

Forecasting Scenario Generation for Multiple Wind Farms Considering Time-series Characteristics and Spatial-temporal Correlation

Qingyu Tu, Shihong Miao, Fuxing Yao, Yaowang Li, Haoran Yin, Ji Han, Di Zhang,
and Weichen Yang

Abstract—Scenario forecasting methods have been widely studied in recent years to cope with the wind power uncertainty problem. The main difficulty of this problem is to accurately and comprehensively reflect the time-series characteristics and spatial-temporal correlation of wind power generation. In this paper, the marginal distribution model and the dependence structure are combined to describe these complex characteristics. On this basis, a scenario generation method for multiple wind farms is proposed. For the marginal distribution model, the autoregressive integrated moving average-generalized autoregressive conditional heteroskedasticity- t (ARIMA-GARCH- t) model is proposed to capture the time-series characteristics of wind power generation. For the dependence structure, a time-varying regular vine mixed Copula (TRVMC) model is established to capture the spatial-temporal correlation of multiple wind farms. Based on the data from 8 wind farms in Northwest China, sufficient scenarios are generated. The effectiveness of the scenarios is evaluated in 3 aspects. The results show that the generated scenarios have similar fluctuation characteristics, autocorrelation, and crosscorrelation with the actual wind power sequences.

Index Terms—Scenario generation, wind farm, regular vine Copula, spatial-temporal correlation, time-series characteristics.

I. INTRODUCTION

TO achieve the clean and low-carbon energy supply, wind power has attracted extensive attention worldwide in the recent few decades. However, at present, the accurate forecasting of wind power generation is not an easy goal to achieve [1]. With the rapid growth of wind power capacity integrated into the grid, the uncertainty of power supply

caused by the ineluctable forecasting error is becoming increasingly prominent. It will affect the reliability of the dispatching plan, which may not only cause severe wind power curtailment, but also bring potential risks for the safe operation of the power grid [2]. Therefore, it is necessary to develop a forecasting method to reflect the uncertainty of wind power generation.

On the other hand, wind farms are often clustered [3]. The site selection of wind farms is usually concentrated in the areas with abundant wind energy. When making operation plans, rather than a single wind farm, more attention is usually paid to the uncertainty of the joint-output of multiple wind farms in the region [4]. Besides, in order to avoid the off-grid events caused by the increase of wind power penetration, it is also necessary to propose forecasting methods applicable to regional wind farms [5].

In view of the uncertainty of wind power generation, the scenario forecasting method, which is an important method of probabilistic forecasting, has been extensively studied. Its basic principle is to establish the probability density function (PDF) of wind power or forecasting error by statistical methods, and then generate the scenarios by sampling methods [6]. Various scenario forecasting methods have been proposed in existing researches [7]–[12]. In [7] and [8], the forecasting error is assumed to follow the Beta distribution and the t location-scale distribution, respectively. In [9], the quantile regression method is used to establish the PDF. The empirical cumulative distribution function (ECDF) is studied in [10]. And in [11] and [12], the kernel density estimation (KDE) model is proposed.

The aforementioned works are proposed to reflect the long-term frequency distribution of wind power generation. Considering the autocorrelation of wind power over a period of time, the studies in recent years begin to focus on the time-series characteristics. In [13]–[16], the scenario generation methods based on the time-series characteristics are put forward. In [13], the Markov method is utilized to simulate the time-varying process of wind power. In [14], the autoregression (AR) model is studied. References [15] and [16] further promote the work of [14] and the AR-moving average (ARMA) model is established. The above researches show that when the time-series characteristics are taken into account, the excessive irregular fluctuations and noise com-

Manuscript received: December 30, 2020; accepted: May 24, 2021. Date of CrossCheck: May 24, 2021. Date of online publication: July 30, 2021.

This work was supported by the National Key Research and Development Program of China (No. 2017YFB0902600).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Q. Tu, S. Miao (corresponding author), F. Yao, H. Yin, J. Han, D. Zhang, and W. Yang are with the State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Hubei Electric Power Security and High Efficiency Key Laboratory, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, China (e-mail: tuqy_d@hust.edu.cn; shmiao@hust.edu.cn; yaofuxing@hust.edu.cn; hr_yin@163.com; han_ji1993@163.com; zdsee@hust.edu.cn; yangwc@hust.edu.cn).

Y. Li is with the Department of Electrical Engineering, Tsinghua University, Beijing, China (e-mail: yaowang_li@126.com)

DOI: 10.35833/MPCE.2020.000935



ponents in wind power will be significantly reduced.

On the other hand, however, with the increase in the number and scale of wind farms, the traditional methods have limitations when applied to regional multiple wind farms. The key problem lies in that, the spatial-temporal correlation should be fully described, while the advantages of the original method are retained. In this respect, a good method is to adopt the Copula model, which can be decomposed into two parts: the marginal distribution model and the dependence structure. The first part is the independent PDF of each single wind power sequence, which ensures good continuity with the existing works mentioned above. The second part describes the spatial-temporal correlation of multiple wind farms. In [17] and [18], the Gaussian Copula model is used to model the PDF of high-dimensional wind power data. In [19] and [20], the C-vine and D-vine Copula models are further proposed. The models have more flexible dependence structures than the Gaussian Copula model and thus achieve higher accuracy.

