A unified multi-step wind speed forecasting framework based on numerical weather prediction grids and wind farm monitoring data

Xingdou Liu [a], Li Zhang [a],[*], Jiangong Wang [b], Yue Zhou [c], Wei Gan [c]

[a] *School of Electrical Engineering, Shandong University, Jinan, 250061, China*
[b] *Integrated Electronic Systems Lab Co., Ltd., Jinan, 250101, China*
[c] *School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK*

ABSTRACT

Wind speed forecasting is the basis of wind farm operation, which provides a reference for the future operation status evaluation of wind farms. For the wind speed forecast of wind turbines in the whole wind farm, a strategy combining unified forecast and single site error correction is proposed in this paper. The unified forecast framework is composed of a unified forecast model and multiple single site error correction models, which combines the forecasted grids of numerical weather prediction (NWP) with the monitoring data of wind farms. The proposed unified forecast model is called spatiotemporal conversion deep predictive network (STC-DPN), which is composed of temporal convolution network (TCN) and 2D convolution long short-term memory network (ConvLSTM). Firstly, the NWP forecasted grids are interpolated to the fan location, and the sequence matrix is composed of the NWP data and the monitored data of each wind turbine according to the time series, which is entered into the TCN network for time sequence feature extraction. Then, the output of the TCN network is converted into a regular spatio-temporal data matrix, which is entered into the ConvLSTM network for joint learning of spatio-temporal features to obtain the wind speed sequence forecasted in the whole wind farm. Finally, an independent TCN-LSTM error correction model is added for each site. Variational modal decomposition (VMD) is used to process data series, and different processing methods are adopted in unified forecast and single site error correction. In the 96 steps forecast test of a wind farm from Jining City, China, the proposed method is superior to several baseline methods and has important practical application value.

1. Introduction

With the continuous growth of global energy demand and the aggravation of fossil energy pollution, the world pays great attention to the application of renewable energy [1]. As an important clean energy, wind energy has developed rapidly. However, the reliability of wind power generation is low, and the daily operation of wind farms is very dependent on accurate wind speed forecast (WPF) [2]. Accurate and efficient wind speed forecast can improve the reliability and safety of wind power grid connection [3], while reducing system operating costs [4]. Multi-step wind speed forecast can provide more information for wind farms [5], and more specifically show the development trend of wind speed in the future [6]. Therefore, this paper focuses on providing multi-step wind speed forecasting for the entire wind farm.

1.1. Current methods to forecast wind speed

After decades of development, wind speed forecast methods mainly form three categories: statistical methods, machine learning and physical methods [7].

Statistical methods generally make forecasts based on the historical wind speed [8], and are mostly used for short-term WPF within a few hours [9]. Traditional statistical models include auto-regressive moving average (ARMA) [10], auto-regressive integrated moving average (ARIMA) [11], persistence method (PM) [12], Kalman filter (KF) [13], principal component analysis (PCA) [14], etc. These statistical methods require low computing resources and have obvious advantages in short-term forecasting [15]. However, statistical models are mainly aimed at linear time series and are suitable for forecasting work with obvious stationary and linear features [16]. Current research points out that the development of wind speed is highly volatile, with obvious

nonlinear and non-stationary characteristics [17], so it is difficult for statistical models to meet the existing prediction accuracy requirements.

Machine learning (ML) method has developed rapidly in recent years, and many researchers have begun to use it for wind speed forecast [18]. ML can effectively fit the complex nonlinearity and uncertainty of wind speed time series through a large number of historical data training [19]. In wind speed forecast, the common ML models include artificial neural network (ANN) [20,21], support regression vector machine (SVR) [22], extreme learning machine (ELM) [23], etc. With the development of neural networks, many neural networks with special architectures have also been applied to wind speed prediction [24], such as Elman neural network (ENN) [23], adaptive wavelet neural network [25], recurrent neural network (RNN) for temporal features [26], long short-term memory network (LSTM) [27], convolutional neural network (CNN) for spatial feature extraction [28]. Liu et al. [29] used the two bidirectional short-term memory (BiLSTM) network as the basic prediction model to provide 10 step forecast results. Bai et al. [24] provided a double-layer staged training echo state network (D-ESN) as a forecast model, showing superior performance in six data sets. Although the performance of machine learning is generally better than that of statistical methods, under the condition of relying only on historical data, the performance declines significantly with the extension of the forecast horizon, which is generally limited to 1–4 h [30].

The physical method carries out modeling according to the geographical environment and atmospheric movement, takes into account the coupling effect of multiple influence factors, and can realize numerical weather prediction (NWP) of meteorological elements including wind speed, temperature and pressure at the same time [31]. Physical methods can achieve 24 h or longer forecast in advance [32]. The common physical numerical models include high-resolution limited area model (HIRLAM) [31,33], the fifth generation mesoscale model (MM5) [34], and weather research and forecasting (WRF) [32,35]. However, the initialization of the physical model has done a lot of approximate processing for the real atmosphere simulation [36], plus the uncertainty in the selection of the simulation scheme, and the deviation between the predicted wind speed and the actual wind speed is always large [37]. Although the data assimilation technology [38] introduces local observation data for model initialization, due to the influence of observation technology and geographical environment [39], the accuracy of physical model in ultra short term prediction is still unsatisfactory. In addition, meteorological simulation based on physical model requires high computing resources, which also makes it difficult for the forecast frequency to meet the needs of ultra short term prediction [40].

1.2. Efforts to improve forecasting accuracy

In order to make up for the defects of a single method and improve the accuracy of wind speed forecast, the current research has proposed many combined methods, including combining the forecast results of multiple models [41], modifying the NWP forecast wind speed [37,39], using the decomposition method to process the input data [42], and adding the error correction (EC) model [29]. Niu et al. [43] used the complementary ensemble empirical mode decomposition (CEEMD) to decompose the historical wind speed, reconstructed the wind speed sequence after removing the high-frequency components, and combined the forecast output of four neural network models and one linear model, which significantly reduced the forecast uncertainty. Zhang et al. [23] used variational modal decomposition (VMD) to decompose the historical wind speed into multiple components, and then integrated nine sub models for prediction. Yan et al. [37] used the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) to process the historical data and the prediction sequence given by the WRF model, and then used the CNN- BiLSTM hybrid model for correction, which significantly improved the wind speed prediction accuracy of the WRF model. Wang et al. [30] proposed a NWP wind speed

sequence transfer correction algorithm. In the input of the correction model, the monitoring wind speed at time t is also introduced as the input variable, and the sequence relationship with the wind speed at time $t+1$ of NWP is established, which improves the forecast accuracy of NWP ultra short term and short term time scales. Li et al. [44] used VMD to decompose historical wind speed, the low-frequency component is used for basic forecast, and the high-frequency component is used to train a special error correction model. The above research involves the combination of multiple models and the re-optimization of forecast results, which poses a challenge to the combination scheme and parameter optimization strategy [45].

The evolution trend of wind speed is determined by the regional atmospheric movement. Introducing more reference data into the model can also reduce the uncertainty of wind speed forecast [46]. Khodayar et al. [47] considered 145 wind stations located in the northern states of the United States and achieved 24h wind speed forecast for the whole region. Du [46] relies on the NWP in Texas and hundreds of weather stations to provide wind speed forecast 3h in advance for wind farms in the region, which can effectively warn of large-scale wind decline events. However, in practice, it is very difficult to obtain a large number of wind speed information and station information in different regions in real time. Spatial information and time delay information will lead to the complexity of the wind forecast process [43]. Liu et al. [48] used wind speeds of 26 × 12 and 20 × 20 wind turbines in a single wind farm, combined ConvGRU and 3D CNN to construct a spatiotemporal neural network (STNN), and completed 3-step forecast 3 h ahead. Zhu et al. [49] proposed a predictive spatiotemporal network (PSTN) based on 2D CNN and LSTM, which was respectively used in a wind farm in Wyoming and California for 10 × 10, and the results show that the forecast is better than NWP within 6h in advance. However, the methods proposed in Refs. [48,49] are only applicable to the regular wind turbine array matrix, and have not been extended to the wind speed forecast of the whole wind farm.

1.3. Summary of main contributions

Based on the analysis above, a multi-step wind speed forecast strategy combining unified forecast and single site error correction is proposed in this paper. The model used for unified forecast is called spatiotemporal conversion deep predictive network (STC-DPN), which is composed of temporal convolution network (TCN) and 2D ConvLSTM. TCN network is used to extract the temporal characteristics of wind farm monitoring data and NWP wind speed forecast data, and transform them into regular spatio-temporal data matrix. Then 2D ConvLSTM performs joint learning of spatiotemporal features to obtain the forecast wind speed sequence of all fan sites. The proposed unified forecast model is not affected by the arrangement of fans [48,49], and the data acquisition from the same supervisory control and data acquisition (SCADA) avoids the delay of information [46,47]. Considering the difference of wake effects at different fan sites, forecast sequence for each fan site is corrected by an independent TCN-LSTM model. The sequence error correction takes into account the change of wind speed over a longer period of time, reducing the randomness of single point wind speed [30, 37]. In addition to being used to reduce noise for monitor wind speed [43], VMD extracts the high-frequency components of NWP wind speed for single site error correction to alleviate the over smoothing of forecast sequence caused by conservative forecast [39]. The main contributions of this study are as follows:

(1) **For a single wind farm, a forecast strategy combining unified forecast and single site error correction is proposed.** The proposed strategy takes into account the integrity of wind speed changes in the wind farm area and the differences of each fan site, which can effectively improve the performance of wind speed forecast.

(2) **The proposed STC-DPN unified forecast model has good versatility.** STC-DPN overcomes the influence of complex terrain of wind farm on the arrangement of wind turbines, and realizes the joint learning of space-time characteristics in the form of virtual wind farm.

(3) **An independent TCN-LSTM hybrid model is used for single site error correction.** In the error correction test, TCN-LSTM considers the continuity of error sequence, which effectively reduces the probability of large error.

(4) **VMD is used for the noise reduction of monitor wind speed sequence, and extract the high-frequency components of NWP wind speed for error correction.** The experimental results show that the two data processing strategies based on VMD reduce the uncertainty of forecast in unified forecast and error correction.

(5) **The effectiveness of the proposed method is verified in a wind farm located in a hilly area of Jining City, China.** In the 96 steps wind speed forecast test for future 24 h, the proposed method is superior to all comparison models.

The rest of this paper is structured as follows: section 2 introduces the relevant methods used. section 3 gives the prediction process of the proposed method. section 4 is a case study to verify the effectiveness of the proposed method. In section 5, the simulation results are further discussed. Finally, section 6 gives the conclusion.

2. Theoretical basis of methods

The proposed forecast methods include the WRF model, various neural network architectures, and VMD decomposition algorithm. This section gives the parameter settings of the WRF model, the basic theories of TCN and 2D ConvLSTM, and the decomposition principle of VMD.

2.1. WRF model parameter setting

WRF model is a new generation of mesoscale numerical prediction model developed by National Center for Atmospheric Research, National Centers for Environmental Prediction and other research institutions [37]. This model adopts highly modular, parallel and hierarchical design technology, and integrates the research results in mesoscale so far. The model provides a variety of physical process schemes for meteorological simulation of real weather, which can meet the multi-scale wind speed prediction needs of wind farms.

In this paper, the WRF model is used to provide NWP for a wind farm in Jining City, Shandong Province, China. Three nested grids are set to $100 \times 100(18 \text{ km}), 100 \times 100(6 \text{ km}), 121 \times 121 (2 \text{ km})$, including 35 vertical layers. The location and simulation domain of the selected wind farm are given in section 3.1. Initialize every 6 h according to the global forecast system (GFS) forecast data, and obtain the NWP grid within the range of the target wind farm from the output, including wind speed, wind direction and other simulation variables. Based on the above nested grid division, the sensitivity test of physical parameters of the WRF model is carried out, and the simulation scheme is selected according to the wind speed prediction effect. Detailed physical parameter settings are given in appendix A Table A1.

