CrossMark

# Wind power forecasting errors modelling approach considering temporal and spatial dependence

Wei HU[1], Yong MIN[1], Yifan ZHOU[1], Qiuyu LU[2]

MPCE

**Abstract** The uncertainty of wind power forecasting significantly influences power systems with high percentage of wind power generation. Despite the wind power forecasting error causation, the temporal and spatial dependence of prediction errors has done great influence in specific applications, such as multistage scheduling and aggregated wind power integration. In this paper, Pair-Copula theory has been introduced to construct a multivariate model which can fully considers the margin distribution and stochastic dependence characteristics of wind power forecasting errors. The characteristics of temporal and spatial dependence have been modelled, and their influences on wind power integrations have been analyzed. Model comparisons indicate that the proposed model can reveal the essential relationships of wind power forecasting uncertainty, and describe the various dependences more accurately.

✉ Wei HU
  huwei@mail.tsinghua.edu.cn

  Yong MIN
  minyong@mail.tsinghua.edu.cn

  Yifan ZHOU
  zhou-yf14@mails.tsinghua.edu.cn

  Qiuyu LU
  luqiuyu22@126.com

[1] The State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing, China

[2] Power Dispatch Control Centre of Guangdong Power Grid Corporation, Guangzhou, China

STATE GRID
STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

 Springer

## 1 Introduction

With mounting concerns over global warming and fossil fuel depletion, in recent years wind power generation has been greatly increased in the worldwide. However, due to the stochastic behaviors of wind resources cannot be fully predicted, the induced uncertainties significantly influence the power system operations, such as reserve deployment, unit commitment, power dispatch, power system's security and reliability assessments, etc [1–6]. Therefore, a thorough uncertainty model for wind power forecasting error is imperative for power system operators and participants, in terms of evaluating the benefits and costs of wind power integration and thus supporting trade-off decisions.

The short-term uncertainty of wind power generation has been typically modelled by the distribution of the wind power forecasting error. Numerous distributions and modelling methods have been proposed, such as the Gaussian distribution [7], Beta distribution [8] and non-parametric approach [9]. These studies mainly focused on the error distribution of a single wind farm based on a per-prediction horizon. However, due to the similar meteorological conditions, the errors of close prediction horizons and nearby farms influence with each other. The temporal dependence of forecast series is crucial for multistage decision-making problems, such as energy storage planning and scheduling. In parallel, the spatial dependence of multiple farms has been particular concerned to cluster operators in terms of transmission congestion and reserve deployment. Thus more detailed models which consider the

temporal or spatial dependence of wind power forecasting errors have drawn wide attentions in wind power integration applications.

Under this background, researchers are inspired to develop feasible models for the correlation structure of wind power forecasting [10–16]. General methods can be roughly classified into three main kinds: time series model, Markov model, and high-dimensional joint probability model. Time series model is able to discover change rules of historical series, the most widely-used of which is auto regression moving average (ARMA) model. ARMA model has advantages of simple structure and easily fitting [10]. However, ARMA model faces difficulty in dealing with non-stationary stochastic process. Additionally, since there exists essential difference in characteristics between temporal dependence and spatial dependence, simply using the same noise signal for temporal and spatial modelling lacks theoretical support [11]. Markov model depicts the state transition distribution of random variables, with advantages of low requirement to sample size and high fitting effect for short-term characteristics [12]. Yet, due to the restriction of model order, it may be difficult to apply Markov model in uncertainty modelling for multiple wind farms [13]. By regarding forecasting errors of different wind farms and different times as dependent multivariate random variables, high-dimensional joint probability model performs mathematical and statistical description for uncertainty characteristics through joint probability distribution function. Copula theory is a typical method in this area, which has been widely applied in financial market analyses, portfolio investments, and risk assessments [14, 15].

Recently copula theory has been applied in power system uncertainty analyses. The spatial dependence of regional wind power outputs using Gaussian Copula was investigated in [17]. Similar methods were used to model the temporal relevance of forecast errors [18–21]. Although Gaussian distribution has been the default Copula choice in numerous studies for practical reasons, its appropriateness for wind power forecasting dependent structure has not been rigorously investigated. Reference [22] evaluated the goodness-of-fit of several bivariate elliptical and Archimedean Copulas for modelling the wind power spatial dependence and concluded that Gumbel Copula was preferred over Gaussian Copula. However, when extending the bivariate model to higher dimensions, Díaz [23] noted that the multivariate Gumbel Copula did not outperform Gaussian Copula due to its monotonous structure and fewer parameters [24]. Thus, although Archimedean Copula family provides various dependent structures, its applicability in high dimensions is significantly restricted due to a lack of feasible extension methods.

In response to these limitations, this paper introduces Pair-Copula theory to model the temporal and spatial dependence of wind power forecasting errors. The paper is organized as follows. Section 2 introduces the Pair-Copula theory and proposes the detailed procedures of modelling wind power forecasting dependence and sampling wind power uncertainty scenarios. Based on the regional wind power dataset from the National Renewable Energy Laboratory (NREL), the stochastic characteristics of temporal and spatial dependence have been analyzed and modelled in Sect. 3. In Sect. 4, different dependent models have been compared to validate the effectiveness of the proposed method, and the effects of temporal and spatial dependence on power system operations are addressed. Finally, concluding remarks have been presented in Sect. 5.

## 2 Pair-Copula theory based uncertainty modelling algorithm

### 2.1 Review of Copula and Pair-Copula theory

#### 2.1.1 Copula theory

Assume $\{x_1, x_2,\ldots, x_n\}$ are correlated random variables. According to the Sklar theorem [25], Copula function links the margin distribution functions to form the joint multivariate cumulative distribution function (CDF), as shown below.

$$F(x_1, x_2, \ldots, x_n) = C_{1,2,\ldots,n}(F(x_1), F(x_2), \ldots, F(x_n)) \qquad (1)$$

where $F(x_i)$ $(i = 1, 2, \ldots, n)$ is the margin distribution of $x_i$; $F(x_1, x_2, \ldots, x_n)$ is the joint distribution function. Thus, $C$ is the Copula function.

