Article:

Zhang, H, Yue, D, Dou, C et al. (2 more authors) (2021) Two-Step Wind Power Prediction Approach With Improved Complementary Ensemble Empirical Mode Decomposition and Reinforcement Learning. IEEE Systems Journal. pp. 1-11. ISSN 1932-8184

https://doi.org/10.1109/jsyst.2021.3065566

Two-step wind power prediction approach with improved complementary ensemble empirical mode decomposition and reinforcement learning

Huifeng Zhang, *Member, IEEE,* Dong Yue, *Fellow, IEEE* Chunxia Dou, *Member, IEEE,* Kang Li, *Senior Member, IEEE,* and Gerhard P. Hancke, *Fellow, IEEE*

***Abstract*—The strong stochastic nature of wind power generation makes it extremely challenging to accurately predict and support the planning and operation of modern power systems with significant penetration of renewable energy. This paper proposes a two-step wind power prediction method, which consists of two phases: long time-scale coarse prediction and short time-scale fine correction. In the long time-scale phase, a complementary ensemble empirical mode decomposition based sigma point Kalman filter (CEEMDSPKF) approach is proposed to coarsely predict wind power merely with historical data. In the short time-scale phase, a deep deterministic policy gradient (DDPG) approach learns from real-time weather information to correct the coarse prediction result, which results in an improved prediction accuracy. A real-life case study confirms that the proposed method can properly predict wind power generation and have a better prediction accuracy than existing techniques, thus offering a viable and promising alternative for predicting wind power generation.**

***Index Terms*—wind power prediction, empirical mode decomposition, coarse prediction, fine correction, weather information.**

I. Introduction

DRIVEN by the increasing concern over the global climate change and environmental pollution due to the intensive consumption of fossil fuels, wind power generation technologies have developed so quickly in the past few years that wind power has now become a mature energy resource worldwide [1]. The increasing wind power generation leads to a high penetration in power systems, which inevitably has a significant impact on the operation and reliability of electrical power systems due to its intermittent characteristics. Accurate wind power prediction has become an indispensable task for system operators worldwide for power system scheduling and operation [2].

H. Zhang is with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Jiangsu Province, China, 210023, e-mail: zhanghuifeng_520@163.com

D. Yue is with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Jiangsu Province, China, 210023, e-mail: medongy@vip.163.com(Corresponding author)

C. Dou is with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications,Jiangsu Province, China, 210023, e-mail: cxdou@ysu.edu.cn

K. Li with the school of Electronic and Electrical Engineering University of Leeds, Leeds, LS2 9JT, United Kingdom, e-mail: K.Li1@leeds.ac.uk

G. P. Hancke is with the Nanjing University of Posts and Telecommunications, Jiangsu Province, China, 210023, and also with the University of Pretoria, Pretoria 0002, South Africa, e-mail: g.hancke@ieee.org(Corresponding author)

The traditional wind power prediction approaches [3] proposed so far in the literature mainly include multiple linear regression [4], the auto-regressive integrated moving average (ARIMA) [5], and the Kalman filter [6]. In [4], a vector-autoregressive-to-anything process is proposed to model wind speeds in multiple locations with a time-dependent interception, which provides both long term and short term forecasting results from 21 locations in Finland. In [6], a stochastic wind power model is proposed by using the ARIMA process, while considering the non-stationary and physical limits of stochastic wind power generation, and the model is validated in terms of temporal correlation and probability distribution. Since these methods rely on the historical data, the challenges are the high complexity of the forecasting data along with the increasing nonlinearity during the power forecasting process [7].

Intelligent methods use artificial intelligence to forecast wind power generation, such as artificial neural network (ANN) models [8], [9], support vector machines [10], particle swarm optimization [11], a Gaussian process [12], neuro-fuzzy systems [13], empirical mode decomposition [14], [15], deep learning methods [16], [17], [18]. In [19], a statistical scenario forecasting method is proposed for wind power ramp event probabilistic forecasting, and the obtained results confirm that it can accurately estimate the characteristics of wind power ramp events. In [10], the support vector machine based enhanced Markov model is developed for short-term wind power forecasting based on the measurement data with specific patterns of wind ramps, and the numerical test results demonstrate its improved accuracy in wind power forecasting. In [12], a hybrid deterministic-probabilistic method with a Gaussian process is proposed to estimate the forecasting errors, and their experimental results show that it can reduce the forecasting error due to its capacity to handle the time-varying characteristics of wind power generation. In summary, the above methods forecast wind power generation mainly based on the historical measurements from wind farms, while numerical weather prediction forecasts the wind power without sufficient analysis of the law of wind power generation [20]. It is highly desirable that the forecasting process should embed the distributed energy resource characteristics at different time scales and time intervals [21]. However, a single algorithm cannot mine and capture the intermittent and volatility of wind power generation properly, which can lead to poor forecasting accuracy, especially when it comes to the nonlinearity problems [22]. Therefore, hybrid models have emerged that aggregated the advantages of different forecasting methods, optimization algorithms and signal processing tools. Reference [23] utilizes wavelet transformation to decompose the historical data into a set of components with different frequencies, and combines this with a genetic algorithm and back-propagation network to forecast short-term wind speed. In literature [24], an accurate prediction method composed two-dimensional convolution neural network, it decomposed the original signals to subsignals and selected best candidate inputs with minimum forecasting error. Reference [25] proposes four modules, including a data process, optimization, forecasting and evaluation. A signal processing method is employed to decompose, construct and mine an electricity power system, a whale optimization algorithm is utilized to optimize those parameters of individual models, and an evaluation module includes hypothesis testing. A proper approach is thus to blend different methods within different time scales, which could compliment the drawbacks of the different methods and adjust the forecasting error in real-time.

In compliance with the above analysis, this paper proposes a multiple time-scale forecasting method with a CEEMD-SPKF correction approach to overcome the challenge of effectively integrating different techniques for dealing with the intermittent characteristics of wind power generation. The main idea of the approach is summarized as follow: (1) A two-step prediction framework with different time scales is used to handle the strong stochastic nature of wind power generation. It includes a long time-scale coarse forecasting phase and short time-scale correction phase. (2) The long time-scale coarse prediction is made by a CEEMD-SPKF approach using historical data. With consideration of boundary effects [26], Sigma points are firstly utilized to approximate boundary for overcoming boundary effect of following complementary ensemble empirical mode decomposition (CEEMD) method, which is used to decompose the data into several relative stable sequences. Then the sigma point based Kalman filter is used to update each stable sequence to reconstruct the original data. (3) In the short-time scale phase, a DDPG method is utilized for a short time-scale correction with real-time weather information, which can further improve the prediction accuracy.

The remainder of the paper is organized as: The forecasting framework is elaborated in section II, and long time-scale prediction with CEEMDSPKF is detailed in section III. Section IV presents short time-scale correction and a case study respectively, and Section V concludes the paper.

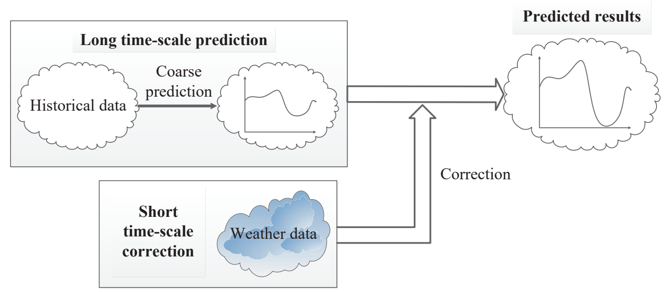

Fig. 1. The main structure of the two-step prediction approach

II. The framework for the multiple time-scale prediction of wind power generation

Wind power generation is significantly influenced by environmental factors, such as wind, temperature, humidity, pressure, precipitation, even lightning, may impact on the power generation, directly or indirectly. To properly deal with these challenges, a multiple time-scale prediction method is proposed, taking the following two steps within different time-scales: (1) A long time-scale coarse prediction based on historical data with more than 6 hours ahead prediction; (2) A short time-scale correction based on real-time weather data with less than 6 hours ahead prediction [6]. The framework of a two-step prediction for wind power generation is illustrated in Fig.1. The long-term and short-term (or real-time) time-scale here are both two relative definitions, the long term time-scale can be one month, one week or one day, and short-term time-scale can be one week, one day or real-time information. In the simulation section, 6-hour interval is taken as short time-scale and one-day interval is taken as long time-scale. Since it is not enough for the prediction results obtained with merely historical data, it requires more detailed real-time weather information to improve the prediction quality. A reliable wind power generation prediction depends heavily on the weather data, but a reliable prediction period with a high confidence level can be very short. To adequately use the weather data to correct the coarse prediction results with deep reinforcement learning method, this paper proposes a two-step prediction approach, and the details are elaborated on in the following section.

