Wind Power Prediction Based on Multi-class Autoregressive Moving Average Model with Logistic Function

Yunxuan Dong, *Student Member*, *IEEE*, Shaodan Ma, *Senior Member*, *IEEE*, Hongcai Zhang, *Member*, *IEEE*, and Guanghua Yang, *Senior Member*, *IEEE*

***Abstract*—The seasonality and randomness of wind present a significant challenge to the operation of modern power systems with high penetration of wind generation. An effective short-term wind power prediction model is indispensable to address this challenge. In this paper, we propose a combined model, i.e., a wind power prediction model based on multi-class autoregressive moving average (ARMA). It has a two-layer structure: the first layer classifies the wind power data into multiple classes with the logistic function based classification method; the second layer trains the prediction algorithm in each class. This two-layer structure helps effectively tackle the seasonality and randomness of wind power while at the same time maintaining high training efficiency with moderate model parameters. We interpret the training of the proposed model as a solvable optimization problem. We then adopt an iterative algorithm with a semi-closed-form solution to efficiently solve it. Data samples from open-source projects demonstrate the effectiveness of the proposed model. Through a series of comparisons with other state-of-the-art models, the experimental results confirm that the proposed model improves not only the prediction accuracy, but also the parameter estimation efficiency.**

***Index Terms*—Wind power prediction, wind generation, time series analysis, logistic function based classification.**

NOMENCLATURE

$\boldsymbol{\alpha}$	Mapping vector of classification
$\varphi_{k,i}$	Parameter of AR process at time i in the k^{th} class
$\boldsymbol{\varphi}_k$	Matrix form of parameter $\varphi_{k,i}$
$\pi_{k,j}$	Parameter of MA process at time j in the k^{th} class
$\boldsymbol{\pi}_k$	Matrix of parameter $\pi_{k,j}$
ε	Bias of predicting process
$\varepsilon_{k,t-j}$	Prediction error at time $t-j$ in the k^{th} class
$\theta_{k,i}$	Parameter of fitting AR process
$\sigma(\boldsymbol{x})$	Smoothness objective
$\mathbb{I}$	Indicative function of classifier
C_k	Wind power set in the k^{th} class
$\boldsymbol{d}_{x_t}$	Wind power observations in the past
e	Euler's number
$\boldsymbol{E}$	Identity matrix
$\boldsymbol{f}$	Feature mapping function
i	Parameter indicator of autoregressive (AR) process
I	The maximum order of AR process
j	Parameter indicator of moving average (MA) process
J	The maximum order of MA process
k	Class indicator
K	The maximum class number
P_{x_t}	Classifier score of data set
t	Time point indicator
T	The maximum number of time index
v_t	Value of observation error
$\bar{x}_t$	Value of wind power observation
x_t	Actual wind power value
$\hat{x}_t$	Estimated wind power value
$\boldsymbol{x}$	Matrix of wind power generation
$\hat{x}_{k,t}$	Prediction results of the k^{th} class at time t

I. INTRODUCTION

WIND power can provide a strong driving force for the future economic development, and combat energy crisis as well as climate change. Investment in wind power generation experienced an explosive growth in the last decade. By the end of 2020, over 910 GW wind generation was in-

Manuscript received: October 21, 2021; revised: March 2, 2022; accepted: April 26, 2022. Date of CrossCheck: April 26, 2022. Date of online publication: June 16, 2022.

This work was supported by the Guangdong-Macau Joint Funding Project (No. 2021A0505080015), Science and Technology Planning Project of Guangdong Province (No. 2019B010137006), and Science and Technology Development Fund, Macau SAR (No. SKL-IOTSC(UM)-2021-2023).

Y. Dong, S. Ma, and H. Zhang (corresponding author) are with the State Key Laboratory of Internet of Things for Smart City and the Department of Electrical and Computer Engineering, University of Macau, Macao S. A. R. 999078, China (e-mail: dyxiscool@outlook.com; shaodanma@um.edu.mo; hczhang@um.edu.mo).

G. Yang is with the Institute of Physical Internet and the School of Intelligent Systems Science and Engineering, Jinan University, Zhuhai 519070, China (e-mail: ghyang@jnu.edu.cn).

DOI: 10.35833/MPCE.2021.000717

stalled worldwide according to the World Wind Energy Association [1]. However, as wind is stochastic and intermittent, it is challenging to integrate high penetration of wind generation into power systems that require real-time balance between load and generation [2]. To overcome this challenge, highly accurate wind power prediction (WPP) is indispensable, which is currently a research hot spot in both academia and industry [3].

WPP models and estimates the future wind power generation. Based on the prediction time, they are normally divided into four cases, i. e., long-term, medium-term, short-term, and very-short-term predictions [4]. In addition, the short-term case and very-short-term case are the most popular ones in the operation of wind generation farm and corresponding power systems [5], [6]. They are important components in wind farm control [7] and power system operation [8], [9].

In the technical literature, the major models used for WPP can be divided into five categories.

1) Linear models with simple smooth assumption. This kind of models assume the future wind power to be equal to a value in the past [10], e.g., persistence [11]. Linear models with simple smooth assumption are the simplest and most economical models for predicting wind power, so that they are widely adopted by wind farms in history. However, they are only reliable for very-short-term predictions (e.g., a couple of minutes ahead), and when the prediction time range is extended, their prediction accuracy decreases rapidly.

2) Numerical weather models. These models predict wind power based on comprehensive weather parameters (e. g., temperature, pressure, and obstacles) observed from radiosondes, satellites, etc. [12]. They can provide accurate long-term macroscopic weather prediction. The main disadvantage of numerical weather models is the high computational time and the storage complexity of computation. When these models encounter abnormal conditions during the prediction period, the models would collapse. Besides, they are also unreliable for microscopic short-term and very-short-term wind predictions.

3) Statistical models. Statistical models include the support vector machine (SVM) [13], the dynamic models (e.g., autoregressive moving average (ARMA) model) [14], the Bayesian models, etc. [15]. These models need to ensure the data to be stationary. However, because there are randomness and seasonality in the wind power generation, it is hard to apply statistical models to the wind power generation [16]. In the meanwhile, they are difficult to give an accurate estimate for wind power generation with implied volatility due to the scale hypothesis and distribution hypothesis [17]. For example, [18] employed multiple variables to predict tuples of wind power generation with the ARMA model. Although this method can model wind generation more effectively with additional variables, the much more variables greatly improve the difficulty in solving the model.

4) Machine learning (ML) based models. ML-based models are widely used because it is capable of capturing the nonlinear relationship between input data (e.g., historical records) and WPPs. There are shallow architecture in ML-based models for WPP which includes extreme learning machine model [19] and combined ensemble models [20]. Such models are plagued by sample imbalance and huge computational time [21]. Therefore, they usually need complex feature selection methods [22]. Some of the ML-based models employ deep architecture (e.g., recurrent neural networks) to achieve the multi-layer hidden computing units with high generalization ability [23], [24]. It is capable to implement complex feature selection automatically in its own hidden layers.

