent tree causality is better than simple equivalent tree causality. Of course, the training of Model 5 requires more computational resources and time than that of Model 8. The prediction results of Task V are shown in the Fig. 11. The histogram in the figure is the ordering of these models on the evaluation metrics RMSE and MAPE. The comparison of the prediction accuracy of Model 5 and 8 can be seen in the Fig. 12(e) and (h). In fact, the equivalent tree causality after pruning is an approximation of the original network causality. The more complex the equivalent tree causality is, the closer it is to the true causality, and the prediction accuracy will be higher.

6. Conclusions

Accurately predicting wind speed is a key task in utilization of wind energy and operation of the power system. In this study, the problems faced by the traditional wind speed prediction method in dealing with complex causality are first introduced. In response to this problem, the causality is divided into five categories: center, chained, ring, tree and network causality, according to the topological structure of causality. Then, all causalities are unified into an equivalent tree structure by decomposition, virtual nodes and pruning. NLSTM is proposed to accurately construct this equivalent tree causality. Finally, NLSTM is applied to predict wind speed and compared in these aspects: different feature, different methods and different equivalent tree causality. Experimental results show that: (1) The feature considered all relevant historical meteorological factors is superior to the feature only considered historical wind speed; (2) NLSTM is very competitive and very suitable for processing complex causality in wind speed prediction; (3) The closer the equivalent tree causality after pruning is to the original network causality, the higher the prediction accuracy.

The method proposed in this study is versatile and universal, which can be used to solve the prediction problems of time series data, such as the complementary system integrated of hydro, wind and solar power.

Declaration of interests

None.

Acknowledgments

This work is supported by the National Key R&D Program of China (2017YFC0405900), the National Natural Science Foundation of China (No. 91647114, 51779013, 51509009), the National Public Research Institutes for Basic R & D Operating Expenses Special Project (CKSF2017061/SZ) and special thanks are given to the anonymous reviewers and editors for their constructive comments.