The corresponding Copula probability density function (PDF) is defined as follows:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) = &\, c_{1,2,\cdots,n}(F(x_1), F(x_2), \ldots, F(x_n)) \\ &\cdot f(x_1)f(x_2)\ldots f(x_n) \end{aligned} \qquad (2)$$

where $f(x_i)$ is the margin density function; $f(x_1, x_2,\ldots, x_n)$ is the joint PDF; $c$ is the Copula density function.

The Copula modelling approach has several attractive properties [26]: 1) it allows the marginal distributions and dependent structure to be modelled separately; 2) it does not require the margins to be normally or uniformly distributed; 3) it contains various Copula functions, which makes it flexible in modelling complex dependences, including nonlinear and asymmetric relationships.

#### 2.1.2 Pair-Copula theory

In practical applications, high-dimensional dependence requires feasible multivariate Copulas and even flexible

combinations of different Copula functions. To resolve this problem, hierarchical and nested approaches were proposed, which used sub-models of lower dimensions to construct the joint Copula function [27, 28]. However, these techniques faced the compatibility problem. Strict limitations were imposed on the selected types and parameters to satisfy the nested condition [28]. On this basis, the Pair-Copula construction method was proposed to build flexible multivariate distributions.

Based on the conditional probability theory, the joint PDF can be formulated as follows:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2|x_1) \ldots f(x_n|x_1, x_2, \ldots, x_{n-1})$$
$$= \prod_{t=2}^{n} f(x_t|x_1, x_2, \ldots, x_{t-1}) \cdot f(x_1)$$

(3)

According to the definition of Copula density in (2), $f(x_t|x_1, x_2, \ldots, x_{t-1})$ can be calculated recursively as:

$$f(x_t|x_1, x_2, \ldots, x_{t-1}) = c_{1,t|2,3,\ldots,t-1}f(x_t|x_2, x_3, \ldots, x_{t-1}) = \cdots$$
$$= \left(\prod_{s=1}^{t-2} c_{s,t|s+1,\ldots,t-1}\right)c_{t-1,t}f(x_t)$$

(4)

Thus the joint density of (3) can be further decomposed as [29]:

$$f(x_1, x_2, \ldots, x_n) = \left(\prod_{i=1}^{n-1}\prod_{j=1}^{n-i} c_{j,i+j|j+1,\ldots,i+j-1}\right)\left(\prod_{k=1}^{n} f(x_k)\right)$$

(5)

where $c_{j,i+j|j+1,\ldots,i+j-1}$ is an abbreviated expression for the conditional bivariate Copula density $c_{j,i+j|j+1,\ldots,i+j-1}(F(x_j|x_{j+1},\ldots,x_{i+j-1}), (x_{i+j}|x_{j+1},\ldots,x_{i+j-1}))$. Equation (5) indicates that the joint density consists of the conditional densities of all pairs and the marginal densities, which explain why this form of decomposition is called Pair-Copula.

Pair-Copula theory maintains the merits of Copula theory, and further provides a flexible and intuitive way to build the multivariate Copula using bivariate Copula blocks. Compared with the other multivariate Copula extension methods, Pair-Copula imposes no restrictions on the selected types or parameters and is thus more flexible and practical for modelling and analysing complex dependences.

## 2.2 Model construction for wind power forecasting errors

To specify the Pair-Copula model for the wind power forecasting uncertainty, the temporal dependence model

are derived as an example, and the spatial dependence can be constructed in the same manner. Suppose that $\{e_1, e_2, \ldots, e_n\}$ are the forecast errors of prediction horizons. According to (5), the model construction includes two parts: 1) margin distribution estimation and 2) dependent model construction, i.e., the estimation of every bivariate pair density $c_{j,i+j|j+1,\ldots,i+j-1}$ evaluated at conditional CDF $F(e_j|e_{j+1},\ldots,e_{i+j-1})$ and $F(e_{i+j}|e_{j+1},\ldots,e_{i+j-1})$. As the modelling methods for the marginal distribution have been widely investigated, the estimation of dependent model is proposed in the following.

### 2.2.1 Parameter estimation and Copula selection

For the given candidate bivariate Copula, the corresponding parameters are estimated by maximum log-likelihood estimate (MLE) algorithm [30] in this paper. The parameter $\theta_{j,i+j}$ of pair $c_{j,i+j}$ is obtained as:

$$\max \ln l(\theta_{j,i+j})$$
$$= \max \sum_{k=1}^{m} \ln c_{j,i+j}(F(e_j^k|e_{j+1}^k, \ldots, e_{i+j-1}^k),$$
$$F(e_{i+j}^k|e_{j+1}^k, \ldots, e_{i+j-1}^k))$$

(6)

where $e_j^k$ is the $k^{\text{th}}$ observation of historical data $e_j$; $m$ is the number of observations.

To find the best-fitted family for a pair, several potential Copula functions are selected and parameters are estimated for each candidate. Then the empirical CDF, as formulated in (7), is regarded as the criterion to evaluate the goodness-of-fit for candidates. The Copula with the closest Euclidean distance to empirical CDF will be selected.

$$C_n(x, y)$$
$$= \frac{1}{m}\sum_{k=1}^{m} I(F(e_j^k|e_{j+1}^k, \ldots, e_{i+j-1}^k) \leq x)$$
$$I(F(e_{i+j}^k|e_{j+1}^k, \ldots, e_{i+j-1}^k) \leq y)$$

(7)

where $(x, y)$ is the sample point in distributed space; $I(\cdot)$ is the indicative function. When the defined condition is satisfied in parenthesis, $I(\cdot) = 1$, otherwise $I(\cdot) = 0$.

