IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 26, 2021, accepted June 10, 2021, date of publication June 22, 2021, date of current version June 30, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3091460*

# GMM-HMM-Based Medium- and Long-Term Multi-Wind Farm Correlated Power Output Time Series Generation Method

**YUFEI LI, BO HU, (Member, IEEE), TAO NIU, (Member, IEEE), SHENGPU GAO, (Graduate Student Member, IEEE), JIAHAO YAN, (Student Member, IEEE), KAIGUI XIE, (Senior Member, IEEE), AND ZHOUYANG REN, (Senior Member, IEEE)**
School of Electrical Engineering, Chongqing University, Chongqing 400044, China

Corresponding author: Bo Hu (hboy8361@163.com)

This work was supported in part by the National Key Research and Development Program of China (Technology and Application of Wind Power/Photovoltaic Power Prediction for Promoting Renewable Energy Consumption) under Grant 2018YFB0904200, and in part by Eponymous Complement S&T Program of State Grid Corporation of China under Grant SGLNDKOOKJJS1800266.

**ABSTRACT** Medium- and long-term wind power output time series are required in stochastic programming model for power system planning. Hidden Markov model (HMM) is a common method to generate wind power output time series, which can simultaneously consider the temporal and spatial correlation of multiple wind farms. However, the existing HMM methods use discrete matrix or Gaussian distribution to describe the output distribution of multiple wind farms, which usually leads to a relatively large error in statistical indices between the generated time series and the historical time series. Therefore, this paper proposes a method for generating medium- and long-term correlated output time series of multiple wind farms based on the Gaussians mixture model-Hidden Markov model (GMM-HMM). The discrete state variable in the hidden Markov model is used to describe the meteorological state. The Markov chain between discrete state variables is used to describe the temporal correlation of wind power output. The wind power output vector of multiple wind farms is used as the observation variable, and the mixed Gaussian probability distribution mapping relationship between the state variable and the multidimensional wind power output vector is established. Based on the Monte Carlo sampling method, the multi-wind farm output series satisfying the spatiotemporal correlation of historical output series are generated monthly. In the calculation example, the monthly wind power output series generated by five wind farms in Jilin Province are analyzed. The results show that the main statistical characteristics of the multi-wind power output time series generated by the proposed method are generally superior to those obtained with the traditional wind power output modeling method, which proves the superiority of the proposed method.

**INDEX TERMS** Multiple wind farms, spatiotemporal correlation, Gaussians mixture model-hidden Markov model (GMM-HMM), time series generation.

## NOMENCLATURE

Most of the symbols and notations used throughout this paper are defined below for quick reference.

### A. ABBREVIATIONS

| | |
|---|---|
| HMM | Hidden Markov model |
| GMM | Gaussians mixture model |

### B. INDICES

| | |
|---|---|
| $t$ | Index of time in HMM |
| $N$ | Index of the hidden state |
| $W$ | Index of the observation variable |
| $M$ | Index of wind farms |

### C. PARAMETERS

| | |
|---|---|
| $\mathbf{A}$ | The state transition probability matrix in HMM |
| $\mathbf{B}$ | The observation probability matrix in HMM |

The associate editor coordinating the review of this manuscript and approving it for publication was Weixing Li.

| | |
|---|---|
| $\boldsymbol{\pi}$ | The initial state probability vector in HMM |
| $\lambda$ | An HMM consisting of **A**, **B**, and $\boldsymbol{\pi}$ |
| $q_i$ | The $i$-th value of the hidden state |
| $v_i$ | The $i$-th value of the observed variable |
| $a_{ij}$ | The transition probability from state $i$ to state $j$ |
| $b_i$ | The joint probability distribution of the observation variable when hidden state is $i$ |
| $N_i(\mu_k, \sum_k)$ | The $k$-th Gaussian distribution in the Gaussian mixture model when the state is $i$ |
| $\mu_k$ | The mean vector of the $M$-dimensional Gaussian distribution in the $k$-th multidimensional Gaussian distribution |
| $\sum_k$ | The covariance matrix of the $M$-dimensional Gaussian distribution in the $k$-th multidimensional Gaussian distribution |
| $\alpha_k$ | The weight coefficient of the $k$-th Gaussian distribution |
| **D** | The cumulative state transition probability matrix |
| $d_{ij}$ | The sum of the probabilities between the current state $i$ and the next state between state 1 and $j$-1 |
| $Q$ | The number of hidden state |
| $R$ | The number of mixed Gaussian distribution |

### D. VARIABLES

| | |
|---|---|
| $i_t$ | The state value of the hidden Markov chain at time $t$ |
| $o_t$ | The observed value of HMM at time t |
| $P_M(t)$ | The wind power output of the $M$-th wind farm at time $t$ |
| $I$ | The hidden state time series |
| $O$ | The observed variable time series |

## I. INTRODUCTION

With the massive access to renewable energy sources in power systems, a large number of wind power grid connections now have a noticeable impact on the safe and stable operation and scheduling of power systems [1]. The fluctuation and intermittency of wind power output brings about great difficulties in renewable energy planning and medium- and long-term scheduling. Due to the lack of sufficient wind power output data, it is difficult to effectively quantify the impact of wind power integration on the power system [2]. It is necessary to mine the statistical characteristics of wind power output from historical output data to generate medium- and long-term wind power output time series [3]. Short-term wind power forecasting for under 7 days usually uses physical methods based on numerical weather forecast (NWP) data to achieve high accuracy. As there are large errors in weather forecasts beyond 7 days, long-term wind power forecasting (monthly and annual) does not pursue point-by-point accuracy but focuses on the extraction and reconstruction of statistical characteristics. With the large-scale access of the regional wind power cluster, there is a strong correlation between the wind farms in the region, and the output uncertainty mechanism is more complex than the one of single wind farm, which will affect the security and stability of the power grid. Therefore, it is necessary to explore time series of medium- and long-term multi-wind farms considering the spatiotemporal correlation of wind power cluster output [4]. The existing power system planning models and medium- and long-term operation models are mainly all equivalent annual cost models, which require an annual power output time series. However, since there are many years of historical wind power data, there is a need to extract a typical annual output time series from the historical years.

Wind power output time series generation mainly extracts and reconstructs the statistical characteristics of historical wind power output data based on statistics, and then uses a simulation method to obtain output time series with similar statistical characteristics [5], [6]. There are two kinds of stochastic simulation methods for medium- and long-term wind power output series [7]: 1) modeling the wind speed and calculating the wind power output by the wind power curve or 2) modeling the wind power output directly. The first kind of method [8] requires a large amount of historical wind speed data, and the errors commonly occur in the process of wind-power conversion, which cannot effectively capture the temporal and spatial correlation of wind power output. The purpose of the second kind of method is to characterize the temporal correlation of the output time series of a single wind farm, the spatial correlation between different wind farms and other statistical features conveniently. The mainstream statistical models include autoregressive integrated moving average (ARIMA), Markov chain Monte Carlo (MCMC) methods, etc.

In references [9]–[11], the MCMC method is used to simulate the time series of wind power output directly, which can relatively accurately simulate the main statistical characteristics of historical data, including the probability density function and autocorrelation function. However, the wind power output time series generated by the MCMC method usually stay on a certain output state. Furthermore, the outputs extracted within the corresponding power range of the same state fluctuate frequently. In addition, this method only models the output time series for a single wind farm.

In references [12], [13], a stochastic wind power model based on the ARIMA model is used, which can describe the nonstationary characteristics of wind power output with less model parameters and accurately reconstruct the main statistical characteristics of historical data. However, since the mean and covariance matrix of time series generated by ARIMA are time-invariant, it is impossible to accurately describe the volatility of time series at a long time scale. Moreover, for the large-scale wind farm cluster, it is difficult

to describe the spatial correlation between different wind farms using the high-dimensional ARIMA model.

In view of the shortcomings of the MCMC model and ARIMA model in terms of simulation accuracy for long time series, reference [14] uses the hidden Markov model (HMM) to discretely simulate the total power output of a single turbine and multi-turbines in a single wind farm. The upper quantiles (90%, 95%, 99%) of the distribution of the wind farm total power output are calculated. This reference proves that although the structure of the HMM is relatively simple, it can still reproduce important statistical characteristics of wind power output. However, the discrete HMM cannot be applied to the joint output distribution of multiple wind farms directly.