2.2. Temporal convolutional networks

TCN is stacked by multi-layer 1-D CNN to extract features from time series [50]. TCN has two remarkable characteristics: 1) when performing convolution operation, the future information acquisition only depends on the past information; 2) the input can be a sequence of any length and can be mapped to an output sequence of the same length.

Fig. 1 shows a TCN architecture case. The convolution operation of upper layer features and lower layer features has time causality. Causal convolution can ensure that the output at time t is only related to the
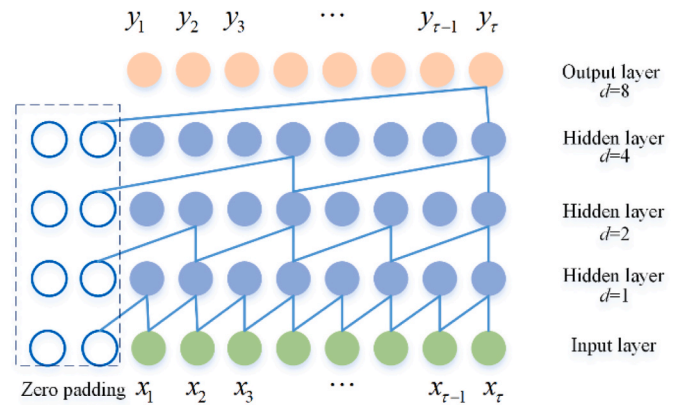

Fig. 1. Four layer 1-D convolution TCN architecture case with $k = 2$, $d = 1,2,4,8$.

input information at time t and before, and strictly follow the forward flow of feature information along the time series. In order to ensure that the length of the input and output sequences is consistent, a 1-D full CNN structure is used. Before the input sequence, a zero value with a length of $k-1$ is filled, and k is the size of the convolution kernel. A single-layer 1-D causal convolution layer is defined as follows:

$$F(s) = (x * f)(s) = \sum_{j=0}^{k-1} f(j)x(s-j) \tag{1}$$

$x \in R^X$ is the input. f is a filter. $\{0, \cdots, k-1\} \rightarrow R$, $F(\cdot)$ represents a convolution operation.

For the problem of increasing the training cost of the depth layer, TCN adds dilated convolution and introduces an dilation factor d. The receptive field of each layer is $(k-1)d$, and the convolution method is as shown in formula (2). In order to ensure that the filter in the deep network obtains a very large effective history without input omission, the selection of d in the r-th hidden layer is generally 2^{r-1}.

$$F(s) = (x * f)(s) = \sum_{j=0}^{k-1} f(j)x(s-dj) \tag{2}$$

The residual network [51] is used for layer hopping to connect the depth network to solve the gradient optimization problem [52], and the specific residual block design process references [50].

2.3. ConvLSTM

LSTM is specially designed for time series [53] and performs well in time series related prediction tasks. However, the input of LSTM only considers the time dimension and feature dimension, so ConvLSTM [54] was proposed to conduct joint learning of time and space features at the same time. The input array of a 2D ConvLSTM cell (Fig. 2) is length × width × channel. The number of cells in 2D convlstm layer is equal to the time sequence length. The specific calculation process is as follows:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f)$$

$$o_t = \sigma(W_{oi} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_{t-1} + b_o)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c)$$

$$H_t = o_t \circ tanh(C_t) \tag{3}$$

where * represents convolution operation and ∘ represents dot product operation; X_t is the input sequence at time t; H_t is the output of the hidden layer; C_t is the preliminary information transmitted to the unit
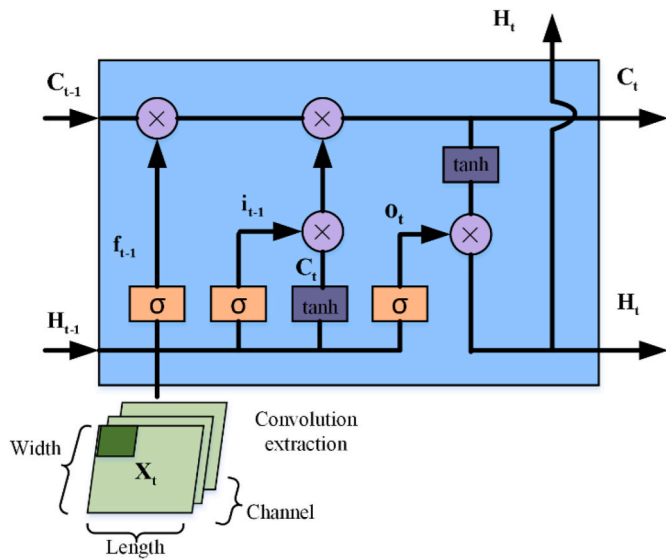
Fig. 2. 2DConvLSTM unit of one time step.

layer; i_t is the input gate, which controls the amount of information input to the current unit state; f_t is the forgetting gate, which is responsible for the selective forgetting of the last unit status information; o_t stands for output gate, which can select how much information of current time unit status is adopted as output; σ and *tanh* are activation functions; W and b are the weights and deviations between neurons in each cell, respectively.

2.4. Variational modal decomposition

VMD [55] is an adaptive and completely non recursive modal variation and signal processing method, which decomposes the original signal into a set of discrete intrinsic mode functions (IMF). The detailed decomposition steps are as follows:

$$\frac{min}{\{u_k\}\{w_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-jw_kt}\right\|_2^2\right\}$$

$$s.t.\sum_k u_k(t)=f(t) \tag{4}$$

where $u_k(t)$ is the modal function of the input signal; $\{u_k\}$ represents modal set; w_k is the center frequency corresponding to the k-th mode of the input signal; (w_k) represents a group of center frequencies corresponding to the decomposed modes; $f(t)$ is the input signal; $\delta(t)$ is the unit pulse function.

By introducing Lagrange multiplier λ and quadratic penalty factor α, equation (4) can be rewritten as:

$$L(\{u_k\},\{w_k\},\lambda)=\alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-jw_kt}\right\|_2^2$$

$$\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle \tag{5}$$

Using the alternating direction method of multiplication algorithm to solve (5), a group of modal components and their respective center frequencies are obtained. Each mode can be estimated from the solution in the frequency domain, expressed as:

$$\widehat{u}_k^{n+1}(w)=\frac{\widehat{f}(w)-\sum_{i=1}^{k-1}\widehat{u}_i^{n+1}(w)-\sum_{i=k+1}^{k}\widehat{u}_i^n(w)+\frac{\widehat{\lambda}^n(w)}{2}}{1+2\alpha(w-w_k^n)^2} \tag{6}$$

where n is the number of iterations; $\widehat{f}(w)$, $\widehat{u}_i^{n+1}(w)$, $\widehat{u}_i^n(w)$ and $\widehat{\lambda}^n(w)$ are

Fourier transformed forms.

In equation (6), the mode in Fourier domain is directly updated. In addition, these modes can be obtained in the time domain by extracting the real part of the inverse Fourier transform of the filtered analysis signal.

$$w_k^{n+1}=\frac{\int_0^\infty w\left|\widehat{u}_i^{n+1}(w)\right|^2 dw}{\int_0^\infty\left|\widehat{u}_i^{n+1}(w)\right|^2 dw} \tag{7}$$

Using equation (7), the center frequency w_k^{n+1} of the modes can be obtained, which indicates that the new center frequency is placed at the center of gravity of the power spectrum of their respective modes.

3. Proposed forecasting strategy

As mentioned in the introduction, this study combines ML method and physical method, and proposes a wind speed forecast strategy of unified forecast and single site error correction to enhance the forecast accuracy of wind speed in the whole wind farm. The forecast process can be divided into three parts: the model input data organization framework, the unified forecast based on STC-DPN and the single site error correction of TCN-LSTM. Next, the forecast process is introduced in detail, and the forecast evaluation indicators are given.

3.1. Data organization framework

Fig. 3 shows the data acquisition, processing and final data shape. The data includes the forecast grid provided by WRF model, and the SCADA system provides the recorded data of each monitoring station.

First, the NWP of wind field is realized by using the forecast grid of the innermost layer of WRF. The case wind field is from Jining City, Shandong Province, which is located in the hilly area. The size of the wind farm is about 8×10 km, 33 fans and 1 wind measuring tower are arranged irregularly due to the influence of hilly terrain. The inverse distance weighting method (IDW) is used to interpolate the four forecast grid points around the station to the height of the fan hub (80m), and the formula is as follows:

$$z=\frac{\sum_i^q\frac{z_i}{(D_i)^p}}{\sum_i^q\frac{1}{(D_i)^p}} \tag{8}$$

z is the final NWP forecast value of the site, z_i is the forecast value of WRF grid points, q is the number of grid points participating in interpolation, D_i is the distance between the difference point and the i-th site, and p is the power of the distance, which can be adjusted according to the interpolation effect. After the interpolation is completed, the NWP forecast sequence of 34 sites in the next 96 steps can be obtained.

During the real-time operation of the wind farm, the SCADA system can collect the recorded data of the station according to a certain frequency, including wind speed, wind direction and other meteorological elements. However, the wind speed has strong randomness, which makes the recorded data series show obvious volatility. The obvious high-frequency components are regarded as noise, VMD is used to decompose the recorded wind speed sequence into k IMF components, and then reconstruct after removing the high-frequency components. The trend of reconstructed wind speed series is obviously flat, which can reduce the negative impact of random volatility on model training and prediction. It should be noted that we only process the recorded wind speed sequence as the model input with VMD, and do not process the real wind speed used to evaluate the prediction effect.

