Modelling and Simulating the Spatio-Temporal Correlations of Clustered Wind Power Using Copula

Ning Zhang†, Chongqing Kang*, Qianyao Xu*, Changming Jiang**, Zhixu Chen** and Jun Liu**

Abstract – Modelling and simulating the wind power intermittent behaviour are the basis of the planning and scheduling studies concerning wind power integration. The wind power outputs are evidently correlated in space and time and bring challenges in characterizing their behaviour. This paper provides a methodology to model and simulate the clustered wind power considering its spatio-temporal correlations using the theory of copula. The sampling approach captures the complex spatio-temporal connections among the wind farms by employing a conditional density function calculated using multidimensional copula function. The empirical study of real wind power measurement shows how the wind power outputs are correlated and how these correlations affect the overall uncertainty of clustered wind power output. The case study validates the simulation technique by comparing the simulated results with the real measurements.

Keywords: Clustered wind power, Copula, Probabilistic modelling, Spatio-temporal correlations, Wind power output simulation

1. Introduction

The integration of large-scale wind power puts forward significant challenges to power system coping with uncertainty sources. The stochastic and fluctuating nature of wind power brings about a series of issues on power system reserve, reliability and operating cost, etc [1].

The uncertainty modelling of wind power is the basis of the related studies [2]. For large-scale clustered wind farms, the spatio-temporal correlations of wind have significant impacts on the overall uncertainty of wind power outputs [3], and in turn, their impacts on the power system planning and operation [4]. From the spatial point of view, a weaker correlation would make the clustered wind power have less disturbance to the power system and a more reliable generating output, i.e., less variance of power flows [5] and a higher capacity credit [6]. This is known as the “smoothing effect” [4]. Such effect also appears from the temporal point of view: the dispersion of wind farms would result in a smaller fluctuation. The smoothing effect brings down the requirement of flexibility, i.e., spinning reserve requirement, ramping rate demand and transmission capacity margin and therefore reduces the system operation cost [7], [8], and the demand for energy storage [9]. Such effect is also found in the forecast errors of wind power [10].

† Corresponding Author: State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing, China. (ningzhang@mail.tsinghua.edu.cn)
* State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing, China. (cqkang@tsinghua.edu.cn, xuqy06@mails.tsinghua.edu.cn)
** Power Control Center of North China, State Grid, China. Received: September 13, 2013; Accepted: September 24, 2013
functions with analytical formulas could be used to model different types of dependencies, i.e., Archimedean copulas, elliptical copulas and extreme-value copulas. Combined with varieties of marginal distributions, the theory enables the modelling of much more complex joint distributions. More description of copula theory can be found in [25].

2.2 Using copula to model wind power with spatio-temporal correlations

Using the theory of copula to model the spatio-temporal correlations of wind power from historical measurements can simply follow the two steps below.

(1) Model the marginal distribution of wind power output of a single wind farm. A variety of published techniques can be used in this step, i.e., regressing using Beta distribution or recording the distribution function numerically.

(2) Model the spatio-temporal correlations using copula. The historical wind power outputs are then transformed into uniform distributed time series by the marginal distribution obtained in step 1. They are used as basic observation for identifying the copula function. Classical fitting procedure can be utilized in choosing the appropriate copula functions and parameters. Usually the common used copulas are competent for such modelling, i.e., $t$ copula and Gaussian copula in the elliptical copula category. Especially, the parameters of these copula has a fixed relationship with the Kendall rank correlation coefficient between the time series and are easy to estimate. For the selection of optimal copula function, various criteria are possible, i.e., P-P plot [27], Akaike Information Criterion (AIC), or maximum likelihood estimation (MLE) [25]. For the detail of copula function modelling, readers can refer to [25]-[27].

The time series involved in the copula modelling can be the synchronized wind power outputs from different sites (spatial correlation) or the same site with different time lags (temporal correlation), or a mixture of above (spatio-temporal correlation). Fig. 1 illustrates the modelling of two wind farms located 100 km apart using scatter plots.