It can be observed from the literature review that, some works have introduced the Copula model in analyzing the correlation of multiple wind farms. However, the existing works have three limitations: ① the time-series characteristics and spatial-temporal correlation are not effectively combined, resulting in frequent unreasonable fluctuations in the scenarios; ② the models are reliable only when dealing with low-dimensional data; for high-dimensional wind power data, the dependence structures can not fully describe the spatial-temporal correlation; ③ the complexity of joint-distribution between two arbitrary wind farms is underestimated. Specifically, the pair Copulas, as the basic units of the high-dimensional Copula model, are too simple to capture the tail characteristics, thus reducing the accuracy of the whole model.

Therefore, this paper aims to propose a scenario generation method that can effectively describe both the time-series characteristics and the spatial-temporal correlation of the power output of multiple wind farms. The contributions of this work are briefly summarized as follows.

1) The time-varying regular vine mixed Copula (TRVMC) model is established to fit the joint probability distribution of the output of regional multiple wind farms. By making the probability distribution capture the characteristics of the joint-frequency distribution, the TRVMC model can reflect the correlation between the wind farms. Compared with the Gaussian, C-vine, and D-vine Copula models in the existing research, the TRVMC model does not need to make strict assumptions about the correlation of input data, but fits the appropriate model structure for different input data. Consequently, the TRVMC model has higher fitting accuracy for the joint-probability distribution.

2) The AR integrated moving average-generalized autoregressive conditional heteroskedasticity- t (ARIMA-GARCH- t) model is established to fit the marginal distribution model of the output of each wind farm. The model can capture the time-series characteristics of wind power output. Compared with the commonly-used static models such as the KDE model, the ECDF model, and the student- t model, the ARIMA-GARCH- t model has higher fitting accuracy, which can provide more reliable input data for the Copula model.

3) The time-varying mixed Copula (TMC) model is established as the pair Copulas, i. e., the basic units of the TRVMC model. On one hand, the TMC model integrates the advantages of multiple basic bivariate Copula models. On the other hand, by introducing the dynamic correlation calculation models (including the dynamic conditional correlation (DCC) model and the Patton model) to fit the parameters of the TMC model, the model can capture the time-varying characteristics of the dependence structure of two-dimensional wind power sequences, so as to improve the accuracy of the entire model.

4) Based on the established ARIMA-GARCH- t model and the TRVMC model, the forecasting scenario generation method for the output of regional multiple wind farms is proposed. The scenarios present similar time-series characteristics and spatial-temporal correlation with the actual wind power sequence. Compared with the methods which ignore the above characteristics, the scenarios generated by the proposed method have a more reasonable fluctuation range and frequency. Besides, the scenarios can better envelop the actual wind power output sequence.

The rest of the paper is organized as follows. In Sections II and III, the ARIMA-GARCH- t model and the TRVMC model are described in detail, respectively. The forecasting scenario generation method is proposed in Section IV. Section V provides an overall description of the modeling and scenario generation process. The evaluation framework is introduced in Section VI. In Section VII, the forecasting scenarios are generated and evaluated based on the data from 8 wind farms. Finally, some concluding remarks are provided in Section VIII.

II. ARIMA-GARCH- T MODEL AS MARGINAL DISTRIBUTION MODEL

The marginal distribution model is the independent probability distribution model of each wind power sequence. By calculating the cumulative probability, the original wind power sequence is transformed into a uniform sequence bounded by $[0, 1]$. The converted sequence will be used as input data for the Copula model.

The cumulative distribution function (CDF) of the power output of each wind farm can be expressed as:

$$F_i(x_i) = x_i^{\text{fc}} + F_i^{\text{err}}(x_i^{\text{err}}) \quad (1)$$

where x_i , x_i^{fc} , and x_i^{err} are the measured power, forecasting power, and forecasting error of the i^{th} wind farm, respectively; and F_i and F_i^{err} are the CDFs of x_i and x_i^{err} , respectively. Then the forecasting error is taken as the input data of the ARIMA-GARCH- t model. For convenience, denote $x_i^{\text{err}} = \{y_i\}$.

The ARIMA model can be expressed as [21]:

$$\begin{cases} \Phi(B)\nabla^d y_t = \Theta(B)\varepsilon_t + \mu_{\text{AM}} \\ E(\varepsilon_t) = 0 \\ \text{var}(\varepsilon_t) = \sigma_\varepsilon^2 \\ E(\varepsilon_t \varepsilon_s) = 0 \mid s \neq t \\ E(y_s \varepsilon_t) = 0 \mid \forall s < t \end{cases} \quad (2)$$

where y_t is the forecasting error sequence; B denotes the lag operator, namely $B^d y_t = y_{t-d}$; $\nabla^d = (1-B)^d$ denotes the d -order difference computation process, which transforms the input data into a stationary sequence, and the augmented Dickey-fuller (ADF) test can be used to evaluate the stationarity of the sequences [22]; E and var are the expectation and variance functions, respectively; ε_t is the residual error; σ_ε^2 is the variance of ε_t ; and $\Phi(B)$ and $\Theta(B)$ are the AR polynomial and moving average (MA) polynomial, respectively, which can be expressed as [21]:

$$\begin{cases} \Phi(B) = 1 - \sum_{rp=1}^{P_{AR}} k_{AR}^{rp} B^{rp} \\ \Theta(B) = 1 - \sum_{rq=1}^{Q_{MA}} k_{MA}^{rq} B^{rq} \end{cases} \quad (3)$$

where μ_{AM} , k_{AR}^{rp} , and k_{MA}^{rq} are the constant parameter, AR coefficient, and MA coefficient obtained by the fitting, respectively; P_{AR} and Q_{AR} are the orders of the AR and MA coefficients, respectively; and rp and rq are the count variables.

Through the ARIMA model, the estimation of the PDF/CDF of $\{y_t\}$ is converted into that of the residual error sequence $\{\varepsilon_t\}$.

The function of the GARCH- t model is to fit the PDF/CDF of $\{\varepsilon_t\}$. According to the existing research [23], $\{\varepsilon_t\}$ presents the conditional-heteroscedasticity characteristics. On this basis, the GARCH- t can be established as [24]:

$$\begin{cases} \varepsilon_t = v_t \sqrt{h_t} | v_t \sim N(0, 1) \\ \varepsilon_t | I_{t-1} \sim N(0, h_t) \\ h_t = \mu_{GA} + \sum_{lp=1}^{P_G} k_G^{lp} B^{lp} h_t + \sum_{lq=1}^{Q_{AC}} k_{AC}^{lq} B^{lq} \varepsilon_t^2 \end{cases} \quad (4)$$

where N is the normal distribution; I_{t-1} denotes all historical information before time $t-1$; h_t is the conditional variance; v_t is an independent and identically-distributed variable that obeys the standard normal distribution $N(0, 1)$; μ_{GA} , k_{AC}^{lq} , and k_G^{lp} are the fixed parameter, the autoregressive conditional heteroskedasticity (ARCH) coefficient, and the GARCH coefficient obtained by the fitting, respectively; P_G and Q_{AC} are the orders of the GARCH and ARCH coefficients, respectively; and lp and lq are the count variables.

By combining the calculation results of the ARIMA model and the GARCH- t model, the PDF of the original wind power output sequence can be obtained as:

$$\begin{cases} y_t = y_t^{ARIMA} + \varepsilon_t | \varepsilon_t \sim y_t^{GARCH} \\ x_{i,t} = x_{i,t}^{fc} + y_t \end{cases} \quad (5)$$

where y_t^{ARIMA} and y_t^{GARCH} are the calculation results of the ARIMA model and GARCH- t model, respectively; and $x_{i,t}$ and $x_{i,t}^{fc}$ are the measured power and forecasting power of the i th wind farm at time t , respectively.

Taking the historical output data of a wind farm for 1 day as an example, the calculation results of ARIMA and GARCH- t models are shown in Fig. 1.

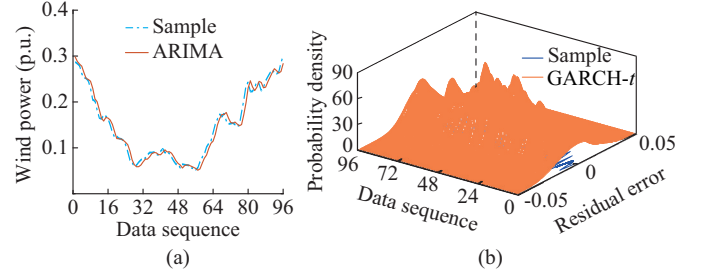


Fig. 1. Calculation results of ARIMA and GARCH- t models. (a) ARIMA model. (b) GARCH- t model.

The parameters of the ARIMA model and the GARCH- t model can be estimated by optimizing the log-likelihood function as:

$$\begin{cases} \hat{\Psi}_{ARIMA}^* = \arg \max (\ln f_{ARIMA}(\{y_t\}, \Psi_{ARIMA}^*)) \\ \hat{\Psi}_{GARCH}^* = \arg \max (\ln f_{GARCH}(\{\varepsilon_t\}, \Psi_{GARCH}^*)) \end{cases} \quad (6)$$

where Ψ_{ARIMA}^* and Ψ_{GARCH}^* are the parameter sets of the ARIMA model and the GARCH- t model, respectively; f_{ARIMA} and f_{GARCH} are the functions of the ARIMA model and the GARCH- t model, respectively; and $\hat{\Psi}_{ARIMA}^*$ and $\hat{\Psi}_{GARCH}^*$ are the estimation results, respectively.

III. TIME-VARYING R-VINE MIXED COPULA MODEL AS DEPENDENCE STRUCTURE

A. R-vine Copula Model for Multiple Wind Farms

An R-vine Copula model $R^* = \{T_i^* | i=1, 2, \dots, M\}$ can be defined as follows [25].

1) The R-vine Copula R^* is a nested set of $M-1$ layers of tree structures. Each tree T_i^* is composed of the node set N_i^* and the edge set E_i^* . The nodes and edges are the input data and pair Copulas, respectively.

