IOP Conference Series: Earth and Environmental Science

Time Series Generation and Complex Correlation Assessment for Multiple Wind Farms

To cite this article: Shenzhi Xu and Bo Xu 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **295** 012068

IOP Conf. Series: Earth and Environmental Science **295** (2019) 012068 doi:10.1088/1755-1315/295/2/012068

Time Series Generation and Complex Correlation Assessment for Multiple Wind Farms

Shenzhi Xu[1*], Bo Xu[1]

[1]State Grid Energy Research Institute co., LTD, Beijing, 102209, China

[*]Corresponding author's e-mail: xushenzhi@sgeri.sgcc.com.cn

Abstract. Analysis and calculation of power system often requires a large number of medium and long-term wind power output series as data foundation. However, for most wind farms that have not been built or put into operation for a long time, the current output data is limited and it is difficult to support related research. It is necessary to use a limited amount of measured data to generate lots of wind power output time series that are similar to the actual data. Considering the concentrated development mode of wind power, this paper proposes a time series generation method for multiple wind farms based on the Markov Chain and Monte Carlo (MCMC) method and combined with high-dimensianal Markov process. The mixed Copula function fitting model is used to describe the spatial correlation of wind farms accurately. The measured wind power ouput is used to verify the results. It shows that the proposed method can simulate the output characteristics of single wind farm while maintaining the spatial correlation among different wind farms with geographical proximity

1. Introduction

In order to deal with environmental pollution and energy supply security issues, developing renewable energy has become the consensus of all countries in the world. With increasing proportion of wind power accessed to the grid, the randomness and volatility of its output have also significantly increased impacts on the power system[1][2]. On the other hand, the wind activity at the same time is similar for wind farms with geographical proximity. Thus, their output will have spatial correlation[3]. Spatial correlation appears to be that the output of each wind farm having the same trend. This will make the cumulative fluctuation of total output of wind farms more intense, which will have more serious impacts on safe and stable operation of the power system.

Assessing impacts of wind power on the system often requires large amounts of wind power output data as a basis[4]. However, most existing wind power output data have problems such as limited data volume or insufficient recording time. It is especially important to use a suitable method to generate time series of wind farms. That is, use limited measured wind speed or wind power series to generate a large number of wind power series with similar statistical features to the original series. At the same time, how to ensure that the generated wind power time series has similar correlation characteristics with actual data of different wind farms is a problem that needs to be considered. At present, most of the research at home and abroad consider the correlation among wind speeds of different wind farms[5][6]. There are few time series generation methods that directly consider the correlation among output of multiple wind farms.

Based on the time series generation method for single wind farm[7], this paper further proposes time series generation for multiple wind farms with multivariate first-order Markov property. This method can simultaneously simulate complex spatially related structures among multiple wind farms.

IOP Conf. Series: Earth and Environmental Science **295** (2019) 012068 doi:10.1088/1755-1315/295/2/012068

Copula function is used to describe the correlation among output of wind farms. The Euclidean distance between the fitting Copula function and the empirical Copula function is used to evaluate the quality of generated series simulating the correlation structure of multiple wind farms.

2. Time series generation method for multiple wind farms based on Markov process

2.1 Multivariate first-order Markov chain

Time series generation for multiple wind farms is equivalent to considering stochastic process of multiple variables simultaneously. Multivariate first-order Markov processes can be compared to univariate high-order Markov processes. That is, the state of random variable at the next moment is not only related to its state at the previous moment, but also related to the states of other random variables at the previous moment. If the conditional probability of *n* random variables satisfy the equation (1), the *n* random variables are said to satisfy the first-order Markov property of multivariate.

$$\begin{aligned} &P\left\{X_1^{t+1}=I_1^{t+1},\cdots,X_n^{t+1}=I_n^{t+1}\middle| X_1^0=I_1^0,\cdots,X_n^0=I_n^0,\cdots,X_1^t=I_1^t,\cdots,X_n^t=I_n^t\right\} \\ &=P\left\{X_1^{t+1}=I_1^{t+1},\cdots,X_n^{t+1}=I_n^{t+1} \mid X_1^t=I_1^t,\cdots,X_n^t=I_n^t\right\} \end{aligned} \tag{1}$$

In the equation, t represents time, $t \in T$. T is a discrete time set, $T=\{0,1,2,\ldots\}$. I_n^t represents the state value of the random variable n at time t.