Accurate reproduction of the spatiotemporal correlation can yield a reasonable investment in power system planning. In the study of describing the output correlation of multiple wind farms, the copulas function [15], [16] and nonparametric kernel density function [17], [18] are usually used to describe the nonlinear correlation of wind power output. However, when the number of wind farms is large, high-dimensional modeling is difficult, and it is difficult to effectively describe the complex nonlinear relationship between wind farms.

The dominant existing statistical learning models are difficult to consider spatial correlation, so it is difficult to directly apply it to model multiple wind farm outputs time series. To simultaneously consider the spatiotemporal correlation of multi-wind farm output, reference [19] uses Gaussian HMM to simulate the monthly output time series of multi-wind farms. Since the actual wind power output distribution does not meet the common probability distribution such as Gaussian distribution, it is obviously insufficient to describe the actual output distribution of multiple wind farms by using a discrete observation transfer matrix in the discrete HMM or multidimensional Gaussian distribution in the Gaussian HMM.

The Gaussians mixture model-Hidden Markov model (GMM-HMM) is a typical statistical learning model, which has been used in many other time series modeling fields [20]–[22] and is proved to be better than Gaussian HMM in some research [22]. Since the actual wind power output does not conform to the Gaussian distribution [9], the Gaussian HMM uses the Gaussian distribution as the emission distribution in each hidden state, which does not conform to the actual wind power output distribution. In order to improve the modeling accuracy of wind power output with HMM, the mixed Gaussian distribution is used to fit the characteristics of any shape distribution. Using the Gaussian mixture distribution in HMM to represent the emission distribution in each hidden state can simulate the actual wind power output more accurately.

Aiming to address the shortcomings of the above, this paper proposes a multi-wind farm correlated power output time series generation method based on GMM-HMM. First, the HMM is used to model the time correlation of multi-wind farm output. The transfer relationship between discrete hidden states is used to represent the time-varying characteristics of wind power output at different moments. After taking the output vector of a multi-wind farm as the observation variable of the HMM, the probability distribution of multi-wind farm output in each hidden state is established by using the Gaussian mixture model. Then, the probability mapping relationship between the hidden state and multi-wind farm output is established. Then, the model parameters are trained with the historical wind power output data, and the transition probability matrix between the hidden states, the probability distribution between the hidden states and the multidimensional wind power output vectors are obtained. Finally, the Monte Carlo sampling method is used to generate a time series of output that considers the spatiotemporal correlations of multiple wind farms. The proposed model can simultaneously consider the temporal and spatial correlation between the output of multi-wind farms and use a multidimensional mixed Gaussian distribution to fit the actual output distribution of multiple wind farms. Parameter learning with this method is easy to implement, and the results are more consistent with the actual distribution characteristics of wind power output [23]. The example analysis uses five wind farms to learn the two-year output data monthly and generate the output time series. The main statistical characteristics of the proposed method are compared with those of the other 3 typical wind power output modeling methods, and the superiority of the proposed method is proven.

The structure of the remainder of this paper is as follows. Section 2 introduces the modeling method of multi-wind farm output correlation based on the GMM-HMM. Section 3 proposes the generation method of multi-wind farm correlated output time series based on the GMM-HMM. Section 4 provides case studies. Section 5 gives the conclusion.

## II. OUTPUT CORRELATION MODELING OF MULTI-WIND FARMS BASED ON THE GMM-HMM

### A. HIDDEN MARKOV MODEL (HMM)

The HMM is the simplest dynamic Bayesian network generation model; it is a typical statistical machine model dealing with annotation problems in natural language that has been widely used in many fields of pattern recognition and machine learning [24]–[26].

The HMM describes the process of randomly generating unobservable state random sequences from a hidden Markov chain [27] and then generating an observable random sequence from each hidden state. The HMM includes a state sequence and an observation sequence, in which the state sequence refers to the sequence of states randomly generated by the hidden Markov chain, and the observation sequence refers to the random sequence of observations generated corresponding to each state.

The HMM is determined by initial state probability vector $\pi$, state transition probability matrix $\mathbf{A}$ and observation probability matrix $\mathbf{B}$. The HMM is represented by a ternary

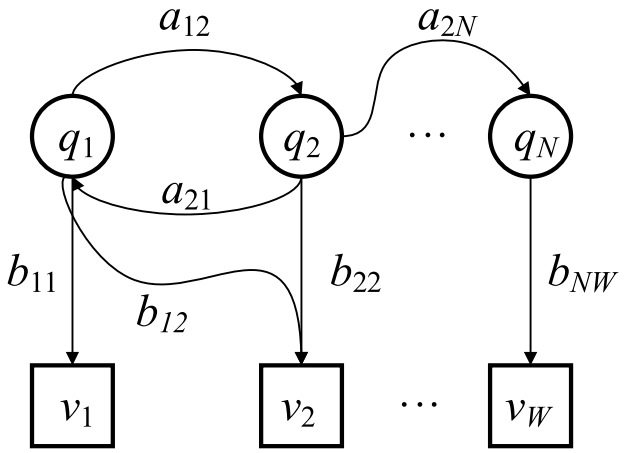


**FIGURE 1.** State jump and output mapping diagram of the hidden Markov model.

symbol $\lambda$:

$$\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}) \tag{1}$$

Two basic assumptions of HMM are as follows:

#### 1) HOMOGENEOUS MARKOV HYPOTHESIS

The state of the hidden Markov chain at any time $i_t$ only depends on its state at the previous moment $i_{t-1}$, which is independent of the state and observation of other time:

$$P(i_t \mid i_{t-1}, o_{t-1}, \cdots, i_1, o_1) = P(i_t \mid i_{t-1}), \quad t = 1, 2, \cdots, T \tag{2}$$

#### 2) OBSERVATION INDEPENDENCE HYPOTHESIS

The observation $o_t$ at any time only depends on the state $i_t$ of the Markov chain at that time and is independent of other observations and states at other times:

$$P(o_t \mid i_T, o_T, \cdots, i_{t-1}, o_{t-1}, \cdots, i_1, o_1) = P(o_t \mid i_t) \tag{3}$$

Fig 1 shows the HMM state transition diagram, where the state variable set is $\{q_1, q_2, \ldots, q_N\}$, and the observation variable set is $\{ v_1, v_2, \ldots, v_W \}$. The transition probability between state variables satisfies the state transition probability matrix $\mathbf{A}_{N\times N}$, and the emission probability between state variables and observation variables satisfies the observation probability matrix $\mathbf{B}_{N\times W}$ The observed variables in the HMM can be discrete or continuous. When the observation variables are continuous, the emission probability is described by continuous probability distribution.

### B. STATE VARIABLES AND OBSERVATION VARIABLES OF HMM OUTPUT IN MULTIPLE WIND FARMS

Because the wind power output is determined by the meteorological conditions and other factors, the wind power output can be regarded as the observation state. Due to the chaotic characteristics of factors, such as the meteorological evolution, that determine the wind power output, we cannot directly determine the hidden state, but the discrete value 1 to $N$ can be used to represent the hidden state [14]. Therefore, this paper sets the hidden state $i_t$ of each wind power output time series at each time as an integer between 1 and $N$ and uses the Markov chain to describe the transfer relationship $\mathbf{A}$ between the hidden states.

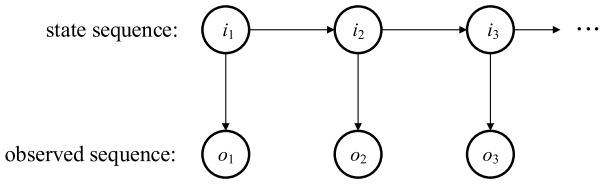


**FIGURE 2.** Schematic diagram of the hidden Markov model for multi-wind farms.

In the output time series of multiple wind farms, the observed variable is the output vector of each wind farm at each time, namely, $o_t = [p_1(t), p_2(t), \ldots, p_M(t)]$, and $p_M(t)$ is the wind power output of the $M$-th wind farm at time $t$, as shown in Fig 2.