![Fig. 1. The joint distributions of the output of two wind farms (left) and their copula (right)](image-url)
The left figure shows the joint distribution of the outputs while the right one shows its copula obtained using Eq. (2). According to the work of Zhang [24], the copula within the wind power could be well fitted by $t$ copula.

For the wind farms that are lack of historical measurements (newly built or under planning), the parameters in their copula functions could be estimated empirically according their geographical locations. The characteristics of how the parameters of copula vary with the distances between wind farms are discussed in detail in Section 4.

3. Wind Power Output Simulation Using Copula

3.1 Conditional modelling using copula

Taking the second derivative of both side of Eq. (2), we get:

$$f_{xy}(x,y) = F_{xy}(x,y) = c\left(F_x(x), F_y(y)\right) f_x(x) f_y(y) \quad (3)$$

where $f_x$ and $f_y$ are the marginal density of $x$, $y$ respectively, $c$ is the density function of $C$ and is named copula density function. According to the definition of conditional probabilistic density function (CPDF), the probability density of $x$ conditional to $y$ can be formulated using Eq. (3) as:

$$f(x | y) = \frac{f_{xy}(x,y)}{f_y(y)} = c\left(F_x(x), F_y(y)\right) f_x(x) \quad (4)$$

Eq. (4) shows that the CPDF of $x$ can be calculated by multiplying its marginal density and the according copula density function. This formula facilitates the calculation of conditional density for higher dimension cases, i.e., for multiple wind farms.

3.2 Methodology

The simulation methodology of wind power time series is proposed using the above mentioned conditional density formulation and the spatio-temporal modelling of wind power proposed in Section 2. The key of this methodology is to take into consideration the spatial and temporal correlation simultaneously through each sampling of the wind power output. The conditional density sampling method is used here to simulate the output of each wind farm successively and the conditional densities used in the sampling are calculated using the copula.

Suppose $F(X', X'^{-1})$ and $f(X', X'^{-1})$ are the joint distribution and the density of $N$ clustered wind farm outputs in two adjacent time intervals, i.e., one hour. $X'=(x'_1, x'_2, ..., x'_N)$, $X'^{-1}=(x'^{-1}_1, x'^{-1}_2, ..., x'^{-1}_N)$ denote the output of wind farms at $t$ and $t-1$. The copula function of $F(X', X'^{-1})$ can be modelled and marked as $C(U,V)$, where $U = \left[F(x'_1), F(x'_2), ..., F(x'_N)\right]$ and $V = \left[F(x'^{-1}_1), F(x'^{-1}_2), ..., F(x'^{-1}_N)\right]$. The dependence structure in $C(U,V)$ is:

$$\Omega = \begin{bmatrix} \Omega_s & \Omega_{s,t} \\ \Omega_{s,t}^\top & \Omega_t \end{bmatrix} \quad (5)$$

where $\Omega_s$ denotes the spatial correlation and $\Omega_{s,t}$ denotes the spatio-temporal correlation between time $t$ and $t-1$. Such formulation of copula links the wind power outputs from both spatial and temporal domains.

Knowing the wind power outputs on time $t-1$, the joint density of wind power outputs on time $t$ could be calculated using Eq. (4) as:

$$f(X' | X'^{-1}) = \frac{f(X', X'^{-1})}{f(X'^{-1})} = \frac{c(U,V) \prod_{i=1}^{N} f(x'_i)}{c(V) \prod_{j=1}^{N} f(x'^{-1}_j)} \quad (6)$$

where $c(U,V)$ is the copula density function of $C(U,V)$ and $c(V)$ is the copula density function that only considers the spatial correlation.