III. Long time-scale prediction with the CEEMDSPKF approach

A. *Sigma point based Kalman filter approach*

Generally speaking, the wind power output is largely a function of the wind speed, wind direction, effective contact area with the wind turbine blades and the wind turbines' wake. Since above factors can be formulated as a time sequence process, wind power generation can also be considered as a time sequence process, the wind power at next period can be estimated with previous power information, which can be expressed with discrete nonlinear dynamic system as follows:

$$\begin{cases} x_{t+1} = F(x_t, u_t) + v_t \\ y_t = H(x_t) + n_t \end{cases} \tag{1}$$

where x_t represents the system state at the t time instant, u_t, y_t denote the input and output vector at the tth time instant, and v_t, n_t are the input noise and measurement noise respectively, and $F(\cdot)$, $H(\cdot)$ are some non-linear functions. The above nonlinear dynamic system can be approximated by a linear system, the extended Kalman filter is a popular method by linearizing the state equation and observation equation with state estimations. However, it has two drawbacks: (1) the complexity of deducing the Jacobian matrix makes it difficult to be applicable to real-life cases; (2) If the sampling interval is not short enough, the linearization can generate large deviations, which makes the dynamic system unstable. To overcome these problems, a sigma point Kalman filter is used to capture the statistical values of the selected sigma points, and to linearize the dynamic system by a weighed statistical linear regression approach to capture smaller truncation errors. As it is introduced in literature [6], Kalman filter model can also be described with linear sequence model, it can also be employed to update CEEMD sequence and thus predict wind power sequence recursively. For a given dynamic system, the sigma point based Kalman filter needs to find sigma points before making predictions, differing from other Kalman filters. This approach is used to obtain the statistical values arising from the nonlinear systems. In particular, a weighted statistical linear regression is used to identify the sigma points. For a given sample set x_0 , the sigma points can be calculated as [27]:

$$\begin{cases} \chi_{k,i} = \hat{x}_{k-1} + (\sqrt{(I+\lambda)P_{k-1}})_i, i = 1, 2, \ldots, I \\ \chi_{k,i} = \hat{x}_{k-1} - (\sqrt{(I+\lambda)P_{k-1}})_{i-I}, i = I+1, I+2, \ldots, 2I \end{cases} \tag{2}$$

Where $\chi_{k,i}$ represents the ith sigma point at the tth step ($\chi_{k,i} = \hat{x}_{k-1}$ especially when $i = 0$), $\hat{x}_{k-1}$ is the estimated value at the $k-1$th step (especially $\hat{x}_0 = E[x_0]$), I is the dimension of x_0, P_{k-1} is the covariance matrix of $\hat{x}_{k-1}$, $(\sqrt{(I+\lambda)P_{k-1}})_i$ represents the ith column of the square root of matrix $(I+\lambda)P_{k-1}$, and $\lambda = \alpha^2(I+\beta) - I$, where α is a distribution statistic of those sigma points, which is taken in $0.0001 \leq \alpha \leq 1$, and β is a scalar parameter, it is generally taken in [3, I]. The sigma point based Kalman filter takes the mean value and variance as two representative statistical measures, and it approximates the mean value, extreme value and variance to avoid boundary effect of following CEEMD process at certain degree.

B. The proposed CEEMDSPKF approach

On the basis of the above sigma points, empirical mode decomposition can be utilized to extract the embedded characteristics from the original data, and it has an excellent self-adaptability for prediction, while independent of subjective experience [28]. It has shown to be an efficient way for mining non-linear and non-stationary data and it is also easy to implement. The main procedure of empirical mode decomposition is to decompose the data/sequence into several intrinsic mode function components. Each IMF is defined as a function that satisfies the following requirements: 1) In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by one. 2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. By definition, an intrinsic mode function is any function with the same number of extrema and zero crossings, whose envelopes are symmetric with respect to zero. For a given sequence, the sequence must have at least one maximum and minimum value, and the data between these extremes cannot be invalid. If the sequence can be differentiable, the maximum and minimum value can be deduced with differentiable information. When the above conditions are properly satisfied, the sequence can be represented by the summation of several intrinsic mode functions and residue components. The extraction of intrinsic mode functions generally takes following steps:

Step 1: Check the condition of the original data/sequence $\chi_0(t)$, ensure that the decomposition conditions can be satisfied, set $l = 1$ and $j = 1$, define $\chi_{11}(t)$ as the original set $\chi_0(t)$;

Step 2: Identify all the maximum and minimum values, and use the cubic spline interpolation to create the lower and upper envelope, which can be represented as $a_{jl}^{low}(t)$, $a_{jl}^{up}(t)$. Then, calculate the mean value:

$$m_{jl}(t) = \frac{a_{jl}^{low}(t) + a_{jl}^{up}(t)}{2} \tag{3}$$

Step 3: Calculate component $h_{jl}(t)$ as follows:

$$\chi_{jl}(t) - m_{jl}(t) = h_{jl}(t) \tag{4}$$

Step 4: Use the criterion (5) to judge whether $h_{jl}(t)$ is an intrinsic mode function, if $sd(j,l) \in [0.2, 0.3]$ is satisfied [29], then $h_{jl}(t)$ is the jth intrinsic mode function, and if $j < N, \chi_{j+1,l}(t) = \chi_{jl}(t) - h_{jl}(t)$, $j = j+1$, go to **Step 2;** if $j \geq N$, go to **Step 5.** If it is not satisfied, $h_{j,l+1}(t) = h_{jl}(t) - m_{jl}(t)$, $l = l+1$, go to **Step 2.**

$$sd(j,l) = \sum_{t=0}^{T} \frac{|h_{jl}(t) - h_{j,l+1}(t)|^2}{h_{j,l+1}^2(t)} \tag{5}$$

Step 5: If $j \geq N$, then all the intrinsic mode functions have been identified, and the residue component $r(t)$ can be calculated as:

$$r(t) = \chi_0(t) - \sum_{j=1}^{N} h_j(t) \tag{6}$$

Based on the above extracted IMFs, the state variable of each IMF can be estimated as follows:

Step 6: Initialize $k = 1$, $h_{j,0} = h_j$;

Step 7: Calculate the following parameters:

$$h_{j,k|k-1} = F\left(h_{j,k-1}, u_{j,k-1}\right) \tag{7}$$

$$\hat{h}_{j,k}^{-} = \sum_{i=0}^{2I} \omega_i^{(m)} \left(h_{j,k|k-1}\right)_i \tag{8}$$

$$P_k^{-} = \sum_{i=0}^{2I} w_i^{(c)} \left[\left(h_{j,k|k-1}\right)_i - \hat{h}_{j,k}^{-}\right] \left[\left(h_{j,k|k-1}\right)_i - \hat{h}_{j,k}^{-}\right]^T + R_v \tag{9}$$

where R_v is the covariance matrix of input noise v_k , $\hat{h}_{j,k}^{-}$ and P_k^{-} are the covariance parameters of $\hat{h}_{j,k}$ and P_k, $\omega_i^{(m)}$ is the weight value which can be obtained by:

$$\omega_0^{(m)} = \frac{\lambda}{I+\lambda} \tag{10}$$

$$\omega_0^{(c)} = \frac{\lambda}{I+\lambda} + \left(1 - \alpha^2 + \gamma\right) \tag{11}$$

$$\omega_i^{(m)} = \omega_i^{(c)} = \frac{1}{2(I+\lambda)}, i = 1, 2 \ldots, 2I \tag{12}$$

$$\lambda = \alpha^2(I+\beta) - I \tag{13}$$

Step 8: Calculate P_k^{-} according to (7), (8) and (9), while the predicted output $\hat{y}_k^{-}$ can also be derived according to the output parameter $Y_{k|k-1}$ as follows:

$$Y_{k|k-1} = H\left[h_{j,k|k-1}\right] \tag{14}$$

$$\hat{y}_k^{-} = \sum_{i=0}^{2I} w_i^{(m)} \left(Y_{k|k-1}\right)_i \tag{15}$$

Step 9: Calculate the following covariance matrices:

$$\hat{P}_{y_k y_k} = \sum_{i=0}^{2I} w_i^{(c)} \left[\left(Y_{k|k-1}\right)_i - \hat{y}_k^{-}\right] \left[\left(Y_{k|k-1}\right)_i - \hat{y}_k^{-}\right]^T + R_n \tag{16}$$

$$\hat{P}_{x_k y_k} = \sum_{i=0}^{2l} w_i^{(c)} \left[\left(\chi_{k|k-1}\right)_i - \hat{x}_k^{-}\right] \left[\left(Y_{k|k-1}\right)_i - \hat{y}_k^{-}\right]^T \tag{17}$$

where R_n is the covariance matrix of the measurement noise n_k.