### C. GAUSSIAN MIXTURE MODEL (GMM)

The Gaussian mixture model is a model to fit the probability density distribution of any non-Gaussian by weighting the sum of several Gaussian distribution functions. Its probability density function is expressed by the sum of the product of each single Gaussian model and the corresponding weight [28], [29]. It is also a typical clustering method that projects the data in the sample onto a class represented by a single Gaussian model. Assuming that the mixed Gaussian model consists of $K$ Gaussian models (namely, the data contain $K$ classes), the probability density function of GMM is as follows:

$$P(y \mid \theta) = \sum_{k=1}^{K} \alpha_k \phi(y \mid \theta_k) \tag{4}$$

where $\alpha_k$ is the weight coefficient, $\alpha_k \geq 0$, and $\sum_{k=1}^{K} \alpha_k = 1$, $\theta_k = (\mu_k, \sigma_k^2)$, $\phi(y \mid \theta_k)$ is the Gaussian distribution probability density function:

$$\phi(y \mid \theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right) \tag{5}$$

The advantage of the GMM is that if the number of Gaussian models fused by the mixed Gaussian model is large enough, and the weight between them is set reasonably enough, it can fit any distribution of samples. When describing the joint distribution of multi-wind farm output under each hidden state in the HMM, the GMM is sure to be more consistent with the actual distribution characteristics of wind power output than the Gaussian distribution used in the general HMM. Compared with other methods that represent the correlation of multi-wind farms, such as the copula function and nonparametric kernel density, it is simple in form and convenient in parameter learning. It is also applicable under the high-dimensional joint probability distribution of

multi-wind farm output when the number of wind farms is large.

### D. CORRELATED OUTPUT TIME SERIES MODEL OF MULTI-WIND FARMS BASED ON THE GMM-HMM

A multi-wind farm output time series model based on the GMM-HMM with consideration of spatiotemporal correlations is established in this section. Through the known wind power output observation sequence, the hidden state transition matrix in the HMM and the probability distribution between hidden state variables and observation variables are learned. Model parameters include the initial state probability vector $\boldsymbol{\pi}$, state transition probability matrix $\mathbf{A}$ and joint probability distribution of observation vector $\mathbf{B}$ [30]. Furthermore, the hidden state variable of wind power output at each time is represented by discrete variable $i_t$, and the observed variable at each time is represented by the output vector of multiple wind farms $o_t$.

#### 1) INITIAL STATE PROBABILITY VECTOR $\pi$

At the initial time $t = 1$, the hidden state probabilities of wind power output are expressed by vector $\boldsymbol{\pi} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \cdots, \boldsymbol{\pi}_N]$, where

$$\boldsymbol{\pi}_i = P(i_1 = i), 1 \le i \le N \tag{6}$$

#### 2) STATE TRANSITION PROBABILITY MATRIX $\mathbf{A}$

The state transition probability matrix A describes the transition probability between two hidden states at adjacent moments. The element $a_{ij}$ in row $i$ and column $j$ of $\mathbf{A}$ represents the probability that the state at time $t$ is $i$ and the state at time $t + 1$ is $j$:

$$a_{ij} = P(i_{t+1} = j | P(i_t = i), \quad 1 < i, j < N \tag{7}$$

#### 3) MULTIDIMENSIONAL JOINT PROBABILITY DISTRIBUTION $\mathbf{B}$ OF OBSERVATION VARIABLES

When the observed variable is discrete, the probability matrix of the observed variable corresponding to the state $i$ at time $t$ is expressed as $\mathbf{B} = [b_i(p)]_{N \times W}$:

$$b_i(p) = P(o_t = v_p | i_t = i), \quad 1 < i < N, 1 < p < W \tag{8}$$

Since the observation variable in this paper is continuous, the emission probability from the state variable to the observed variable is no longer the form of the observation probability matrix but is described by the continuous probability distribution, namely, the joint probability distribution of the observation variable. In this paper, a multidimensional mixed Gaussian distribution is used to describe the joint probability distribution of the observation variables. Therefore, when the state at time $t$ is $i$, the probability distribution of the observed variable is expressed as follows:

$$b_i = \sum_{k=1}^{K} \alpha_k N_i(\mu_k, \sum_k), \quad 1 < i < N \tag{9}$$

where $N_i(\mu_k, \sum_k)$ represents the $K$-th Gaussian distribution in the mixed Gaussian model when the state is $i$ at time $t$. $\mu_k = \{\mu_k^1, \mu_k^2, \ldots, \mu_k^M\}$ and $\sum_k$ represent the mean vector and covariance matrix, respectively, of the $M$-dimensional Gaussian distribution in the $K$-th multidimensional Gaussian distribution, and $M$ is the number of wind farms. Equation (9) indicates that when the state at time $t$ is $i$, the observation variable $o_t$ is subject to a mixed Gaussian distribution with a weight coefficient of $\alpha_k$, a mean value of Gaussian distributions of all dimensions of $\mu_k$, and a covariance matrix of $\sum_k$.

## III. GENERATION METHOD OF MULTI-WIND FARM CORRELATED OUTPUT TIME SERIES BASED ON THE GMM-HMM

The generation method of the multi-wind farm correlated output time series based on the GMM-HMM is mainly divided into two steps. The first step is to learn the parameters of the GMM-HMM by using the historical observation time series of wind power output in which the maximum likelihood method is used to estimate the parameters, and the Expectation-Maximum (EM) algorithm is used as the optimization algorithm, so that the probability $P(O|\lambda)$ of the historical observation time series $O$ is the in the model parameter $\lambda$ is the largest. The second step is to generate the time series according to the model parameters given the length of the observation sequence and to continuously generate the hidden state and observation state of the next moment. The observation time series is the time series of wind farm output.

### A. MODEL PARAMETER LEARNING

Given the Gaussian distribution number $K$ of the GMM and the hidden state number $N$ of the HMM, the model parameter $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ needs to be learned. The main method to learn parameters in the GMM and HMM is the EM algorithm [31], which also makes the process of parameter learning easier. The EM algorithm is a maximum likelihood estimation method that can solve the model parameters with hidden variables from the known data set when the maximum likelihood estimation method or Bayesian estimation method cannot be used directly due to the hidden variables in the model.

When the model variable includes hidden variable $I = (i_1, i_2, \ldots, i_T)$, the input of the EM algorithm is the observation data $O = (o_1, o_2, \ldots, o_T)$, and the output is the model parameter $\lambda$ that needs to be estimated. Then, suppose that the likelihood function of incomplete data $O$ is $P(O|\lambda)$, the joint probability distribution of $O$ and $I$ is $P(O, I|\lambda)$, and the logarithmic likelihood function of complete data $(O, I)$ is $\log P(O, I|\lambda)$. Solving the maximum likelihood estimation $\hat{\lambda}$ of model parameter $\lambda$ is to maximize the logarithmic likelihood function $L(\lambda)$ of $O$ with respect to $\lambda$:

$$\hat{\lambda} = \arg\max_{\lambda} \log L(\lambda) \tag{10}$$

$$L(\lambda) = \log P(O|\lambda) = \log \sum_{I} P(O, I|\lambda)$$

$$= \log\left(\sum_{I} P(O|I,\lambda)P(I|\lambda)\right) \tag{11}$$

Since equation (11) contains hidden data $I$, it is difficult to optimize directly [24]; therefore, the EM algorithm approximates the maximization of $L(\lambda)$ by iteratively solving the maximum Q function step by step. The Q function is the expected value of the logarithmic likelihood function $\log P(O, I|\lambda)$ of complete data $(O, I)$,with respect to the current conditional distribution $P(I|O, \lambda^{(i)})$ of $I$ given $O$ and the current estimates of the parameters $\lambda^{(i)}$ and is also known as the lower boundary function:

$$\begin{aligned} Q(\lambda, \lambda^{(i)}) &= E_I\left[\log P(O, I|\lambda)|O, \lambda^{(i)}\right] \\ &= \sum_{I} \log P(O, I|\lambda) P(I|O, \lambda^{(i)}) \end{aligned} \tag{12}$$

The steps of the EM algorithm are shown in algorithm 1:

**Algorithm 1** EM Algorithm

**Step 1**: Select the initial value of the parameter $\lambda^{(0)}$, set $i = 0$ and start the iteration;

**Step 2**: step E: The estimated value of the $i$-th iteration parameter $\lambda$ is denoted as $\lambda^{(i)}$ and the Q function of the $i+1$-th iteration is calculated.