### 2.2.2 Dependent structure estimation

As mentioned above, the Copula densities are fitted based on conditional CDFs, which can be calculated by [31]:

$$F(e_j|D) = \frac{\partial C_{e_j,e_k|(D\backslash e_k)}[F(e_j|(D\backslash e_k)), F(e_k|(D\backslash e_k))]}{\partial F(e_k|(D\backslash e_k))}$$

(8)

where $D$ is the conditional variable set, $e_k \in D$; $D\backslash e_k$ means removing element $e_k$ from $D$; $\partial$ is the partial differential signal.

Equation (8) indicates the conditional CDFs used in high dimensions can be calculated by the fitted bivariate Copulas of lower dimensions. Therefore, the $n(n-1)/2$ underestimated pairs $c_{j,i+j|j+1,\ldots,i+j-1}$ can be divided into $(n-1)$ trees according to the value of $i$, as shown in Fig. 1. The pair densities are estimated in the sequence of the tree number. As the Copula functions in lower trees have been selected and parameterized, the fitted Copula function can be applied to calculate conditional CDFs for the next tree by (8). Thus the entire model is achieved in a recursive manner.

Through the construction process, it can be seen that the Pair-Copula method allows mixing different Copula types and sets no limitations on types or parameters, thus it can select the best-fitted distributions and gain a higher accuracy. When the Gaussian Copula is nominated for all pairs, the Pair-Copula model is proven to be equal to the multivariate Gaussian Copula. Therefore, the Pair-Copula model can replace the conventional multivariate models without sacrificing the accuracy of the model.

### 2.3 Wind power uncertainty scenario simulation

The most important application of the wind power uncertainty model is to evaluate the benefits and costs of wind power integration. For end-users, Monte Carlo simulated scenarios provide uncertainty information in a discrete form, which are practical for risk assessments and optimization programming. The simulation process is also based on conditional probability theory.

First, the independent uniform variables $\{w_1, \ldots, w_n\}$ is randomized to follow the stratified sampling principle. Next the correlated uniform forecast errors $\{z_1, \ldots, z_n\}$ for horizon 1 to $n$ are calculated by (9). The conditional CDFs of (9) have been developed when fitting the dependent model; thus, they can be used directly in the simulation process. The uniform errors then will be transferred back to the original domain by inverse marginal CDFs, and finally form the wind power scenarios.
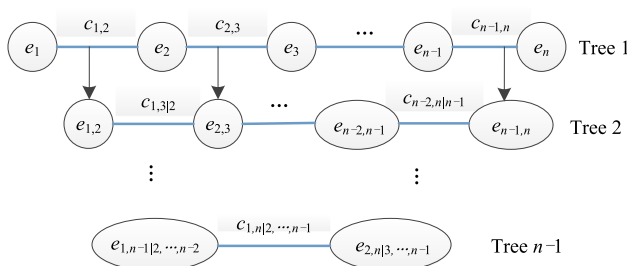
$$\begin{cases} z_1 = w_1 \\ z_2 = F^{-1}(w_2|z_1) \\ \vdots \\ z_n = F^{-1}(w_n|z_1, z_2, \ldots, z_{n-1}) \end{cases} \tag{9}$$

Therefore, the entire procedure for modelling and generating wind power forecasting scenarios considering the temporal dependence is described as follows, and the mechanism sketch of each step is drawn in Fig. 2 for further clarification.

*Step 1* Form the n-dimensional dataset $E = \{e_1, e_2, \ldots, e_n\}$ of prediction errors for multiple forecast horizons or wind farms using the historical data.
*Step 2* Develop the marginal distribution $F(e_i)$ ($i = 1, 2,\ldots, n$) and transform the data from the actual domain to the uniform domain.
*Step 3* Construct the multivariate dependent model based on the uniform data.
*Step 4* Sample correlated uniform error $\{z_1, z_2, \ldots, z_n\}$ using (9) based on the model estimated in Step 3 .



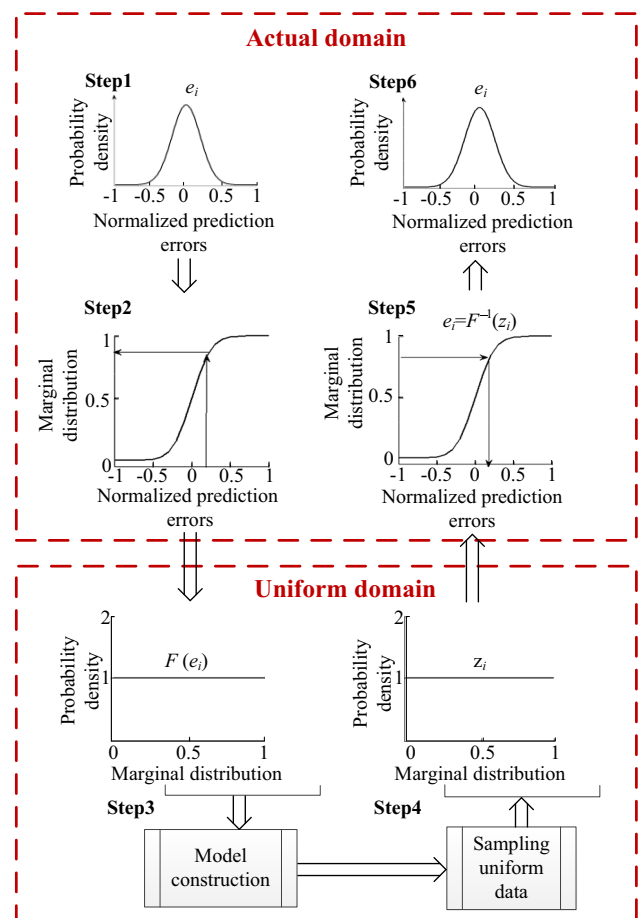Fig. 1 Mechanism sketch of dependent model construction process



Fig. 2 Mechanism sketch of entire modelling and simulation process

*Step 5* Transform samples back to the actual domain using the inverse margin $e_i = F^{-1}(z_i)$.