5) Combined models. The combined models employ the data processing model and the optimization model to jointly improve the performance of WPP [25]. Since the joint optimization goal of combined models is usually non-convex, the parameters of combined models are usually optimized by decoupling the problem [21], [26]. The combined models are also employed to solve the seasonality and randomness in wind power generation. Specially for WPP problems, they are usually necessary to establish multiple statistical models for different conditions or to employ the ML-based models with large parameter scales [27]. The data processing model of combined models is also an important part, which can ensure that the input for prediction model meets the corresponding model assumptions, and the expected data distribution can help the prediction model control the parameter scale [22]. Although the combined models need to solve the parameters of multiple models, they always have a clear mathematical structure [28], therefore, there is always a better model interpretability of combined models than ML-based models [29].

Because a combined model needs to calculate the parameters of multiple models, its training difficulty is generally higher than that of a single model [30], [31]. To reduce the difficulty in training a combined model and ensure the quality of prediction, a simple and effective cost function is necessary. In the meanwhile, effectively optimizing the parameters of the model helps improve prediction efficiency [32]. Reference [33] designed the cost function by weighted combination of the reliability and sharpness measurement and it enhanced the convergence ability of the model. However, its cost function is not differentiable, which makes the optimization algorithm unable to achieve global optimum. Besides, it also relies on computational effectiveness of classification or feature extraction in the data preprocessing model. Some combined models (e.g., [34]) optimize parameters based on non-convex optimization, which are solved by heuristic algorithms. However, these algorithms may fall into local optimum [22], [35]. Non-convexity also causes a prediction model to be highly dependent on the initial parameters so that the results of independent operations are inconsistent [36]. Besides, the parameter scales of heuristic algorithms may explode with the increase of problem complexity [37]. To solve the above problems, [26] proposed a two-layer structure, where the first-layer model is responsible for optimizing the weights of data preprocessing model, so that in the

second layer, a classic algorithm can be used to optimize parameters in the prediction process. However, this model is still highly dependent on the initial parameters [36].

Taking into account the above challenges, this paper proposes a combined WPP model, i.e., the WPP model based on multi-class ARMA. There are two advantages in the proposed model.

1) It employs a two-layer structure to decrease the scale of parameters. In the first layer, the wind power data are classified into multi-class classification based on logistic function. In the second layer, an ARMA-based algorithm is trained for each class of data for prediction. This structure helps effectively tackle seasonality and randomness of wind power with moderate model complexity.

2) We model the WPP problem as a solvable optimization problem, and the non-dominated sorting genetic algorithm II (NSGA-II) is employed to find the optimal parameters. This algorithm effectively guarantees the convergence speed and training efficiency of the training process.

The rest of this paper is organized as follows. In Section II, the system model of the WPP problem is presented, and the proposed cost functions are described. In Section III, the process of proposed model is presented. Section IV conducts the numerical experiments and Section V present the conclusion.

II. System Model of WPP Problem

Taking into account various fluctuations and different types of seasonality and randomness in wind power generation, we propose a WPP model based on multi-class ARMA. The key idea of this model is to divide the data samples into multiple classes based on logistic function, and the corresponding prediction models would be trained independently. Then, in application, the proposed model first classifies the new data, and then estimates the future data based on the corresponding WPP model. The structure of the proposed model is presented in Fig. 1. As shown in Fig. 1(a), the model training consists of two layers, the first layer for classifier training can map the original series into discontinuous statistical features and use logistic function for classification. Then the second layer for model training would train different prediction models based on different classes, if the iteration termination condition is not met, the parameters of the classifier would be updated and the two-layers process would be employed again. After the training process, the achieved classifier and the prediction model would be employed in the WPP process, as shown in Fig. 1(b).

The typical classes of wind power time series are shown in Fig. 2, where the sampling point is obtained every 30 min. Obviously, different classes have quite different wind power profiles.

Without loss of generality, we will use formula derivation to show how the proposed model predicts time series through ARMA model. Mathematically, denoting the vector form of wind power generation as $\boldsymbol{x}=(x_t), t=1,2,...,T$, the system model of WPP problem is given as:

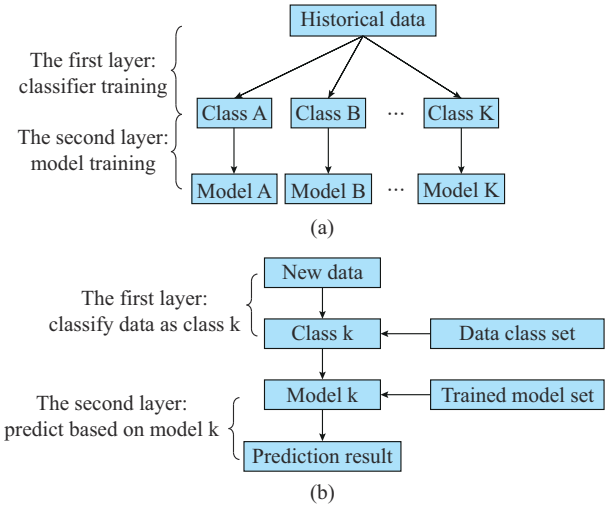

Fig. 1. Structure of proposed model. (a) Model training. (b) WPP.

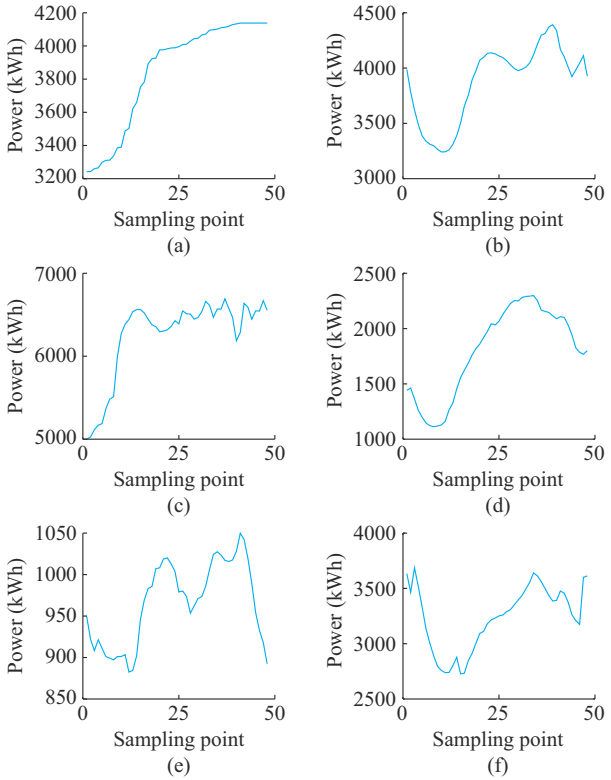

Fig. 2. Typical classes of wind power time series. (a) Typical class 1. (b) Typical class 2. (c) Typical class 3. (d) Typical class 4. (e) Typical class 5. (f) Typical class 6.

$$\hat{x}_t=\sum_{k=1}^{K}\mathbb{I}(\boldsymbol{d}_{x_t}\in C_k)\left(\sum_{i=1}^{I}\varphi_{k,i}x_{t-i}+\sum_{j=1}^{J}\pi_{k,j}\varepsilon_{k,t-j}\right) \tag{1}$$