**Step 3**: step M: The parameter estimate of the $i+1$th iteration is $\lambda^{(i+1)}$, which is $\lambda$ that maximizes $Q(\lambda, \lambda^{(i)})$:

$$\lambda^{(i+1)} = \arg\max_{\lambda} Q(\lambda, \lambda^{(i)}) \tag{13}$$

**Step 4**: Repeat steps 1 and 2 until convergence condition $\|\lambda^{(i+1)} - \lambda^{(i)}\| < \epsilon_1$ is satisfied

When the EM algorithm is used to train parameters of the GMM-HMM, the observation data $o_t$, $t = 1, 2, \ldots, N$ can be regarded as generated by selecting the $k$-th Gaussian distribution model according to the weight coefficient. When $I$ and $O$ are known, variable $\gamma_{t,j,k}$ is defined as the joint probability of $i_{t=}j$, and $o_t$ comes from the $k$-th Gaussian distribution under $i_t$. $\xi_{t,i,j}$ is defined as the joint probability of transition from hidden state $i$ at $t$ to hidden state $j$ at $t+1$:

$$\gamma_{t,j,k} = P\left(i_t = j, \phi_{t,j} = \psi_{j,k}|O, \lambda\right) \tag{14}$$

$$\xi_{t,i,j} = P\left(i_t = i, i_{t+1} = j|O, \lambda\right) \tag{15}$$

where, $\phi_{t,j}$ represents the Gaussian distribution from the hidden state $j$, which is used to obtain the observation data $o_t$. $\psi_{j,k}$ represents the $k$-th Gaussian distribution corresponding to the hidden state $j$. In step E(step 2) of the EM algorithm, to reduce the calculation time, $\gamma_{t,j,k}$ and $\xi_{t,i,j}$ can be obtained according to the forward-backward algorithm[30].

In step M(step 3), according to the probability distribution $\gamma_{t,j,k}$ of the hidden variables under the current model parameters, the parameters are re-estimated, including the probability of hidden state $j$ at the initial time $\hat{\pi}_j$, state transition probability $\hat{a}_{i,j}$ from hidden state $i$ to hidden state $j$, weight coefficient $\alpha_{j,k}$, mean vector $\mu_{j,k}$ and covariance matrix $\Sigma_{j,k}$ of the $k$-th Gaussian distribution corresponding to hidden state $j$:

$$\hat{\pi}_j = \sum_{k=1}^{K} \gamma_{1,j,k} \tag{16}$$

$$\hat{a}_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_{t,i,j}}{\sum_{t=1}^{T-1}\sum_{k=1}^{K} \gamma_{t,j,k}} \tag{17}$$

$$\hat{\alpha}_{j,k} = \frac{\sum_{i=1}^{T} \gamma_{t,j,k}}{\sum_{j=1}^{T}\sum_{k=1}^{K} \gamma_{t,j,k}} \tag{18}$$

$$\hat{\mu}_{j,k} = \frac{\sum_{t=1}^{T} \gamma_{t,j,k} o_t}{\sum_{t=1}^{T} \gamma_{t,j,k}} \tag{19}$$

$$\sum_{j,k} = \frac{\sum_{t=1}^{T} \gamma_{t,j,k} \left(o_t - \mu_{j,k}\right)\left(o_t - \mu_{j,k}\right)^{\mathrm{T}}}{\sum_{t=1}^{T} \gamma_{t,j,k}} \tag{20}$$

### B. MULTI-WIND FARM CORRELATED OUTPUT TIME SERIES GENERATION METHOD

After the monthly correlated output model of multiple wind farms for 12 months is established based on historical wind power output data, the monthly wind power output time series of multi-wind farm is generated by the Monte Carlo random sampling method. Assuming that the number of wind farms is $M$, the number of hidden states is $N$, and the required output time series length is $T$, the output time series generation process is as follows:

(1) According to the state transition probability matrix $\mathbf{A} = (a_{ij})_{N\times N}$, the cumulative state transition probability matrix $\mathbf{D} = (d_{ij})_{N\times(N+1)}$ is calculated where $d_{ij}$ represents the sum of the probabilities between the current state $i$ and the next state between state 1 and $j$-1 (when $j = 1$, $d_{ij}$ is 0). The calculation formula is as follows:

$$d_{ij} = \begin{cases} 0, & j = 1 \\ \sum_{k=1}^{j-1} a_{ik}, j = 2, \ldots, N+1, & i = 1, 2, \ldots, N \end{cases} \tag{21}$$

(2) According to the probability distribution $\pi$ of each hidden state at the initial time in the HMM model parameters, sampling generates the common hidden state $q_1 = o$ for M wind farms at time $t = 1$;

(3) After the hidden state $q_t$ at the current time $t$ is obtained, the mean vector and covariance matrix of

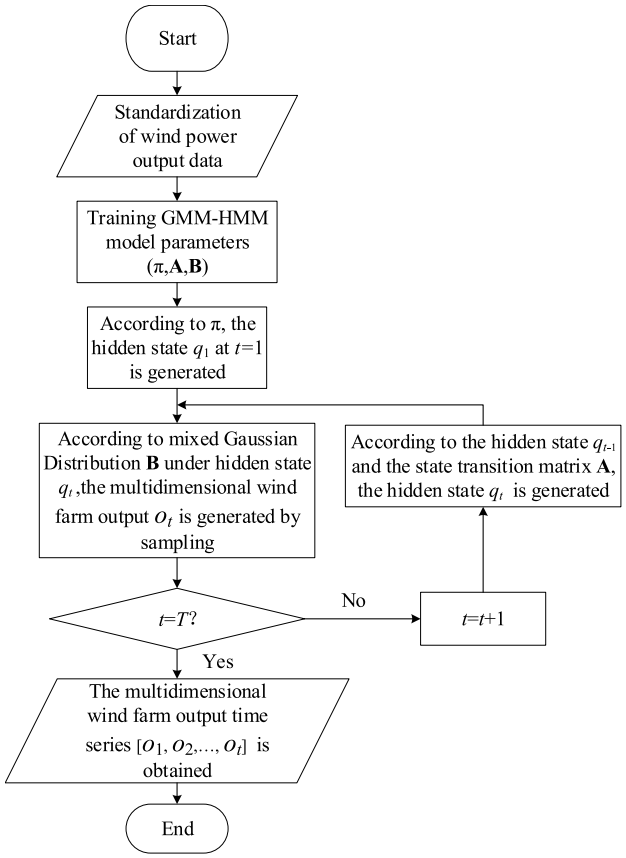


**FIGURE 3.** Flow chart of the GMM-HMM-based output time series generation method for multi-wind farms.

k multidimensional Gaussian distributions that constitute the output vector of M-dimensional wind farm are obtained according to the parameter $N_c(\mu_k, \sum_k)$ of the Gaussian mixture model. At this time, k M-dimensional wind power output subvectors are obtained by random sampling and then weighted by the weight coefficient $\alpha_k$ of each Gaussian distribution in GMM; thus, the M-dimensional wind power output vector $o_t = [p_1(t), p_2(t), \ldots, p_M(t)]$ is obtained.

(4) Since the current time is $t$ $(1 \leq t \leq T)$ and the hidden state of the current time is $q_t$, the random number $s$ is generated according to the uniform distribution of (0,1); $s$ is compared one by one with each column element in the $q_t$-th row in the cumulative state transition probability matrix **D** calculated in the first step. When $s$ is larger than the element in column $c$ of the $q_t$ row and smaller than the element in column $c+1$, the hidden state $q_{t+1}$ of the next moment $t$ is set as $c$;

(5) Judge whether time $t$ is the last time $T$, if not, set $t = t+1$, and repeat step (3)-(4); if so, the algorithm is terminated. and $M$-dimensional wind power output time series $[o_1, o_2, \ldots, o_T]$ are obtained.

The flow chart of a multi-wind farm correlated output time series generation algorithm based on the GMM-HMM is shown in Figure 3. First, the wind power output data are normalized, and the model parameters are trained according to the historical wind power output data. Then, the hidden state of the initial time $q_1$ is sampled according to the probability distribution of each hidden state at the initial time. The parameters of multidimensional mixed Gaussian distribution are determined according to the hidden state value at that time $q_t$, and the multidimensional wind farm output vector at that time $O(t)$ is then generated by sampling. Whether the desired length of the time series is produced is then determined. If not, the hidden state at the next moment $q_{t+1}$ is generated according to the hidden state at the moment $q_t$, and the state transfer matrix **A** and the wind farm output vector at the next moment continue to be generated until the multidimensional wind farm output time series with sequence length T is generated, and the algorithm ends.