*Step 6* Obtain wind power uncertainty scenarios by adding the given point forecasts and simulated errors.

Further extension to spatial-temporal dependence modelling also follows a similar principle, but the model dimension is higher.

## 3 Modelling temporal and spatial dependence of wind power forecasting errors

### 3.1 Dataset

In this section, the temporal and spatial dependences of wind power forecasting errors are modelled based on the Wind Integration datasets of the NREL [32]. The wind speeds in the dataset are generated by MASS v.6.8 mesoscale model, which is initialized with input from Atmospheric Research Global Reanalysis and assimilates both surface and rawindsonde data. The wind power outputs are computed by composite turbine power curves, and the forecasts are produced by a statistical forecast tool named SynForecast. Though the data are simulated, they are verified to contain similar stochastic features of actual wind farms and reflect the geographic diversity of wind power outputs. Thus they are suitable for researches of wind power temporal and spatial relevance [22, 23, 33].

In this paper, regional wind plants in Vermont are selected. The data consist of 17 onshore wind farms varying in capacity from 100 to 180 MW and distributed spatially from 10 to 250 km. Detailed spatial locations of these wind farms are presented in Table 1. The wind power outputs and day-ahead forecasts span 3 years and have a 1 h resolution. All data have been normalized by the rated capacity.

### 3.2 Temporal dependence model

The Gaussian model for temporal dependence is first built to analyse the linear dependence. Figure 3a shows the fitted parameters of the Gaussian model, where the matrix element in position (*j, i*) represents the linear correlation coefficient of errors of horizon *j* and *i*. High correlations are observed for close prediction horizons, whereas the relevance decreases dramatically as the intervals increase. Furthermore, Fig. 3b presents the colour histogram of the forecast errors in horizons 1 and 2 in the uniform domain. The dependent structure of adjacent errors is generally elliptical. Moreover, the tail is steep and symmetric, which highlights a higher coherence when large prediction errors occur. This characteristic should be addressed in multistage

**Table 1** Detailed spatial locations of concerning 17 wind farms

| Index | Latitude | Longitude | Altitude |
| --- | --- | --- | --- |
| A | 44.97 | −72.52 | 503 |
| B | 44.87 | −72.51 | 806 |
| C | 44.81 | −72.69 | 687 |
| D | 44.92 | −71.82 | 700 |
| E | 44.71 | −71.79 | 744 |
| F | 44.54 | −72.18 | 681 |
| G | 44.2 | −72.35 | 720 |
| H | 44.29 | −72.89 | 949 |
| I | 44.23 | −73 | 627 |
| J | 43.99 | −72.8 | 646 |
| K | 43.68 | −72.66 | 673 |
| L | 43.56 | −72.8 | 691 |
| M | 43.52 | −72.62 | 623 |
| N | 43.35 | −72.76 | 641 |
| O | 43.42 | −73.08 | 754 |
| P | 43.22 | −73.14 | 853 |
| Q | 42.9 | −72.8 | 578 |



**(a)** Parameters of Gaussian model

**(b)** Temporal dependent structure of adjacent error
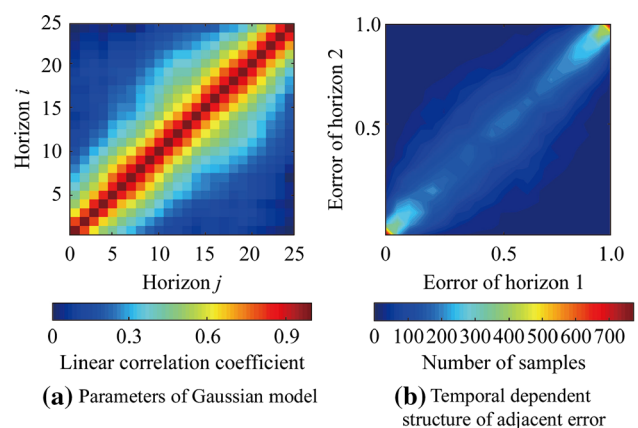
**Fig. 3** Data analysis for temporal dependence

decision-making problems, as it would significantly affect the risk index.

To develop the Pair-Copula model, five typical Copula function, i.e. Gaussian, Student t, Clayton, Gumbel and Frank Copulas are selected as candidates for estimating the pair densities. The independent function is also included here, and its bivariate CDF is the product of the marginal distributions. When the CDF of independent type owns the nearest distance to the empirical CDF, the concerning variables are judged to be independent. As parameters of different Copula families vary in different ranges, to facilitate model analyses and comparisons, the fitted parameters are translated into Kendall's rank correlation coefficients, which can be identically formulated according

to Copula class and parameters [29]. Figure 4 shows the selected Copula types and corresponding parameters (translated into Kendall's rank correlations) for all pairs. The element in position $(j, i)$ denotes the $j^{th}$ pair in the $i^{th}$ Tree, i.e., the conditional dependence of the errors of horizon $j$ and $i + j$. Figure 4a denotes the best fitted Copula type with different colours, and Fig. 4b demonstrates the estimated parameter for each Copula function.

As the dimension is 24 (the forecast errors of 24 h), there are 23 trees in the temporal model. Tree1 has 23 pairs, each pair represents the temporal correlation of horizon $j$ and $j + 1$. In parallel, Tree 2 has 22 pairs, each pair represents the conditional temporal correlation of horizon $j$ and $j + 2$. The rest trees can be analysed in the same way. The fitted results show that the independent function dominates in Trees 2–23, which implies that non-adjacent errors are conditionally independent, i.e., the wind power forecasting error has a one-order Markov property. This result is not surprising, as wind speed and wind power outputs have been proven to be Markov chains in many studies [12, 34]. Using the Pair-Copula approach in this paper, the day-ahead wind power forecasting errors are verified to possess a Markov property with an optimal order of 1.