The two terms in the bracket in (1) correspond to the AR and MA terms in the ARMA model, respectively. The indicative function $\mathbb{I}(\boldsymbol{d}_{x_t}\in C_k)$ models the multi-class classification defined as:

$$\mathbb{I}(\boldsymbol{d}_{x_t}\in C_k)=\begin{cases}1 & \boldsymbol{d}_{x_t}\in C_k\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

In this paper, we employ the logistic function through sta-

tistical feature mapping for multi-class classification problems. In practice, wind power is affected by numerous factors and the complete wind power features for each class C_k are hard to be obtained so that the multi-class classification is quite challenging. The logistic function using statistical feature mapping effectively expands the basis of the classification space. Inspired by this feature, we extend the original logistic classification to multi-class case by adopting the following logistic function:

$$p_{x_t} = \frac{1}{1+\mathrm{e}^{-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{d}_{x_t}}} \tag{3}$$

Similar to original logistic classification, the bounded range [0, 1] is uniformly partitioned into K subregions, i.e., $[0, 1/K)$, $[1/K, 2/K)$, ..., $[(K-1)/K, 1]$. The output of the logistic function p_{x_t} is compared with the thresholds $k/K, k=1, 2, ..., K$ to determine the class that $\boldsymbol{d}_{x_t}$ belongs to. Essentially, the multi-class classification based on logistic function leverages the logistic function to transform the multi-dimensional data $\boldsymbol{d}_{x_t}$ into one-dimensional data in the classification space. The mapping is nonlinear and involves a new classification parameter $\boldsymbol{\alpha}$, which offers certain degree of freedom to control the performance of the classifier. The design of the mapping function will be discussed in the next section.

Incorporating the multi-class classification based on logistic function into (1), the WPP problem can be reformulated as:

$$\hat{x}_t = \sum_{k=1}^{K} \mathbb{I}\left(p_{x_t} \in \left[\frac{k-1}{K}, \frac{k}{K}\right)\right)\left(\sum_{i=1}^{I}\varphi_{k,i}\, x_{t-i} + \sum_{j=1}^{J}\pi_{k,j}\,\varepsilon_{k,t-j}\right) \tag{4}$$

Though we expect that the proposed model can effectively tackle the seasonality and randomness features in the original data, its training process, i.e., optimizing the classification parameter $\boldsymbol{\alpha}$ and the ARMA parameters $\boldsymbol{\varphi}_k = [\varphi_{k,1}, \varphi_{k,2}, \ldots, \varphi_{k,I}]^{\mathrm{T}}$ and $\boldsymbol{\pi}_k = [\pi_{k,1}, \pi_{k,2}, \ldots, \pi_{k,J}]^{\mathrm{T}}$, is computationally challenging due to the corresponding optimization problem which is non-convex. In this paper, we decouple the original non-convex optimization problem with three parameters into two sub-optimization problems, one of which is solved by the NSGA-II, and the other of which is convex and can be solved by an analytical method. In this way, we reduce the complexity of solving the original problem and improve the convergence speed and training efficiency of the parameter estimation process.

III. Process of Proposed Model

In this section, we present the proposed data processing method and the process of solving the objective function.

A. Data Preprocessing

In this subsection, we employ a dual-criterion optimization method to pre-process the wind power time series, which can remove obvious outliers in the original data. Usually, wind power generations are observed through specific instruments (e.g., a windmill anemometer), and the observations may be contaminated by certain errors. Considering the errors, the wind power observations can be generally modeled as:

$$\bar{x}_t = x_t + v_t \tag{5}$$

Moreover, the wind power time series may contain some outliers due to extreme weather, malfunction of the measurement instruments and maintenance activities. The presence of outliers and errors in the data will degrade the prediction accuracy. Hence, preprocessing on the wind power data is necessary [38].

Different from the normalization preprocessing, which is frequently adopted by neural network based prediction, we aim to minimize the random observation errors as well as ensure the smoothness of the data [39]. The data preprocessing can then be formulated as a regularized least squares optimization problem [40]. The objective function with two items is $\sum_{t=1}^{T}|x_t - \bar{x}_t|^2 + \sigma(\boldsymbol{x})$ and the limitation is $\sigma(\boldsymbol{x}) = \sum_{t=1}^{T-1}(x_{t+1} - x_t)^2$. The first item in the objective function denotes the sum of squared errors, and the second item $\sigma(\boldsymbol{x})$ refers to the smoothness objective. Defining a difference matrix $\boldsymbol{D}$ for $\boldsymbol{x}$ as (6), the smoothness item $\sigma(\boldsymbol{x})$ can be represented as (7).

$$\boldsymbol{D} = \begin{bmatrix} 1 & -1 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} \tag{6}$$

$$\sigma(\boldsymbol{x}) = \|\boldsymbol{D}\boldsymbol{x}\|_2^2 \tag{7}$$

Based on the convex optimization theory [40], the corrected data $\boldsymbol{x}$ can be straightforwardly derived as:

$$\boldsymbol{x} = (\boldsymbol{E} + \boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})^{-1}\bar{\boldsymbol{x}} \tag{8}$$

The original wind power generation usually contains missing values ranging from 1 to 6 hours. This may be due to a system overhaul or measurement failure. Therefore, the wind power data finally used are different from the real data. In this paper, we employ quadratic smooth function to treat this kind of error as a dual-criterion problem, i.e., $\sum_{t=1}^{T}|x_t - \bar{x}_t|^2$ and $\sigma(\boldsymbol{x})$, under the assumption that the magnitude of error is very small. In this dual criteria optimization problem, the data achieved by the first item would be smooth when the value of the first item is large, and the second item $\sigma(\boldsymbol{x})$ is set to be the penalty term to avoid the over-fitting problem when the first item is optimized. Therefore, we usually expect the second item to be small.

B. Objective Function Design

In this paper, we employ the mean square error (MSE) as the training objective metric because it can be differentiated. The objective of WPP is to minimize the squared prediction errors. Assuming that a total of N samples are tested, MSE can be expressed as:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2 \tag{9}$$

Based on (3) and (4), the WPP problem can thus be for-

mulated as the following minimum squared error problem:

$$\begin{cases} \min\limits_{\boldsymbol{\varphi},\boldsymbol{\pi},\boldsymbol{\alpha}} S(\boldsymbol{\varphi},\boldsymbol{\pi},\boldsymbol{\alpha}) = \sum\limits_{t} |x_t - \hat{x}_t|^2 \\ \text{s.t. } \hat{x}_t = \sum\limits_{k=1}^{K} \mathbb{I}\left(p_{x_t} \in \left[\frac{k-1}{K}, \frac{k}{K} \right) \right) \hat{x}_{k,t} \\ \hat{x}_{k,t} = \sum\limits_{i=1}^{p} \varphi_{k,i} x_{t-i} + \sum\limits_{j=1}^{q} \pi_{k,j} \varepsilon_{k,t-j} \\ p_{x_t} = \dfrac{1}{1+\mathrm{e}^{-\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{d}_{x_t}}} \\ \boldsymbol{\varphi} = \left[\boldsymbol{\varphi}_1 \quad \boldsymbol{\varphi}_2 \quad \cdots \quad \boldsymbol{\varphi}_K \right] \\ \boldsymbol{\pi} = \left[\boldsymbol{\pi}_1 \quad \boldsymbol{\pi}_2 \quad \cdots \quad \boldsymbol{\pi}_K \right] \end{cases} \tag{10}$$