**TABLE 1.** The basic information of the wind farms used in the case section.

| Wind farm serial number | Installed capacity (MW) | Geographic longitude | Geographic latitude |
|---|---|---|---|
| 1 | 98.8 | 122°25′56″ | 45°23′49″ |
| 2 | 197.9 | 122°59′27″ | 45°47′57.5″ |
| 3 | 45 | 122°24′ | 45°49′30″ |
| 4 | 200.5 | 123.582544 | 45.27323 |
| 5 | 99 | 122°56'0.00" | 45°45'18.00" |

## IV. CASE STUDIES

Since the wind power output beyond 7 days cannot be predicted accurately, this paper does not pursuit point-by-point accuracy for the generated medium- and long-term wind power output time series but focuses on the extraction and reproduction of the statistical characteristics of the historical output series.

In this paper, the historical power output data of five wind farms in Jilin Province of China every 15 minutes in 2017 and 2018 were adopted. Because the monthly meteorological model is more stable and consistent than the annual meteorological model. Hence, the monthly output model is adopted in the paper. Table 1 shows the installed capacity, geographical longitude and latitude of the five wind farms. In the use of historical wind power output data, we conducted a normalization process. The 5 wind farms and multiple wind farms mentioned below refer to the 5 wind farms in Table 1.

### A. COMPARISON OF MAIN STATISTICAL CHARACTERISTICS OF POWER OUTPUT TIME SERIES GENERATED BY EACH METHOD

In Case 1, the proposed GMM-HMM method is used to generate the output time series of 5 wind farms at a total of 70080 moments in two years, in which the number of hidden states of HMM is set to 7, and the number of Gaussian distributions of GMM is set to 11. At the same time,

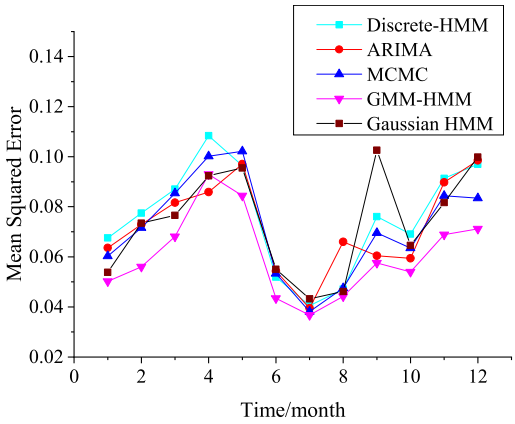


**FIGURE 4.** Mean square errors of monthly output series of five wind farms produced by different methods.

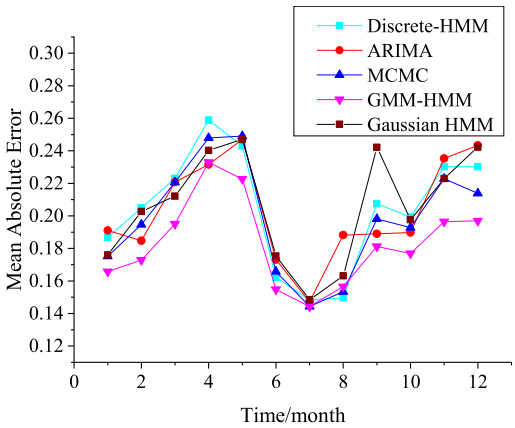


**FIGURE 5.** Mean absolute errors of monthly output series of five wind farms produced by different methods.

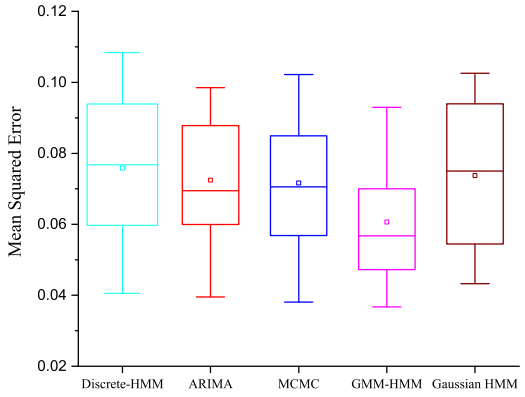


**FIGURE 6.** Boxplot of the MSEs of monthly output series of five wind farms produced by different methods.

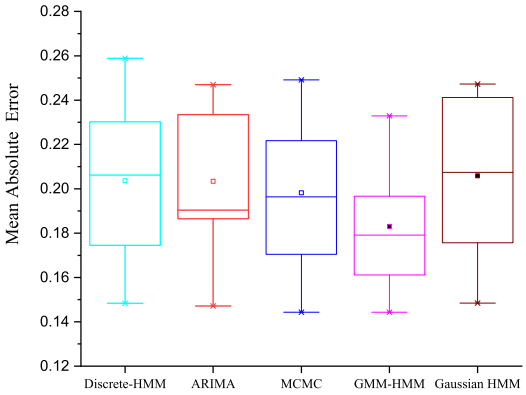


**FIGURE 7.** Boxplot of the MAEs of monthly output series of five wind farms produced by different methods.

the MCMC, ARIMA and discrete HMM (the discrete matrix is used as the observation emission matrix) [14] methods are used to generate the output time series of each single wind farm. In the MCMC model, the number of discrete intervals is 15. In the ARIMA model, $p = 1$, $d = 1$, $q = 2$. In the Discrete HMM, the number of discrete intervals is 15, and the number of hidden states is 17. The Gaussian HMM method is used to generate the output time series of 5 wind farms, in which the number of hidden states of HMM is set to 20. The proposed method is compared with the output time series generated by the MCMC, ARIMA, discrete HMM and Gaussian HMM method. In this paper, the final probability indicators are the mean of the probability indicators of 100 different simulation series.

### 1) COMPARISON OF ERRORS AND STATISTICAL INDICATORS

First, the MSE (mean square error) and MAE (mean absolute error) [32]–[34] of the power output of five wind farms in 12 months under the five methods are plotted, as shown in Figure 4 and Figure 5. The MSE and MAE of each method in the following refer to the MSE and MAE between the output time series generated by this method and the historical output time series.

The index calculation results show that the MSE of the GMM-HMM method is the smallest in all months except the fourth month, when the MSE of the ARIMA method is slightly smaller than that of the GMM-HMM.

The index calculation results show that the MAE of the ARIMA method is slightly smaller than that of the GMM-HMM in the fourth month. At the 8th month, the MAE of the discrete HMM and MCMC methods are smaller than that of the GMM-HMM. In other months, the MAE of the GMM-HMM method is the smallest.

According to the MSE and MAE data of different methods, the box-line diagrams of MSE and MAE of the power output of five wind farms in 12 months under five methods are drawn, as shown in Figure 6 and Figure 7. In Figure 6, the upper limit value, lower limit value, mean, upper quartile, median, and lower quartile of errors when using GMM-HMM method are 0.09294, 0.03672, 0.06063, 0.0700, 0.0568 and 0.0472, respectively. In Figure 7, the upper limit value, lower limit value, mean, upper quartile, median, and lower quartile of errors when using GMM-HMM method are 0.23287, 0.1443, 0.1830, 0.1967, 0.1791 and 0.1611, respectively. They are all lower than the indices of the other methods. Thus,

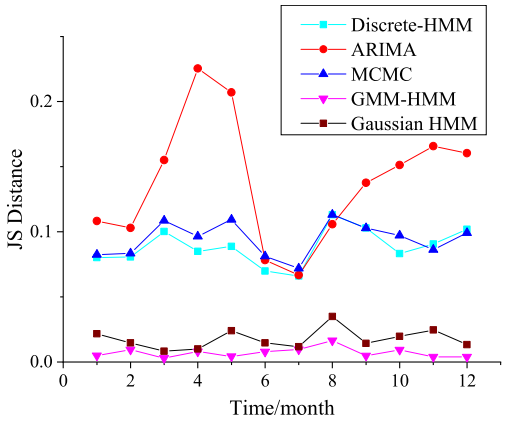


**FIGURE 8.** JS distance of the joint probability density of the annual output of five wind farms with different methods.