The marginal density of a single wind farm i.e., the first wind farm, is:

$$f(x'_1 | X'^{-1}) = \int_{x'_1} \cdots \int_{x'_2} f(X' | X'^{-1}) \, dx'_1 \cdots dx'_2$$

$$= \int_{x'_1} \cdots \int_{x'_2} c(U,V) \prod_{i=1}^{N} f(x'_i) dx'_1 \cdots dx'_2 \quad (7)$$

In Equation (7), each term of $f(x'_i) \, dx'_i$ equals to $dF(x'_i)$, $c(V)$ is constant, $c(U,V)$ can be rewritten as:

$$c(U,V) = \frac{C(F(x'_1), ..., F(x'_N), F(x'^{-1}_1), ..., F(x'^{-1}_N))}{\partial F(x'_1) \cdots \partial F(x'_N) \cdots \partial F(x'^{-1}_1) \cdots \partial F(x'^{-1}_N)} \quad (8)$$

Therefore Eq. (7) can be rewritten as:

$$f(x'_1 | X'^{-1}) = \frac{1}{c(V)} \int_{x'_1} \cdots \int_{x'_2} \frac{C(F(x'_1), ..., F(x'_N), F(x'^{-1}_1), ..., F(x'^{-1}_N))}{\partial F(x'_1) \cdots \partial F(x'_N) \cdots \partial F(x'^{-1}_1) \cdots \partial F(x'^{-1}_N)} \cdot f(x'_1) \cdots f(x'_2) \cdots \partial F(x'_1) \cdots \partial F(x'_N) \cdots \partial F(x'^{-1}_1) \cdots \partial F(x'^{-1}_N) \quad (9)$$
According to the characteristic of copula, the following equation holds:

\[
C(F(x_1^t), 1,...,1, F(x_{i-1}^t),...,F(x_N^t)) = C(F(x_1^t), F(x_{i-1}^t),...,F(x_N^t))
\]

Eq. (9) can be further written as:

\[
f(x_i^t|X^{t-1}) = \frac{f(x_i^t) C(F(x_i^t), F(x_{i-1}^t),...,F(x_N^t))}{c(V) \frac{\partial F(x_i^t)}{\partial x_i} \frac{\partial F(x_{i-1}^t)}{\partial x_{i-1}}\frac{\partial F(x_N^t)}{\partial x_N}}
\]

\[
= \frac{c(F(x_i^t), V)}{c(V)} f(x_i^t)
\]

Eq. (11) suggests that the density of a single wind farm output conditional to the outputs of all the clustered wind farms can be calculated by its own marginal density with a multiplier formulated by copula density function. \(c(V)\) is the \(N\)-dimensional copula density function that only considers the spatial correlation of \(X^{t-1}\), while \(c(F(x_i^t), V)\) is \(N+1\)-dimensional copula density function that considers the spatio-temporal correlation of \([x_i^t, X^{t-1}]\). It should be noted that \(f(x_i^t|X^{t-1})\) is unconditional to \(x_i^t, x_j^t, ..., x_N^t\). If \(x_i^t, x_j^t, ..., x_{i-1}^t\) are known, the conditional density \(x_i^t\) could be calculated by:

\[
f(x_i^t|x_i^t, x_j^t, ..., x_{i-1}^t, X^{t-1}) = \frac{c(F(x_i^t), F(x_j^t), ..., F(x_{i-1}^t), F(x_N^t), V)}{c(F(x_i^t), F(x_j^t), ..., F(x_{i-1}^t), V)} f(x_i^t)
\]

The derivation procedure of Eq. (12) is similar with that of Eq. (11).

Given an initial values of wind power outputs on \(t-1\), the outputs of \(t\) can be sampled recursively using the conditional density in Eqs. (11) and (12).

### 3.3 Simulation procedure

The procedure of simulating the spatio-temporal correlated wind power output can be summarized as follows:

1. Model the marginal distribution of the output of each wind farm.
2. Model the spatio-temporal dependency of the clustered wind farm outputs in two adjacent time intervals using copula.
3. Set the initial output of all wind power on \(t=1\), i.e., supposing all the \(x_1^t, x_2^t, ..., x_N^t\) equal to 20% of the installed capacity.
4. Sample the output of the first wind farm on \(t=2\) with the conditional density in Eq. (11), namely \(x_i^t\). Any probability density based sampling technique can be used, i.e., inverse transformation method.
5. Sample \(x_i^t\) using Eq. (12) and the known \(x_i^t, x_j^t, ..., x_{i-1}^t\). Repeat this procedure while \(t = 2 \sim N\), recursively.
6. Repeat (4) and (5) until the required length of time.