Step 10: Update the gain factor and estimation with the covariance parameter K_k:

$$K_k = \hat{P}_{h_{j,k} y_k} \hat{P}_{y_k y_k}^{-1} \tag{18}$$

$$\hat{h}_{j,k} = \hat{h}_{j,k}^{-} + K_k\left[y_k - \hat{y}_k^{-}\right] \tag{19}$$

Step 11: Update the covariance matrix P_k, if $k < maxcount1$ (where $maxcount1$ denotes the maximum iteration number), $k = k+1$ and go to **Step 7.**

$$P_k = P_k^{-} + K_k \hat{P}_{y_k y_k} K_k^T \tag{20}$$

The complementary ensemble empirical mode decomposition approach is then used to solve the mode mixing problem by homogenizing its scale in the time-frequency space with added noise. The principle of the complementary ensemble empirical mode decomposition approach can automatically project the composed components in different scales onto an appropriate uniform reference frame by adding positive and negative white noise, which can properly deal with the mode mixing problem in a certain degree.

(1) Before mode decomposition, add positive and negative white noise to the calculated sigma data as follows:

$$\begin{cases} S_i(t)^{+} = S(t) + n_i(t) \\ S_i(t)^{-} = S(t) - n_i(t) \end{cases} \tag{21}$$

(2) Combined with formula (3)-(6), decompose the processed data $S_i(t)^{+}$, $S_i(t)^{-}$ into the summation of intrinsic mode functions: $IMF_{i,j}^{+}$, $IMF_{i,j}^{-}$ and residue component with Ne trials, it can be described as:

$$IMF_j(t) = \frac{1}{2Ne} \sum_{i=1}^{Ne} \left(IMF_{i,j}^{+}(t) + IMF_{i,j}^{-}(t)\right) \tag{22}$$

where the positive intrinsic mode functions and negative intrinsic mode functions are taken together to recover the information of the targeted data.

(3) Update the above IMFs with a sigma point based Kalman filter, and reconstruct the original data with the above positive and negative intrinsic mode functions, which can be generally presented as follows:

$$S'(t) = \sum_{j=1}^{N} IMF_j(t) \tag{23}$$

where the obtained data $S'(t)$ can be reconstructed to approximate the targeted data.

C. The CEEMDSPKF approach for coarse wind power prediction

Since empirical mode decomposition can transform the sequence into several intrinsic mode functions with steady state features together with a residual component, the sigma point Kalman filter can then be effectively used to conduct non-linear prediction, and if the sequence like the intrinsic mode functions has a steady state feature, the predicted results can be more accurate. Due to nonlinear and intermittent characteristics of wind power generation, this paper proposes the elaborated CEEMDSPKF method for wind power prediction (See Algorithm 1). The complementary ensemble empirical mode decomposition decomposes the original sequence into several steady sequences, and then the decomposed sequences are used to reconstruct the original data for wind power prediction based on the extended Kalman filter, which can be seen in Fig.2.

IV. SHORT TIME-SCALE CORRECTION WITH A DEEP REINFORCEMENT LEARNING APPROACH

Since the long time-scale approach can merely coarsely predict the power generation process, those predicted results cannot be accurate enough for further scheduling. With consideration of the real-time weather's impact on wind power generation, the DDPG approach is employed to learn the wind power generation process to correct those coarse

Algorithm 1

procedure The CEEMDSPKF algorithm for long time-scale coarse prediction
    Select sigma points with formula (2), $count = 1$;
    Fit those sigma points with spline interpolation;
    while $sd(j,l) \notin [0.2, 0.3]$ and $count < maxcount1$ **do**
        Decompose interpolated curve with formula (3),(4),(5),(6);
        Calculate different IMFs and residue;
        $count = count + 1$;
    end while
    Update IMFs by formula (19) with those updated parameters in formula (16),(17),(18) and (20);
    Reconstruct the data with formula (21),(22) and (23);
end procedure

prediction results. Some definitions can be presented as follows:

Definition 1 (State set): The weather information $V(t) = [Tem(t), Spd(t), Dir(t), Hum(t)]$ is defined as state set, where $Tem(t)$,$Spd(t)$, $Dir(t)$ and $Hum(t)$ represent the temperature, speed, wind direction and humidity variables, and then it can be utilized to obtain a corrected prediction result $S'^{*}(t)$.

Definition 2 (Action set): The deviation $S'^{*}(t) - S'(t)$ between the corrected set and the coarse prediction set is taken as action set $A(t)$, which can be deduced with weather information.

Definition 3 (Reward set): The correction error $\delta(t)$ can be described as $|S^{*}(t) - S'^{*}(t)|$ ($S^{*}(t)$ represents real data), and the reward set $R(V, A)$ can be defined as $-\delta(t)$.

Definition 4 (Optimal state-action value): The optimal state-action value function $Q^{*}(V, A)$ can also be expressed as the maximum value of $Q(V, A)$ with the Bellman theory as follows:

$$Q^{*}(V', A') = \max_{A'}[R(V, A) + \eta_V Q^{*}(V, A)] \tag{24}$$

where V' and A' represent the next state of V and A, and $\eta_V \in [0, 1)$ denotes the discount factor. The actor-critic network learns the relationship between weather information and wind power with training network weights θ_μ and θ_Q, and two target networks are also copied to calculate the target value $z_i = R(V^{(i)}, A^{(i)}) + \eta_{V^i} Q(V^{(i)}, A^{(i)})$ and on-line approximation loss $L = 1/N_Q \sum_{i \in N_Q}(Q(V^{(i)}, A^{(i)}) - z_{(i)})^2$, where N_Q denotes the sample number. For seeking the optimal value, a policy gradient approach is utilized to update the actor policy as follows:

$$\bigtriangledown J = 1/N_Q \sum_{i \in N_Q} \bigtriangledown_A Q(V^{(i)}, A^{(i)}) \bigtriangledown_{\theta_\mu} \mu(V^{(i)}) \tag{25}$$

where $\mu(V^{(i)}) = argmax_A Q(V^{(i)}, A^{(i)}) + \xi_i$, ξ_i represents the designed noise for avoiding local convergence. After the