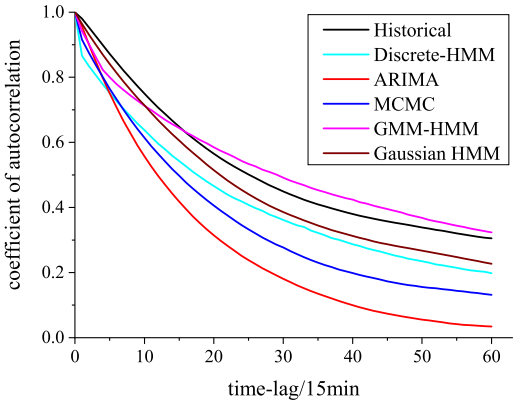


**FIGURE 9.** ACF of annual output of No. 4 wind farm generated by different methods.

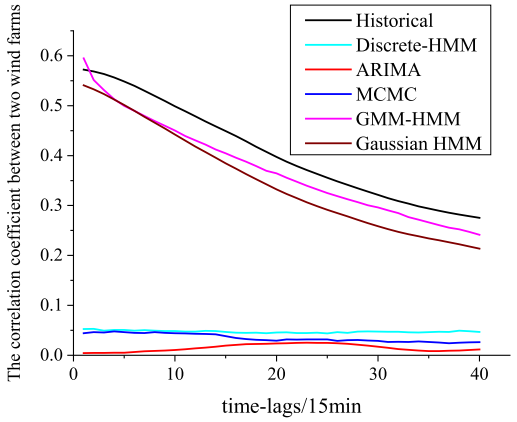


**FIGURE 10.** CCF of annual output between the No. 3 and No. 4 wind farms generated by different methods.

the statistical indices based on the GMM-HMM method perform well.

### 2) COMPARISON OF PROBABILITY DISTRIBUTIONS OF OUTPUT TIME SERIES

First, based on the historical monthly output data of five wind farms and the monthly output time series generated by five methods (GMM-HMM, MCMC, ARIMA, discrete HMM, Gaussian HMM), the kernel density estimation method is used to estimate the joint probability density function of five wind farms in 12 months of the year. Then, the JS (Jensen-Shannon) distance between the output joint probability density distribution function generated by these five methods and the historical output joint probability density distribution function is calculated and compared, as shown in Fig. 8. The JS distance of each method below refers to the JS distance between the joint probability density distribution function of output generated by each method and the joint probability density distribution function of historical output.

The index calculation results show that in each month, the GMM-HMM method has the best effect. Its JS distance is only 0.002-0.01, which is the closest to 0, indicating that the joint distribution of output generated by this method is closest to the joint distribution of historical output. The ARIMA method has the worst performance, with a JS distance up to 0.23.

### 3) COMPARISON OF SPATIOTEMPORAL CORRELATION INDEXES OF OUTPUT TIME SERIES

To compare the time correlation of output time series generated by different methods, the autocorrelation function (ACF) of the annual output series generated by the five methods and the historical annual output series of the No. 4 wind farms are calculated, as shown in Fig. 9.

The calculation results of the ACF index shows that the correlation of wind farm output decreases with the increase in time delay. When the time delay is order 1 to order 15, the ACF of the GMM-HMM method is lower than the ACF of the historical wind power output. After order 15, they are all slightly higher than the ACF of historical wind power output, but they are closer to the ACF of the historical wind power output than the other three methods. The sum of the absolute values of the errors between the ACF of the five methods (discrete HMM method, Gaussian HMM method, ARIMA method, MCMC method, and GMM-HMM method) and the historical ACF of wind power output are 6.1,3.5, 14.5, 9.6, and 2.2 respectively. Thus, the proposed method can more accurately describe the variation characteristics of time autocorrelation of single wind power output series with a time delay. The ACF of the other three methods at each time delay is less than that of the historical series, which will lead to overestimation of the wind power output ramp rate and volatility, resulting in excessive investment in flexibility resources.

To compare the spatial correlation of wind power output among different wind farms, the cross-correlation function (CCF) of the annual output series between No. 3 and No. 4 wind farms generated by the five methods and the historical annual output series are calculated, as shown in Figure 10.

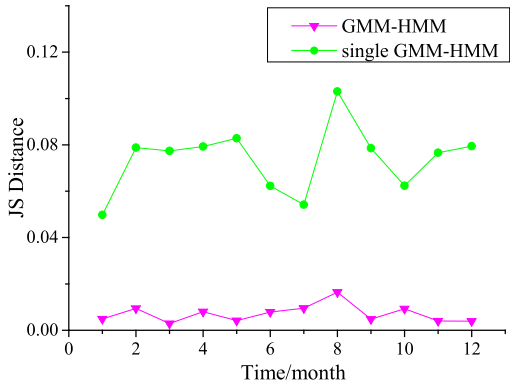


**FIGURE 11.** JS distance of the No. 4 wind farm generated by the GMM-HMM method for a single wind farm and multiple wind farms.

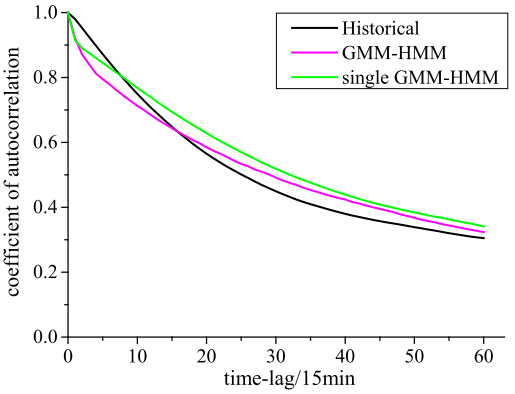


**FIGURE 12.** ACF of annual output of the No. 4 wind farm generated by the GMM-HMM method for a single wind farm and multiple wind farms.

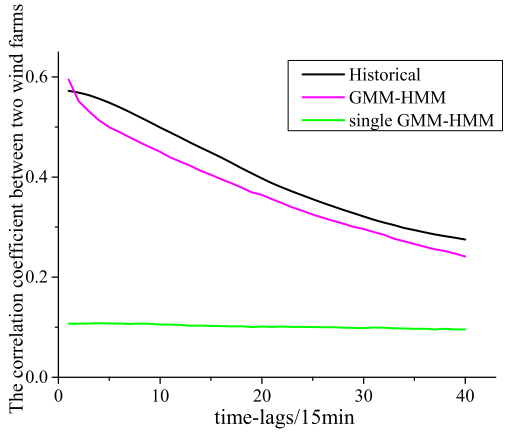


**FIGURE 13.** CCF of annual output between the No. 3 and No. 4 wind farms generated by the GMM-HMM method for a single wind farm and multiple wind farms.

According to the calculation results of the CCF index, the CCF of the annual output series generated by the GMM-HMM is slightly higher than that of the historical wind power output when the time delay is order 1 and has been slightly lower than that of the historical wind power output after the time delay is order 2. Since the other three methods do not consider spatial correlation, the CCF of the annual output series generated by them are all close to zero. The sum of the absolute values of the errors between the CCF of the five methods (discrete HMM method, Gaussian HMM method, ARIMA method, MCMC method, and GMM-HMM method) and the historical CCF of wind power output are 14.4, 2.4, 15.7, 14.9, and 1.4 respectively. Thus, the proposed method can accurately describe the variation characteristics of spatial cross-correlation between wind power output series with a time delay. Insufficient consideration of spatial correlation will lead to underestimation of volatility and overestimation of complementarity during planning, resulting in insufficient investment in flexibility resources.

### B. COMPARISON OF MAIN STATISTICAL CHARACTERISTICS OF THE GMM-HMM METHOD FOR A SINGLE WIND FARM AND MULTIPLE WIND FARMS

The statistical characteristics of the wind power output time series generated by the proposed GMM-HMM in the case of a single wind farm and multiple wind farms are compared. In Case 2, the proposed GMM-HMM method is used to generate the output series of a single wind farm for the five wind farms in Case 1. The HMM hidden state number is also set to 7, and the Gaussian distribution number of GMM is set to 11.