### 4. Evaluating the Spatio-temporal Correlations: An Empirical Study

This section carries out the empirical study using the wind power data of the real world. The purpose of this section is to demonstrate that, for clustered wind farms, to what extent the outputs are correlated and how this correlation affects the overall uncertainty. The findings and results in this section will act as the basic parameters for the simulation case study in Section 5.

#### 4.1 Source of the data and basic settings

The wind power and the according forecast data is from the wind integration datasets of NREL [28], which contains a consistent set of wind data of more than 30000 sites over the US, with ten-minute time resolution from 2004 to 2006. As shown in Fig. 2, we chose a series of sites along the east coast of US. In these sites the wind speeds are not actual measurements and is actually calculated by downscaling the reanalysis data of the mesoscale model in MASS v.6.8 with 2 km² resolution. Assuming 170 MW rated wind power for each sites, the wind speeds were transformed into wind power using IEC standard wind turbine curves. The day-ahead forecast is performed using a forecast tool named SynForecast based on a statistical model. The detail of this data set can be found in [29]. Though the wind data chosen in this case study is somehow simulated, it is derived from the real meteorology measurements and contains a similar uncertainty characteristics with actual wind farms.

#### 4.2 The spatio-temporal correlations

Since the modelling of the marginal distributions of Fig. 2. Sites chosen for the measurement of wind power and the forecast power
single wind farms has been widely discussed, only the spatio-temporal correlations are focused. $t$ copula is used in this paper, which is able to capture the correlations around the extreme value. The key parameter of $t$ copula is its correlation coefficient matrices. Its elements are in the range from -1 to 1 weighting the extent of correlation. It can be determined by the Kendall rank correlation coefficient of the input data [26]. It should be noted that this correlation coefficient is not exactly the same as the well known linear correlation coefficient, since $t$ copula represents a non-linear dependence.

Fig. 3 shows the spatial correlation coefficients of each two-site-pair for both wind power outputs and day ahead forecast errors. These coefficients are plotted with the accordant distances between the sites. It is clear from the results that the spatial correlations of both the outputs and forecast errors decrease as wind farms locate further away from each other. Especially, the wind farms within the distance of 50 km would have such strong correlation (around 0.9) that the smoothing effect could be ignored and could be considered integrally as one single wind farm. Beyond this distance, the wind farms should be considered separately. The spatial correlation of the forecast errors is weaker than that of the outputs for the same sites pair. This suggests that the smoothing effect of forecast errors would be more evident which could benefit the system in terms of declining the overall forecast errors.

The relationship between the correlation coefficient and the distance was regressed using the exponential function. The damping ratio of the correlation coefficient for the outputs and forecast errors are 0.0018 and 0.0035 km$^{-1}$, respectively. Such relationship could be used empirically to estimate the spatial correlations of planned wind farms based on their locations.

Fig. 4 shows the temporal correlation coefficients of the wind power outputs for both single site and aggregation of all sites. Similarly, coefficients are plotted with the time lags increases and the damping ratio of the first 20 hours is regressed using exponential function. As expected, the temporal correlation coefficient of the aggregate output damps slower than that of the single site, which reveals in part the smoothing effect of clustered wind power.

Fig. 5 shows the joint spatio-temporal correlation coefficients. The results show that the outputs of wind farms with adjacent locations and time periods have relatively strong correlations. This results suggest that the similarities of the outputs might be used as redundant information to cross-check the short term forecast, which might contribute to its accuracy. However, the effectiveness of such connection is limited within a certain range of space and time. For instance, taking the benchmark of the correlation coefficients at 0.5, the limit would be 400 km in space and 4 hours in time.

4.3 The overall uncertainty characteristic

The purpose of this subsection is to evaluate the impact of spatio-temporal correlation on the overall uncertainty of clustered wind power. Three common indices were chosen, namely the aggregate output, day ahead forecast error and the hourly variation. For each index, statistics were performed at four different scales of area along the coast of US, which are of the length of 2 km (single wind farm), 20
km, 150 km and 500 km, respectively.

Results are shown in Figs. 6-8, respectively. The results confirm the argument in Section 3.1 that the wind farms within 20 km area can hardly see the smoothing effect, because the distributions of the three indices of the wind power from 2 km and 20 km are very close.