After obtaining NWP forecast data of all sites and monitoring data of SCADA system, and then data splicing needs to be completed. In Fig. 3, x represents the input value and y represents the output value; s represents the number of variable sequences, which is determined by the number of sites and the type of selected variables; t represents the current time, and
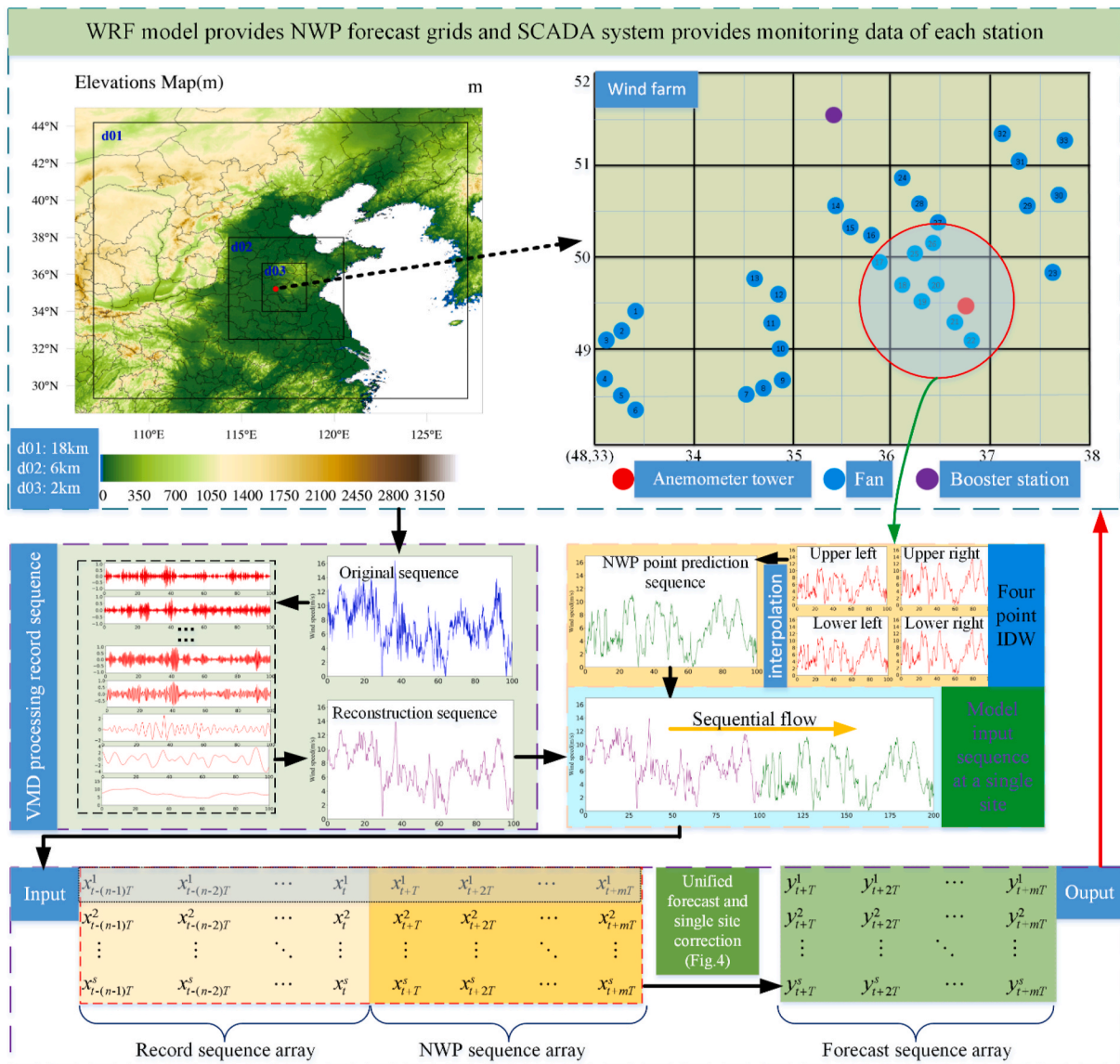
Fig. 3. Data organization process based on WRF mode and SCADA system.

T is the time interval; n represents the sequence length of intercepting SCADA system monitoring variables, and m represents the sequence length of NWP forecast and final wind speed forecast. After data organization, 96 step forecast sequence of all sites can be obtained through unified forecast and single site error correction.

3.2. Unified forecast and single site correction

The data organization of the forecast model is completed in section 3.1. Next, based on the case wind farm Fig. 4 shows the detailed process of unified forecast and single site correction. Wind speed and direction are regarded as input variables. T in the data set is 15 min. The sequence length of SCADA system monitoring variables is $n = 16$, and the sequence length of NWP forecast and final wind speed forecast is $m = 96$.

First, the TCN network is used to extract the time series features of the wind speed and wind direction sequence matrices respectively. The input dimensions are the sequence length and the sequence number (112×34). The TCN network does not change the sequence length of the input. By setting the number of filters, the number of features is increased after the feature extraction is completed. By adjusting the size of the convolution kernel, it is suitable for feature learning of short

sequences. In order to convert the output of TCN into a regular spatio-temporal matrix, the number of output features of TCN is set to 256. After the feature extraction is completed, two 112×256 feature matrices can be obtained and converted into a $112 \times 16 \times 16 \times 1$ four-dimensional matrix to meet the input requirements of the 2D ConvLSTM network layer. Finally, the two matrices are merged according to the channel dimension, and the input type into the 2D ConvLSTM network layer is $112 \times 16 \times 16 \times 2$.

After completing the spatiotemporal transformation of the input data, a three-layer encoding-forecasting architecture is completed based on the 2D ConvLSTM to generate the forecast of the virtual wind speed map. A 2×2 max-pooling operation is used after each ConvLSTM layer, and the output size of the ConvLSTM unit before pooling is recorded as an index for de-encoding in the forecast stage. In the encoding stage, the array dimension is finally $112 \times 4 \times 4 \times 128$, and it is restored to the original array shape in the forecast stage, and finally the output is converted into a 112×34 forecasted wind speed matrix. In the data flow of STC-DPN, end-to-end [56] training of the same loss function for the entire architecture can be achieved without changing the time length during the process.

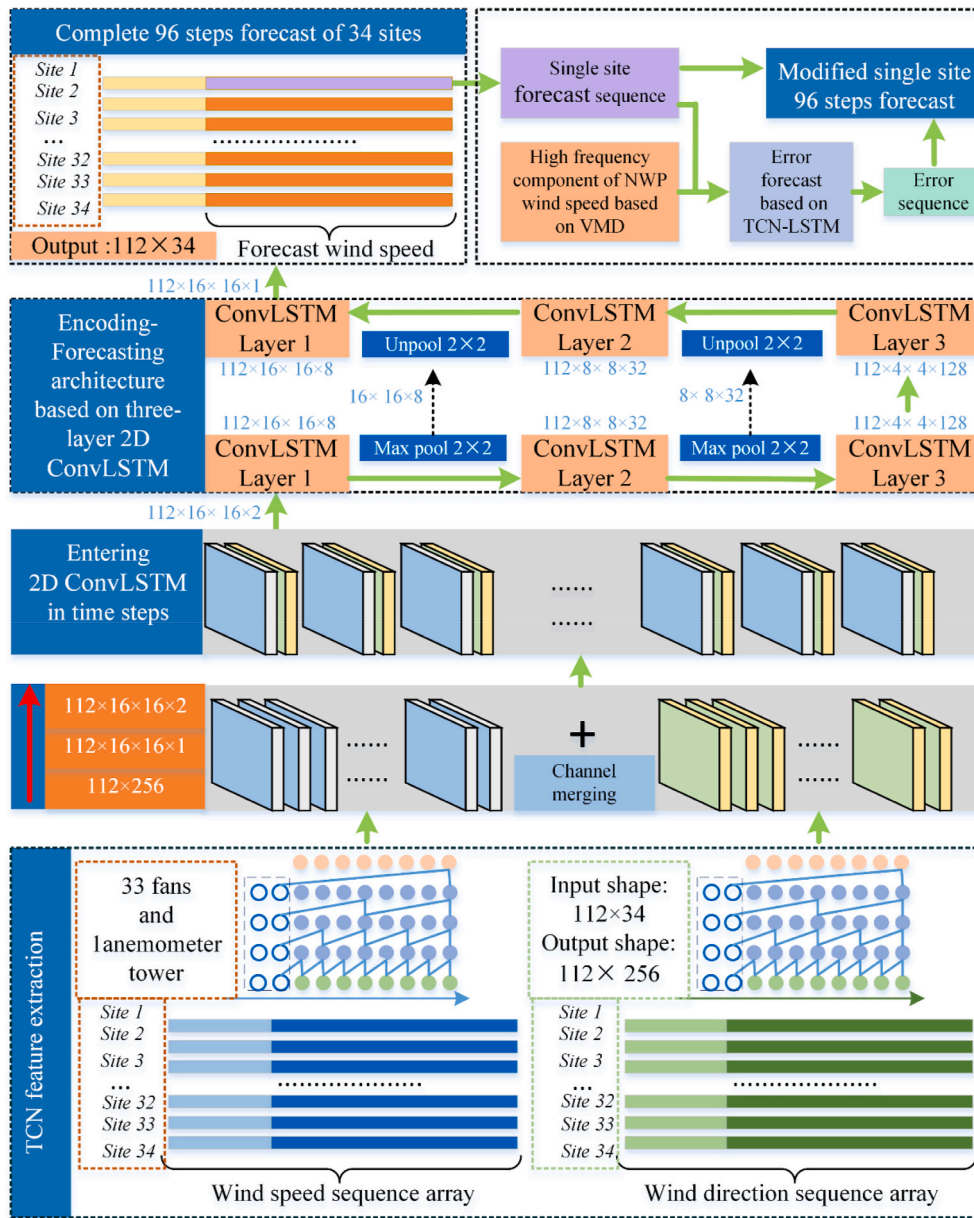After the unified prediction is completed, 96 step prediction

Fig. 4. Unified forecasting and single site error correction framework.

sequence can be obtained for each position, and then correct its error. After the model training, the error sequence can be calculated based on the predicted sequence and the real sequence. VMD is used to decompose the NWP wind speed sequence of the station. After retaining the high-frequency components, we use it and the single point prediction sequence obtained from the unified prediction as the input of the TCN-LSTM correction model, and the prediction error sequence as the output. After the error prediction of this position is realized by the modified model, combined with the unified wind speed sequence, the 96 step prediction of this point can be obtained.

In the forecast framework, the selection of model super parameters is combined with references and trial and error method, and the settings are shown in appendix A Table A2.

3.3. Evaluation indices

MAE and RMSE are widely used in the error evaluation of point forecast, which can reflect the long-term forecast reliability of the model. The formula is as follows:

$$MAE = \frac{1}{K} \sum_{i=1}^{K} \left| \widehat{y}(i)_{t+\tau} - y(i)_{t+\tau} \right| \tag{9}$$

$$RMSE = \sqrt{\left(\frac{1}{K} \sum_{i=1}^{K} \left(\widehat{y}(i)_{t+\tau} - y(i)_{t+\tau} \right)^2 \right)} \tag{10}$$

$\widehat{y}(i)_{t+\tau}$ and $y(i)_{t+\tau}$ are the forecasted wind speed and real wind speed, t is the current time, τ is the forecast baseline, and K is the number of samples in the test set.

$$COR = \frac{COV(\widehat{Y}_{t+\tau}, Y_{t+\tau})}{\sigma_{\widehat{Y}_{t+\tau}} \sigma_{Y_{t+\tau}}} \tag{11}$$

$COV(A, B)$ represents the covariance of A and B, σ_A and σ_B represent the standard deviation of A and B, $\widehat{Y}_{t+\tau}$ represents the predicted wind speed sequence, and $Y_{t+\tau}$ represents the real wind speed sequence of the test set. The larger the COR, the closer the change trend of the predicted wind speed series to the real wind speed series.

In order to comprehensively evaluate the forecast effect of different methods, a comprehensive improvement index (CII) is proposed based on MAE, RMSE and COR, which is defined as:

$$CII = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\widehat{I}_i - I_i}{I_i} \right| \tag{12}$$

N represents the number of indicators, $\widehat{I}_i$ is the ith indicator value of a method, and $\widehat{I}_i$ represents the i-th indicator value of the baseline algorithm.

The forecast of high wind speed is more difficult than that of low wind speed [47], especially in the high wind speed range with large fluctuations. With reference to the forecast interval coverage rate (PICP) [44] in the probability forecast, the hit rate (HR) is used as the evaluation index of different wind speed intervals for auxiliary error analysis, which is defined as follows:

$$HR = \frac{1}{Q} \sum_{i=1}^{Q} H_i, H_i = \begin{cases} 1, \left| \widehat{y}(i)_{t+\tau} - y(i)_{t+\tau} \right| \leq W \\ 0, \left| \widehat{y}(i)_{t+\tau} - y(i)_{t+\tau} \right| > W \end{cases} \tag{13}$$

W represents the threshold value set by HR, and H_i indicates whether the predicted absolute value error meets the condition of less than W. A relatively loose w will be set for the high wind speed range, and a strict w will be required for the low wind speed range. Q represents the number of samples in the wind speed range.

4. Case study

This section shows in detail the forecast performance of the proposed method in the case wind farm, provides the comparison of multiple baseline methods, and carries out error analysis.