First, the JS distances of No. 4 wind farm generated by the GMM-HMM method of a single wind farm and multiple wind farms are calculated and compared as shown in Figure 11. The results show that the JS distance of No. 4 wind farm output generated by the GMM-HMM method for a single wind farm is larger than that generated by the GMM-HMM method for multiple wind farms.

The ACF of annual output of No. 4 wind farm generated by the GMM-HMM method for a single wind farm and multiple wind farms are calculated and compared with the ACF of historical series, as shown in Fig. 12. The sum of the absolute error of the ACF generated by the GMM-HMM method for a single and multiple wind farms and the ACF of historical series are 3.0 and 2.2, respectively.

The CCF of annual output between No. 3 and No. 4 wind farms generated by the GMM-HMM method for a single wind farm and multiple wind farms are calculated and compared with the CCF of historical series, as shown in Fig. 13. The GMM-HMM method for a single wind farm does not consider the spatial correlation among wind farms, so the CCF of the annual output series generated is close to 0.

### C. SENSITIVITY ANALYSIS OF PARAMETERS IN THE GMM-HMM

The number of hidden state and the number of Gaussian distribution are parameters in the model. Case 3 compares the statistical characteristics of wind power output series generated by the GMM-HMM under a different hidden state number $Q$ and mixed Gaussian distribution number $R$. Appropriate parameters are selected by comparing the

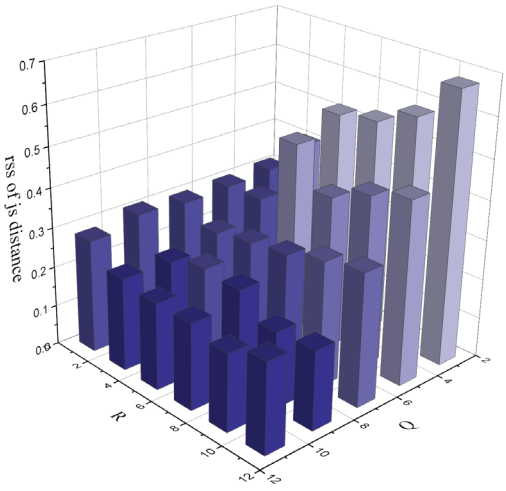


**FIGURE 14.** Residual sum of squares of the JS distance changing with Q and R.

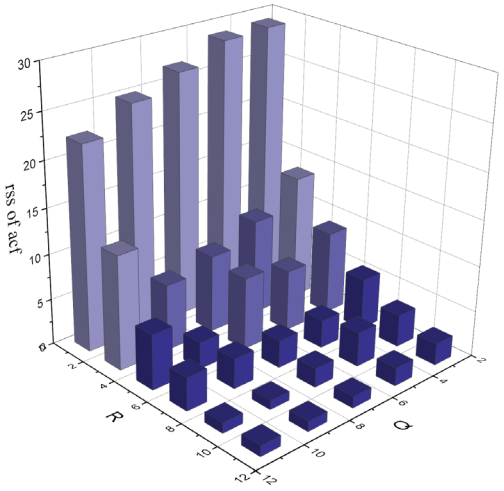


**FIGURE 15.** Residual sum of squares of ACF changing with Q and R.

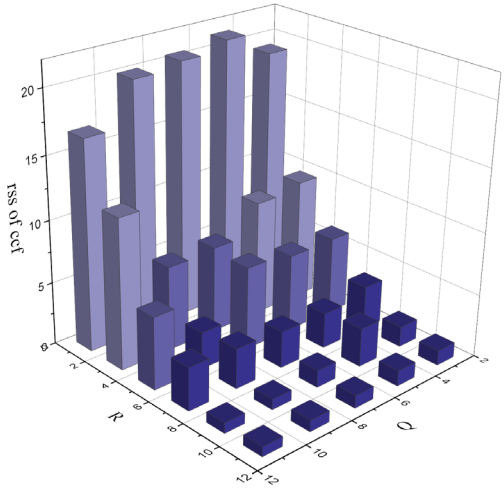


**FIGURE 16.** Residual sum of squares of CCF changing with Q and R.

residual sum of square (RSS) of the JS distance, ACF and CCF output of five wind farms in 12 months. Figure 14 compares the JS distance of the GMM-HMM under different $Q$ and $R$ values. Thus, at the same $R$, the JS distance decreases with an increase in $Q$, and the JS distance reaches a minimum when $Q$ is greater than 7 and $R$ is greater than 9.

Figure 15 and Figure 16 compare the residual sum of the ACF and CCF values of the GMM-HMM under different $Q$ and $R$ values. When $Q$ is the same, the larger $R$ is, the smaller the ACF and CCF values are, and the JS distance is smaller when $R$ is greater than 7.

As shown in the above three figures, the changes in $Q$ and $R$ have great influence on the statistical indicators; thus, $Q$ and $R$ have great influence on the performance of the proposed method. To achieve the best performance, a sensitivity analysis of parameters should be performed before using this method. According to the calculation results, the error of the above three indicators is minimal when $Q = 9$ and $R = 9$.

## V. CONCLUSION

This paper proposes a method for generating medium- and long-term correlated output time series of multiple wind farms based on the Gaussians mixture Hidden Markov model. The meteorological state is used as the hidden state variable, and the multidimensional wind farm output is used as the observation variable. The output time series is generated in two steps. First, the meteorological state is continuously sampled to generate a hidden state sequence, which is a Markov chain. After that, the output time series is finally obtained by sampling based on the GMM corresponding to different hidden states. At the same time, this method takes into account the spatiotemporal correlation of multiple wind farms and uses the GMM to fit the joint output distribution of multiple wind farms so that the generated output series is more consistent with the actual statistical characteristics. The generated typical annual wind power output time series with several years of historical data statistical features can be used for power system planning and medium- and long-term operation problems.

The example part is based on the simulation analysis of five wind farms in Jilin Province, China. The proposed method is compared with the common MCMC, ARIMA, discrete HMM and Gaussian HMM methods. The results show that the output series generated by the proposed method are closer to the historical output series in terms of the mean square error, mean absolute error, distance of probability distribution, autocorrelation and cross-correlation, which proves the advantages of the proposed method in reproducing these statistical characteristics. The sensitivity analysis of the effects of different hidden state numbers $Q$ and mixed Gaussian distribution numbers $R$ on the statistical characteristics of the series is also carried out. With increases in $Q$ and $R$, the error of the statistical index decreases.

In this paper, the discrete state transfer model is used for the chaotic meteorological state. Although this model has the advantage of simplicity in modeling, it may have the disadvantage of not being able to accurately describe the

nonlinear state transfer relationship. In future work, the nonlinear transfer relationship between factors such as the meteorological state and meteorological conditions in nonlinear dynamic systems can be studied.

## REFERENCES


[1] E. M. Gouveia and M. A. Matos, "Evaluating operational risk in a power system with a large amount of wind power," *Electr. Power Syst. Res.*, vol. 79, no. 5, pp. 734–739, May 2009, doi: 10.1016/j.epsr.2008.10.006.

[2] J. A. Carta, S. Velázquez, and J. M. Matías, "Use of Bayesian networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site," *Energy Convers. Manage.*, vol. 52, no. 2, pp. 1137–1149, Feb. 2011, doi: 10.1016/j.enconman.2010.09.008.

[3] H. Holttinen, P. Meibom, A. Orths, B. Lange, M. O'Malley, J. O. Tande, A. Estanqueiro, E. Gomez, L. Söder, G. Strbac, J. C. Smith, and F. van Hulle, "Impacts of large amounts of wind power on design and operation of power systems, results of IEA collaboration," *Wind Energy*, vol. 14, no. 2, pp. 179–192, Mar. 2011, doi: 10.1002/we.410.

[4] Y. Han and L. Chang, "A study of the reduction of the regional aggregated wind power forecast error by spatial smoothing effects in the maritimes Canada," in *Proc. IEEE Electr. Power Energy Conf.*, Aug. 2010, pp. 1–6, doi: 10.1109/EPEC.2010.5697199.

[5] A. Sturt and G. Strbac, "Time series modelling of power output for large-scale wind fleets," *Wind Energy*, vol. 14, no. 8, pp. 953–966, Nov. 2011, doi: 10.1002/we.459.