The transition probability matrix P_M is used to quantify the probability characteristics of the Markov chain transition between different states. Taking the first-order Markov process of bivariate as an example, the size of corresponding transition probability matrix P_M is $(N_1N_2)\times(N_1N_2)$. N_1 and N_2 are the number of possible values of random variables state. The probability that variable 1 is state i and variable 2 is state j at the time t, and then variable 1 is state k and variable 2 is state m at the time $t+1$ corresponds to the element of the $[(i-1)N_2+j]$th row and the $[(k-1)N_2+m]$th column of P_M, as shown in the equation (2).

$$P_M\left((i-1)N_2+j,(k-1)N_2+m\right)=p\left(X_1^{t+1}=k,X_2^{t+1}=m \mid X_1^t=i,X_2^t=j\right) \tag{2}$$

2.2 Time series generation method for multiple wind farms

Time series generation method for multiple wind farms based on Markov process is similar to method for single wind farm. Taking two wind farms as an example, the specific steps are as follows:

1) Determine the state number of each wind farm by using the principle of optimal state number selection in [7].

2) Define the values range for state of each wind farm, convert the original measured series into state series, and statistically calculate the empirical distribution function of measured data in the corresponding range of each state.

3) Generate transition probability matrix P_M from the original state series.

4) Generate cumulative transition probability matrix.

5) Generate the initial state of each series randomly, and generate simulated state series of each wind farm with the Monte Carlo method according to the cumulative transition probability matrix.

6) Convert the simulated state series into specific wind power output series by using random power generation method based on distribution function within the range of values corresponding to each state.

3. Description method for spatial correlation of wind farms with Copula function

3.1 Introduction to Copula function

Sklar theorem[8]: If F is a joint distribution function of q dimensional random variables, and the edge distribution function corresponding to each variable is $F_1(x_1),\ldots, F_q(x_q)$, there exists the Copula function C which satisfies:

$$F\left(x_1,x_2,\cdots,x_n\right)=C\left(F_1\left(x_1\right),F_2\left(x_2\right),\cdots,F_n\left(x_n\right)\right) \tag{3}$$

Taking partial derivative of the equation and let $u_i=F_i(x_i)$, the joint density function is

$$f\left(x_1,x_2,\cdots,x_q\right)=c\left(u_1,u_2,\cdots,u_q\right)\prod_{i=1}^{q} f_i\left(x_i\right) \tag{4}$$

In the equation, $f_i(x_i)$ is the edge density function and c is the Copula density function. That is, any multiple dimensional joint distribution density function can be expressed as the product of several edge density functions and Copula density function.

3.2 Description method based on Copula function

Taking two wind farms as an example, in order to obtain the Copula density function corresponding to their relevant structure, the output series are first transformed into series uniformly distributed within the interval [0,1] and then needs to calculate the joint probability density function. In this case, the Copula density function is equal to the joint probability density function.

Finding the distribution density function problem of random variables from sample data is the first problem to be solved. The kernel density estimation method is used to study the probability distribution of wind power output in this paper. There is no need to make a distribution hypothesis when performing kernel density estimation, and the estimate of any point only depends on sample data. Estimation accuracy depends on the choice of kernel function and computation window width. When the calculation window width is constant, the type of the kernel function has little effect on the result[9]. In this paper, the calculation window width is taken as 0.01 and the Gaussian kernel function is selected, as shown below.

$$K\left(x\right)=\frac{1}{\left(2\pi\right)^{1/2}}\exp\left(-\frac{x^2}{2}\right) \tag{5}$$

The Corresponding kernel distribution is

$$F\left(x\right)=\int_{-\infty}^{x}\hat{f}\left(\tau\right)d\tau=\frac{1}{l}\sum_{i=1}^{l}\Phi\left(\frac{x-x_i}{h}\right) \tag{6}$$

In the equation, Φ is the standard normal distribution, l is sample size, h is calculation window width, x_i is sample data and $\hat{f}\left(\tau\right)$ is kernel density estimation for sample data. The equation (6) can be used to calculate the edge distribution of wind farm output, that is, the series uniformly distributed within the interval [0, 1].