For the aggregate output shown in Fig. 6, the wind power outputs are more concentrated when the wind farms are geographically wider dispersed, i.e., in 500 km area. The most evident changes are the reduction of the probabilities of the minimum output and maximum output. This suggests that the dispersion of wind farms would largely reduce the risks caused by the extreme wind farm outputs.

Figs. 7 and 8 show the benefits that the dispersion brings, in terms of reducing the relative forecast errors and variations. For the 500 km case, the relative average forecast error declines from 13.88% to 9.74% while the relative average hourly variation declines from 6.59% to 4.18%. Such benefits will shrunk when the dispersed area is 150 km or less.

5. Simulating Clustered Wind Power Outputs: Illustrative Example and Case Study

This section uses the proposed simulation methodology to reproduce the wind power outputs studied in Section 4. The parameters of copula and the according marginal distributions obtained by Section 4 were used as the basic data. Firstly a simple example of two wind farms is carried out to illustrate how this approach is performed. Secondly the simulation of all the ten wind farms is carried out. Finally the approach is tested on four wind farms located in North China Grid. The stochastic behaviours of simulated data and actual measurements are compared to validate the simulation methodology. The simulation results are also compared with those obtained by two conventional simulation techniques: ARMA model [17] and Markov chain [18].

5.1 Illustrative example

The outputs of two wind farms were simulated according to the stochastic characteristics of the northernmost two sites. The simulation runs for one year with hourly resolution and the outputs on hour $t$ are named as $P_{1}^{t}$ and $P_{2}^{t}$, respectively. Figs. 9 and 10 show the CPDF calculated during the simulation using copula. Fig. 9 shows the CPDF of $P_{1}^{t}$ with various level of $P_{1}^{t-1}$ and fixed $P_{2}^{t-1}$ (27.9 MW), while Fig. 10 shows the same index with various level of $P_{2}^{t-1}$ and fixed $P_{1}^{t-1}$ (30.2 MW). It should be noted that the bars in these two figures do not represent two dimensional joint density, but represent one CPDF per row. Comparison between Figs. 9, 10 and 11 demonstrate that the sampling takes into account the
spatio-temporal correlations of sites. The spatial correlation of \( t \) copula between the simulated two sites is 0.8863, which is closed to the observed value of 0.8994.

### 5.2 Case study 1: US data

The 8760 h outputs of all the 10 sites were simulated using the proposed methodology. At the same time, simulations using ARMA model and Markov chain were also carried out for comparison. In the ARMA approach, ARMA(1,1) model was used to generate dependent time series. These time series were then transformed into the marginal distribution that coincides with the output distribution of each wind farm. In the Markov chain approach, the 20 state transition probability matrix was built for each wind farm. Since the Markov chain approach is hard to handle the spatial correlation between wind farms, the wind farms were considered independent in this approach.

A one-week fragment of both the measured output and simulated output using the proposed methodology are shown in Fig. 12. The simulated time series shows a similar time variance and the spatial correlation with the measured output.
measurements. Figs. 13 and 14 compare the marginal distribution and the autocorrelation coefficients between simulated and measured time series respectively. The left sub-graph of each figure shows the statistics of the site 1 (northernmost site in Fig. 2) while the right one shows those of the aggregate output. The simulation results of ARMA model and Markov chain are also compared. The results show that the statistics indices of simulated time series coincide with the observed ones, which implies that the stochastic behaviours of the simulated time series is similar with the actual wind power outputs. Comparing the simulation results among these approaches shows that the stochastic behaviours of the single wind farm are very similar while the stochastic behaviours of the aggregate output have great differences. The marginal distribution of aggregate output obtained by Markov chain deviates greatly from the reality since it does not consider the spatial correlation, while the result of ARMA model which do consider the spatial correlation also has little deviations, this is because the ARMA model only captures the linear part of the correlation and may miss the nonlinear part. The autocorrelation coefficient of aggregate output obtained by both ARMA model and Markov chain declines faster than reality, which suggests that the output has greater variability. This is because these two approaches do not take into account the spatial-temporal correlations among different wind farms between adjacent time intervals (as shown in Fig. 5) and may underestimate the smoothing effect of multiple wind farms. The overall results from Fig. 12 to 14 suggest that the proposed methodology could better capture the nonlinear correlation of wind farm output in both spatial and temporal domain.