4.1. Data set

The time range of collected data is from January 1, 2019 to December 31, 2020, with a total of 70176 time points. The data set includes wind speed and direction recorded for 15min at 80m height of 33 wind turbines and a wind measuring tower in the whole wind farm, as well as NWP forecast data with the same time resolution. Fig. 5 shows the annual statistical distribution of wind speed and direction, which can reflect the actual wind energy characteristics of the wind farm. The recorded wind speed and NWP wind speed are basically between 0 and 15 m/s, but the trend of NWP frequency fitting curve is smoother, and the wind direction shows the characteristics of higher south wind frequency.

4.2. Baseline algorithms

In order to reflect the advantages of the method proposed in this paper, considering the shape of input data, SVR [22], ANN [20], RNN [26] and LSTM [27] are used as single site forecast models, which are compared with unified forecast. TCN-3DCNN [48] is used as a unified forecast model to reflect the advantages of STC-DPN space-time joint learning. SVR, TCN, LSTM and TCN-LSTM are used as single site error correction models to compare the error correction effects.

As an application of SVM to regression problem, SVR can realize forecast task. ANN is composed of basic full connection layer, which can extract explicit information and nonlinear relationship between input and output data. The design of RNN and LSTM has the ability to capture time series features, and the learning performance of the evolution trend of temporal data is better. The combination of TCN and 3DCNN can deal with the four-dimensional data matrix after TCN output conversion. All models are implemented in Python.
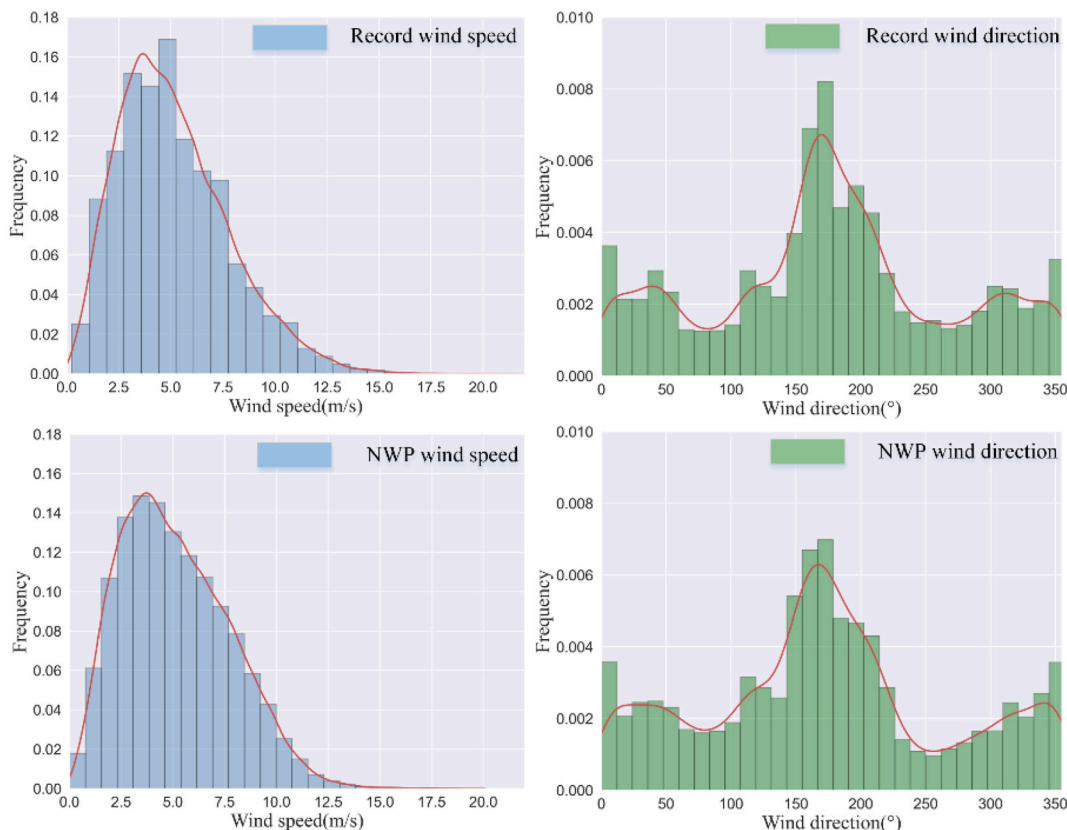

Fig. 5. Annual statistical distribution frequency of wind speed and direction.

(1) SVR is tested in four parameter kernel functions: linear, poly, sigmoid and rbf. The penalty factor C is adjusted in steps of 0.1 based on the default 1.0. Implemented by Python's sklearn package.

(2) For ANN, RNN and LSTM neural network models, the optimization setting standard for the number of layers and neurons is to gradually increase the number of network layers. Before the output layer, the number of neurons in each layer increases exponentially by2. Python's keras package implementation is used.

(3) TCN performs grid optimization for expansion factor d, convolution kernel size and number of neurons, which is also implemented by Python's keras package.

(4) 3DCNN and 2DConvLSTM in STC-DPN are the same three-layer coding forecast architecture, which is implemented using keras package.

In this paper, all models are trained and tested on the same computer. The hardware includes i7-12700h CPU, NVIDIA geforce RTX 3060 GPU and 16gb memory. The python package includes tensorflow 1.13.1, keras 2.2.4, and sklearn 0.23.2. The training and test data are divided by 4:1. Table 1 gives the case number of models(34 sites), training information, training time, model loading time and the time for doing1000 forecasts. Obviously, SVR has the highest training efficiency in forecast and error correction, and the time used is much lower than other models. In the forecast model, the training time of ANN, RNN and LSTM models for single site forecast is also higher than that of the two unified forecast models. Moreover, the training time of STC-DPN plus modified model is also lower than that of RNN and LSTM. In the time statistics of model loading and 1000 forecasts, the completion time of all models is less than 1 min, and the model loading time is greater than the forecast time. This shows that in the actual forecast, the main time for completing a rolling forecast is to load the forecast model, and a higher forecast frequency can be achieved by increasing the frequency of data acquisition.

4.3. Performance comparison

Using the 10 methods presented in section 4.2, 96 steps point forecast for comparison is completed. Tables 2–4 give the statistics of MAE, RMSE and COR which including 6 time point forecasts (indicators statistics at a certain time point) and 3 segmented range average forecasts (average indicators statistics within 6 h).

Among the four models for single site forecast, the forecast results of SVR and ANN are poor, which is predictable. The reason is that the input historical monitoring data and NWP forecast data have strong time correlation. LSTM and RNN have the ability to learn time series characteristics, which are more suitable for the wind speed forecast task in this paper. LSTM demonstrates better forecast performance than RNN, which has been verified many times in other literatures [27,49]. This is because the unique working mechanism of LSTM has better sequence feature learning ability, and also solves the optimization problem of RNN gradient training. Unified forecast is significantly better than single site forecast. Considering that the inputs of TCN-3DCNN and STC-DPN models include the information of the whole wind farm, this reduces the impact of the randomness of data collection at a single site to a certain extent. Obviously, using more location data can more effectively forecast the future wind speed changes of the wind farm. STC-DPN achieves better forecast results than TCN-3DCNN. The difference between the two models comes from the processing of the time dimension of the data matrix. 2DConvLSTM has the time sequence learning ability of LSTM, and can convolute the two-dimensional matrix at each time. The data dimensions of 3DCNN convolution include length, width and height. In the actual data organization, the time dimension can only be used as the height dimension for model training. After adding the error correction model, even SVR correction can significantly reduce the forecast error and improve the correlation of the forecasted wind speed series. This shows that the error correction strategy is very effective after the unified forecast is completed. Among the four correction models, TCN-LSTM obtained the best correction effect. TCN can control the amount of historical information contained in the output sequence at each time by adjusting the convolution kernel size and dilation factor d, and the LSTM layer can realize the long-term evolution feature learning of the whole sequence. The combination of the two architectures is more suitable for the forecast of wind speed error series.

With the extension of the forecast horizon, the change characteristics of the forecast performance are interesting. In the first 6h, the forecast effect decreased significantly at any time, and the 15min forecast error of all models was less than 60% of that of 6h. In terms of the average forecast performance after 6h, the change of each index is no more than 5%. This is worthy of our further error analysis.

Fig. 6 shows the forecast curve and error scatter distribution of 200 continuous wind speed points at the anemometer tower, which are forecasted in advance from different forecast horizons. The forecast curve given by NWP is smoother than the real wind speed curve, and the unified forecast trend does not deviate much during this period. It can be seen that there are obvious large error points, especially when the real wind speed fluctuates rapidly, the trend will also lag or advance. Monitoring data and initial NWP forecast data are used as inputs to the model proposed in this paper, and 96 steps 15min rolling forecast can be realized on the basis of NWP. After the big data training, the forecast curves of the proposed model with 15min and 30min in advance are very close to the real wind speed curve. Not only the absolute error is limited within 1 m/s, but also the trend fitting is achieved in the peak range of 150–175. With the extension of the forecast horizon, the distribution of error points is gradually dispersed, and obvious outliers will appear first in the large fluctuation region. The trend following ability of the forecast curve to the real wind speed gradually decreases and the curve gradually smoothes from 6h in advance.

Figs. 7 and 8 show the error distribution box diagram of NWP and the proposed method, including the error distribution change in 96 consecutive steps. The box chart gives the distribution statistics of errors

Table 1
Test information statistics.

Forecast model	EC model	Number of case	Training parameter setting	Training time(min)	Loading time and forecast time (1000 times) (s)
SVR	–	34	The kernel is rbf; the penalty parameter C is 1.2	48.2	7.4 + 0.4
ANN	–	34	Adam 100epochs	508.1	15.7 + 3.3
RNN	–	34	Adam 100epochs	1277.4	21.7 + 6.8
LSTM	–	34	Adam 100epochs	1942.9	31.6 + 9.7
TCN-3DCNN	–	1	Adam 100epochs	124.8	3.9 + 1.9
STC-DPN	–	1	Adam 100epochs	168.2	5.4 + 2.6
STC-DPN	SVR	1 + 34	Adam 100epochs/ Same as forecast model	168.2 + 32.6	12.3 + 3.6
STC-DPN	TCN	1 + 34	Adam 100epochs	168.2 + 427.6	24.6 + 12.3
STC-DPN	LSTM	1 + 34	Adam 100epochs	168.2 + 809.7	32.8 + 16.7
STC-DPN	TCN-LSTM	1 + 34	Adam 100epochs	168.2 + 637.2	28.3 + 13.9

Table 2
MAE wind speed (m/s) statistics of 10 models.

Forecast model	EC model	Forecast horizon								
		15min	30min	1h	2h	4h	6h	6–12h	12–18h	18–24h
SVR	–	0.939	1.045	1.186	1.358	1.516	1.655	1.673	1.696	1.715
ANN	–	0.919	1.031	1.172	1.339	1.495	1.635	1.653	1.683	1.694
RNN	–	0.904	1.026	1.166	1.328	1.491	1.626	1.644	1.676	1.695
LSTM	–	0.882	1.002	1.124	1.306	1.453	1.594	1.612	1.643	1.662
TCN-3DCNN	–	0.793	0.915	1.028	1.206	1.335	1.477	1.493	1.523	1.539
STC-DPN	–	0.771	0.891	1.016	1.157	1.307	1.451	1.472	1.479	1.502
STC-DPN	SVR	0.703	0.786	0.896	1.055	1.179	1.309	1.374	1.389	1.418
STC-DPN	TCN	0.714	0.798	0.910	1.071	1.197	1.329	1.395	1.410	1.440
STC-DPN	LSTM	0.682	0.763	0.870	1.023	1.145	1.271	1.334	1.349	1.377
STC-DPN	**TCN-LSTM**	**0.669**	**0.749**	**0.854**	**1.004**	**1.124**	**1.248**	**1.312**	**1.324**	**1.352**

Table 3
RMSE statistics of 10 models (m/s).