2) For the first tree T_1^* , the node set N_1^* is the calculation results of the marginal distribution models. For other trees, $N_i^* = E_{i-1}^*$, i.e., each edge in tree T_i^* corresponds to a node in tree T_{i+1}^* .

3) A constraint is that two edges in E_i^* only share one node in N_i^* .

Except for the first tree, the nodes and edges are all composed of the conditioning and conditioned sets. For example, suppose that an edge e is defined as $\alpha(e), \beta(e) | \gamma(e)$, then $\alpha(e)$ and $\beta(e)$ are the conditioned sets, $\gamma(e)$ is the conditioning set. Further, suppose $e = (a, b) \in E_i^* | a, b \in N_i^*$, the edges (in tree T_{i-1}^*) corresponding to node a and node b are $e_a = \alpha(e_a), \beta(e_a) | \gamma(e_a)$ and $e_b = \alpha(e_b), \beta(e_b) | \gamma(e_b)$, respectively, then the correlation between the edges and the nodes can be expressed as [25]:

$$\gamma(e) = \{\alpha(e_a), \beta(e_a), \gamma(e_a)\} \cap \{\alpha(e_b), \beta(e_b), \gamma(e_b)\} \quad (7)$$

$$\begin{aligned} \{\alpha(e), \beta(e)\} = & \left(\{\alpha(e_a), \beta(e_a), \gamma(e_a)\} \setminus \gamma(e) \right) \cap \\ & \left(\{\alpha(e_b), \beta(e_b), \gamma(e_b)\} \setminus \gamma(e) \right) \end{aligned} \quad (8)$$

Based on (7) and (8), the pair Copula corresponds to edge e can be written as $C_{\alpha(e),\beta(e)|\gamma(e)}$. The input data for the pair Copula are two conditional cumulative probability sequences, which can be written as $F_{\alpha(e)|\gamma(e)}$ and $F_{\beta(e)|\gamma(e)}$. Then the joint-PDF of an M -dimensional data set can be expressed as [25]:

$$f_{\text{int}}(x^*) = \left[\prod_{j=1}^{M-1} \prod_{k=1}^{M-j} c_{\alpha(e_k),\beta(e_k)|\gamma(e_k)} \left(F_{\alpha(e_k)|\gamma(e_k)} F_{\beta(e_k)|\gamma(e_k)} \right) \right] \prod_{i=1}^M f_i(x_i) \quad (9)$$

where $x^* = \{x_i | i=1, 2, \dots, M\}$; f_{int} is the joint-PDF of x^* ; f_i is the marginal distribution model of x_i ; and $c_{\alpha(e_k),\beta(e_k)|\gamma(e_k)}$ is the k^{th} pair Copula in the j^{th} tree.

For convenience, the parentheses in the pair Copulas are omitted in the rest of this paper.

B. Time-varying Mixed Copula Model as Pair Copulas

The pair Copulas are the basic units of the R-vine Copula model. The function is to fit the joint-PDF of the binary data sequences. In this paper, the TMC model is established as the pair Copulas, which can be expressed as:

$$\begin{cases} C_{\text{mix}}(u, v, \Psi_C^*) = \sum_{i=1}^n \omega_i C_i(u, v, \theta_i^*) \\ \Psi_C^* = \left\{ \omega_i \mid i \in [1, n], \sum_{i=1}^n \omega_i = 1 \right\} \cup \left\{ \theta_i^* \mid i \in [1, n] \right\} \end{cases} \quad (10)$$

where u and v are the input data of the pair Copula model; C_{mix} is the TMC model; C_i is the basic bivariate Copula model selected to compose C_{mix} ; n is the number of basic bivariate Copula models; ω_i is the weight of C_i ; and Ψ_C^* and θ_i^* are the parameter sets of C_{mix} and C_i , respectively.

In this paper, the commonly-used t Copula model, Clayton Copula model, and Gumbel Copula model are selected as C_i . The expressions are as follows [26]:

$$C_{\text{tc}}(u, v, \rho_i, k_{\text{CT}}) = \int_{-\infty}^{t_k^{-1}(u)} \int_{-\infty}^{t_k^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho_i^2}} \left[1 + \frac{x^2 + y^2 - 2\rho_i xy}{k_{\text{CT}}(1-\rho_i^2)} \right]^{-\frac{(k_{\text{CT}}+2)}{2}} dx dy \quad (11)$$

$$C_{\text{Cl}}(u, v, \rho_i^{\text{Cl}}) = \max \left(\left(u^{-\rho_i^{\text{Cl}}} + v^{-\rho_i^{\text{Cl}}} - 1 \right)^{-\frac{1}{\rho_i^{\text{Cl}}}}, 0 \right) \quad (12)$$

$$C_{\text{G}}(u, v, \rho_i^{\text{G}}) = \exp \left\{ - \left[(-\ln u)^{\rho_i^{\text{G}}} + (-\ln v)^{\rho_i^{\text{G}}} \right]^{\frac{1}{\rho_i^{\text{G}}}} \right\} \quad (13)$$

where C_{tc} , C_{Cl} , and C_{G} are the t Copula, Clayton Copula, and Gumbel Copula models, respectively; ρ_i , ρ_i^{Cl} , ρ_i^{G} , and k_{CT} are the model parameters; and t_k^{-1} is the inverse function of the t Copula model.