Fig. 15 calculates the spatial correlation coefficients of simulated and measured wind power outputs.

Fig. 16. Simulated (lines) and measured (bars) marginal distributions of wind farm output in North China Grid, single site (left) and aggregate output (right).

Fig. 17. Simulated (lines) and measured (bars) autocorrelation coefficients of wind farm output in North China Grid, single site (left) and aggregate output (right).
5.3 Case study 2: wind farm of north china grid

The proposed methodology is further tested using the data from North China Grid. Four wind farms were selected, namely, Danjinghe, Hanjiazhuang, Batou and Hufeng. The measured wind power outputs are available from Oct. 10, 2011 to May 7, 2012. All of the four wind farms are located in a county so that spatial correlation of their outputs is strong. The same indices in Section 5.2 were calculated and are shown in Figs. 16 and 17. Due to the feature of wind resources in North China, the probability of the wind farm to have a output in the range of 90% -100% is much less than that of US costal wind farms. The results show that the proposed copula based methodology is also competent in such circumstances.

6. Conclusion

This paper addresses the issue of how to model and simulate the spatio-temporal correlated outputs of clustered wind farms. Copula theory is used to model the stochastic dependency existed in space and time, irrespective of how the marginal distribution of wind power would have. The chronological simulation of wind power exploits the conditional distribution calculated by multidimensional copula so that the spatial and temporal correlations could be considered simultaneously.

The empirical study based on historical measurements shows that the correlation of wind power outputs decreases as the connection of wind farms gets looser in terms of both space and time. The wind power forecast errors show the same trend. The outputs of wind farms within 20 km area show high degree of correlation, and the variations of their outputs are so similar that they can be treated as one single wind farm. The spatial smoothing effect is not evident until the wind power spreads over a significantly wide area, i.e., 500 km.

The case study using the proposed simulation methodology confirms that the simulated time series retain the main statistical properties of the wind power output in terms of marginal distribution, the spatial and the temporal correlations.

The findings and methodology in this paper are useful for both planning and operation studies involving large-scale cluster wind power, i.e., stochastic unit commitment, probabilistic load flow, generation and transmission expansion in systems with a high penetration of wind power. However, the parameters of spatio-temporal correlations might be site specific. It is also a potential technique for modelling higher frequency wind power outputs, i.e., on minute or second resolution.

Acknowledgements

This work was supported by National High-Tech Research and Development Program of China (863 Scheme) (2011AA05A101) and China Postdoctoral Science Foundation funded project (2012M520272, 2013T60113).

References


Ning Zhang He received both B.S. and Ph.D. from the Electrical Engineering Department of Tsinghua University in China in 2007 and 2012, respectively. His research interests include stochastic characteristic analysis and simulation of renewable energy, power system planning and scheduling with renewable energy.

Chongqing Kang He received his Ph.D. from the Dept. of Electrical Engineering of Tsinghua University in 1997. He is now a Professor at the same university. His research interests include low-carbon electricity, power system planning, power markets and power system load forecasting.
Qianyao Xu  He received his bachelor’s degree from the Department of Electrical Engineering of Tsinghua University in China in 2010. He is currently working towards the Ph.D. in the same University. His research interests is wind power planning and accommodation assessment.

Changming Jiang  He received Ms.c in Department of Electrical Engineering at Chongqing University in China in 1995. He is now the director of the Power Control Center of North China in State Grid. His research interests include power systems dispatch with wind power.

Zhixu Chen  He received Ph.D. in the Electrical Engineering Department of North China Electric Power University in 2007. He is a member of staff at the Power Control Center of North China in State Grid. His research interests include electric market, power systems economic dispatch.

Jun Liu  He is a member of staff at the Power Control Center of North China in State Grid. His research interests include EMS system and power system ideal scheduling.