Forecast model	EC model	Forecast horizon								
		15min	30min	1h	2h	4h	6h	6–12h	12–18h	18–24h
SVR	–	1.258	1.405	1.589	1.819	2.031	2.217	2.241	2.267	2.298
ANN	–	1.231	1.381	1.570	1.794	2.003	2.190	2.215	2.259	2.269
RNN	–	1.211	1.375	1.562	1.779	1.997	2.178	2.202	2.245	2.271
LSTM	–	1.182	1.342	1.506	1.750	1.947	2.135	2.160	2.201	2.227
TCN-3DCNN	–	1.062	1.226	1.377	1.616	1.788	1.979	2.013	2.040	2.052
STC-DPN	–	1.033	1.194	1.361	1.550	1.751	1.944	1.962	1.997	2.026
STC-DPN	SVR	0.942	1.053	1.200	1.413	1.580	1.754	1.841	1.861	1.893
STC-DPN	TCN	0.957	1.069	1.219	1.435	1.604	1.780	1.869	1.889	1.929
STC-DPN	LSTM	0.914	1.022	1.166	1.371	1.534	1.703	1.787	1.807	1.845
STC-DPN	**TCN-LSTM**	**0.896**	**1.003**	**1.144**	**1.345**	**1.506**	**1.672**	**1.758**	**1.774**	**1.811**

Table 4
COR statistics of 10 models.

Forecast model	EC model	Forecast horizon								
		15min	30min	1h	2h	4h	6h	6–12h	12–18h	18–24h
SVR	–	0.850	0.790	0.766	0.732	0.645	0.598	0.586	0.575	0.573
ANN	–	0.869	0.801	0.771	0.738	0.651	0.602	0.590	0.581	0.576
RNN	–	0.873	0.808	0.779	0.742	0.659	0.604	0.592	0.586	0.578
LSTM	–	0.875	0.814	0.782	0.746	0.662	0.608	0.596	0.590	0.582
TCN-3DCNN	–	0.903	0.831	0.806	0.759	0.685	0.611	0.599	0.593	0.585
STC-DPN	–	0.908	0.835	0.809	0.765	0.694	0.618	0.606	0.598	0.591
STC-DPN	SVR	0.927	0.853	0.822	0.781	0.712	0.636	0.623	0.617	0.609
STC-DPN	TCN	0.922	0.848	0.817	0.777	0.709	0.629	0.616	0.613	0.602
STC-DPN	LSTM	0.931	0.857	0.833	0.785	0.718	0.641	0.628	0.622	0.614
STC-DPN	**TCN-LSTM**	**0.938**	**0.863**	**0.839**	**0.790**	**0.724**	**0.647**	**0.634**	**0.628**	**0.619**

of 25%, 50% and 75%, as well as the position of the average value and the distribution of outliers, and enlarges the range of 25%–75%. Obviously, the performance of NWP does not decline significantly with the extension of the forecast horizon within 24h, but the error in the 0–6h ultra short term forecast cannot meet the actual demand. Also based on the outlier discrimination rule of box graph, when the absolute error value is greater than 4 m/s, it will be judged as outliers, and even the outlier discrimination condition is greater than 7 m/s. In addition to the loose criteria for outliers, the distribution of outliers also shows the inadequacy of NWP forecast performance. A large number of outliers are not near the critical line, but are randomly distributed in a larger error range, and even there is a forecast error of more than 15 m/s. In the 96 step error statistics of the proposed method, compared with the original NWP, many aspects have been significantly improved. First, the interval absolute value of 25% and 75% loci in 0–6h forecast is limited to 2 m/s, and does not exceed 3 m/s in 6–24h forecast. The criterion of outliers is more strict, and the distribution of outliers is more concentrated on the critical line.

Based on the analysis in Figs. 6–8, it can be believed that the results reflect the respective characteristics of ML method and physical method.

ML method can achieve small forecast error depending on the persistence of wind speed in the near forecast, but it is difficult to effectively predict the wind speed after a few hours. Due to the lack of accuracy of modeling data, the forecast error of physical methods in the first few hours is too large, but it can be robust in the longer-term forecast.

4.4. Error analysis

Ten models are used for forecast experiment in section 4.3, and MAE, RMSE and COR were compared. Some subgraphs in Fig. 6 show that the error performance has trend lag [15], the forecast sequence is excessively smooth [39], and the forecast between high and low wind speeds is conservative [30]. In this section, with the help of CII and HR, we take the forecasted wind speed of the original NWP as the baseline to conduct auxiliary analysis on the errors of different models.

Table 5 shows the MAE, RMSE and COR segment statistics of the four initial NWP ranges. In Table 6, in addition to the CII of the original ten models for NWP, the impact of the wind speed recorded in the VMD processing model input on the forecast effect is also evaluated. Among the four ranges, 0–6h forecast has the greatest improvement effect, and
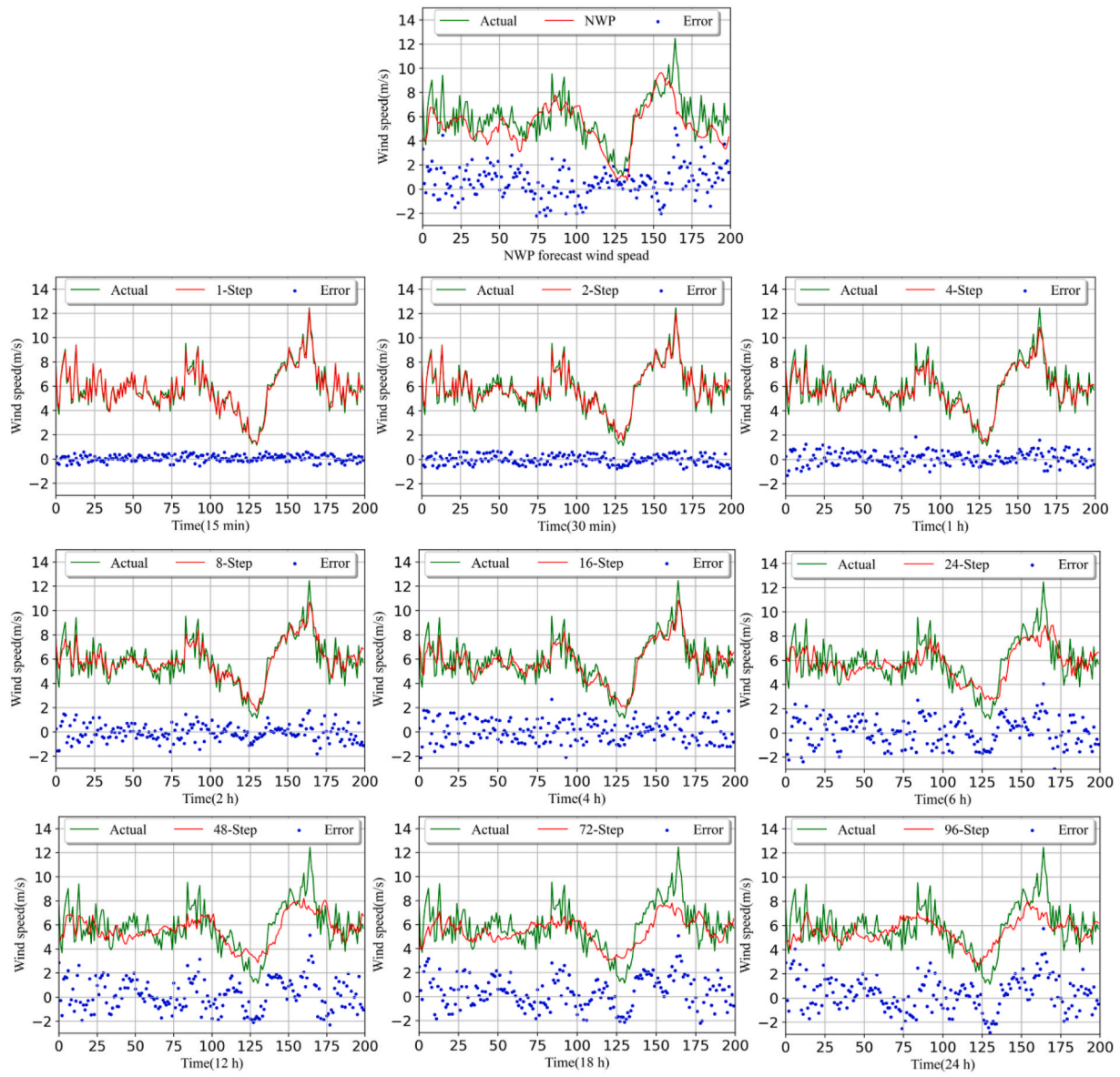
Fig. 6. NWP and the proposed method for continuous 200 point forecast curve and error distribution.

only SVR is used in the forecast model to obtain more than 20% CII improvement effect. In the other three time ranges, it also achieved an increase of more than 8%. By ranking the forecast results of the model, it can be seen that there have been three obvious performance improvements: (1) from single site forecast to unified forecast, the improvement performance of the four ranges is about 6%; (2) After adding the EC model, the increase range reached 7%; (3) After the recorded wind speed is treated with VMD, the improvement performance of 0–6h is nearly 6%, but the forecast effect of 6–24h is hardly affected. Under the same conditions, it is difficult to achieve a 5% improvement by model replacement alone, which shows that the application of the new strategy is more effective than the simple model replacement.

Fig. 9 presents the NWP and the proposed method forecasted-real wind speed distribution. The wind speed points are near the central axis, and the degree of divergence can reflect the forecast effect. The divergence degree of wind speed points given by NWP is obviously higher than that of the proposed method, and there are a large number of outliers. The proposed method performs well in the interval of 0–6h, and the degree of divergence gradually increases with the extension of the forecast horizon. Compared with NWP, the proposed method can

significantly correct some large error points. It is worth noting that the outliers in the NWP mainly come from the higher wind speed range. Moreover, in the high wind speed range, there are many samples with the predicted wind speed lower than the actual wind speed. The forecast given by the proposed method also has the same problem in 6–24h, at least it can be shown that in the high wind speed range, the conservative wind speed forecast is a factor that causes large errors.