Since the K subregions, i. e., $[0, 1/K)$, $[1/K, 2/K)$, ..., $[K-1/K, 1]$, are mutually exclusive, the objective function can be rewritten as:

$$S(\boldsymbol{\varphi},\boldsymbol{\pi},\boldsymbol{\alpha}) = \mathbb{I}\left(p_{x_t} \in \left[\frac{k-1}{K}, \frac{k}{K} \right) \right) \sum_{t} |x_t - \hat{x}_{k,t}|^2 \tag{11}$$

Then the MSE problem can be decoupled into a two-level optimization problem, i.e.,

$$\min_{\boldsymbol{\alpha}} \sum_{k=1}^{K} \mathbb{I}\left(p_{x_t} \in \left[\frac{k-1}{K}, \frac{k}{K} \right) \right) \min_{\boldsymbol{\varphi}_k, \boldsymbol{\pi}_k} \left(\sum_{t} |x_t - \hat{x}_{k,t}|^2 \right) \tag{12}$$

This optimization problem is non-convex and challenging. For the classification parameter $\boldsymbol{\alpha}$, because the objective function is non-convex, the original problem is easy to fall into local optimization when using the traditional optimization algorithm. Even if the enumeration method is only used to solve the classification parameter $\boldsymbol{\alpha}$, the time complexity of solving the original problem will increase at least $O(10^{4|\boldsymbol{D}\boldsymbol{x}|})$, which is unacceptable. In fact, we notice that the inner minimization problem with respect to $\boldsymbol{\varphi}_k$ and $\boldsymbol{\pi}_k$ corresponds to the MSE design for each ARMA model and its optimal solution can be derived in semi-closed form as detailed in Section III-C. In regards to the outer minimization problem with respect to $\boldsymbol{\alpha}$, we can resort to a heuristic optimization algorithm with guaranteed convergence, i.e., the NSGA-II [41]. Section III-D will introduce the process of NSGA-II and further details can be found in [42].

C. Closed-form Solution of ARMA Model

In this subsection, we present the analytical solution of the ARMA model. ARMA model is not only a widely used statistical model, but also one effective regression model with local interpretability for prediction [43]. In particular, the local interpretability of ARMA model can lay a foundation for future research on how to improve the performance of prediction model. As shown in [44], the ARMA model can be approximated by an infinite AR model. We employ this approximation and conduct certain truncation on the infinite AR model. The estimated value in the k^{th} ARMA model can be approximated as:

$$\hat{x}_{k,t} = \sum_{i=1}^{I} \varphi_{k,i} x_{t-i} + \sum_{j=1}^{J} \pi_{k,j} \varepsilon_{t-j} \approx \sum_{i=1}^{n} \theta_{k,i} x_{t-i} \tag{13}$$

where n is a much larger value than the orders in ARMA model, i.e., I and J.

Substituting the approximation in (13) into the inner minimization in (12), the WPP problem can be represented as:

$$\begin{cases} \min\limits_{\{\theta_{k,i}\}} \sum\limits_{t} |x_t - \hat{x}_{k,t}|^2 \\ \text{s.t. } \hat{x}_{k,t} = \sum\limits_{i=1}^{n} \theta_{k,i} x_{t-i} \end{cases} \tag{14}$$

This is a least square problem. By rewriting the problem with matrix form, the optimal solution of (14) is given as [40]:

$$\boldsymbol{\Theta}_k = \left(\boldsymbol{x}_{t-1}^{\mathrm{T}} \boldsymbol{x}_{t-1} \right)^{-1} \boldsymbol{x}_{t-1}^{\mathrm{T}} \hat{\boldsymbol{x}}_{t-1} \tag{15}$$

$$\hat{\boldsymbol{x}}_{t-1}^{\mathrm{T}} = \begin{bmatrix} \hat{x}_{t-n} & \hat{x}_{t-n+1} & \cdots & \hat{x}_{t-1} \\ \hat{x}_{t-n-1} & \hat{x}_{t-n} & \cdots & \hat{x}_{t-2} \\ \vdots & \vdots & & \vdots \\ \hat{x}_{t-2n} & \hat{x}_{t-2n+1} & \cdots & \hat{x}_{t-n-1} \end{bmatrix} \tag{16}$$

$$\boldsymbol{x}_{t-1}^{\mathrm{T}} = \begin{bmatrix} x_{t-n} & x_{t-n+1} & \cdots & x_{t-1} \\ x_{t-n-1} & x_{t-n} & \cdots & x_{t-2} \\ \vdots & \vdots & & \vdots \\ x_{t-2n} & x_{t-2n+1} & \cdots & x_{t-n-1} \end{bmatrix} \tag{17}$$

$$\boldsymbol{\Theta}_k^{\mathrm{T}} = \begin{bmatrix} \theta_{k,n+1} & \theta_{k,n} & \cdots & \theta_{k,1} \\ \theta_{k,n+2} & \theta_{k,n+1} & \cdots & \theta_{k,2} \\ \vdots & \vdots & & \vdots \\ \theta_{k,2n} & \theta_{k,2n-1} & \cdots & \theta_{k,n} \end{bmatrix} \tag{18}$$

Considering the solution of $\theta_{k,i}$ in (15), the solutions of $\boldsymbol{\varphi}_k$ and $\boldsymbol{\pi}_k$ to the inner MSE problem in (12) can be derived in analytical forms. Specifically, substituting $\varepsilon_{k,t-j} = \hat{x}_{t-j} - x_{t-j}$ and $\hat{x}_{k,t} = \sum\limits_{i=1}^{n} \theta_{k,i} x_{t-i}$ into (13), we can obtain:

$$\boldsymbol{\varphi}_k = \begin{bmatrix} \pi_{k,1} & \pi_{k,1} & \cdots & \pi_{k,p} \end{bmatrix}^{\mathrm{T}} + \boldsymbol{\Pi}_k \begin{bmatrix} \theta_{k,1} & \theta_{k,1} & \cdots & \theta_{k,p} \end{bmatrix}^{\mathrm{T}} \tag{19}$$

$$\bar{\boldsymbol{\Theta}}_k \boldsymbol{\pi}_k = \begin{bmatrix} -\theta_{k,p+1} & -\theta_{k,p+2} & \cdots & -\theta_{k,p+n} \end{bmatrix}^{\mathrm{T}} \tag{20}$$

$$\boldsymbol{\Pi}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \pi_{k,1} & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \pi_{k,p-1} & \pi_{k,p-2} & \cdots & 1 \end{bmatrix} \tag{21}$$

$$\bar{\boldsymbol{\Theta}}_k^{\mathrm{T}} = \begin{bmatrix} \theta_{k,p} & \theta_{k,p+1} & \cdots & \theta_{k,p+n-1} \\ \theta_{k,p-1} & \theta_{k,p} & \cdots & \theta_{k,p+n-2} \\ \vdots & \vdots & & \vdots \\ \theta_{k,p-q+1} & \theta_{k,p-q+2} & \cdots & \theta_{k,n} \end{bmatrix} \tag{22}$$