To track the time-varying process of the correlation between the wind farms, the parameters of the above 3 models need to be calculated dynamically. For this purpose, the DCC model and the Patton model are introduced in this paper.

1) t Copula Model

For the t Copula model, the DCC model is used to calculate the parameters [27], which is presented as:

$$\begin{cases} \rho_t = (1 - \alpha_{\text{DCC}} - \beta_{\text{DCC}}) \bar{Q}_{\text{cor}} + \alpha_{\text{DCC}} (\varepsilon_{t-1} \varepsilon'_{t-1}) + \beta_{\text{DCC}} \rho_{t-1} \\ \alpha_{\text{DCC}} + \beta_{\text{DCC}} < 1 \mid \alpha_{\text{DCC}}, \beta_{\text{DCC}} \in (0, 1) \end{cases} \quad (14)$$

where α_{DCC} and β_{DCC} are the parameters of the DCC model; \bar{Q}_{cor} is the covariance coefficient of the input data; ε_{t-1} is the input data at time $t-1$, i.e., $\varepsilon_{t-1} = [u_{t-1}, v_{t-1}]$; and the symbol ' denotes the transposition.

2) Clayton and Gumbel Copula Models

For the Clayton Copula model and the Gumbel Copula model, the Patton model is used to calculate the parameters [28], which is presented as:

$$\begin{cases} R_t = A_{\text{ptt}} \left(\omega_{\text{ptt}} + \beta_{\text{ptt}} R_{t-1} + \alpha_{\text{ptt}} \frac{1}{P_{\text{ptt}}} \sum_{p=1}^{P_{\text{ptt}}} |u_{t-p} - v_{t-p}| \right) \\ A_{\text{ptt}}(x) = (1 + e^{-x})^{-1} \end{cases} \quad (15)$$

where ω_{ptt} , α_{ptt} , and β_{ptt} are the parameters of the Patton model; P_{ptt} is the length of the historical data sequence used to fit the correlation coefficient at time t , which is usually set to be 10; and A_{ptt} is the logistic function which keeps the calculation results of the Patton model within the required range.

Moreover, the calculation results R_t of the Patton model need to be further converted into the parameters of the Copula models. The calculation method is [26]:

$$\rho_t^{\text{Cl}} = \frac{2R_t}{1-R_t} \quad (16)$$

$$\rho_t^{\text{G}} = (1-R_t)^{-1} \quad (17)$$

Having adopted the DCC and Patton models, the parameter set θ_i^* in (10) can be expressed as:

$$\begin{cases} \theta_{\text{tc}}^* = \{\alpha_{\text{DCC}}, \beta_{\text{DCC}}, k_{\text{CT}}\} \\ \theta_{\text{Cl}}^* = \{\omega_{\text{ptt}}^{\text{Cl}}, \alpha_{\text{ptt}}^{\text{Cl}}, \beta_{\text{ptt}}^{\text{Cl}}\} \\ \theta_{\text{G}}^* = \{\omega_{\text{ptt}}^{\text{G}}, \alpha_{\text{ptt}}^{\text{G}}, \beta_{\text{ptt}}^{\text{G}}\} \end{cases} \quad (18)$$

where θ_{tc}^* , θ_{Cl}^* , and θ_{G}^* are the parameter sets of the t Copula, Clayton Copula, and Gumbel Copula, respectively; α_{DCC} , β_{DCC} , and k_{CT} are the parameters of the t Copula; $\omega_{\text{ptt}}^{\text{Cl}}$, $\alpha_{\text{ptt}}^{\text{Cl}}$, and $\beta_{\text{ptt}}^{\text{Cl}}$ are the parameters of the Clayton Copula; and $\omega_{\text{ptt}}^{\text{G}}$, $\alpha_{\text{ptt}}^{\text{G}}$, and $\beta_{\text{ptt}}^{\text{G}}$ are the parameters of the Gumbel Copula. And the parameters of the TMC model can be estimated by optimizing the log-likelihood function as:

$$\hat{\Psi}_C^* = \arg \max (\ln C_{\text{mix}}(u, v, \Psi_C^*)) \quad (19)$$

where $\Psi_C^* = \{\omega_{\text{tc}}, \omega_{\text{Cl}}, \omega_{\text{G}}, \theta_{\text{tc}}^*, \theta_{\text{Cl}}^*, \theta_{\text{G}}^*\}$, and ω_{tc} , ω_{Cl} , and ω_{G} are the weights of the t Copula, Clayton Copula, and Gumbel Copula, respectively; and $\hat{\Psi}_C^*$ is the estimation result.

C. Sequential Generation Method of R-vine Structure

In this subsection, the structure generation method of the TRVMC model is proposed based on the MST algorithm in [29]. The specific calculation process is as follows.

Step 1: establish the TMC model as the pair Copula model.