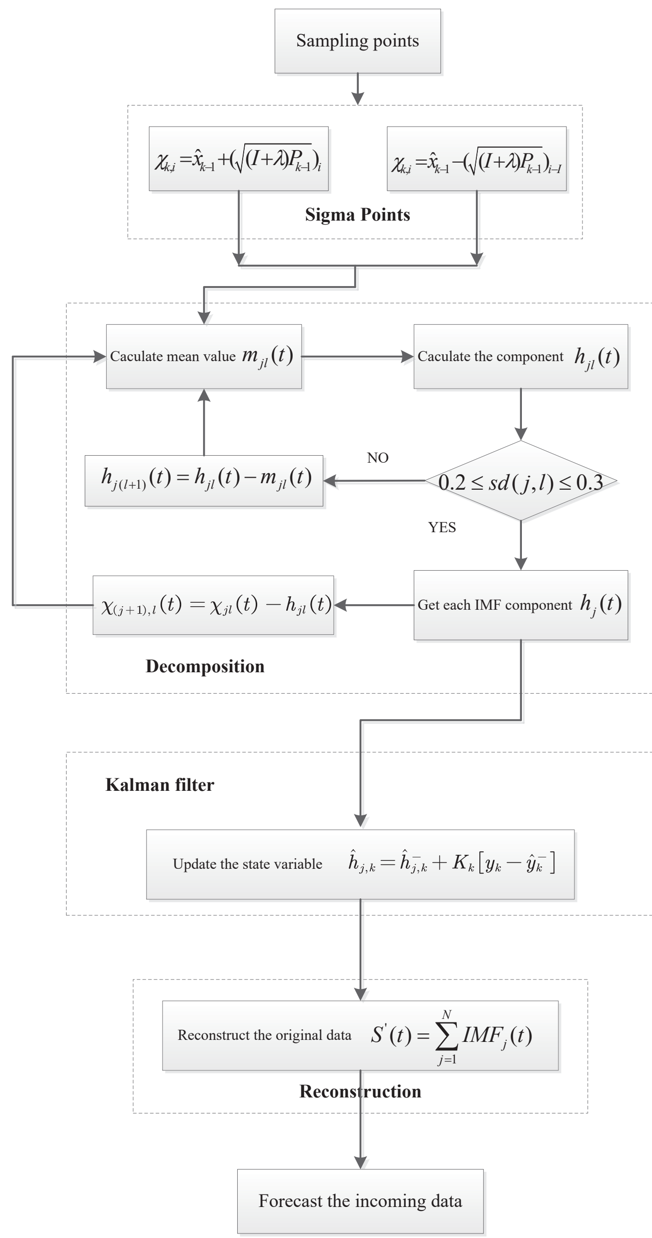

Fig. 2. The flowchart of CEEMDSPKF approach

above procedures, two network weights θ_μ and θ_Q can be updated as follows:

$$\begin{cases} \theta'_\mu = \tau\theta_\mu + (1-\tau)\theta'_\mu \\ \theta'_Q = \tau\theta_Q + (1-\tau)\theta'_Q \end{cases} \tag{26}$$

where $\tau \in [0, 1)$ represents updating control parameter. Then the optimal weights can further deduce the predicted results combined with weather information, the flowchart is shown in Fig.3, and the procedures of whole algorithm is presented in **Algorithm 2**, where $maxcount2$ denotes the maximum iteration number.

Algorithm 2

procedure The DDPG algorithm for short time-scale correction
Initialization: Collect weather information V and coarse prediction result S', initialize critic network Q and actor network μ,, and their copied target networks $Q^{'}$ and $\mu^{'}$, $ReplayBuffer = \phi$, $episode = 1$;
  while $episode < maxcount2$ **do**
    Calculate action A, R and V';
    Store transition information (V, A, R, V') in $ReplayBuffer$;
    Select random minibatch from N_Q transition samples;
    Set $z_i = \begin{cases} R(V^{(i)}, A^{(i)}), If\ i+1 = N_Q \\ R(V^{(i)}, A^{(i)}) + \eta_{V^i} Q(V^{(i)}, A^{(i)}), else \end{cases}$
    Update critic network by minimizing loss L;
    Update actor network with sampled policy gradient ∇J;
    Update weights $\theta^{\mu^{'}}$ and $\theta^{Q^{'}}$ of two networks;
    $episode = episode + 1$;
  end while
end procedure

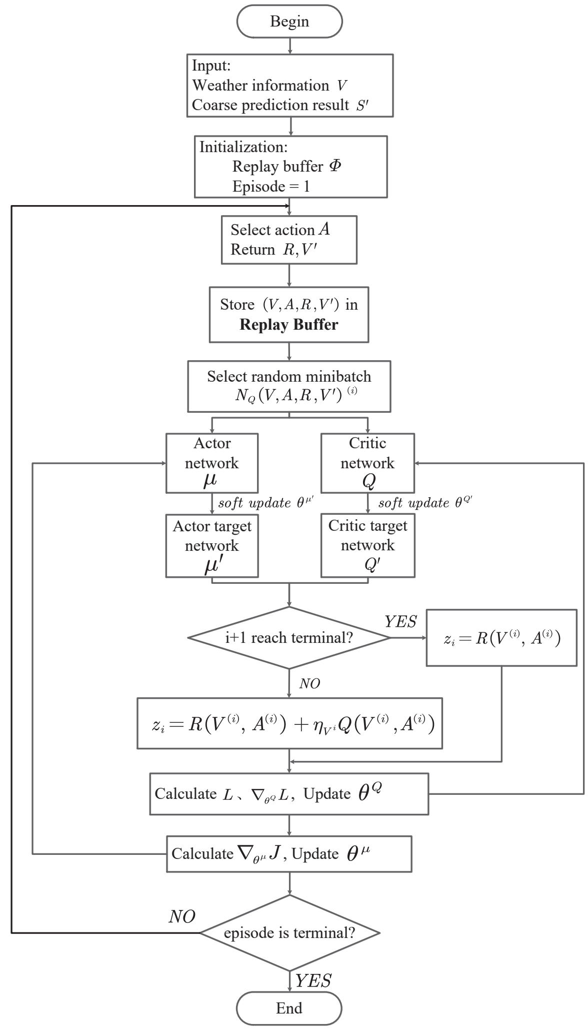

Fig. 3. The flowchart of DDPG approach

V. CASE STUDY

The proposed method is further applied to predict the wind power generation where the whole period is split into two time series with different time-scales: long time series and short time series. Since the effect of weather information on wind farm can be more representative cases than single wind turbine, and some observable information can be easier to obtain from wind farm rather than single wind turbine, wind farm is taken as for analysis, those related data can be found in [30]. Two cases are studied: 1) The ordinary period is selected from May 1st to May 6th of 2014, the historic data from April 25th to April 30th of 2014 is taken for training. 2) Four seasonal periods are selected for representing Spring, Summer, Autumn and Winter, the data of each season is the typical month of the season. The distribution statistic parameter α is set as 0.8, the initial values of covariance matrix P_0 is the identity matrix, the initial covariance matrix R_v and R_n of input noise and measurement noise are both set as identity matrix, and the trials number of decomposition parameter N_e is set as 10. The network weight θ_μ and θ_Q is initially set as 0, the updating control parameter τ is set as 0.5, the designed noise ξ_i is taken as white noise, it can be generated with function $randn$, and discount factor η_V is set as 0.9.

A. Case 1: Two-step prediction approach for ordinary period

one hour interval is taken as coarse forecasting, the whole period can be divided into 125 time periods, and 15 minutes for each interval is considered as short time-scale correction, which means that the whole period is split into 500 time periods for correcting coarse prediction results. This test case is utilized to verify the efficiency of the proposed multiple time-scale prediction approach, the weather information consists of wind speed, air humidity, environmental temperature, wind direction and effective wind speed on the wind turbine blades. Firstly, sigma points are selected from long time-scale historic information, those sigma points can cover all turning points. Then it can decompose into four IMFs and residual results with different frequency levels, which can be seen in Fig. 4. The predicted results with one hour interval are shown in Fig.5, where proposed CEEMDSPKF can be better than ARIMA, CEEMD,complete ensemble empirical mode decomposition-local linear embedding-extreme learning machine (CEEMD-LLE-ELM) and empirical mode decomposition-local linear embedding-improved extreme learning machine (EMD-LLE-IELM) in literature [14], which reveals that the proposed CEEMDSPKF can be superior to other EMD based approaches. Combined with real-time weather information, the predicted wind power in all time-scale with the proposed two-step prediction approach (TPA) can be obtained in Fig.6,