Solving (20), the analytical-form solution of $\boldsymbol{\pi}_k$ can be derived as (23) and then the solution of $\boldsymbol{\varphi}_k$ directly follows (19). Through this solution, the predictor can be trained more efficiently. In the meanwhile, the order of the AR model is defined by the autocorrelation analysis and the partial correlation analysis [14], which are effective in determining the parameters of the dynamic process.

$$\boldsymbol{\pi}_k = \left(\bar{\boldsymbol{\Theta}}_k^{\mathrm{T}} \bar{\boldsymbol{\Theta}}_k \right)^{-1} \bar{\boldsymbol{\Theta}}_k^{\mathrm{T}} - \begin{bmatrix} \theta_{k,p+1} & -\theta_{k,p+2} & \cdots & -\theta_{k,p+n} \end{bmatrix}^{\mathrm{T}} \tag{23}$$

D. NSGA-II

In the previous subsections, we introduce the logistic function based multi-class classification. Since the classification parameters are too complex to be obtained through enumeration, we employ the NSGA-II to find the optimal parameters. NSGA-II is an iterative optimization algorithm, which is applicable to general optimization problem with guaranteed convergence [41]. It is applied here to achieve the optimal solutions of the parameters $\{\boldsymbol{\alpha}, \boldsymbol{\varphi}, \boldsymbol{\pi}\}$ in the two-level optimization problem (12). Based on [45], the procedures for the NSGA-II in optimizing the objective function can be summarized as follows.

Step 1: randomly generate a set of vectors $\boldsymbol{\alpha}_1^0, \boldsymbol{\alpha}_2^0, ..., \boldsymbol{\alpha}_M^0$ as the initial candidates for the classification parameter $\boldsymbol{\alpha}$, and set the iteration index $r=0$.

Step 2: conduct the multi-class classification based on the logistic function (3) for each candidate $\boldsymbol{\alpha}_m^r$; compute the ARMA parameters based on (15), (19), and (23); compute the MSE $S(\boldsymbol{\varphi}_k^r, \boldsymbol{\pi}_k^r, \boldsymbol{\alpha}_m^r)$ for each ARMA class; and store the MSE and its corresponding parameter $\boldsymbol{\alpha}_m^r$.

Step 3: encode the classification parameters $\boldsymbol{\alpha}_1^r, \boldsymbol{\alpha}_2^r, ..., \boldsymbol{\alpha}_M^r$; increase the iteration index by 1, i.e., $r=r+1$; and conduct the mutation and crossover process to generate a new set of classification parameters $\boldsymbol{\alpha}_1^r, \boldsymbol{\alpha}_2^r, ..., \boldsymbol{\alpha}_M^r$. The details of the mutation and crossover process for NSGA-II can be found in [45].

Step 4: repeat *Step 2* and *Step 3* until the MSE converges, i.e., $\left| S(\boldsymbol{\varphi}^r, \boldsymbol{\pi}^r, \boldsymbol{\alpha}_m^r) - S(\boldsymbol{\varphi}^{r-1}, \boldsymbol{\pi}^{r-1}, \boldsymbol{\alpha}_m^{r-1}) \right| \le \eta$, where η is a sufficiently small number, e.g., 10^{-3}.

The hyper-parameters of NSGA-II are shown in Table I.

TABLE I
HYPER-PARAMETERS OF NSGA-II

Hyper-parameter	Value
Updating rate	0.35
End error	0.01
The maximum iteration	2000
Population scale	100
Number of objective parameters	3
Probability of crossover	0.9

The key contributions of the proposed model are the introduction of a novel multi-class classification based on logistic function into the prediction model and the findings of the classification parameters and ARMA prediction parameters jointly to minimize the squared prediction error. We interpreted the parameter training problem as a solvable optimization problem. Meanwhile, we proposed an efficient iterative algorithm with semi-closed-form solutions for the parameters. The hyper-parameters in Table I are selected in the proposed model by the enumeration algorithm. The prediction accuracy of the proposed model is demonstrated using open wind power data in Section IV. As a summary, the pseudocode of the proposed model is shown in Algorithm 1. The complete process of the proposed model includes four processes: the data preprocessing, the training process of the multi-class classification, the predictor training process based on multi-class classification, and the prediction process.

Algorithm 1: pseudocode of proposed model

Require the original $\boldsymbol{x}$, the initial $\boldsymbol{\alpha}$, and the MSE threshold value (judgement objective)

1. **Return** $\hat{x}_{k,t}$
2. **while** input $\boldsymbol{x}$ **do**
3. Get the modified input vector $\bar{\boldsymbol{x}} \leftarrow \boldsymbol{x}$
4. Get the wind power generation vector $\boldsymbol{d}_{x_t} \leftarrow \boldsymbol{x}_t$
5. Get the current mapping vector of classification $\boldsymbol{\alpha}$
6. Get the initial classification set of training data $C_k \leftarrow (\boldsymbol{x}_t, \boldsymbol{d}_{x_t}, \boldsymbol{\alpha})$, as shown in (10)
7. Train the predictor of each class
8. Get the initial parameters of ARMA model $\boldsymbol{\Phi}$ and $\boldsymbol{\pi}$ from (13) to (23)
9. Do prediction process and judge if the error meet the MSE threshold value
10. Update $\boldsymbol{\alpha}$ through the NSGA-II
11. **end while**
12. Get the estimated prediction time series of the testing data $\hat{x}_{k,t} \leftarrow \boldsymbol{x}$, as shown in (4)

IV. NUMERICAL EXPERIMENTS

A. Data Description and Pretreatment

The proposed model is trained for the wind power generation per half an hour provided by the Australian Energy Market Operator (AEMO) [46], which collects the wind power generation data of 10651 equipment in Australia. Half-hourly wind power data from 17 years (from 1999 to 2016) are available in five states, i.e., Queensland state (QLD), New South Wales state (NSW), South Australia state (SA), Tasmania state (TAS), Victoria state (VIC). The data from 15 years (from 1999 to 2014) are collected for training process, and the data from two years (from 2015 to 2016) are collected for testing process. In this paper, the results have the same granularity of 30 min. The smoothing model in (8) is employed to smooth the training data and the parameters in (8) are automatically adjusted. The statistical features of wind power generation are presented in Table II. The prediction is made every half an hour, and the wind power generation in the next half an hour is predicted each time. In the meanwhile, the order of AR model would be changed according to the data features.

Selecting the data sets with multiple features is helpful to present the performance of WPP models with a comprehensive perspective. The case studies of [47] show that the conditional variance of the power series from AEMO changes over time, i.e., seasonality. Therefore, we employ AEMO data as the data set for testing the performance of the proposed model.