Based on the above considerations, according to the annual distribution frequency of the real wind speed (Fig. 5), it is divided into three wind speed ranges of 0–5 m/s, 5–10 m/s, and above 10 m/s to calculate HR, and W is set to 1 m/s, 2 m/s, 3 m/s. Fig. 10 shows the error distribution statistics of the three wind speed intervals of the NWP. In the high wind speed range above 10 m/s, the central axis of the error is about 2 m/s, indicating that the high wind speed interval is easy to give a smaller predicted wind speed. In the small wind speed range of 0–5 m/s, the central axis of the error is about -1 m/s, indicating that the small wind speed range is easy to give a larger forecasted wind speed. Reference [30] divides the NWP forecast wind speed into ranges and analyzes the error distribution. The results are the same as in this paper, in the high-low wind speed range, the wind speed forecast will tend to be
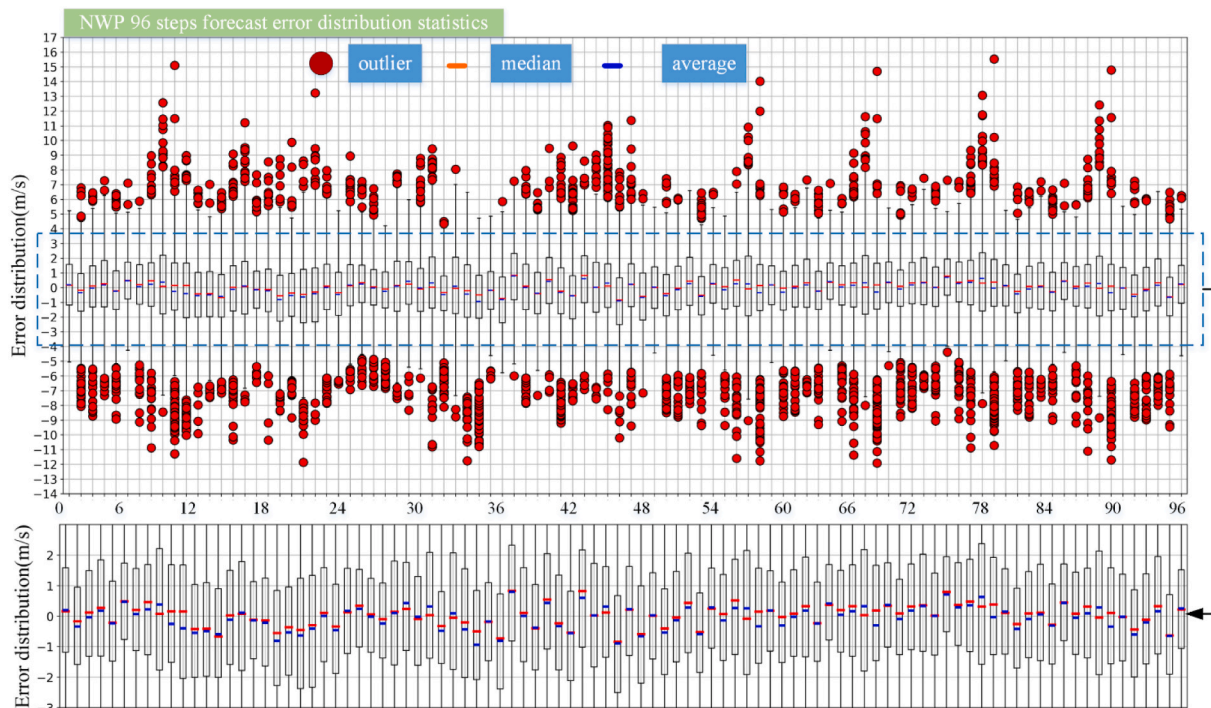
Fig. 7. Statistics of NWP 96 steps forecast error distribution based on box diagram.
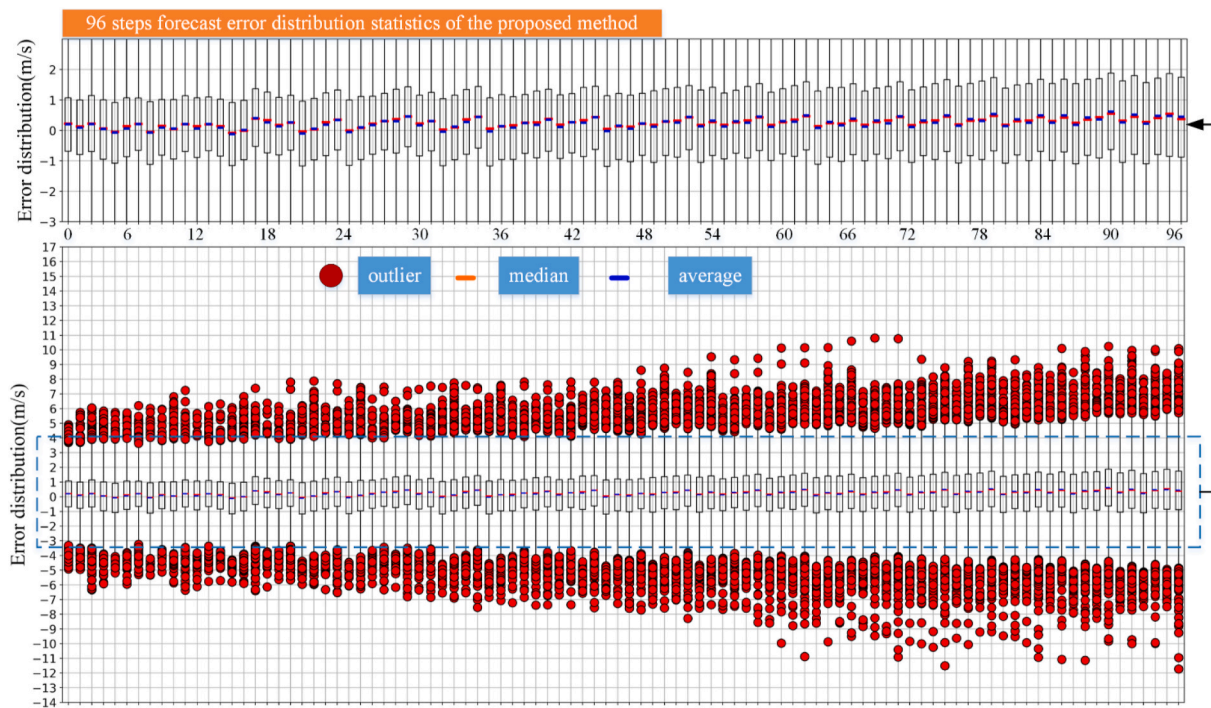

Fig. 8. Statistics of proposed method 96 steps forecast error distribution based on box diagram.

Table 5
Segment statistics of NWP forecast wind speed performance.

Indicator	Forecast horizon			
	0–6h	6–12h	12–18h	18–24h
MAE	1.997	2.032	2.044	2.071
RMSE	2.413	2.418	2.422	2.431
COR	0.590	0.582	0.571	0.568

conservative.

Fig. 11 presents the detailed error statistics of NWP, STC-DPN and STC-DPN-EC in three wind speed ranges. It can be seen from the bar chart that even if a loose *W* is set in the wind speed range above 10 m/s, the HR is still less than the HR of 0–5 m/s, which also verifies the argument that the error and difficulty of high wind speed forecast are greater [47,49]. In different wind speed ranges, with NWP as the base-line, after unified forecast by STC-DPN and error correction, the

Table 6
Models CII statistics based on NWP.

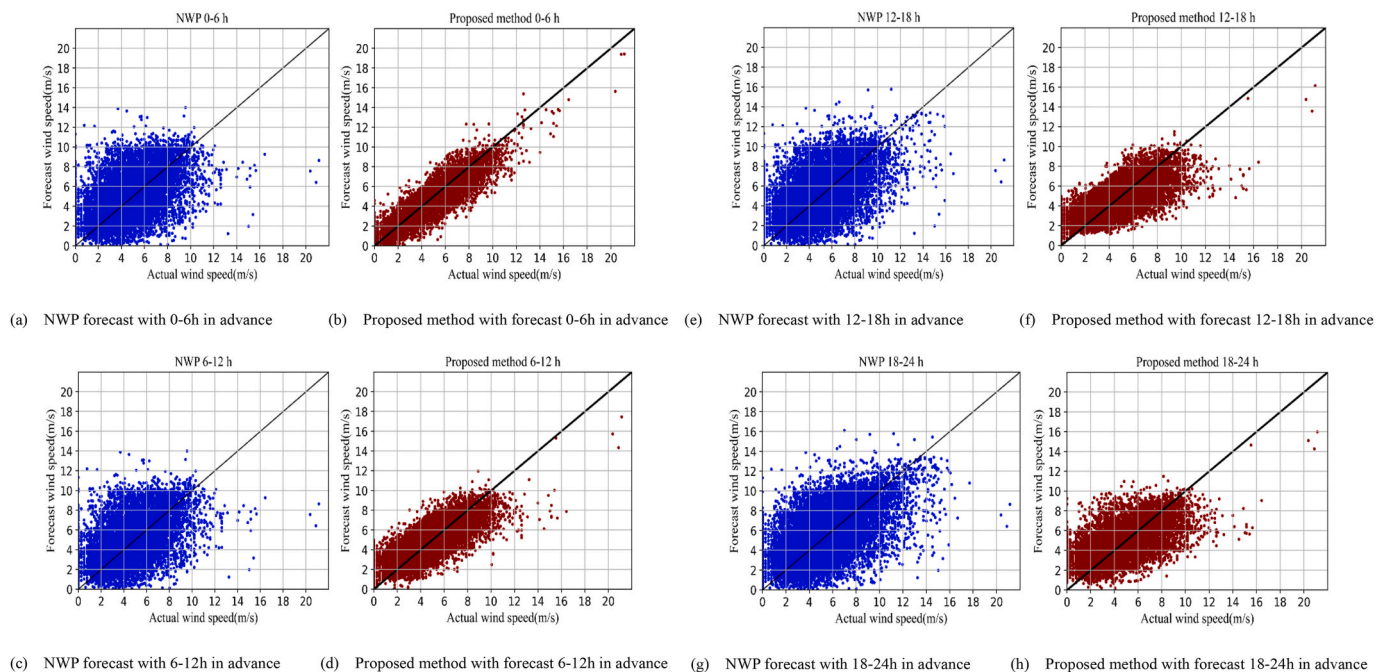| Forecast model	EC model	Forecast horizon			
		0–6 h	6–12 h	12–18 h	18–24 h
SVR	–	20.73	8.56	8.31	8.14
ANN	–	21.57	9.47	9.19	9.01
RNN	–	22.22	10.06	9.76	9.56
LSTM	–	23.09	11.25	10.91	10.69
TCN-3DCNN	–	29.38	17.40	16.88	16.54
STC-DPN	–	32.41	18.85	18.28	17.91
STC-DPN	SVR	38.60	26.10	25.32	24.81
STC-DPN	TCN	37.93	25.97	25.19	24.69
STC-DPN	LSTM	40.92	27.78	26.95	26.41
STC-DPN	**TCN-LSTM**	**42.79**	**29.89**	**28.93**	**28.35**
STC-DPN	TCN-LSTM(without VMD)	36.84	28.93	28.87	28.32

following characteristics are presented:

(1) In 0–6h, after STC-DPN unified forecast and error correction, HR is significantly improved compared to NWP. Especially in the 15min-advanced forecast, the HR of 0–5 m/s and above 10 m/s is better than NWP by more than 30%. With the extension of the forecast horizon, the HR performance in the 5–10 m/s range is the
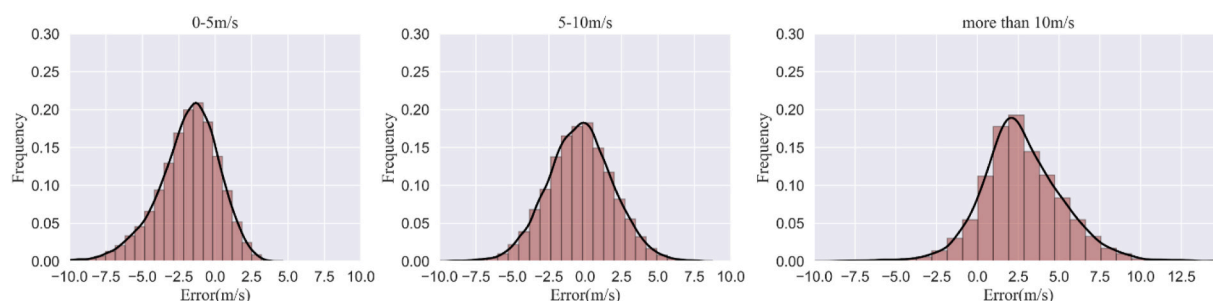
most stable, with the smallest decrease, and the ratio of the 15min-advanced forecast is no more than 20%.

(2) With the extension of the forecast horizon, the HR of the proposed method decreased significantly in the two ranges above 0–5 m/s and 10 m/s. Especially in the range above 10 m/s, the STC-DPN decreased by 40% in the comparison between 6h and 15min. In the analysis of Fig. 9, this is because the model makes the forecast more conservative in order to reduce the average error in the forecast of the high and low wind speed ranges.