3.3 Mixed Copula function and parameter estimation

In order to ensure the comprehensiveness of analysis, mixed Copula function is used instead of single Copula function for analysis. The mixed Copula density function is shown below.

$$c\left(u,v,\boldsymbol{\theta}\right)=\lambda_1 c_C\left(u,v,\theta_1\right)+\lambda_2 c_G\left(u,v,\theta_2\right)+\lambda_3 c_F\left(u,v,\theta_3\right) \tag{7}$$

In the equation, λ_1, λ_2 and λ_3 are weight coefficients satisfying $0\le\lambda_1, \lambda_2, \lambda_3\le 1$ and $\lambda_1+\lambda_2+\lambda_3=1$. $\boldsymbol{\theta}$ is the parameter of mixed Copula function. θ_1, θ_2 and θ_3 are the correlation parameters of Clayton Copula function, Gumbel Copula function and Frank Copula function, respectively.

When performing parameters estimation, penalty function is introduced so that the weight coefficients can be penalized to remove some inconspicuous Copula functions. Constructing penalty likelihood function as follows.

$$PL\left(\boldsymbol{\theta}\right)=L\left(\boldsymbol{\theta}\right)\text{-}l\sum_{i=1}^{S}p_\gamma\left(\lambda_i\right) \tag{8}$$

$$L\left(\boldsymbol{\theta}\right)=\ln\left[\lambda_1 c_C\left(u,v,\theta_1\right)+\lambda_2 c_G\left(u,v,\theta_2\right)+\lambda_3 c_F\left(u,v,\theta_3\right)\right] \tag{9}$$

In the equations, S is the number of Copula functions. $p_\gamma(\ )$ is penalty function and γ is smooth parameter which can be used to control complexity of model. The SCAD penalty function is selected in this paper, its formula as follows.

$$p_\gamma(\lambda)=\gamma \mathrm{I}(\lambda \le \gamma)+\frac{(a\gamma-\lambda)}{a-1}\mathrm{I}(\lambda>\gamma) \tag{10}$$

I( ) is characteristic function and a>2, λ>0. Since the unknown parameters in equation (8) cannot be solved directly with the method of partial bias, the Expectation Maximization (EM) algorithm is used for parameter estimation.

Taking the actual measured data of two adjacent wind farms in Texas, USA as an example. Table 1 shows the basic information of two wind farms. Figure1 is the frequency histogram of their output. Figure 2 is density function image of the fitting mixed Copula function and the fitting function is

$$c(u,v)=0.4197c_C(u,v,4.3264)+0.1448c_G(u,v,1)+0.4355c_F(u,v,19.7851) \tag{11}$$

Table 1. Basic information of data.

Serial number	*Name*	Capacity	Data interval	Data length	*Time span*
1st wind farm	Buff1and2	353.1MW	1 min	259200	180 days
2nd wind farm	Callahan	114MW	1 min	259200	180 days