B. Analysis of Classification Coefficient $\boldsymbol{\alpha}$

The regression structure in the ARMA model to perform prediction tasks is limited with the assumption of stationarity. In practice, the stationarity assumption of the ARMA model cannot always be satisfied. When the training data do not meet the assumption of the ARMA model, estimating the parameters of the optimal solution is a very difficult problem. Hence, we use classification methods to relax the stationarity assumption of the prediction model.

TABLE II
STATISTICAL FEATURES OF WIND POWER GENERATION

State name	Sample scale	Number of training samples	Number of testing samples	The maximum power (kWh)	The minimum power (kWh)	Mean power (kWh)	Variance
QLD	267298	237598	29700	14579.86	4624.03	8340.28	1871263.00
NSW	315586	276138	39448	3385.42	21.89	1462.44	87924.00
SA	314098	282688	31410	3385.45	45.96	1462.44	3384908.00
TAS	192864	160720	32144	1760.23	479.39	1134.71	23555.15
VIC	315586	236690	78896	10414.86	2726.88	5542.48	746433.20

Within each class, the stationarity can be easily assumed. Specifically, we select certain representative features $\boldsymbol{d}_{x_t}$ from the past observations for multi-class classification. In this paper, we utilize 12 statistical features to extract the features of training data, i. e., mean value (y_1), median value (y_2), modal number (y_3), range of data (y_4), variance value (y_5), standard value (y_6), 0.25 quantile (y_7), 0.5 quantitle (y_8), 0.75 quantitle (y_9), 0.95 quantile (y_{10}), the maximum value (y_{11}), and the minimum value (y_{12}). Based on this set of statistical features, an example of the classification parameters $\boldsymbol{\alpha}$, which can also be regarded as the optimal feature weights, is shown in Fig. 3. With a solvable parameter estimation process, the classification parameters are automatically optimized in the training phase.

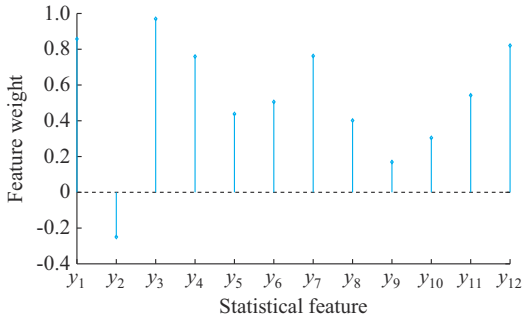

Fig. 3. Optimal feature weights in classification.

C. Evaluation Metrics

In the meanwhile, we use two metrics, i.e., the root mean square error (RMSE) and the mean absolute percentage error (MAPE), to present the performance of the proposed model and the comparison models. Assuming that a total of N samples are tested, RMSE and MAPE can be expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} \tag{24}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \times 100\% \tag{25}$$

The MAPE measures the size of the error in percentage terms, and is generally regarded as the common measure metric used to predict errors. Roughly speaking, it can quantify the prediction performance of the proposed model and the comparison models while eliminating the misleading of magnitude to the evaluation. The RMSE only measures the deviation of the prediction from the actual value [48]. Although RMSE and MAPE are commonly used as prediction error metrics, they are not often used in the training process due to the existence of absolute value operation and open radical operation. Instead, MSE is usually employed as the metric in the training process as introduced in Section III.

D. Evaluation of Prediction Performance

In this paper, we compare the performance of the proposed model with two individual prediction models and two combined models, which include: ① the support vector regression (SVR) model, which is a popular basis model in WPP and is a statistical model as well [47]; ② the extreme learning machine (ELM) [19], which is an effective ML-based model for modeling WPP problems (the parameters of ELM in this paper are optimized by back-propagation method); ③ the neural network-based ARMA (NN-ARMA) model [49], where the ARMA model is employed as the predictor and the parameters of the predictor are solved by neural networks; and ④ the *k*-nearest neighbor based ARMA (KNN-ARMA) model [50], where the *k*-nearest neighbor (KNN) is a famous method in classification task. This combined model classifies data sets with KNN, while employing ARMA as the predictor, in which the parameters of ARMA are solved by the maximum likelihood estimation. The KNN-ARMA model can serve as a good benchmark for comparison to verify the effectiveness of the classifier of the proposed model.

The comparison results of RMSE is presented in Table III and the results of MAPE are shown in Table IV. The results in the five individual states as well as the average results across all states are shown in the tables as well. The RMSE results show that the prediction accuracy of different states in the proposed model is better than that in the KNN-ARMA model. Specifically, we can observe from Table IV that the prediction error of the proposed model is the lowest among the five models, specially reaching 0.95% in the SA state. Moreover, the average MAPE error is 1.25%. From Table III, we can observe that the proposed model provides the lowest RMSE in the five states. Together with the results in Table IV, this means that the proposed model can not only track the changes of the wind power accurately, but also keep the prediction values very close to the actual values in a point basis. In terms of the average RMSE across all the five states, the proposed model is superior to the other four models. The superiority of the proposed model can also be further demonstrated by jointly considering the results of MAPE and RMSE in Tables III and IV. In addition, it can be observed from the results in Table III and Table IV that

there are significant differences in the prediction results of different states. According to Table II, wind power data in different states have different statistical features. In the meanwhile, we note that Table III shows that the RMSE results in the TAS state are much smaller than those in other states, this is because the wind power generation in the TAS state has a smaller magnitude and less variation. From Table II, it can be observed that the range of wind power generation variation in the TAS state is significantly smaller than that in other states.

TABLE III
COMPARISON RESULTS OF RMSE

Model	RMSE of different states					Average result
	QLD	NSW	SA	TAS	VIC	
Proposed	115.7	167.6	126.6	35.7	189.7	127.10
SVR [47]	189.7	176.4	145.8	40.0	199.2	150.22
ELM [19]	154.9	185.7	140.2	39.5	210.1	146.10
NN-ARMA [49]	183.6	231.4	138.7	48.8	204.4	161.40
KNN-ARMA [50]	177.9	226.3	136.3	46.2	201.5	157.60

TABLE IV
COMPARISON RESULTS OF MAPE

Model	MAPE of different states (%)					Average result (%)
	QLD	NSW	SA	TAS	VIC	
Proposed	1.14	1.40	0.95	1.11	1.66	1.25
SVR [47]	1.41	2.51	2.97	1.26	2.42	2.11
ELM [19]	1.38	1.69	1.15	1.34	1.77	1.47
NN-ARMA [49]	1.47	1.72	1.43	1.91	1.87	1.68
KNN-ARMA [50]	1.35	1.57	1.04	1.52	1.79	1.45

The interval prediction results (5%, 10%, 25%, 50%, and 95%) of wind power generation in 240 hours in the QLD state are presented in Fig. 4 as a special case.

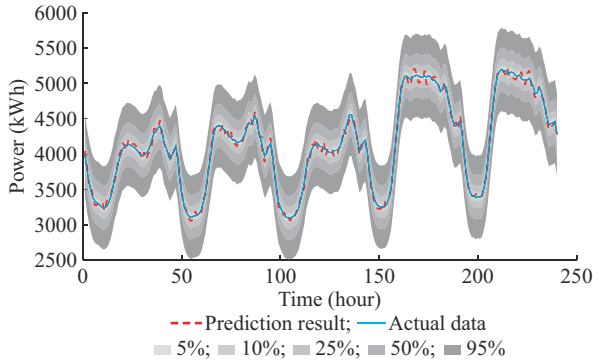

Fig. 4. Interval prediction results of wind power generation in 240 hours in QLD state.