For the dependence of adjacent errors in Tree 1, the Student t is the most favored Copula function due to its elliptical structure and strong tail dependence, and the Gumbel function is also selected for some pairs. For parameter results, strong correlations are found for adjacent errors, and the variation of parameters in Tree 1 is small, indicating that the correlations are not significantly affected by the prediction horizons 4 figures and tables.

### 3.3 Spatial dependence model

Seventeen wind farms in Vermont are included in the spatial dependence study and are sequenced by distance. The analysis and modelling process are similar to those described above. The parameters of 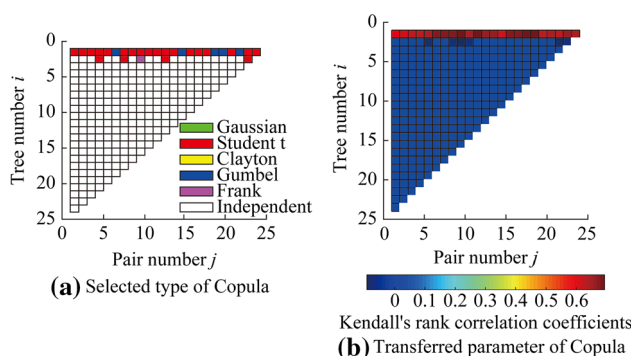the Gaussian model are shown in Fig. 5a. Compared with the temporal dependence, the spatial correlation is remarkably weak, i.e., the smoothing effect of aggregate wind farms is considerable. Fig. 5b displays the colour histogram of wind farm 1 and 2 with a correlation of 0.452. It is discovered that the spatial dependence has an upper tail, which indicates that the probability of simultaneous large positive errors is higher than that of negative errors. The asymmetry property should be taken into consideration in reserve deployments and transmission congestion problems in wind farm cluster integration.

The Copula types and parameters of the Pair-Copula spatial model are shown in Fig. 6, where the element in position $(j, i)$ denotes the conditional dependence of errors of wind farm $j$ and $i + j$. Unlike the temporal model, the spatial dependences are inherent; thus, Markov phenomenon is not observed. Furthermore, the asymmetric characteristic is common among the selected wind farms, thus the Gumbel function is superior to other candidates owing to its upper tail structure; yet the symmetrical Student t and Frank functions are also selected for some pairs of wind farms. Comparing Fig. 6b and the spatial distribution of the wind farms, the spatial correlation generally decreases when distance between wind farms increasing. For instance, farm $a$ and $b$ is nearest, thus the correlation coefficient is the largest (the correlation of pair $c_{1,2}$ is 0.3177). However, the distance of farm $b$ and $c$ is farther than farm $h$ and $i$, but the correlation coefficient is larger (the correlation of pair $c_{2,3}$ and $c_{8,9}$ is 0.2652 and 0.2522 respectively). Therefore, the spatial dependence is affected by factors like distance, elevation, topographical and surface structure, and it is more variable and complex.

## 4 Model analysis and comparisons

To illustrate the effectiveness of the proposed method, the Pair-Copula, Gaussian-dependent and the independent (i.e., temporal or spatial correlation is ignored) models are
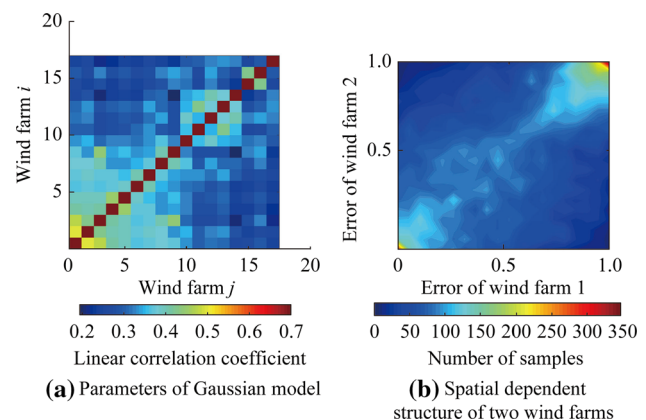
**Fig. 4** Estimated result of Pair-Copula temporal model

**Fig. 5** Data analysis for spatial dependence

(a) Selected type of Copula
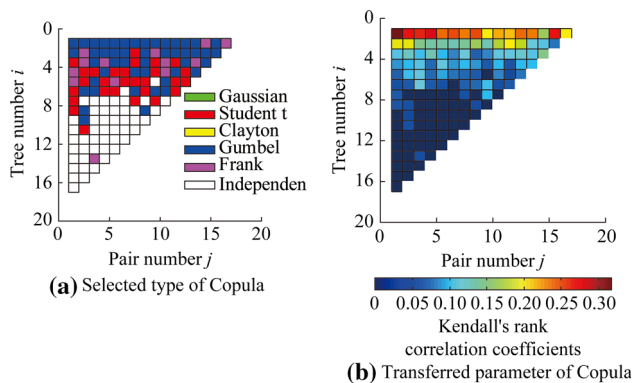
(b) Transferred parameter of Copula

**Fig. 6** Estimated result of Pair-Copula spatial model

compared in this section. The goodness-of-fit evaluation of multivariate distribution needs huge computation and is time-consuming, especially for high dimensional cases. Hence, the typical application scenes in which the temporal or spatial dependence of wind power forecasting errors are required are investigated. And according to the effects of the dependences in application scenes, the corresponding criterion is proposed to map the multivariate distribution to a univariate one, hence facilitating the comparisons between different multivariate models.