Step 2: take the calculation results of the marginal distribution models as the input data of the first tree. Fit the pair Copulas of each two input data sequences with the TMC model.

Step 3: evaluate the accuracy of the pair Copulas in *Step 2* with the quantitative index. In this paper, the Akaike information criterion (AIC) is adopted [30].

$$I_{AIC} = 2n_{ml} - 2\ln L_m \quad (20)$$

where I_{AIC} is the AIC index; n_{ml} is the number of model parameters; and L_m is the value of the maximum likelihood function.

Step 4: generate the structure of each tree. This process is an optimization problem, and the objective is to minimize the sum of the AIC values of all pair Copulas in the tree. For the first tree, the Prim algorithm in [25] is implemented in this paper.

Step 5: when the structure of the k^{th} tree is generated, calculate the input data of the $(k+1)^{\text{th}}$ tree as [25]:

$$F_{a(e)|\{\gamma(e), \beta(e)\}} = \frac{\partial C_{a(e), \beta(e)|\gamma(e)}}{\partial F_{\beta(e)|\gamma(e)}} \quad (21)$$

Step 6: except for the first tree, the structure of each tree is constrained by that of the previous tree, which limits the number of possible structures. When the structure of the k^{th} tree is generated, list all possible structures of the $(k+1)^{\text{th}}$ tree. The validity of the structures can be judged as [25]:

$$\forall e_a^{k+1}, e_b^{k+1} \in E_{k+1}^* \rightarrow \#(e_a \cap e_b) \leq 1 \quad (22)$$

where e_a^{k+1} and e_b^{k+1} are the edges in the $(k+1)^{\text{th}}$ tree; E_{k+1}^* is the set of all edges in the $(k+1)^{\text{th}}$ tree; and $\#$ denotes the cardinality of the set.

Step 7: for each possible structure in *Step 6*, fit the pair Copulas with the TMC model.

Step 8: repeat *Step 4* to *Step 7* until the structures of all trees are generated, which together make up the structure of the R-vine Copula model.

D. Model Applicability Analysis

The TRVMC model mainly aims at the wind farms located in the same region. More specifically, the model is more suitable for wind farms with a strong correlation. The reasons are as follows. The main function of the TRVMC model is to establish the joint-probability distribution model of the output of multiple wind farms. By making the probability distribution of the TRVMC model capture the characteristics of the statistical joint-frequency distribution of the wind power output sequences, the model can reflect the correlation of multiple wind farms.

Take the power output of 2 wind farms in the same region for example. The Kendall correlation coefficient is 0.761. The joint-frequency distribution of the power output sequences is shown in Fig. 2(a). The probability distribution of the corresponding TRVMC model is shown in Fig. 2(b). As shown in Fig. 2, the probability distribution can well capture the characteristics of the statistical joint-frequency distribution.

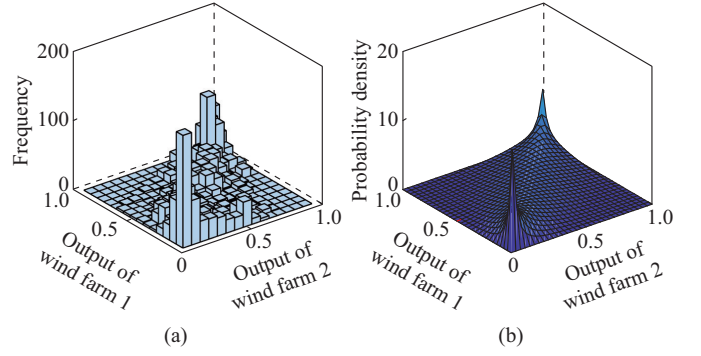


Fig. 2. Joint-frequency distribution and probability distribution of power output of wind farms. (a) Joint-frequency distribution. (b) Probability distribution.

However, there is no clear boundary between strong correlation and weak correlation. Since the model is data driven, the data of any wind farm can be selected as the input data. If the wind farms are far away from each other and the power output sequences are independent, according to the Bayesian formula, the joint-PDF can be expressed as [31]:

$$f_{\text{jnt}}(x^*) = \prod_{k=1}^M f_k(x_k) \quad (23)$$

In (23), the joint-PDF is equal to the multiplication of each independent PDF. In other words, the models considering the correlation are the same as those ignoring the correlation. Therefore, the calculation results of the models will become the same. Due to the above reasons, the TRVMC model is mainly applicable to regional wind farms.

IV. SCENARIO GENERATION METHOD FOR MULTIPLE WIND FARMS

Suppose that N_{sc} scenarios are generated for N_{wf} wind farms, and each scenario contains N_{pt} sampling points. Having fitted parameters of the ARIMA-GARCH- t model and the TRVMC model, the forecasting scenarios can be generated through the following steps.

Step 1: generate an $N_{sc} \times N_{pt} \times N_{wf}$ random matrix **Rnd**, in which all elements obey the uniform distribution $U(0,1)$.

Step 2: decompose the joint-PDF and reorder the input data sequences to $\{x'_1, x'_2, \dots, x'_{N_{wf}}\}$ according to the generated R-vine structure, as shown in (9).