(3) After the unified forecast of STC-DPN, compared with the HR of NWP, the two ranges above 0–5 m/s and 10 m/s did not get significant improvement in the forecast of 6–18h, and the improvement rate was only about 2%. This phenomenon shows that although STC-DPN can reduce the large error amplitude of NWP, it is still difficult to reduce to below 3 m/s in the high wind speed range, and the improvement degree in the low wind speed range is also limited. Due to the regularity of error generation, after adding the error correction model, the HR is improved by about 6% in the 6–18h forecast. This shows that the revised model can improve the forecast effect in the high and low wind speed range.

5. Further discussions

In the simulation results of the case wind farm, the unified forecast

(a) NWP forecast with 0-6h in advance
(b) Proposed method with forecast 0-6h in advance
(c) NWP forecast with 6-12h in advance
(d) Proposed method with forecast 6-12h in advance
(e) NWP forecast with 12-18h in advance
(f) Proposed method with forecast 12-18h in advance
(g) NWP forecast with 18-24h in advance
(h) Proposed method with forecast 18-24h in advance

Fig. 9. NWP and Proposed method forecast - real wind speed distribution.

Fig. 10. Frequency step of NWP wind speed prediction error in different wind speed ranges.
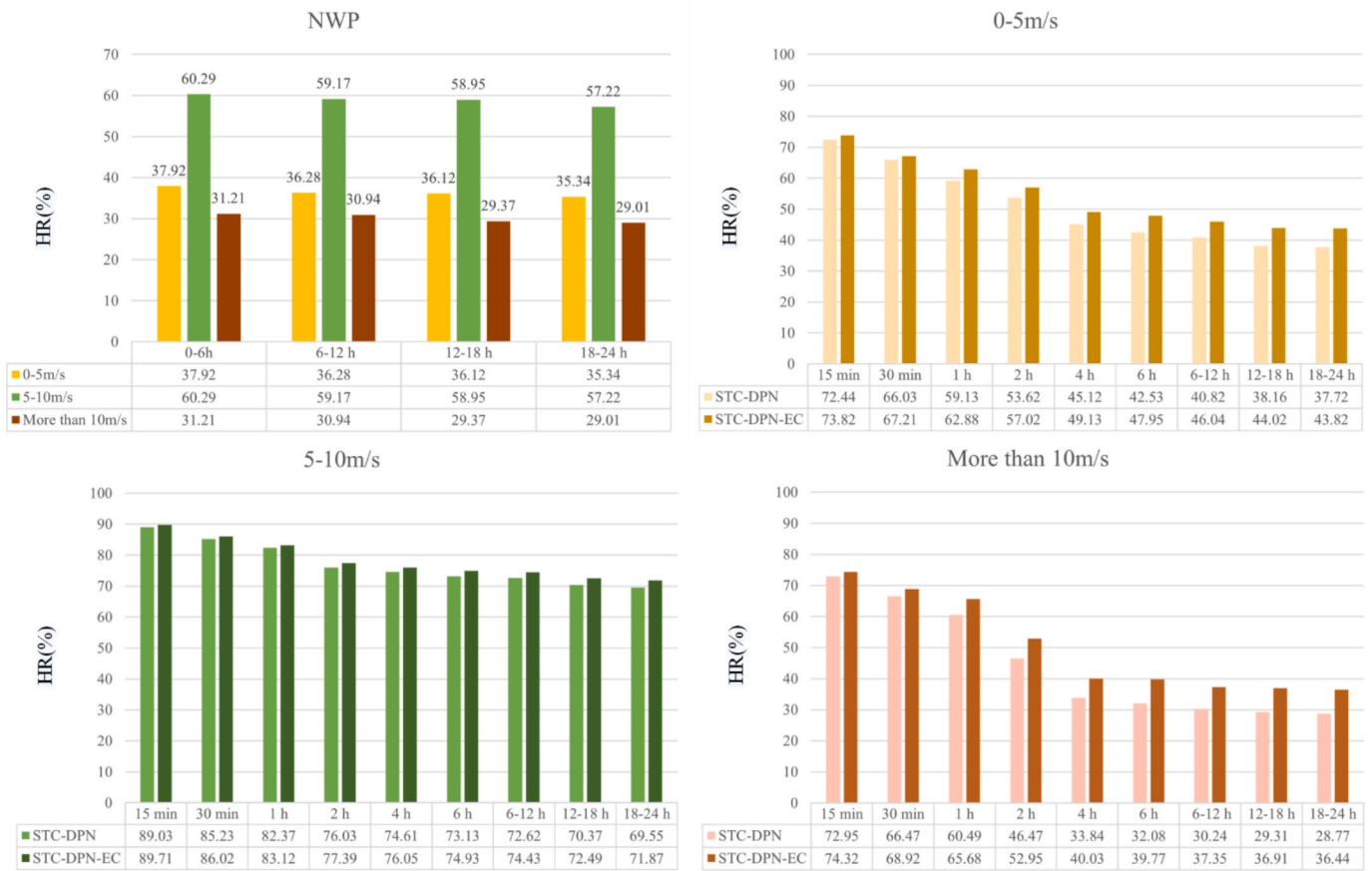
Fig. 11. HR statistics for NWP, STC-DPN and STC-DPN-EC.

combined with the single site error correction forecast strategy proposed in this paper achieved the best forecast results. In order to gain a deep understanding of the effectiveness of the proposed method for wind speed forecasting in the whole area of wind farms, this section will discuss the following aspects.

5.1. Advantages of the STC-DPN-EC: combination of ML and physics model , NWP and monitoring data

As mentioned in the introduction, the unified forecaste model STC-DPN combines ML and physical models. The ML model can be trained using a large amount of historical data, and a small average error can be achieved in the prediction within 6 h, which guarantees the prediction performance of STC-DPN in the first 6 h. The WRF-based physical model provides a grid of NWP forecasts for the next 24 h as part of the input to STC-DPN. Due to the rough initial modeling data, the physical model has insufficient near-prediction accuracy and may have obvious spatial and temporal deviations. However, the calculation of the physical model considers a wide range of atmospheric motions, and it still has stable prediction performance after 6 h, which effectively improves the prediction performance of the ML model that only relies on historical data.

The forecast grid of NWP and the wind turbine monitoring data of the whole wind farm have obvious spatiotemporal characteristics, and STC-DPN effectively combines them in the process of data preprocessing and model training. In Fig. 3, the WRF mode divides the NWP grid in a regular manner, and the grid and geographic resolution size can be adjusted by parameter settings. However, the distribution of wind turbines is affected by hilly terrain, and the locations of wind speed monitoring points are extremely irregular. After obtaining the NWP grid interpolation wind speed sequence, it cannot be directly entered into ConvLSTM for joint learning of spatiotemporal features. The use of TCN

solves the problems of data combination and dimension conversion. Monitoring data and NWP grid data are spliced into an input matrix according to time series, and the number of features is increased through TCN to convert into a regular four-dimensional space time series matrix. It can be believed that this establishes a virtual wind farm that satisfies the ConvLSTM model's requirements for data shape and is not affected by the arrangement of fans in the wind farm.

5.2. Consideration on error cause and correction

After realizing the unified wind speed forecast of the wind farm, this paper provides an independent error correction model for the forecasted wind speed of each wind turbine site. Analyzing the error is helpful to adopt a targeted correction strategy. However, it can only reduce the size of the error and the influence of the wind speed forecast error on the operation of the wind farm.

The historical monitoring data used as input has obvious volatility during the collection process, especially the rapid change of wind speed. The input of STC-DPN includes the monitoring data of all wind turbine positions, which reduces the influence of randomness of data collection to a certain extent, and better reflects the wind speed change trend of the wind farm over a period of time. However, the acquisition frequency of 15 min affects the continuity of the data sequence, and an excessively high acquisition frequency will increase storage and computing costs. This paper uses VMD to process historical wind speed data, aiming to reduce the stochastic volatility of recorded wind speeds. For the NWP provided by the physical model, the defect of insufficient initial modeling accuracy is unavoidable. In practice, various methods is tried to improve the accuracy of the initial NWP wind speed forecast, including optimizing the combination of physical parameters, adjusting the grid size and the near-surface vertical layer height. However, after

STC-DPN forecast and error correction, the error will always have a larger decrease. Therefore, it can be believed that for the wind speed forecast of a certain wind farm, the forecast error of NWP is specific, and the forecast model can be trained by combining this large amount of historical data to improve the forecast effect.

6. Conclusions

Aiming at the wind speed forecast of the entire wind farm, this paper proposes a strategy combining unified forecast and single site error correction. A unified forecasting framework of STC-DPN is proposed, which combines the historical monitoring data and the forecasting grid of NWP, and corrects the forecasting error of each forecasting site individually, providing a more accurate multi-step wind speed forecasting. In data processing, VMD is used to denoise the recorded wind speed, and the high frequency components of NWP forecast wind speed are retained to participate in single-point error correction. Compared with the baseline methods, the proposed method shows better forecast performance in an actual operating wind farm. The proposed method has good versatility and is suitable for wind speed forecast in complex wind farm terrain such as hills, without being affected by the arrangement of wind turbines. The main conclusions are summarized as follows:

1) The STC-DPN architecture is very flexible. Combined with the monitoring data and the NWP forecast grids, on the basis of the original NWP forecast wind speed, multi-step high-frequency rolling wind speed forecast can be achieved. On the basis of STC-DPN, according to the size of the wind farm, the arrangement of wind turbines and the forecast demand, the forecast wind speed can be obtained by adjusting the forecast horizon and the TCN output data conversion dimension.
2) The error generated by the forecast model has obvious regularity, and adding the error correction model can effectively reduce the uncertainty of the forecasted wind speed. After completing the error correction in this paper, the performance of the CII index has been improved by more than 10%, and the HR index has been improved in different wind speed ranges, especially in the high wind speed range. The increase in the HR index indicates that error correction reduces the probability of large errors.
3) When the data input and processing methods are the same, compared with the replacement of the forecast model, adding a new forecast strategy can greatly reduce the forecast error. In the case study,

taking NWP as the baseline, single site forecasting, unified forecasting, adding error correction, and processing of VMD all significantly reduce forecasting errors. Among them, only replacing the forecast model, the CII index increases by less than 3%, while adding new forecast strategies can increase by more than 5%.
4) On the basis of NWP, the proposed method improves the comprehensive forecast performance of the 0–6h by more than 40%, and the comprehensive forecast performance of the 6–24h by nearly 30%. Also it can achieve high-frequency rolling forecast, and simultaneously perform ultra-short-term forecast within a few hours and short-term forecast within a day, which provides a reliable reference for the real-time grid connection and scheduling tasks of wind farms.

In the future research, we plan to apply the strategies and models proposed in this paper to more research fields, such as wind power forecast or wind turbine fault early warning. According to the actual situation of the wind field, taking into account the temperature, air pressure, humidity and other factors, the TCN model can be introduced into the 2D ConvLSTM model to jointly predict the wind speed by increasing the number of channels, so as to further improve the accuracy of wind speed forecast. In the future research, it needs more actual wind fields to verify the effectiveness and superiority of the proposed method.

CRediT authorship contribution statement

Xingdou Liu: Conceptualization, Methodology, Software, Writing – original draft. **Li Zhang:** Supervision, Investigation. **Jiangong Wang:** Data curation. **Yue Zhou:** Validation. **Wei Gan:** Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Li Zhang reports financial support was provided by the National Key Research and Development Program of China.

Acknowledgement

Appendix A

Table A1
Adopted physical parameter scheme of WRF model.