### 4.1 Temporal dependence model comparison

The scheduling and planning of an energy storage system (ESS) is a typical multistage decision-making problem. When the ESS is applied to balance the wind power forecasting deviations, the absorption or production energy of the ESS is determined by both the magnitude and sequence of the prediction errors, as formulated in (10). Therefore, the temporal dependence needs to be considered.

$$
\begin{aligned}
E_b &= \sum_{i \in T_b} e_i \Delta t \\
T_b &= \{e_i \geq 0, \forall i \in T_b\} \text{ or } \{e_i \leq 0, \forall i \in T_b\}
\end{aligned}
\tag{10}
$$

where $E_b$ is the charge or discharge energy of ESS; $\Delta t$ is the time interval; $T_b$ is the charging or discharging period during which only positive or negative prediction errors occur.

Therefore, the distribution of $E_b$ is regarded as the criterion to evaluate the accuracy of different temporal models. To develop the univariate distribution, temporal-related prediction errors are generated by the Monte Carlo algorithm based on different temporal models, $E_b$ can then be calculated by (9), and the empirical distribution can be further obtained using the simulated data.

Figure 8 shows the probability density frequency of the ESS energy based on the measurements and the three models studied, and the data are normalized by the wind

farm rated capacity. The measured data exhibit a fat-tailed characteristic due to the presence of continuous errors with the same sign, which agrees with the analysis in temporal modelling. The independent model does not reflect this property and therefore will greatly underestimate the necessary ESS energy capacity. On the other hand, the Pair-Copula and Gaussian models are alike because the selected Student t Copula has similar stochastic characteristics with Gaussian Copula, and they both fit the data well.

To further qualify the goodness-of-fit of the estimated model, Cramer-von Mises (CvM) and Kolmogorov-Smirnov (KS) indices [35], which respectively measure the average and maximum Euclidean distances between the estimated and empirical CDF, are calculated and shown in Table 2. Both indices are negatively oriented, meaning that lower values are preferable. The Pair-Copula model is better than the Gaussian model in terms of both indices.

### 4.2 Spatial dependence model comparison

Due to the weak spatial correlation, the prediction errors of disperse wind farms is offset thus reserve demand can be reduced. Thus the spatial dependence is desired for the assessment of power system operation reliability and security. The distribution of the aggregated prediction error of regional wind farms is selected as the criterion to evaluate the accuracy of different spatial models.

The probability density frequency of aggregated forecast error of 17 wind farms in Vermont based on the measurements and three estimated models are shown in Fig. 7. Apparently, the independent model significantly underestimates the error range and reserve demands. And the measurements are slightly right-skewed due to the asymmetry spatial dependent structure, which is reflected in the Pair-Copula model but not in the Gaussian model. The goodness-of-fit indices are listed in Table 3, and Pair-Copula model is still better than the Gaussian model for both evaluated indices.

Furthermore, it should be highlighted that the asymmetry distribution would result in different requirements on reserves. Suppose the upward and downward reserves providing for the wind power uncertainty are 0.2 p.u. In this case, when negative error exceeds upward reserve, load shedding will occur. Similarly, if the positive error is greater than the downward reserve, the surplus wind power must be curtailed, as shown in Fig. 8. Table 4 lists the load shedding and wind curtailing probabilities. The wind curtailing probability is clearly larger than the load shedding probability due to the asymmetry distribution. The probabilities of the Pair-Copula model are the closest to the measured data, while the Gaussian model overestimates the load shedding probability, which will significantly increase

**Table 2** Goodness-of-fit indices of three temporal models

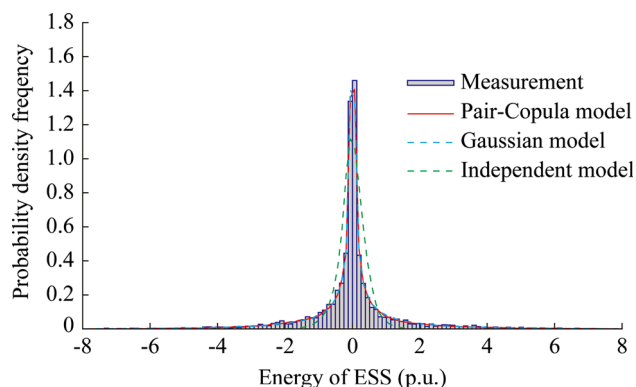| Index | Pair-Copula model | Gaussian model | Independent model |
|---|---|---|---|
| CvM index | 0.0059 | 0.0100 | 0.0626 |
| KS index | 0.0153 | 0.0191 | 0.1095 |



**Fig. 7** Probability density frequency of ESS energy of measurements and three estimated models

**Table 3** Goodness-of-fit indices of three spatial models

| Index | Pair-Copula model | Gaussian model | Independent model |
|---|---|---|---|
| CvM index | 0.0119 | 0.0123 | 0.0769 |
| KS index | 0.0376 | 0.0478 | 0.2351 |
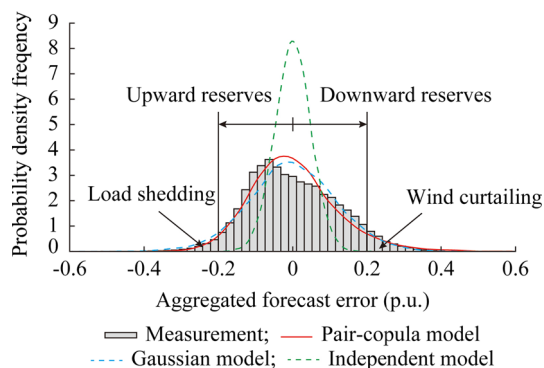


**Fig. 8** Probability density frequency of aggregated forecast error of measurements and three estimated models

### 4.3 Feasibility evaluation of model

More comparisons for different wind farms are carried out to test the feasibility of the proposed method. Figure 9