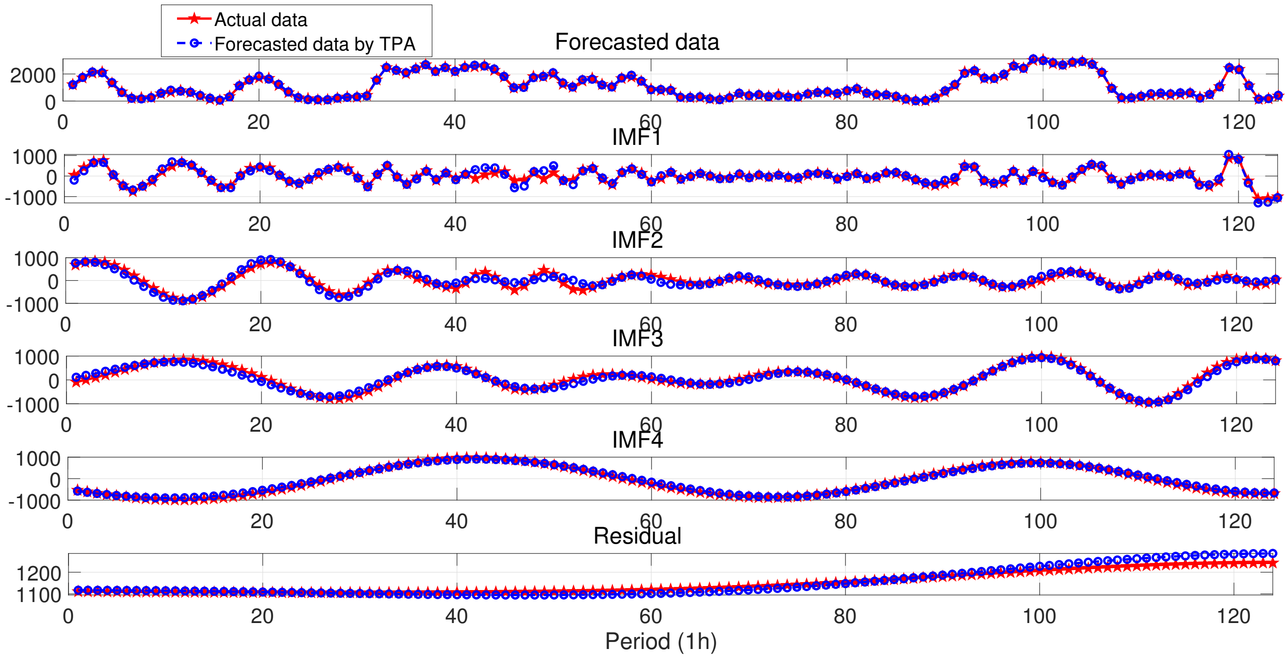

Fig. 4. The obtained IMFs and residual results

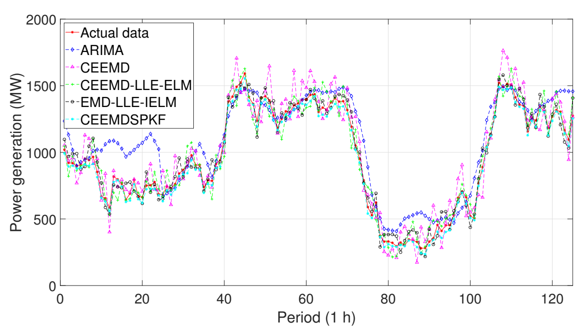

Fig. 5. The obtained results in comparison to other alternatives with 1 hour interval

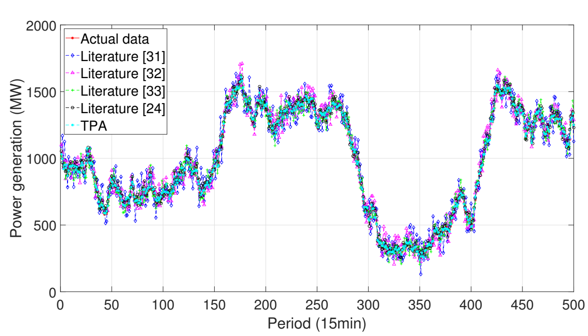

Fig. 6. The obtained results in comparison to other alternatives in all time-scale

TABLE I
THE COMPARISON OF DIFFERENT ALTERNATIVES ON FOUR METRICS WITH 1 HOUR INTERVAL

Methods	MAE	SDE	RMSE	MAPE
ARMA	211.31	101.56	173.25	29.71
ARIMA	203.12	98.43	154.27	25.4173
RBF	193.56	78.77	94.56	20.12
BP	159.25	62.35	79.29	12.1124
CEEMD	164.26	70.74	85.97	14.7611
CEEMD-LLE-ELM	155.17	64.38	85.22	12.85
EMD-LLE-IELM	151.25	61.75	83.75	12.93
CEEMDSPKF	144.76	60.32	81.77	12.79

where those methods in [31], [32], [33], [24] are taken for verifying the priority of the proposed TPA. It can be seen that the predicted results by the multiple time-scale prediction method converge to real data better in comparison to other alternatives. Here, the mean absolute error (MAE), standard deviation of error (SDE), root mean squared error (RMSE) and mean absolute percent error (MAPE) metrics are utilized to testify the prediction efficiency. All those obtained results are classified into two parts: long time-scale and all time-scale. The obtained results with one hour interval are shown in Table I, where the unit of MAPE is %, CEEMD-LLE-ELM and EMD-LLE-IELM are also taken for comparison, it can see that those traditional methods such as ARMA, ARIMA, RBF and BP have worse performance, and CEEMDSPKF has better prediction results in comparison to other improved EMD methods. The obtained results with all time-scale information are presented in Table II, where it can be seen that TPA has better performance on four metrics in comparison to other alternatives in literatures [31], [32], [33], [24], [34].

TABLE II
THE COMPARISON OF DIFFERENT ALTERNATIVES ON FOUR METRICS IN ALL TIME-SCALE

Methods	MAE	SDE	RMSE	MAPE
Literature [31]	111.37	54.77	62.54	12.11
Literature [32]	99.32	49.28	58.27	11.59
Literature [33]	85.76	41.35	49.92	10.26
Literature [24]	81.39	37.22	44.76	9.98
Literature [34]	73.46	36.58	38.93	9.14
TPA	53.96	30.783	34.3332	8.4152

B. Case 2: Two-step prediction approach for seasonal periods

The four seasonal sets of data of wind power generation has been taken to test the forecasting accuracy, the data of January is taken as representative data of Spring, April is taken as the representative data of Summer, July is taken as the representative data of Autumn, and October is taken as the representative data of Winter. A six-hour interval is taken as short time-scale and a one-day interval is taken as long time-scale, short time-scale information is taken to correct the predicted data of long time-scale, the analysis and comparison are taken in both long time-scale and all time-scale, further analysis can be done on those results in Spring, Summer, Autumn and Winter. In Fig.7, it shows that ARIMA, BP and RBF has worse performance than other alternatives, and CEEMDSPKF can approximate to actual data better than other alternatives. In all time-scale, comparisons with other alternatives in literatures [31], [32], [33], [24] are taken in Fig.8, it can see that all predicted results are better than those results in mere long time-scale, and the proposed TPA is also better than those existing alternatives. For better comparison with global prediction efficiency, the obtained results of the whole year are presented in Fig.9 and Fig.10. In long time-scale, CEEMDSPKF has less prediction on MAE, SDE, RMSE and MAPE metrics in comparison to other alternatives. In comparison to alternatives in literatures [31], [32], [33], [24], the proposed TPA has better accuracy than other alternatives on MAE, SDE, RMSE and MAPE metrics. The comparison of those results can be found in Table III and Table IV, it can be seen that the performance on four metrics can be ordered as BP<ARMA<RBF<ARIMA<CEEMD<CEEMD-LLE-ELM<EMD-LLE-IELM<CEEMDSPKF in long time-scale, and the proposed TPA has better performance than other methods in the whole year.

VI. CONCLUSION

Considering the strong stochastic characteristic of intermittent wind power generation, a two-step prediction method with multiple time-scales is proposed to improve the prediction performance. The proposed prediction approach consists of two phases: the long time-scale coarse prediction and short time-scale correction. The CEEMDSPKF approach can better predict wind power generation in comparison to other improved EMD approaches, and the proposed TPA can be superior to other two-step methods. The results can reveal that the proposed two-step prediction method can offer a promising way for predicting wind power generation.

REFERENCES

[1] J. Shi, Z. Ding, W. Lee, Y. Yang, Y. Liu, and M. Zhang, “Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 521–526, 2014.

[2] C. Peng, S. Lei, Y. Hou, and F. Wu, “Uncertainty management in power system operation,” *CSEE Journal of Power and Energy Systems*, vol. 1, no. 1, pp. 28–35, 2015.