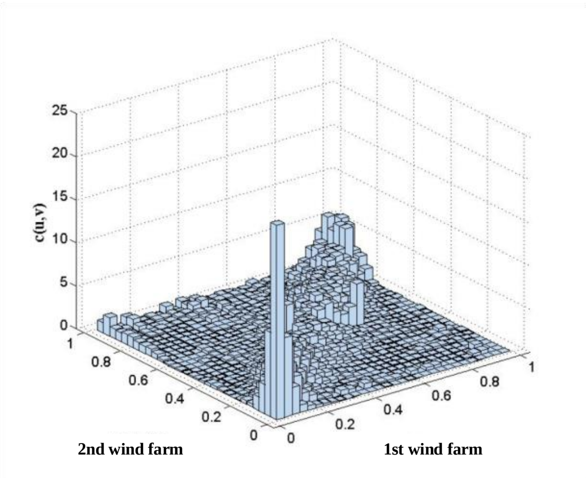

Figure 1. Binary frequency histogram of the measured series

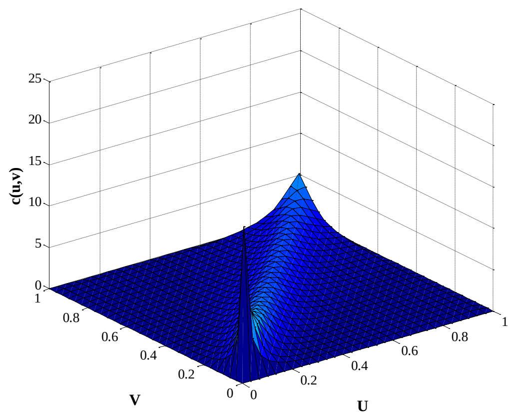

Figure 2. Image of mixed Copula density function.

It can be seen from the figures that the mixed Copula density function image is in good agreement with the binary frequency histogram of the actual wind farms output. Introducing empirical Copula function to evaluate the pros and cons of mixed Copula function model. If (x_i, y_i) is sample of two dimensional random variables and $\hat{F}(x)$, $\hat{G}(y)$ are the empirical distribution function of X and Y, the empirical Copula function values are calculated as below.

$$\hat{C}(u,v)=\frac{1}{l}\sum_{i=1}^{l}\left[I\left(\hat{F}(x_i)\le u\right)\cdot I\left(\hat{G}(y_i)\le v\right)\right] \tag{12}$$

After obtaining values of empirical Copula function, the fitting effect can be evaluated by examining the Euclidean square distance between the fitting Copula function and the empirical Copula function. The smaller the distance, the more accurate the description of correlation structure of the binary random variable with Copula function.

4. Case verification

In order to verify the effectiveness of time series generation method for multiple wind farms proposed in this paper, also taking the two wind farms in table 1 as an example. The optimal state numbers for series generation of 1st and 2nd wind farm are 15 and 22 respectively with the method mentioned in [7]. The result of series generation for single wind farm without considering correlation is taken as a comparative analysis.

Table 2 gives a comparison of the statistical characteristics of the series generated for multiple wind farms by the method considering correlation and not. The evaluation index of mean and standard deviation are relative error. The evaluation index of Probability density function (PDF) and Autocorrelation function (ACF) are the sum of squared residuals of the corresponding curves and the characteristic curves of original series. It can be seen from table 2 that the wind power series generated by the multiple wind farms time series generation method considering correlation is similar to the results obtained by the single wind farm time series generation method in the statistical properties such as mean, standard deviation and PDF. But it is slightly worse than the results of single wind farm in the statistical characteristics of ACF.

Table 2. Comparison of statistical characteristics

Statistical characteristics	Considering correlation			Not considering correlation		
	1st farm	2nd farm	mean	1st farm	2nd farm	mean
Mean	2.04%	2.21%	2.13%	2.80%	3.28%	3.04%
Standard deviation	1.80%	1.57%	1.69%	1.18%	1.94%	1.56%
PDF	6.93e-5	7.12e-4	3.91e-4	8.95e-5	6.95e-4	3.92e-4
ACF	1.17	1.03	1.10	0.52	0.41	0.47

Figure 3 and figure 4 show the binary frequency histogram corresponding to the time series generated for multiple wind farms by the method considering correlation and not. It can be seen intuitively from the figure that the correlation structure between the generated series is more similar to the original series when it is considering correlation. But there is almost no correlation between the generated series when correlation is not considered. Evaluating the pros and cons of time series results for multiple wind farms by examining the Euclidean square distance between the fitting Copula function of the measured series and the empirical Copula function of the generated series. When the correlation is considered, the value is 1.08×10^3. And when the correlation is not considered, the value is 4.22×10^3. It can be seen that the time series generation method considering correlation is more effective in simulating the correlation of multiple wind farms.