**Table 4** Load shedding and wind curtailing probabilities

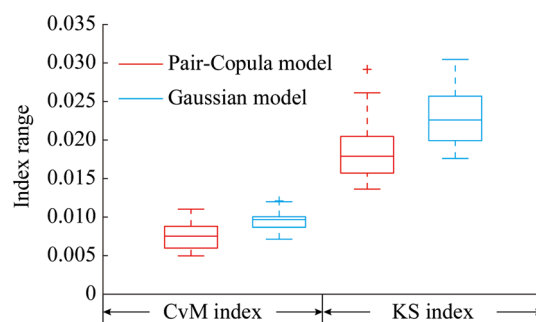| Probability | Measured data (%) | Pair-Copula model (%) | Gaussian model (%) | Independent model (%) |
|---|---|---|---|---|
| Load shedding probability | 2.79 | 2.44 | 4.45 | 0 |
| Wind curtailing probability | 5.36 | 5.69 | 4.64 | 0.01 |



**Fig. 9** Temporal goodness-of-fit indices for different wind farms

draws the CvM and KS index of Pair-Copula model and Gaussian temporal model for 17 wind farms in Vermont, where the box shows the median, 25% and 75% percentiles of the indices, while the whiskers are the extreme indices. The Pair-Copula model has generally lower indices than Gaussian model for different wind farms.

Figure 10 shows the goodness-of-fit indices of spatial models in different aggregated scale. Each column represents the index range of 60 wind farm clusters with the specified scale, which are randomly selected from the regional wind plants in Vermont. For both compared
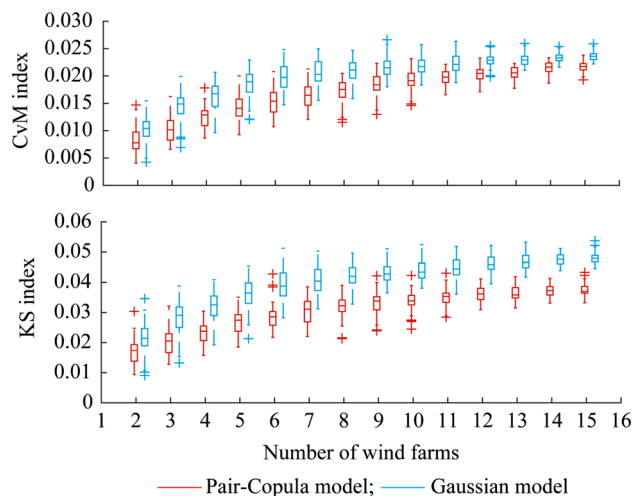
the operation costs to maintain the required reliability level.



**Fig. 10** Spatial goodness-of-fit indices for different scales of wind farm

models, the goodness-of-fit index worsens as the model dimension increases. The Pair-Copula model yields greater accuracy for all dimensions, and the improvement is more evident than that of the temporal case. It can be concluded that the Pair-Copula model is flexible in different dependent relationships, and the superiority is more prominent for nonlinear and asymmetric structures when compared to the traditional Gaussian model.

## 5 Conclusion

Because of the highly variable forecast errors of wind power generation, a thorough uncertainty model describing both the predictive distribution and dependent structure of forecast errors is essential. In this paper, a Pair-Copula modelling approach has been proposed to construct joint distribution and generate scenarios that mimic wind power stochastic behaviours. The feasibility of the Pair-Copula approach for wind power temporal and spatial dependence is verified, and the advantages of the Pair-Copula are noted. First, owing to the basic of conditional dependence, the Pair-Copula model can reflect the essential relationships between random variables, which the other multivariate models fail to reveal, e.g., the one-order Markov characteristic of the temporal model. Furthermore, the Pair-Copula approach is more adaptive and practical in various applications, as it allows the combination of different types of Copulas and thus yields the best-fitted model. In the illustrated example, the symmetrical Student t function is selected for the temporal model, while the asymmetrical Gumbel Copula is chosen for the spatial model.

Based on the developed models, the effects of the temporal and spatial dependence of wind power forecasting errors on power system operations are analysed. For the temporal model, the tail dependence indicates that large prediction errors have stronger correlations, resulting in a fat-tail distribution of the ESS energy. Thus, a large ESS capacity is required to balance the prediction errors. For the spatial model, the upper tail makes the probability of excess generation higher than that of deficits, causing different demands for upward and downward reserves. Furthermore, goodness-of-fit indices of different models are calculated and comparison results show that Pair-Copula model gives greater accuracy in various application scenes.

The main contribution of this paper is to significantly advance the uncertainty model of wind power and to provide a systematic procedure for modelling and generating wind power scenarios, which can then be applied to a range of wind power integration problems. Although only wind power uncertainty modelling is discussed in this paper, the Pair-Copula approach provides a powerful tool for constructing flexible multivariate distributions that it can also be applied to a wide range of statistical analyses and stochastic decision fields of power systems.

## References

[1] Yuan XM (2013) Overview of problems in large-scale wind integrations. J Mod Power Syst Clean Energy 1(1):22–25. doi:10.1007/s40565-013-0010-6

[2] Halamay DA, Brekken TKA, Simmons A et al (2011) Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation. IEEE Trans Sustain Energy 2(3):321–328

[3] Zhou BR, Geng GC, Jiang QY (2016) Hierarchical unit commitment with uncertain wind power generation. IEEE Trans Power Syst 31(1):94–104

[4] Li YZ, Wu QH, Jiang L et al (2016) Optimal power system dispatch with wind power integrated using nonlinear interval optimization and evidential reasoning approach. IEEE Trans Power Syst 31(3):2246–2254

[5] Ding Y, Singh C, Goel L et al (2014) Short-term and medium-term reliability evaluation for power systems with high penetration of wind power. IEEE Trans Sustain Energy 5(3):896–906