[3] J. Zhao, Z. H. Guo, Z. Y. Su, Z. Y. Zhao, X. Xiao, and F. Liu, “An improved multi-step forecasting model based on wrf ensembles and creative fuzzy systems for wind speed,” *Applied Energy*, vol. 162, no. JAN.15, pp. 808–826, 2016.

[4] M. Koivisto, J. Seppanen, I. Mellin, J. Ekstrom, J. Millar, I. Mammarella, M. Komppula, and M. Lehtonen, “Wind speed modeling using a vector autoregressive process with a time-dependent intercept term,” *International Journal of Electrical Power & Energy Systems*, vol. 77, no. May, pp. 91–99, 2016.

[5] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, “Arima-based time series model of stochastic wind power generation,” *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 667–676, 2010.

[6] K. Chen and J. Yu, “Short-term wind speed prediction using an unscented kalman filter based state-space support vector regression approach,” *Applied Energy*, vol. 113, no. jan., pp. 690–705, 2014.

[7] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, “A new fuzzy-based combined prediction interval for wind power forecasting,” *IEEE Transactions on Power Systems*, pp. 1–9, 2015.

[8] Z. Men, E. Yee, F. S. Lien, D. Wen, and Y. Chen, “Short-term wind speed and power forecasting using an ensemble of mixture density neural networks,” *Renewable Energy*, vol. 87, no. MAR.PT.1, pp. 203–211, 2016.

[9] L. Zjavka, “Wind speed forecast correction models using polynomial neural networks,” *Renewable Energy*, vol. 83, pp. 998–1006, 2015.

[10] L. Yang, M. He, J. Zhang, and V. Vittal, “Support-vector-machine-enhanced markov model for short-term wind power forecast,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 791–799, 2017.

[11] C. Ren, N. An, J. Wang, L. Li, B. Hu, and D. Shang, “Optimal parameters selection for bp neural network based on particle swarm optimization: A case study of wind speed forecasting,” *Knowledge-Based Systems*, vol. 56, no. jan., pp. 226–239, 2014.

[12] J. Yan, K. Li, E. W. Bai, J. Deng, and A. M. Foley, “Hybrid probabilistic wind power forecasting using temporally local gaussian process,” *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 87–95, 2017.

[13] G. J. Osório, J. C. O. Matias, and J. P. S. Catal. O, “Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information,” *Renewable Energy*, vol. 75, no. mar., pp. 301–307, 2015.

[14] C. Zhang, M. Ding, W. Wang, R. Bi, L. Miao, H. Yu, and L. Liu, “An improved elm model based on ceemd-lzc and manifold learning for short-term wind power prediction,” *IEEE Access*, vol. 7, pp. 121 472–121 481, 2019.

[15] O. Abedinia, M. Lotfi, M. Bagheri, B. Sobhani, M. Shafie-khah, and J. P. S. Catalão, “Improved emd-based complex prediction model for wind power forecasting,” *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2790–2802, 2020.

[16] M. A. Ghorbani, R. Khatibi, M. H. Fazelifard, L. Naghipour, and O. Makarynskyy, “Short-term wind speed predictions with machine learning techniques,” *Meteorology and Atmospheric Physics*, vol. 128, no. 1, pp. 57–72, 2016.

[17] Q. Hu, R. Zhang, and Y. Zhou, “Transfer learning for short-term wind speed prediction with deep neural networks,” *Renewable Energy*, vol. 85, no. JAN., pp. 83–95, 2016.

[18] M. Khodayar, O. Kaynak, and M. E. Khodayar, “Rough deep neural architecture for short-term wind speed forecasting,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 2770–2779, 2017.

[19] M. Cui, D. Ke, D. Gan, and Y. Sun, “Statistical scenarios forecasting method for wind power ramp events using modified neural networks,” *Journal of Modern Power Systems and Clean Energy*, vol. 3, no. 3, pp. 371–380, 2015.

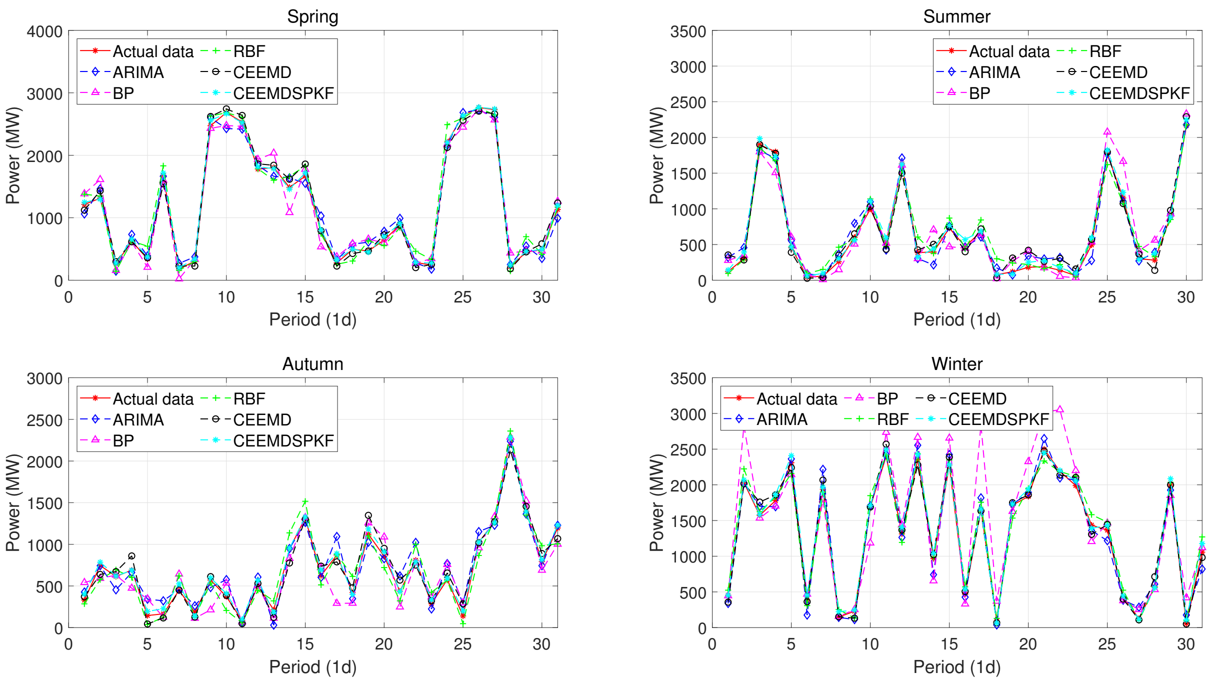

Fig. 7. The predicted seasonal results in comparison to other alternatives in long time-scale