Physical scheme	Mode scheme selection
Microphysics	Goddard microphysics scheme
Longwave radiation	New Goddard scheme
Shortwave radiation	Dudhia scheme
Surface layer	MM5 scheme
Land surface	5-layer thermal diffusion
Planetary boundary layer	Shin-Hong scheme
Cumulus parameterization	Grell-Freitas scheme

Table A2
Parameter setting of the proposed forecast framework.

Model		parameter
VMD [16] (data processing)	Process monitoring data sequence	$k = 9$, $\alpha = 1000$, Reconstruction sequence retention IMF1-6
	Process NWP wind speed sequence	$k = 5$, $\alpha = 1000$, Reconstruction sequence retention IMF3-5
STC-DPN (unified forecast)	TCN [50]	$d = [1,2,4,8,16,32]$, number of kernels and kernel size are 256 and 3

(continued on next page)

Table A2 (*continued*)

Model		parameter
TCN-LSTM (single site correction)	ConvLSTM [54]	ConvLSTM layer (1,2,3): number of kernels and kernel size are (8, 32,128) and (3 × 3, 3 × 3,3 × 3)
	TCN	$d = [1,2,4,8]$, number of kernels and kernel size are 32 and 3
	LSTM [49]	LSTM layer (1,2): number of neurons is (32, 64)

References

[1] A. Agga, A. Abbou, M. Labbadi, Y.E. Houm, Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models, Renew. Energy 177 (2021) 101–112.

[2] Y.Y. Gao, J.Z. Wang, H.F. Yang, A multi-component hybrid system based on predictability recognition and modified multi-objective optimization for ultra-short-term onshore wind speed forecasting, Renew. Energy 188 (2022) 384–401.

[3] J. Heng, Y.G. Hong, J.M. Hu, S.Y. Wang, Probabilistic and deterministic wind speed forecasting based on non-parametric approaches and wind characteristics information, Appl. Energy 306 (2022), 118029.

[4] J.Z. Wang, S. Wang, B. Zeng, H.Y. Lu, A novel ensemble probabilistic forecasting system for uncertainty in wind speed, Appl. Energy 313 (2022), 118795.

[5] L. Liu, J.J. Wang, Z.D. Tian, Super multi-step wind speed forecasting system with training set extension and horizontal–vertical integration neural network, Appl. Energy 292 (2021), 116908.

[6] M.A. Costa, R. Ruiz-Cardenas, L.B. Mineti, M.O. Prates, Dynamic time scan forecasting for multi-step wind speed prediction, Renew. Energy 177 (2021) 584–595.

[7] J.M. Hu, J.N. Heng, J.M. Wen, W.G. Zhao, Deterministic and probabilistic wind speed forecasting with denoising-reconstruction strategy and quantile regression based algorithm, Renew. Energy 162 (2020) 1208–1226.

[8] Q. Zheng, P. Yan, Z. Hamidreza, C. Niya, A review and discussion of decomposition-based hybrid models for wind energy forecasting applications, Appl. Energy 235 (2019) 939–953.

[9] H. Liu, Z. Duan, C. Chen, Wind speed big data forecasting using time-variant multi-resolution ensemble model with clustering auto-encoder, Appl. Energy 280 (2020), 115975.

[10] W.Z. Zhang, L.F. Zhang, J.Z. Wang, X.S. Niu, Hybrid system based on a multi-objective optimization and kernel approximation for multi-scale wind speed forecasting, Appl. Energy 277 (2020), 115561.

[11] Aasim, S.N. Singh, A. Mohapatra, Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting, Renew. Energy 136 (2019) 758–768.

[12] H. Bludszuweit, J. Dominguez-Navarro, A. Llombart, Statistical analysis of wind power forecast error, IEEE Trans. Power Syst. 23 (2) (2008) 983–991.

[13] O.B. Shukur, M.H. Lee, Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA, Renew. Energy 76 (2015) 637–647.

[14] Q. Hu, P. Su, D. Yu, J. Liu, Pattern-based wind speed prediction based on generalized principal component analysis, IEEE Trans. Sustain. Energy 5 (3) (2014) 866–874.

[15] C.Y. Zhang, C.L. Philip Chen, M. Gan, L. Chen, Predictive deep Boltzmann machine for multiperiod wind speed forecasting, IEEE Trans. Sustain. Energy 6 (4) (2015) 1416–1425.

[16] H.S. Cai, X.D. Jia, J.S. Feng, Q.B. Yang, W.Z. Li, F. Li, J. Lee, A unified Bayesian filtering framework for multi-horizon wind speed prediction with improved accuracy, Renew. Energy 178 (2021) 709–719.

[17] J. Wang, Z.S. Yang, Ultra-short-term wind speed forecasting using an optimized artificial intelligence algorithm, Renew. Energy 171 (2021) 1418–1435.

[18] J.J. Song, J.Z. Wang, A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting, Appl. Energy 215 (2018) 643–658.

[19] W.L. Fu, P. Fang, K. Wang, Z.X. Li, D.Z. Xiong, K. Zhang, Multi-step ahead short-term wind speed forecasting approach coupling variational mode decomposition, improved beetle antennae search algorithm-based synchronous optimization and Volterra series model, Renew. Energy 179 (2021) 1122–1139.

[20] Z. Qu, W. Mao, K. Zhang, W. Zhang, Z. Li, Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network, Renew. Energy 133 (2019) 919–929.

[21] Y. Noorollahi, M.A. Jokar, A. Kalhor, Using artificial neural networks for temporal and spatial wind speed forecasting in Iran, Energy Convers. Manag. 115 (May) (2016) 17–25.

[22] P. Jiang, H. Yang, J. Heng, A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting, Appl. Energy 935 (2019) 786–801.

[23] C.Y. Wu, J.Z. Wang, X.J. Chen, P. Du, W.D. Yang, A novel hybrid system based on multi-objective optimization for wind speed forecasting, Renew. Energy 146 (2020) 149–165.

[24] G. Lopez, P. Arboleya, Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the Andes Mountains, Ecuador, Renew. Energy 183 (2022) 351–368.

[25] D.R. Chandra, M.S. Kumari, M. Sydulu, F. Grimaccia, M. Mussettaet, Adaptive wavelet neural network based wind speed forecasting studies, J. Electr. Eng. Technol. 9 (6) (2014) 1812–1821.

[26] C. Park, C. Lee, L. Hong, Y.Y. Hwang, T. Yoo, J. Jang, et al., S2-Net: machine reading comprehension with SRU-based self-matching networks, ETRI J. 41 (3) (2019).

[27] M.D. Liu, L. Ding, Y.L. Bai, Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction, Energy Convers. Manag. 233 (2021), 113917.

[28] H. Liu, X.W. Mi, Y.F. Li, Z. Duan, Y.N. Xu, Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression, Renew. Energy 143 (2019) 842–854.

[29] H. Liu, C. Chen, Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction, Appl. Energy 254 (2019), 113686.

[30] H. Wang, S. Han, Y.Q. Liu, J. Yan, L. Li, Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system, Appl. Energy 237 (2019) 1–10.

[31] S. Alessandrini, S. Sperati, P. Pinson, A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data, Appl. Energy 107 (2013) 271–280.

[32] V. Hoolohan, A.S. Tomlin, T. Cockerill, Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data, Renew. Energy 126 (2018) 1043–1054.

[33] J. Zhao, Z. Guo, Y. Guo, W. Lin, W. Zhu, A self-organizing forecast of day-ahead wind speed: selective ensemble strategy based on numerical weather predictions, Energy 210 (2021), 119509.

[34] S. Salcedo-Sanz, A.M. P'erez-Bellido, E.G. Ortiz-García, A. Portilla-Figueras, L. Prieto, D. Paredes, Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction, Renew. Energy 34 (2009) 1451–1457.

[35] J. Zhao, Y. Guo, X. Xiao, J. Wang, D. Chi, Z. Guo, Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method, Appl. Energy 197 (2017) 183–202.

[36] M. Mu, B. Che, F. Zhou, Y. Yu, Methods and uncertainties of meteorological forecast, Meteorol. Mon. 37 (2011) 1–13.

[37] H.S. Cai, X.D. Jia, J.S. Feng, Q.B. Yang, Y.M. Hsu, Y.D. Chen, J. Lee, A combined filtering strategy for short term and long term wind speed prediction with improved accuracy, Renew. Energy 136 (2019) 1082–1090.

[38] J.L. Williams Jl, R.M. Maxwell Rm, L. Delle Monache, Development and verification of a new wind speed forecasting system using an ensemble Kalmanfilter data assimilation technique in a fully coupled hydrologic and atmospheric model, J. Adv. Model. Earth Syst. 5 (2013) 785–800.

[39] J. Zhao, Z.H. Guo, Z.Y. Su, Z.Y. Zhao, X. Xiao, et al., An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed, Appl. Energy 162 (2016) 808–826.

[40] W. Lin, J. Wang, W. Zhang, Z. Guo, D. Chi, Y. Zhang, Program of wind speed prediction based on numerical simulation with intelligent optimization algorithm, Climatic Environ Res 17 (2012) 646–658.

[41] H. Liu, R. Yang, T.T. Wang, L. Zhang, A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections, Renew. Energy 165 (2021) 573–594.

[42] Y.G. Zhang, G.F. Pan, B. Chan, J.Y. Han, et al., Short-term wind speed prediction model based on GA-ANN improved by VMD, Renew. Energy 156 (2020) 1373–1388.

[43] X.S. Niu, J.Y. Wang, A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting, Appl. Energy 241 (2019) 519–539.

[44] C.S. Li, G. Tang, X.M. Xue, A. Saeed, X. Hu, Short-term wind speed interval prediction based on ensemble GRU model, IEEE Trans. Sustain. Energy 11 (3) (2020) 1370–1380.

[45] Z.D. Tian, H. Chen, Multi-step short-term wind speed prediction based on integrated multi-model fusion, Appl. Energy 298 (2021), 117248.

[46] P.W. Du, Ensemble machine learning-based wind forecasting to combine NWP output with data from weather statio, IEEE Trans. Sustain. Energy 10 (4) (2018) 509–523.

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

[48] Y.Q. Liu, H. Qin, Z.D. Zhang, S.Q. Pei, Z.Q. Jiang, et al., Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model, Appl. Energy 269 (2020), 114259.

[49] Q.M. Zhu, J.F. Chen, D.Y. Shi, L. Zhu, X. Bai, X.Y. Duan, Y. Liu, Learning temporal and spatial correlations jointly:A unified framework for wind speed prediction, IEEE Trans. Sustain. Energy 11 (1) (2020) 509–523.

[50] S.J. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, preprint (2018) arXiv:1803.01271. [Online]. Available: http://arxiv.org/abs/1803.01271.

[51] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Conf Comput Vis Pattern Recognit. (CVPR) (2016;Jun) 770–778.

[52] J. Gao, D. He, X. Tan, T. Qin, L.W. Wang, T.Y. Liu, Representation degeneration problem in training natural language generation models, preprint (2019) arXiv: 1907.12009v1. [Online]. Available: http://arxiv.org/pdf/1907.12009.pdf.

[53] S. Hochreiter, S. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[54] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, W.-C. Woo, Convolutional LSTM network: a machine learning approach for precipitation nowcasting, Adv. Neural Info. Process. Syst. Preprint (2015) arXiv: 1506.04214. [Online]. Available: https://arxiv.org/pdf/1506.04214.pdf.

[55] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, IEEE Trans. Signal Process. 62 (3) (2014) 531–544.

[56] Z. Wang, C. Liu, D. Cheng, L. Wang, K-T. Tim Yang, Automated detection of clinically significant prostate cancer in mp-MRI images based on an end-to-end deep neural network, IEEE Trans. Med. Imag. 37 (5) (2018) 1127–1139.