[6] Ciupuliga AR, Gibescu M, Pelgrum E et al (2012) Round-the-year security analysis with large-scale wind power integration. IEEE Trans Sustain Energy 3(1):85–93

[7] Holttinen H (2005) Impact of hourly wind power variations on the system operation in the Nordic countries. Wind Energy 8(2):197–218

[8] Bludszuweit H, Dominguez-Navarro JA, Llombart A (2008) Statistical analysis of wind power forecasting error. IEEE Trans Power Syst 23(3):983–991

[9] Pinson P (2006) Estimation of the uncertainty in wind power forecastinging. Ph.D. thesis. Ecole des Mines de Paris, Paris, France

[10] Billinton R, Chen H, Ghajar R (1996) Time-series models for reliability evaluation of power systems including wind energy. Microelectron Reliab 36(9):1253–1261

[11] Miranda MS, Drun RW (2007) Spatially correlated wind speed modelling for generation adequacy studies in the UK. In: Proceedings of the 2007 IEEE Power Engineering Society general meeting, Tampa, FL, USA, 24–28 Jun 2007, 6 pp

[12] Shamshad A, Bawadi MA, Wan Hussin WMA et al (2005) First and second order Markov chain models for synthetic generation of wind speed time series. Energy 30(5):693–708

[13] Bouffard F, Galiana FD (2008) Stochastic security for operations planning with significant wind power generation. IEEE Trans Power Syst 23(2):306–316

[14] McNeil AJ, Frey R, Embrechts P (2005) Quantitative risk management: concepts, techniques, tools. Princeton University Press, Princeton

[15] Cherubini U, Luciano E, Vecchiato W (2004) Copula methods in finance. Wiley, New York

[16] Wan C, Xu Z, Pinson P et al (2014) Probabilistic forecasting of wind power generation using extreme learning machine. IEEE Trans Power Syst 29(3):1033–1044

[17] Papaefthymiou G, Kurowicka D (2009) Using copulas for modeling stochastic dependence in power system uncertainty analysis. IEEE Trans Power Syst 24(1):40–49

[18] Pierre P, Henrik M, Aa NH et al (2009) From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12(1):51–62

[19] Ma XY, Sun YZ, Fang HL (2013) Scenario generation of wind power based on statistical uncertainty and variability. IEEE Trans Sustain Energy 4(4):894–904

[20] Yang M, Lin Y, Zhu SM et al (2015) Multi-dimensional scenario forecast for generation of multiple wind farms. J Mod Power Syst Clean Energy 3(3):1–10. doi:10.1007/s40565-015-0110-6

[21] Haghi HV, Bina MT, Golkar MA (2013) Nonlinear modeling of temporal wind power variations. IEEE Trans Sustain Energy 4(4):838–848

[22] Louie H (2014) Valuation of bivariate archimedean and elliptical copulas to model wind power dependence structures. Wind Energy 17(2):225–240

[23] Díaz G (2014) A note on the multivariate archimedean dependence structure in small wind generation sites. Wind Energy 17(8):1287–1295

[24] Hofert M, Mchler M, McNeil AJ (2011) Likelihood inference for archimedean copulas in high dimensions under known margins. J Multivar Anal 110:133–150

[25] Nelsen RB (1999) An introduction to copulas. Springer, Berlin

[26] Drouet MD, Kotz S (2001) Correlation and dependence. Imperial College Press, London

[27] Savu C, Trede M (2010) Hierarchies of archimedean copulas. Quant Finance 10(3):295–304

[28] McNeil AJ (2008) Sampling nested archimedean copulas. J Stat Comput Simul 78(6):567–581

[29] Czado C (2010) Pair-copula constructions of multivariate copulas. In: Jaworski P, Durante F, Hardle W et al (eds) Copula theory and its applications. Springer, Berlin, pp 93–109

[30] Genest C, Ghoudi K, Rivest LP (1995) A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82(3):543–553

[31] Aas K, Czado C, Frigessi A et al (2009) Pair-copula constructions of multiple dependence. Insur Math Econ 44(2):182–198

[32] Pennock K (2012) Updated eastern interconnect wind power output and forecasts for ERGIS. NREL/SR-5500-56616. National Renewable Energy Laboratory, Golden, CO, USA

[33] Zhang N, Kang CQ, Xu QY et al (2013) Modelling and simulating the spatio-temporal correlations of clustered wind power using copula. J Electr Eng Technol 8(6):1615–1625

[34] Papaefthymiou G, Klockl B (2008) MCMC for wind power simulation. IEEE Trans Energy Conver 23(1):234–240

[35] Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: A review and a power study. Insur Math Econ 44(2):199–213

**Wei HU** received a B.S. and a Ph.D. in Electrical Engineering from Tsinghua University, China, in 1998 and 2002, respectively. Now he is working as an Associate Professor in the Department of Electrical Engineering, Tsinghua University. His research interests are in power system analysis and control, big data technology in power system and Energy Internet.

**Yong MIN** received a B.S. degree and a Ph.D. in Electrical Engineering from Tsinghua University, China, in 1984 and 1990, respectively. He is now a Professor in the Department of Electrical Engineering there. His research interests are power system stability and control and large-scale renewable energy integration.

**Yifan ZHOU** received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2014. She is currently pursuing the Master degree in electrical engineering at the same university. Her research interests include power system modeling, stability analysis and control, and Energy Internet analysis and control.

**Qiuyu LU** received a B.S. degree and a Ph.D. in Electrical Engineering from Tsinghua University, China, in 2010 and 2015. Now she is working in Power Dispatch Control Centre of Guangdong Power Grid Corporation. Her current research interests are uncertainty modeling of renewable energies, wind power integration and wind-storage combined operation optimization.

STATE GRID
STATE GRID ELECTRIC POWER RESEARCH INSTITUTE