It can be observed that the prediction result obtained by the proposed model could be covered by the confidence intervals well. Figure 4 also shows that the stationary of the predicted distribution changes significantly in temporal domain.

E. Computational Efficiency Comparison

Among the comparison models, the KNN-ARMA model is a combined model that employs KNN as the classification method and ARMA as the predictor. It also has a two-layer prediction structure. Therefore, to compare the proposed model with the comparison models from different perspectives, we present the performance of classification and the computational efficiency between the KNN-ARMA model and the proposed model. The whole computational time of both models are presented in Fig. 5.

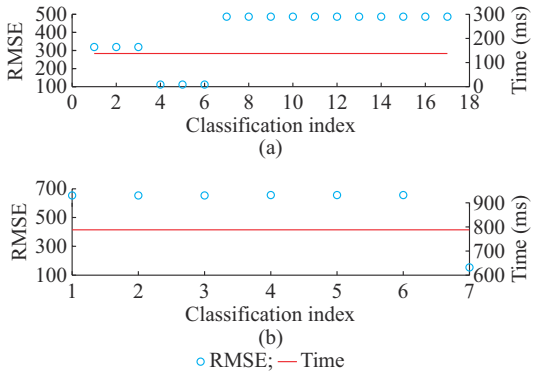

Fig. 5. Comparison of computational time and RMSE of different models. (a) Proposed model. (b) KNN-ARMA model.

The numerical experiments are instantiated on a PC with the Intel Core Duo Intel® Core™ i7-8700 processor (8 MB cache, the maximum frequency is up to 3.20 GHz). The comparison models are realized by MATLAB Parallel Server. As shown in Fig. 5, the proposed model divides the data set into 17 categories, while the KNN-ARMA model divides the data set into 7 categories. In this paper, the enumeration algorithm is used to determine the number of clusters in the classification process. Specifically, when the average prediction error after classification is the smallest, we choose this classification number as the value selected by the proposed model. It can be observed from Fig. 5 that although the proposed model has more categories than the KNN-ARMA model, its computational time is much lower than that of KNN-ARMA model. The advantage of the proposed model over the KNN-ARMA model mainly comes from the efficient logistic function based classification whose complexity is much lower than the KNN classification which requires heavy Euclidean distance calculations. From the results in Tables III and IV and Fig. 5, we can conclude that the proposed model has not only higher prediction accuracy but also less computational time than the KNN-ARMA model.

V. CONCLUSION

This paper proposes a model to conduct the short-time WPP by employing a logistic function based multi-class classification method into the prediction model. The non-convex classification problem is interpreted as a solvable optimization structure, which enables an analytical solution. In the meanwhile, we adopt an iterative algorithm with a semi-closed-form solution to efficiently solve the predictor. In ex-

periments, the proposed model achieves the best accuracy on WPP problems comparing with a series of state-of-the-art models.

Open data sets from AEMO are applied to present the performance of the proposed model comprehensively. The case studies show that the proposed model has superior prediction performance. Experimental results show that the proposed model can achieve more accurate prediction results. Moreover, there is a lower training complexity in the proposed model to ensure prediction accuracy compared with traditional models.

There are still some limitations in the proposed model. From the perspective of the model mechanism, if the input data are non-stationary so that the proposed data preprocessing fails, the proposed model may not be able to obtain accurate prediction results. In the meanwhile, the classification number of the proposed model is selected by enumeration method, which may lead to high computational time. Furthermore, the purposed model considers the case where the input features come from the temporal domain, and it may not be suitable for wind power generation prediction problems involving spatial domain data. In addition, to address some of the data challenges, the proposed model still needs improvement. Some possible research directions of future work include: ① considering more detailed classification models to incorporate time-varying wind power features; ② improving the interpretability of the combined models for further accuracy enhancement; and ③ incorporating the spatial correlation features into the classification and prediction.

REFERENCES

[1] REVE. (2021, Dec.). Wind power capacity is set to rise. [Online]. Available: https://www.evwind.es/topics/news-menu/wind-energy/

[2] M. A. Ahmed, M. R. El-Sharkawy, and Y.-C. Kim, "Remote monitoring of electric vehicle charging stations in smart campus parking lot," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 124-132, Jan. 2020.

[3] A. Cerejo, S. J. P. S. Mariano, P. M. S. Carvalho *et al*., "Hydro-wind optimal operation for joint bidding in day-ahead market: storage efficiency and impact of wind forecasting uncertainty," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 142-149, Jan. 2020.

[4] 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, Jun. 2016.

[5] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 800-810, Nov. 2012.

[6] M. Groch and H. J. Vermeulen, "Modeling high wind speed shutdown events using meso-scale wind profiles and survival analysis," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4955-4963, Nov. 2019.

[7] V. Yaramasu, B. Wu, P. C. Sen *et al*., "High-power wind energy conversion systems: state-of-the-art and emerging technologies," *Proceedings of the IEEE*, vol. 103, no. 5, pp. 740-788, May 2015.

[8] M. Hosseinzadeh and F. R. Salmasi, "Robust optimal power management system for a hybrid AC/DC micro-grid," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 675-687, Apr. 2015.

[9] E. Du, N. Zhang, B. Hodge *et al*., "Operation of a high renewable penetrated power system with CSP plants: a look-ahead stochastic unit commitment model," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 140-151, Jan. 2019.

[10] E. B. Ssekulima, M. B. Anwar, A. A. Hinai *et al*., "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review," *IET Renewable Power Generation*, vol. 10, no. 7, pp. 885-989, Mar. 2016.

[11] T. Hong and S. Fan, "Probabilistic electric load forecasting: a tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914-938, Nov. 2016.

[12] S. Sun, S. Liu, J. Liu *et al*., "Wind field reconstruction using inverse process with optimal sensor placement," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1290-1299, Aug. 2019.

[13] M. Liu, Z. Cao, J. Zhang *et al*., "Short-term wind speed forecasting based on the Jaya-SVM model," *International Journal of Electrical Power & Energy Systems*, vol. 121, pp. 56-106, Oct. 2020.

[14] A. A. Ezzat, M. Jun, and Y. Ding, "Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1437-1447, Jan. 2018.

[15] Y. Zhang and J. Wang, "A distributed approach for wind power probabilistic forecasting considering spatio-temporal correlation without direct access to off-site information," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5714-5726, Sept. 2018.

[16] Y. Wang, Q. Hu, D. Srinivasan *et al*., "Wind power curve modeling and wind power forecasting with inconsistent data," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 9, pp. 16-25, Jan. 2018.

[17] S. V. Medina and U. P. Ajenjo, "Performance improvement of artificial neural network model in short-term forecasting of wind farm power output," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 3, pp. 484-490, May 2020.

[18] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 763-770, Mar. 2015.

[19] C. Wan, C. Zhao, and Y. Song, "Chance constrained extreme learning machine for nonparametric prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3869-3884, Sept. 2020.