References

[1] He Y, Li H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation. Energy Convers Manage 2018;164:374–84.
[2] Tasnim S, Rahman A, Oo AMT, Haque ME. Wind power prediction in new stations based on knowledge of existing Stations: A cluster based multi source domain adaptation approach. Knowl-Based Syst 2018;145:15–24.
[3] Yuan X, Yuan Y, Tan Q, Lei X, Wu X. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 2017;129:122–37.
[4] Allen DJ, Tomlin AS, Bale CSE, Skea A, Vosper S, Gallani ML. A boundary layer scaling technique for estimating near-surface wind energy using numerical weather prediction and wind map data. Appl Energy 2017;208:1246–57.
[5] Soman SS, Zareipour H, Malik O, Mandal P, "A review of wind power and wind speed forecasting methods with different time horizons,", (IEEE, 2010), pp. 1–8.
[6] Al-Yahyai S, Charabi Y, Gastli A. Review of the use of numerical weather prediction (NWP) Models for wind energy assessment. Renew Sustain Energy Rev 2010;14:3192–8.
[7] Zhang J, Draxl C, Hopson T, Monache LD, Vanvyve E, Hodge B. Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. Appl Energy 2015;156:528–41.
[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] Zhang C, Zhang K, Wei H, Zhao X, Liu T. A Gaussian process regression based hybrid approach for short-term wind speed prediction. Energy Convers Manage 2016;126:1084–92.
[10] Nielsen HA, Madsen H, Nielsen TS. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy 2006;9:95–108.
[11] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energ 2010;87:2313–20.
[12] Wang J, Hu J, Ma K, Zhang Y. A self-adaptive hybrid approach for wind speed forecasting. Renew Energy 2015;78:374–85.
[13] Santhosh M, Venkaiah C, Vinod Kumar DM. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. Energy Convers Manage 2018;168:482–93.
[14] Liu H, Tian H, Li Y. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24.
[15] Peng T, Zhou J, Zhang C, Zheng Y. Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. Energy Convers Manage 2017;153:589–602.
[16] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[17] Yamada T, Murata S, Arie H, Ogata T. Representation learning of logic words by an RNN: from word sequences to robot actions. Front Neurorobot. 2017;11:70.
[18] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.
[19] Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. Neural Comput 2000;12:2451–71.
[20] Qin Y, Li K, Liang Z, Lee B, Zhang F, Zhang L, et al. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. Appl Energy 2019;236:262–72.
[21] Li Y, Wu H, Liu H. Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction. Energy Convers Manage 2018;167:203–19.
[22] Han S, Qiao Y, Yan J, Liu Y, Li L, Wang Z. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl Energy 2019;239:181–91.
[23] Hu Y, Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm. Energy Convers Manage 2018;173:123–42.
[24] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manage 2018;159:54–64.
[25] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical. Mach. Transl. 2014.
[26] Yu C, Li Y, Bao Y, Tang H, Zhai G. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Convers Manage 2018;178:137–45.
[27] Wang J, Li Y. Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl Energy 2018;230:429–43.
[28] Carta JA, Cabrera P, Matías JM, Castellano F. Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study. Appl Energy 2015;158:490–507.
[29] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, Blanco-Aguilera A, García-Herrera R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization – Extreme learning machine approach. Energy Convers Manage 2014;87:10–8.
[30] Fan C, Sun Y, Zhao Y, Song M, Wang J. Deep learning-based feature engineering methods for improved building energy prediction. Appl Energy 2019;240:35–45.
[31] Feng C, Cui M, Hodge B, Zhang J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 2017;190:1245–57.
[32] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2014;62:592–7.
[33] Yesilbudak M, Sagiroglu S, Colak I. A novel implementation of kNN classifier based on multi-tupled meteorological input data for wind power prediction. Energy Convers Manage 2017;135:434–44.
[34] Ly A, Marsman M, Wagenmakers EJ. Analytic posteriors for Pearson's correlation coefficient. Stat Neerl 2018;72:4–13.
[35] Boik RJ. Second-order accurate inference on eigenvalues of covariance and

correlation matrices. J Multivariate Anal 2005;96:136–71.
[36] Reshef D, Reshef Y, Mitzenmacher M, Sabeti P. Equitability Analysis of the Maximal Information Coefficient, with Comparisons. 2013.
[37] Gao Y, Liang J, "Noncausal directional intra prediction: theoretical analysis and simulation,", (IEEE, 2012), pp. 2917–2920.
[38] Duan P, Yang F, Chen T, Shah SL. Direct causality detection via the transfer entropy approach. IEEE T Contr Syst T 2013;21:2052–66.
[39] McGraw MC, Barnes EA. Memory matters: a case for granger causality in climate variability studies. J Clim 2018;31:3289–300.
[40] Cudeck R, O'Dell LL. Applications of standard error estimates in unrestricted factor analysis: significance tests for factor loadings and correlations. Psychol Bull 1994;115:475–87.
[41] Velandy J, Mishra P. Prediction of winding faults between non-stationary signals: granger causality analysis for lightning impulse testing of transformers. Int T Electr Energy 2016;26:4–15.
[42] Liu Y, Ye J, Ye L, Qin H, Hong X, Yin X. Monthly streamflow forecasting based on hidden Markov model and Gaussian Mixture Regression. J Hydrol 2018;561:146–59.
[43] De Myttenaere A, Golden B, Le Grand B, Rossi F. Mean absolute percentage error for regression models. Neurocomputing 2016;192:38–48.
[44] Zhang Y, Chen S, Wan Y, "An Intelligent Algorithm Based on Grid Searching and Cross Validation and its Application in Population Analysis,", (IEEE, 2009), pp. 96–99.
[45] Kingma DP, Ba J. Adam: a Method for Stochastic Optimization. 2014.