Step 3: assign the values in **Rnd** to the conditional cumulative probability values as:

$$\begin{cases} F_1^{j_{pt}}(x'_1) = \text{Rnd}[i_{sc}, j_{pt}, 1] \\ F_{2|1}^{j_{pt}}(x'_2 | x'_1) = \text{Rnd}[i_{sc}, j_{pt}, 2] \\ \vdots \\ F_{N_{wf}|1,2,\dots,N_{wf}-1}^{j_{pt}}(x'_{N_{wf}} | x'_1, x'_2, \dots, x'_{N_{wf}-1}) = \text{Rnd}[i_{sc}, j_{pt}, N_{wf}] \end{cases} \quad (24)$$

where i_{sc} is the i_{sc}^{th} scenario; and j_{pt} is the j_{pt}^{th} sampling point.

Step 4: calculate the cumulative probability of the power output of each wind farm at the j_{pt}^{th} sampling point.

Supposing two variables $F_{a_{wf}}^{j_{pt}}(x'_{a_{wf}} | \gamma)$ and $F_{b_{wf}}^{j_{pt}}(x'_{b_{wf}} | \{\gamma, x'_{a_{wf}}\})$

have been given by (24) or calculated by (25) in the last circle, the two variables are connected by a pair Copula model $C_{a_{wf}, b_{wf} | \gamma}$ as:

$$F_{b_{wf}}^{j_{pt}}(x'_{b_{wf}} | \{ \gamma, x'_{a_{wf}} \}) = \frac{\partial C_{a_{wf}, b_{wf} | \gamma}}{\partial F_{a_{wf}}^{j_{pt}}(x'_{a_{wf}} | \gamma)} \quad (25)$$

For the pair Copula $C_{a_{wf}, b_{wf} | \gamma}$, $F_{a_{wf}}^{j_{pt}}(x'_{a_{wf}} | \gamma)$ and $F_{b_{wf}}^{j_{pt}}(x'_{b_{wf}} | \gamma)$ are the corresponding input data. $F_{b_{wf}}^{j_{pt}}(x'_{b_{wf}} | \gamma)$ can be calculated by the interpolation methods [25].

Step 5: repeat *Step 4* until the cumulative probability of the power output of all wind farms at the j_{pt}^{th} sampling point has been calculated, namely $\{F_1^{j_{pt}}(x'_1), F_2^{j_{pt}}(x'_2), \dots, F_{N_{wf}}^{j_{pt}}(x'_{N_{wf}})\}$.

Step 6: taking $\{F_1^{j_{pt}}(x'_1), F_2^{j_{pt}}(x'_2), \dots, F_{N_{wf}}^{j_{pt}}(x'_{N_{wf}})\}$ as the input data, calculate the power output of each wind farm through the inverse function of the marginal distribution model:

$$P_{k_{wf}}^{j_{pt}} = D_{k_{wf}}^{j_{pt}}(F_{k_{wf}}^{j_{pt}}(x'_{k_{wf}})) \quad (26)$$

where $P_{k_{wf}}^{j_{pt}}$ is the power output of the k_{wf}^{th} wind farm at the j_{pt}^{th} sampling point; and $D_{k_{wf}}^{j_{pt}}$ is the inverse function of the marginal distribution model.

Step 7: based on the ARIMA-GARCH- t model, calculate the CDF of each wind power output sequence at the next sampling point.

Step 8: repeat *Step 3* to *Step 7* until all sampling points in the i_{sc}^{th} scenario are generated.

Step 9: repeat *Step 3* to *Step 8* until all scenarios are generated.

V. OVERALL DESCRIPTION OF MODELING AND SCENARIO GENERATION PROCESS

The basic principle of the models in this paper is the Sklar theorem. According to the theorem, the joint-PDF of high-dimensional data can be calculated by (9). And the corresponding joint-CDF is [25]:

$$\begin{cases} u_i = F_i(x_i) |_{i=1,2,\dots,M} \\ F_{jnt}(x_1, x_2, \dots, x_M) = C_{1,2,\dots,M}(u_1, u_2, \dots, u_M, \Psi_{ent}^*) \end{cases} \quad (27)$$

where $\{x_1, x_2, \dots, x_M\}$ is the wind power output sequence; $\{u_1, u_2, \dots, u_M\}$ is the cumulative probability of $\{x_1, x_2, \dots, x_M\}$; F_{jnt} is the joint-CDF; $C_{1,2,\dots,M}$ is the Copula model; and Ψ_{ent}^* is the parameter set of $C_{1,2,\dots,M}$.

According to (27), the joint-CDF can be calculated by combining the marginal distribution models with the Copula model. To this end, the ARIMA-GARCH- t model and the TRVMC model are established in this paper.

The modeling process is shown in part 1 of Fig. 3. The historical data from the wind farms are taken as the input data to fit the model parameters. Firstly, the ARIMA-GARCH- t models are established. Then, the cumulative probability of historical wind power output sequences is calculated and taken as the input data for the TRVMC model. After that, according to (27), the joint-CDF is established based on the

TRVMC model.