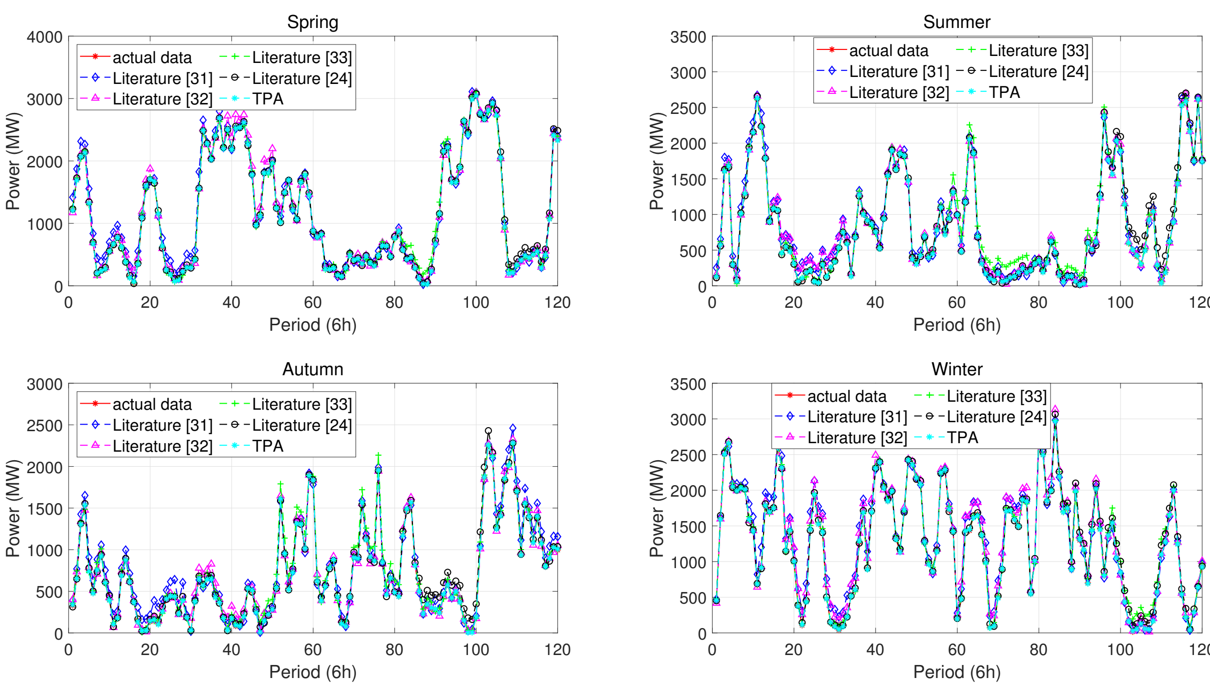

Fig. 8. The predicted seasonal results in comparison to other alternatives in all time-scale

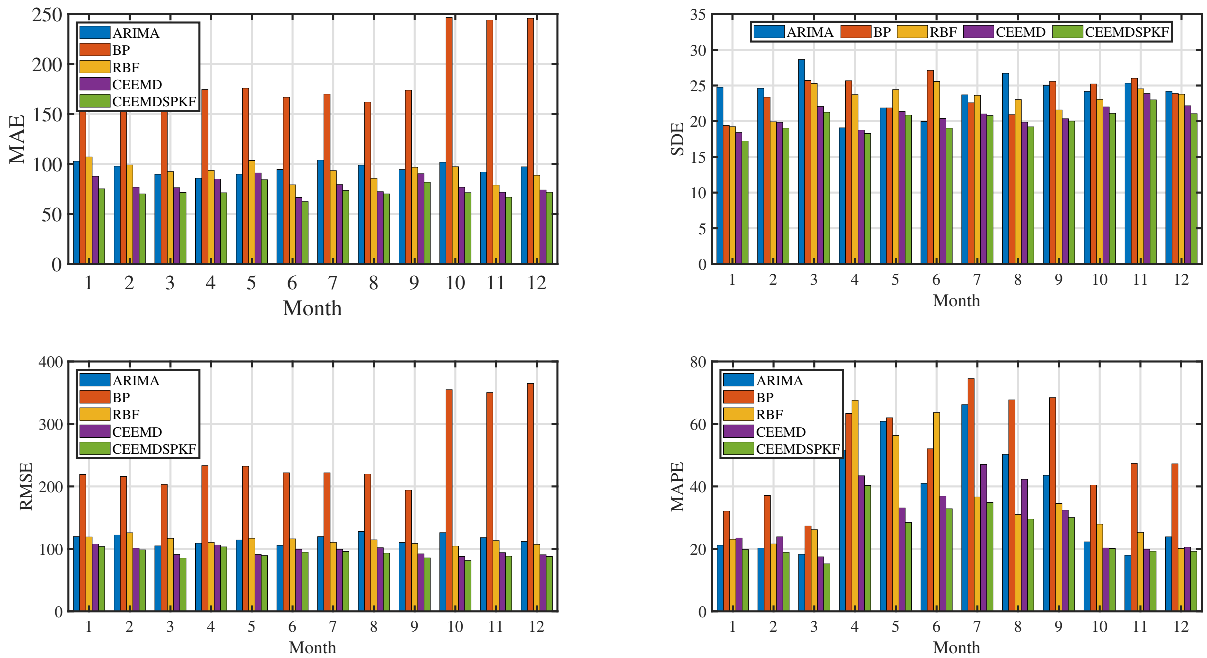

Fig. 9. The prediction results of a one-year period on four metrics in long time-scale

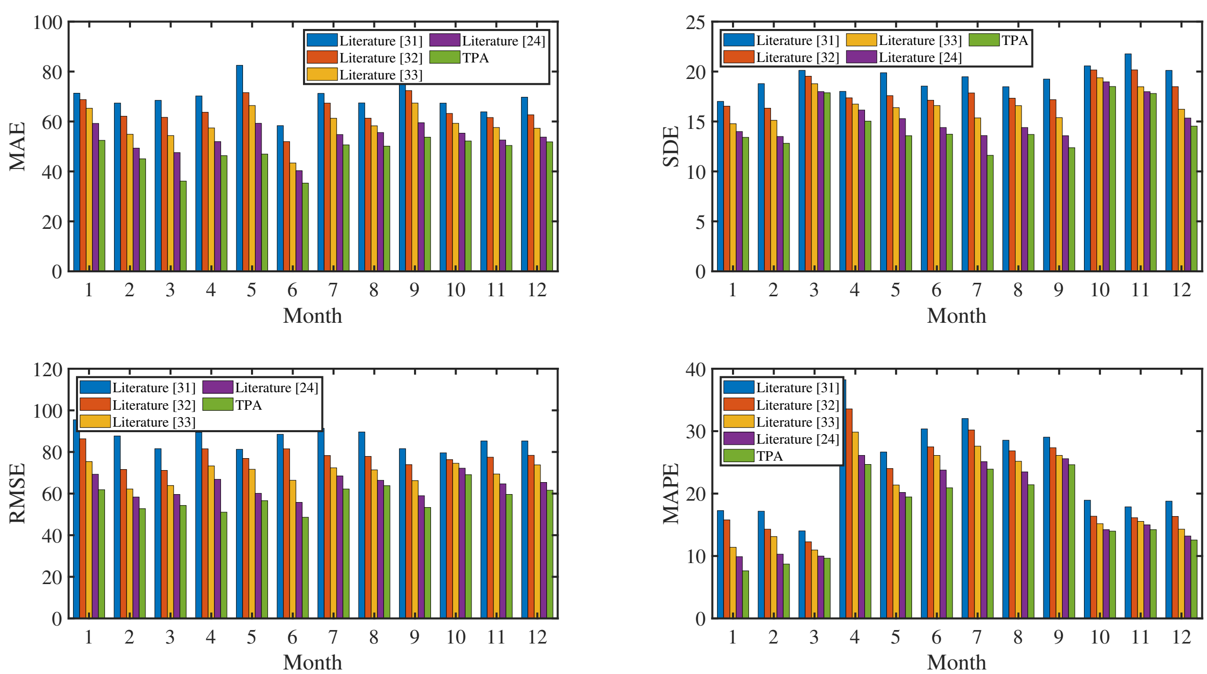

Fig. 10. The prediction results of a one-year period on four metrics in all time-scale

TABLE III
THE COMPARISON OF SEASONABLE RESULTS ON FOUR METRICS IN LONG TIME-SCALE

Seasons	Metrics	ARMA	ARIMA	BP	RBF	CEEMD	CEEMD-LLE-ELM	EMD-LLE-IELM	CEEMDSPKF
Spring	MAE	103.5517	99.1467	161.7125	99.5328	80.3325	69.9392	63.4836	58.1264
	SDE	13.9	12.113	11.701	19.2155	17.4001	16.6532	15.2756	14.8879
	RMSE	125.0485	117.4678	212.7866	120.601	100.0168	90.1383	88.603	80.3785
	MAPE	23.9011	22.1992	32.18576	23.61708	21.6106	18.5863	16.2749	15.5362
Summer	MAE	96.2291	93.2281	172.441	92.1358	80.8852	67.2865	65.0093	58.2219
	SDE	17.4615	15.2674	15.4424	15.1044	14.9973	14.7287	14.1178	13.9356
	RMSE	119.1673	110.1766	229.1526	114.4996	98.9329	87.0549	85.6783	79.3811
	MAPE	55.1463	50.3771	59.1355	62.5189	37.8256	35.1175	34.1298	32.4672
Autumn	MAE	105.4868	97.3812	165.1092	91.9658	80.7007	69.2793	65.3851	63.5325
	SDE	22.9177	20.3376	21.7765	18.5576	19.3287	18.8543	18.3321	17.6322
	RMSE	128.6471	119.7954	211.9081	111.1505	97.7741	93.2937	89.5859	81.7677
	MAPE	53.1157	50.1756	40.2398	32.1179	20.1876	17.3376	17.5622	17.0156
Winter	MAE	103.9573	90.2761	245.5414	88.3632	74.2191	70.923	66.2385	62.7455
	SDE	17.3329	15.0036	16.3477	14.9975	12.7643	12.5123	12.3376	12.2137
	RMSE	128.0657	111.3782	356.596	108.3483	90.8024	88.3214	84.6279	80.6599
	MAPE	25.3495	19.2888	45.0265	24.4516	17.5925	16.2761	16.1188	15.6653