[6] S. Kennedy and P. Rogers, "A probabilistic model for simulating long-term wind-power output," *Wind Eng.*, vol. 27, no. 3, pp. 167–181, May 2003, doi: 10.1260/030952403769016654.

[7] X. Wang, P. Guo, and X. Huang, "A review of wind power forecasting models," *Energy Procedia*, vol. 12, pp. 770–778, Jan. 2011, doi: 10.1016/j.egypro.2011.10.103.

[8] X. Wang, G. Sideratos, N. Hatziargyriou, and L. H. Tsoukalas, "Wind speed forecasting for power system operational planning," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst.*, Sep. 2004, pp. 470–474.

[9] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 234–240, Mar. 2008, doi: 10.1109/TEC.2007.914174.

[10] D. Li, W. Yan, W. Li, and Z. Ren, "A two-tier wind power time series model considering day-to-day weather transition and intraday wind power fluctuations," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4330–4339, Nov. 2016, doi: 10.1109/TPWRS.2016.2531739.

[11] E. A. Denaxas, R. Bandyopadhyay, D. Patino-Echeverri, and N. Pitsianis, "SynTiSe: A modified multi-regime MCMC approach for generation of wind power synthetic time series," in *Proc. Annu. IEEE Syst. Conf. (SysCon)*, Apr. 2015, pp. 668–674, doi: 10.1109/SYSCON.2015.7116827.

[12] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "ARIMA-based time series model of stochastic wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, May 2010, doi: 10.1109/TPWRS.2009.2033277.

[13] S. Tian, Y. Fu, P. Ling, S. Wei, S. Liu, and K. Li, "Wind power forecasting based on ARIMA-LGARCH model," in *Proc. Int. Conf. Power Syst. Technol. (POWERCON)*, Nov. 2018, pp. 1285–1289, doi: 10.1109/POWERCON.2018.8601740.

[14] D. Bhaumik, D. Crommelin, S. Kapodistria, and B. Zwart, "Hidden Markov models for wind farm power output," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 533–539, Apr. 2019, doi: 10.1109/TSTE.2018.2834475.

[15] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu, "Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2239–2247, Mar. 2018, doi: 10.1109/TPWRS.2017.2737580.

[16] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic forecast for multiple wind farms based on regular vine copulas," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 578–589, Jan. 2018, doi: 10.1109/TPWRS.2017.2690297.

[17] S. Chowdhury, A. Messac, and L. Castillo, "Multivariate and multimodal wind distribution model based on kernel density estimation," *Energy Sustainability*, vol. 54686, pp. 2125–2135, Aug. 2011, doi: 10.1115/ES2011-54507.

[18] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 1074–1080, Jul. 2016, doi: 10.1016/j.ijforecast.2015.11.006.

[19] P. Li and C. Liu, "Modeling correlated power time series of multiple wind farms based on hidden Markov model," *Proc. CSEE*, vol. 39, no. 19, pp. 5683–5691, 2019.

[20] C. L. P. Lim, W. L. Woo, S. S. Dlay, and B. Gao, "Heartrate-dependent heartwave biometric identification with thresholding-based GMM–HMM methodology," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 45–53, Jan. 2019, doi: 10.1109/TII.2018.2874462.

[21] P. Swietojanski, A. Ghoshal, and S. Renals, "Revisiting hybrid and GMM-HMM system combination techniques," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6744–6748, doi: 10.1109/ICASSP.2013.6638967.

[22] B. G. Celler, P. N. Le, A. Argha, and E. Ambikairajah, "GMM-HMM-based blood pressure estimation using time-domain features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3631–3641, Jun. 2020, doi: 10.1109/TIM.2019.2937074.

[23] Y. Xiang, T. Wang, and Z. Wang, "Improved Gaussian mixture model based probabilistic power flow of wind integrated power system," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2019, pp. 1–5, doi: 10.1109/PESGM40551.2019.8973751.

[24] B. Schuster-Böckler and A. Bateman, "An introduction to hidden Markov models," in *Current Protocols in Bioinformatics*. Jul. 2007, pp. A.3A.1–A.3A.9, doi: 10.1002/0471250953.

[25] N. Ding and Z. Ou, "Variational nonparametric Bayesian hidden Markov model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 2098–2101, doi: 10.1109/ICASSP.2010.5495125.

[26] S. Adams, P. A. Beling, and R. Cogill, "Feature selection for hidden Markov models and hidden semi-Markov models," *IEEE Access*, vol. 4, pp. 1642–1657, 2016, doi: 10.1109/ACCESS.2016.2552478.

[27] H. Li, *Statistical Learning Methods*. Beijing, China: Tsinghua Univ. Press, 2012.

[28] G. W. Chang, H.-J. Lu, P.-K. Wang, Y.-R. Chang, and Y.-D. Lee, "Gaussian mixture model-based neural network for short-term wind power forecast," *Int. Trans. Electr. Energy Syst.*, vol. 27, no. 6, p. e2320, Jun. 2017, doi: 10.1002/etep.2320.

[29] L. Cao, Y. Shen, T. Shan, Y. Xia, J. Wang, and Z. Lin, "Bearing fault diagnosis method based on GMM and coupled hidden Markov model," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 932–936, doi: 10.1109/PHM-Chongqing.2018.00166.

[30] F. Najar, S. Bourouis, N. Bouguila, and S. Belghith, "A comparison between different Gaussian-based mixture models," in *Proc. IEEE/ACS 14th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2017, pp. 704–708, doi: 10.1109/AICCSA.2017.108.

[31] G. Xuan, W. Zhang, and P. Chai, "EM algorithms of Gaussian mixture model and hidden Markov model," in *Proc. Int. Conf. Image Process.*, vol. 1, Oct. 2001, pp. 145–148, doi: 10.1109/ICIP.2001.958974.

[32] Y. Tao, H. Chen, and C. Qiu, "Wind power prediction and pattern feature based on deep learning method," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Dec. 2014, pp. 1–4, doi: 10.1109/APPEEC.2014.7066166.

[33] H. Liu, H.-Q. Tian, C. Chen, and Y.-F. Li, "A hybrid statistical method to predict wind speed and wind power," *Renew. Energy*, vol. 35, no. 8, pp. 1857–1861, Aug. 2010, doi: 10.1016/j.renene.2009.12.011.

[34] F. Shahid, A. Zameer, A. Mehmood, and M. A. Z. Raja, "A novel wavenets long short term memory paradigm for wind power prediction," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 115098, doi: 10.1016/j.apenergy.2020.115098.


**YUFEI LI** received the B.S. degree from the College of Energy and Electrical Engineering, Hohai University, China, in 2019. She is currently pursuing the M.S. degree with the School of Electrical Engineering, Chongqing University, China.

**BO HU** (Member, IEEE) was born in Henan, China, in 1983. He received the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in 2010. He is currently a Professor with the School of Electrical Engineering, Chongqing University. His research interests include power system reliability and parallel computing techniques in power systems.

**TAO NIU** (Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2014 and 2019, respectively. He is currently an Assistant Professor with Chongqing University, Chongqing, China. His research interests include power system operation and optimization, power system reliability assessment, voltage security region, renewable generation integration, and reactive power analysis of hybrid AC/DC systems.

**SHENGPU GAO** (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering from the Hefei University of Technology, China, in 2018. He is currently pursuing the M.S. degree with Chongqing University, China. His research interests include power system optimization and renewable energy integration.

**JIAHAO YAN** (Student Member, IEEE) received the B.E. degree from the Hefei University of Technology, China, in 2016. He is currently pursuing the Ph.D. degree with Chongqing University, China. His research interests include power system risk assessment and optimization.

**KAIGUI XIE** (Senior Member, IEEE) received the Ph.D. degree in power system and its automation from Chongqing University, Chongqing, China, in 2001. He is currently a Full Professor with the School of Electrical Engineering, Chongqing University. He has authored and coauthored over 200 academic articles and six books. His main research interests include power system reliability, planning, and analysis. He is an IET Fellow. He is awarded the title of the National Science Fund for Outstanding Young Scholar. He is an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS and an Associate Editor of *IET Proceedings-Generation, Transmission and Distribution*.

**ZHOUYANG REN** (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in June 2014. He is currently an Associate Professor with the School of Electrical Engineering, Chongqing University. His research interests include reliability evaluation and planning of power systems, and power system data analysis.