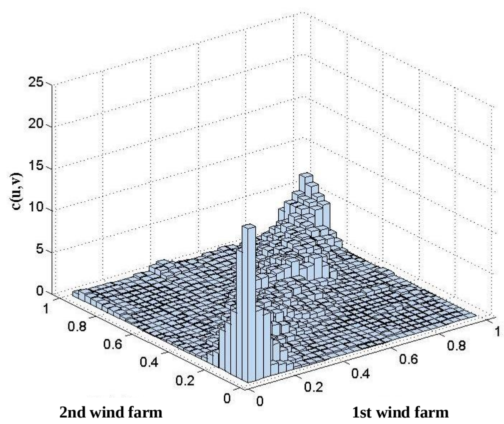

Figure 3. Binary frequency histogram of the generated series (correlation is considered)

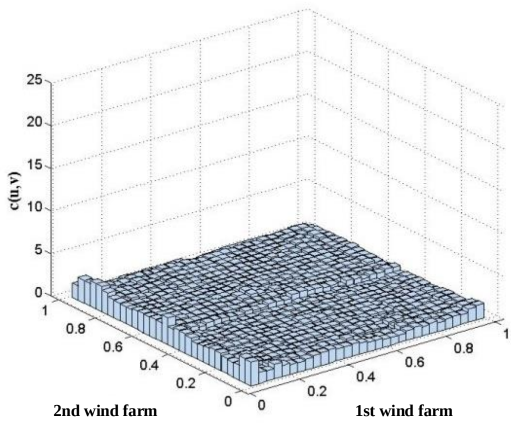

Figure 4. Binary frequency histogram of the generated series (correlation is not considered)

Therefore, the time series generation method for multiple wind farms based on the multivariate first-order Markov chain can account for the complex spatial correlation among wind farms while preserving the advantages of the single wind farm method mentioned in [7].

5. Conclusion

The time series generation method for multiple wind farms based on Markov process proposed in the paper can simulate the statistical characteristics of actual wind power output well. At the same time, this method can accurately simulate the complex spatial correlation structure among wind farms with geographical proximity. It can provide sufficient wind power output data for quantitative assessment of the impacts of large-scale wind power on the power system.

Acknowledgments

This work is supported by Science and Technology Project of State Grid Corporation of China (SGTYHT/17-JS-199): Theory, Model and Road Study on Energy Supply Side Structure Optimization and Adjustment.

References

[1] Bouffard F, Galiana F D. (2008) Stochastic security for operations planning with significant wind power generation. IEEE Transactions on Power Systems, 23(2):306-316.
[2] Bu S Q, Du W, Wang H F, et al. (2012) Probabilistic Analysis of Small-Signal Stability of Large-Scale Power Systems as Affected by Penetration of Wind Generation. IEEE Transactions on Power Systems, 27(2):762-770.
[3] Xue Y S, Chen N, Wang S M, et al. (2017) Review on wind speed prediction based on spatial correlation. Automation of Electric Power System, 41(10): 161-169.
[4] Silva A M L L D, Sales W S, Manso L A D F, et al. (2010) Long-Term Probabilistic Evaluation of Operating Reserve Requirements With Renewable Sources. IEEE Transactions on Power Systems, 25(1): 106-116.
[5] K. Philippopoulos, D. Deligiorgi. (2009) Statistical simulation of wind speed in Athens, Greece based on Weibull and ARMA models. International Journal of Energy Environment & Economics, 3(4): 151-158.
[6] Wu F, Kong W Y, Zhou Y, et al. (2016) A wind speed modeling method for multiple wind farms considering variation regularities. Power System Technology, 40(7): 2038-2044.
[7] Xu S Z, Ai X M, Li J M, et al. (2016) The optimal state number selection of wind power series generated through MCMC method. In: 12th IET International Conference on AC and DC Power Transmission. Beijing. pp: 1-6.
[8] Sklar, A. (1959) Fonctions de repartition a n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8, 229–231.
[9] Epanechnikov V A. (2006) Non-Parametric Estimation of a Multivariate Probability Density. Theory of Probability & Its Applications, 14(1):153-158.