TABLE IV
THE COMPARISON OF SEASONABLE RESULTS ON FOUR METRICS IN ALL TIME-SCALE

Seasons	Metrics	Literature [31]	Literature [32]	Literature [33]	Literature [24]	Literature [34]	TPA
Spring	MAE	55.5536	52.1276	51.0036	48.2756	46.4583	44.543
	SDE	14.5671	14.0219	13.7655	13.1793	11.8766	11.2122
	RMSE	77.9615	71.3368	68.5987	65.1195	60.8799	56.3191
	MAPE	14.3674	13.2874	11.4769	10.0285	9.2379	8.66679
Summer	MAE	56.1785	52.1996	48.8754	46.1189	43.2258	42.8713
	SDE	13.5342	13.0176	12.7549	12.1349	11.9887	11.4988
	RMSE	71.6588	68.2975	61.7699	60.3896	57.9983	52.1073
	MAPE	31.7689	30.0786	28.1187	27.5789	26.7543	25.0225
Autumn	MAE	61.2786	60.1876	58.4988	57.3986	52.1199	51.4962
	SDE	17.2987	17.0019	16.6165	16.5519	14.9998	14.665
	RMSE	77.7653	75.3987	69.8909	66.4895	62.3355	59.7863
	MAPE	16.8864	16.2668	15.0375	14.8879	13.8766	13.3258
Winter	MAE	62.0193	61.3354	57.8476	55.3998	52.2785	51.4824
	SDE	12.0111	11.9674	11.8077	11.6734	11.6688	11.6983
	RMSE	78.3765	75.2788	74.2987	68.4989	64.8695	63.442
	MAPE	15.2876	14.7654	14.2297	13.8657	13.7799	13.5982

[20] S. Buhan, Y. Özkazanç, and I. Çadırcı, "Wind pattern recognition and reference wind mast data correlations with nwp for improved wind-electric power forecasts," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 991–1004, 2016.

[21] J. Liu, H. Gao, Z. Ma, and Y. Li, "Review and prospect of active distribution system planning," *Journal of Modern Power Systems and Clean Energy*, vol. 3, no. 4, pp. 457 – 467, 2015.

[22] S. Buhan and I. Çadırcı, "Multistage wind-electric power forecast by using a combination of advanced statistical methods," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1231–1242, 2015.

[23] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and ga-bp neural network method," *Renewable Energy*, pp. 629–636, 2016.

[24] O. Abedinia, M. Bagheri, M. S. Naderi, and N. Ghadimi, "A new combinatory approach for wind power forecasting," *IEEE Systems Journal*, vol. 14, no. 3, pp. 4614–4625, 2020.

[25] P. Du, J. Wang, W. Yang, and T. Niu, "Multi-step ahead forecasting in electrical power system using a hybrid forecasting system," *Renewable Energy*, vol. 122, pp. 533–550, 2018.

[26] Yamin Wang, L. Wu, and Shouxiang Wang, "Challenges in applying the empirical mode decomposition based hybrid algorithm for forecasting renewable wind/solar in practical cases," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[27] F. U. Jian-Guo, X. T. Wang, L. A. Jin, and M. A. Ye, "Sigma-point kalman filter and its application," *Systems Engineering and Electronics*, vol. 27, no. 1, pp. 141–144, 2005.

[28] J. Zheng, J. Cheng, and Y. Yang, "Partly ensemble empirical mode decomposition: An improved noise-assisted method for eliminating mode mixing," *Signal Processing*, vol. 96, no. pt.B, pp. 362–374, 2014.

[29] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *PROCEEDINGS OF THE ROYAL SOCIETY A-MATHEMATICAL PHYSICAL AND ENGINEERING SCIENCES*, vol. 454, no. 1971, pp. 903–995, MAR 8 1998.

[30] M. O'Flaherty, N. Riordan, N. O'Neill, and C. Ahern, "A quantitative analysis of the impact of wind energy penetration on electricity prices in ireland," *Energy Procedia*, vol. 58, no. 58, pp. 103–110, 2014.

[31] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using gaussian processes and numerical weather prediction," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 656–665, 2014.

[32] Q. Xu, D. He, N. Zhang, C. Kang, Q. Xia, J. Bai, and J. Huang, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1283–1291, 2015.

[33] N. Safari, C. Y. Chung, and G. C. D. Price, "Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 590–601, 2018.

[34] P. Du, "Ensemble machine learning-based wind forecasting to combine nwp output with data from weather station," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 2133–2141, 2019.

Huifeng Zhang (M'16) received Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China, in 2013. From 2014 to 2016, he was a Post-Doctoral Fellow with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. From 2017 to 2018, He was granted as visiting research fellow by China Scholarship Council to study in Queen's University Belfast and University of Leeds, UK. He is currently an Associate Professor at the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interest includes electrical power management, optimal operation of power system, distributed optimization, and multi-objective optimization.

Dong Yue (SM'08-F'21)received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1995. He is currently a Professor and the Dean with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China, and also a Changjiang Professor with the Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, China. His current research interests include analysis and synthesis of networked control systems, multiagent systems, optimal control of power systems, and internet of things. He has published over 100 papers in international journals, domestic journals, and international conferences. Prof. Yue is currently an Associate Editor of the IEEE Control Systems Society Conference Editorial Board and the International Journal of Systems Science.

Chunxia Dou (M'18)received the B.S. and M.S. degrees in automation from the Northeast Heavy Machinery Institute, Qiqihaer, China, in 1989 and 1994, respectively, and the Ph.D. degree in Institute of Electrical Engineering from Yanshan University, Qinhuangdao, China, in 2005. In 2010, she joined the Department of Engineering, Peking University, Beijing, China, where she was a Postdoctoral Fellow for two years. Since 2005, she has been a Professor in Institute of Electrical Engineering, Yanshan University. Her current research interests include multi-agent based control, event-triggered hybrid control, distributed coordinated control, and multi-mode switching control and their applications in power systems, Microgrids and smart grids.

Kang Li (M'05–SM'11) received the B.Sc. degree in Industrial Automation from Xiangtan University, Hunan, China, in 1989, the M.Sc. degree in Control Theory and Applications from Harbin Institute of Technology, Harbin, China, in 1992, and the Ph.D. degree in Control Theory and Applications from Shanghai Jiaotong University, Shanghai, China, in 1995. He also received D.Sc. degree in Engineering from Queen's University Belfast, UK, in 2015. He currently holds the Chair of Smart Energy Systems at the University of Leeds, UK. His research interests cover nonlinear system modelling, identification, and control, and artificial intelligence, with substantial applications to energy and power systems, smart grid, electric vehicles, railway systems, and energy management in energy intensive manufacturing processes.

Gerhard P. Hancke (M'88–SM'00–F'16) received the B.Sc. and B.Eng. degrees, in1970, and the M.Eng. degree in Electronic Engineering, in 1973, from the University of Stellenbosch, South Africa, and the Ph.D. degree from the University of Pretoria, South Africa, in 1983. He is a Professor with the University of Pretoria, South Africa and recognized internationally as a pioneer and leading scholar in industrial wireless sensor networks research. He initiated and co-edited the first Special Section on Industrial Wireless Sensor Networks in the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS in 2009 and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS in 2013. He co-edited a textbook, Industrial Wireless Sensor Networks: Applications, Protocols and Standards (2013), the first on the topic. Prof. Hancke has been serving as an Associate Editor and Guest Editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE ACCESS, and previously the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS. Currently, he is a Co-Editor-in-Chief for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.