[20] 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, Nov. 2019.

[21] J. N. K. Liu, K. M. Kwong, and P. W. Chan, "Chaotic oscillatory-based neural network for wind shear and turbulence forecast with lidar data," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 42, no. 6, pp. 1412-1423, Apr. 2012.

[22] M. Cui, J. Zhang, Q. Wang *et al*., "A data-driven methodology for probabilistic wind power ramp forecasting," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1326-1338, Oct. 2019.

[23] D. Cao, W. Hu, J. Zhao *et al*., "Reinforcement learning and its applications in modern power and energy systems: a review," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1029-1042, Nov. 2020.

[24] Y. Wang, R. Zou, F. Liu *et al*., "A review of wind speed and wind power forecasting with deep neural networks," *Applied Energy*, vol. 304, p. 117766, Dec. 2021.

[25] W. Zou, C. Li, and P. Chen, "An inter type-2 FCR algorithm based T-S fuzzy model for short-term wind power interval prediction," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 4934-4943, Apr. 2019.

[26] N. Safari, S. M. Mazhari, and C. Y. Chung, "Very short-term wind power prediction interval framework via bi-level optimization and novel convex cost function," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1289-1300, Mar. 2019.

[27] J. C. López, M. J. Rider, and Q. Wu, "Parsimonious short-term load forecasting for optimal operation planning of electrical distribution systems," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1427-1437, Sept. 2018.

[28] Y. Teng, Q. Hui, Y. Li *et al*., "Availability estimation of wind power forecasting and optimization of day-ahead unit commitment," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 6, pp. 1675-1683, Nov. 2019.

[29] Y. Zhao, L. Ye, P. Pinson *et al*., "Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5029-5040, Jan. 2018.

[30] 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, Oct. 2015.

[31] P. Lu, L. Ye, Y. Zhao *et al*., "Feature extraction of meteorological factors for wind power prediction based on variable weight combined method," *Renewable Energy*, vol. 179, no. 4, pp. 1925-1939, Dec. 2021.

[32] L.-L. Li, Z.-Y. Cen, M.-L. Tseng *et al*., "Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic-support vector regression machine," *Journal of Cleaner Production*, vol. 279, no. 5, p. 123739, Jan. 2021.
[33] O. Abedinia, M. Lotfi, M. Bagheri *et al*., "Improved EMD-based complex prediction model for wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2790-2802, Feb. 2020.
[34] F. Yao, Z. Y. Dong, K. Meng *et al*., "Quantum-inspired particle swarm optimization for power system operations considering wind power uncertainty and carbon tax in Australia," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 4, pp. 880-888, Jul. 2012.
[35] T. Niu, J. Wang, H. Lu *et al*., "Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: application to offshore wind speed," *Energy Conversion and Management*, vol. 156, no. 1, pp. 597-617, Jan. 2018.
[36] M. A. Munoz, J. M. Morales, and S. Pineda, "Feature-driven improvement of renewable energy forecasting and trading," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3753-3763, Sept. 2020.
[37] C. Wan, M. Niu, Y. Song *et al*., "Pareto optimal prediction intervals of electricity price," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 817-819, Apr. 2017.
[38] B. Zhang and W. Perrie, "Recent progress on high wind-speed retrieval from multi-polarization sar imagery: a review," *International Journal of Remote Sensing*, vol. 35, no. 11-12, pp. 4031-4045, May 2014.
[39] J. Li and M. Li, "Prediction of ultra-short-term wind power based on BBO-KELM method," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 5, pp. 1-9, Sept. 2019.
[40] X. Taylor, *Convex Optimization of Power Systems*. Cambridge: Cambridge University Press, 2015.
[41] Y. Chen, X. Zou, and W. Xie, "Convergence of multi-objective evolutionary algorithms to a uniformly distributed representation of the pareto front," *Information Science*, vol. 181, no. 16, pp. 3336-3355, Aug. 2011.
[42] S. Jiang, K.-S. Chin, L. Wang *et al*., "Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department," *Expert Systems with Applications*, vol. 82, pp. 216-230, Oct. 2017.
[43] C. Molnar, *Interpretable Machine Learning*. New York: Lulu Press, 2022.
[44] G. E. Box, G. M. Jenkins, G. C. Reinsel *et al*., *Time Series Analysis: Forecasting and Control*. New York: John Wiley & Sons, 2015.
[45] U. Lee, S. Park, and I. Lee, "Robust design optimization (RDO) of thermoelectric generator system using non-dominated sorting genetic algorithm II (NSGA-II)," *Energy*, vol. 196, pp. 90-117, Apr. 2020.
[46] A. E. M. Operator, "Integrating renewable energy – wind integration studies report," Australian Energy Market Operator (AEMO), Tech. Rep., 2013.
[47] J. Wu, Z. Cui, Y. Chen *et al*., "A new hybrid model to predict the electrical load in five states of Australia," *Energy*, vol. 166, pp. 598-609, Jan. 2019.
[48] A. de Myttenaere, B. Golden, B. L. Grand *et al*., "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38-48, Jun. 2016.
[49] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network based models," *Applied Energy*, vol. 172, pp. 132-151, Jun. 2016.
[50] H. Demolli, A. S. Dokuz, A. Ecemis *et al*., "Wind power forecasting based on daily wind speed data using machine learning algorithms," *Energy Conversion and Management*, vol. 198, p. 111823, Oct. 2019.

Yunxuan Dong received the bachelor's and master's degrees from Lanzhou University, Lanzhou, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao S.A.R., China. His current research interests include system optimization and spatial-temporal feature learning.

Shaodan Ma received the double bachelor's degrees in science and economics and the master's degree in engineering from Nankai University, Tianjin, China, and the Ph.D. degree in electrical and electronic engineering from The University of Hong Kong, Hong Kong, China, in 1999, 2002, and 2006, respectively. From 2006 to 2011, she was a Postdoctoral Fellow at The University of Hong Kong. Since August 2011, she has been with the University of Macau, Macao S.A.R., China, where she is currently a Full Professor. Her research interests include signal processing and communications, particularly, array signal processing, transceiver design, resource allocation, and performance analysis.

Hongcai Zhang received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2013 and 2018, respectively. He is currently an Assistant Professor with the State Key Laboratory of Internet of Things for Smart City and also with the Department of Electrical and Computer Engineering, University of Macau, Macau, China. From 2018 to 2019, he was a Postdoctoral Scholar with the Energy, Controls, and Applications Lab, University of California, Berkeley, USA, where he also worked as a Visiting Student Researcher in 2016. His current research interests include Internet of Things for smart energy, optimal operation and optimization of power and transportation systems, and grid integration of distributed energy resources.

Guanghua Yang received the Ph.D. degree in electrical and electronic engineering from The University of Hong Kong, Hong Kong, China, in 2006. From 2006 to 2013, he was a Postdoctoral Fellow and a Research Associate with The University of Hong Kong. Since April 2017, he has been with Jinan University, Zhuhai, China, where he is currently a Full Professor in the School of Intelligent Systems Science and Engineering. His research interests include communications and networking.