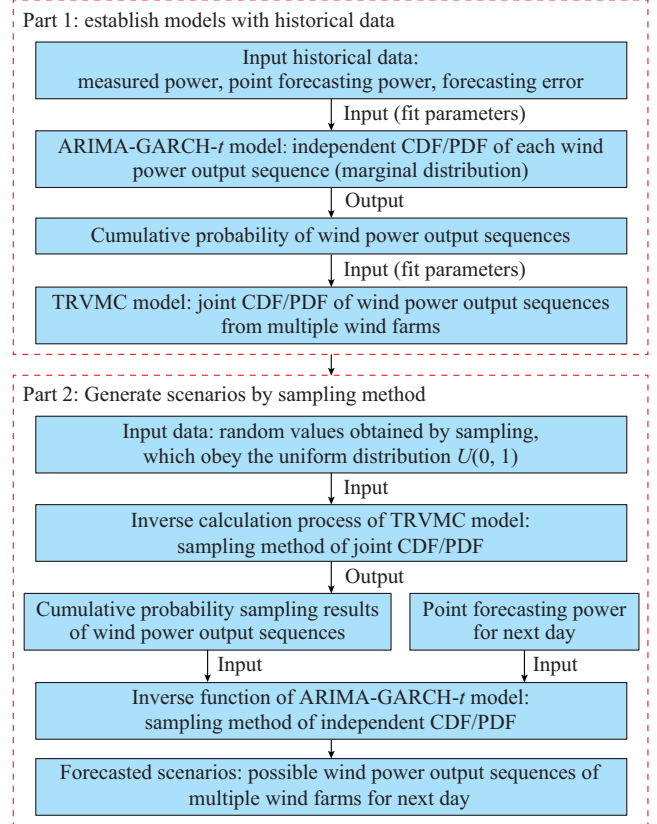


Fig. 3. Overall process of modeling and scenario generation.

The scenario generation process is shown in part 2 of Fig. 3, which can be regarded as the inverse calculation process of the established models. In this part, the random values obtained by sampling are taken as the input data. Firstly, the cumulative probability of the output sequences is calculated by the sampling method of the joint-CDF. Then, the calculation results together with the point forecasting power of the next day are taken as the input data. Through the sampling method of the independent CDFs, the forecasting scenarios, i.e., the possible output sequences of multiple wind farms for the next day, are generated.

VI. FRAMEWORK OF EVALUATION

A. Evaluation of Marginal Distribution Models

The function of the marginal distribution model is to calculate the independent probability distribution of each wind power sequence. The model can be evaluated by the following steps.

Step 1: calculate the model parameters based on the historical output data of a wind farm.

Step 2: taking the wind power output of a period in the future as the test data, fit the probability distribution intervals of the test data under the preset confidence levels with the marginal distribution model.

Step 3: compare the fitted probability distribution intervals with the statistical characteristics of the test data.

According to [32], the marginal distribution model can be evaluated from two aspects: reliability and sharpness.

1) Reliability

The reliability index reflects the deviation between the probability distribution fitted by the model and the frequency distribution of the actual data sequence [32].

$$R_{lb}^{(\lambda)} = \left| \frac{n_{in}}{n_{sp}} - \lambda \right| \times 100\% \quad (28)$$

where $R_{lb}^{(\lambda)}$ is the reliability index; n_{in} is the number of data points in the interval; n_{sp} is the length of the test data sequence; and λ is the preset confidence level.

2) Sharpness

The sharpness index reflects the redundancy of the probability distribution intervals [32].

$$S_{hp}^{(\lambda)} = \frac{1}{n_{sp}} \sum_{i=1}^{n_{sp}} (D_i^{up} - D_i^{lw}) \quad (29)$$

where $S_{hp}^{(\lambda)}$ is the sharpness index; and D_i^{up} and D_i^{lw} are the upper and lower boundaries of the interval of the i^{th} data point, respectively.

The marginal distribution model with smaller reliability and sharpness values has better effectiveness.

B. Evaluation of Copula Models

The AIC and the Bayesian information criterion (BIC) are commonly used to evaluate the accuracy of the Copula models. The expression of the AIC index has been given in (20). The BIC index can be expressed as [33]:

$$I_{BIC} = n_{ml} \ln n_p - 2 \ln L_m \quad (30)$$

where I_{BIC} is the BIC index; n_{ml} is the number of model parameters; n_p is the number of sample points; and L_m is the value of the maximum likelihood function.

Both the AIC and the BIC are based on the maximum likelihood function. The AIC introduces a penalty factor for the complexity of the model. And the BIC further takes into account the influence of the sample size. The Copula model with smaller AIC and BIC has better effectiveness.

C. Evaluation of Generated Scenarios

In this paper, the generated scenarios are evaluated from the following 3 aspects.

1) Energy Score (ES)

The ES index evaluates the difference between the actual wind power sequence and the generated scenarios [34]. The scenarios with smaller ES values have better effectiveness.

$$I_{ES} = \frac{1}{N_{sc}} \sum_{i=1}^{N_{sc}} \|x_{sc}^{Ac} - x_{sc}^i\| - \frac{1}{2N_{sc}^2} \sum_{i=1}^{N_{sc}} \sum_{j=1}^{N_{sc}} \|x_{sc}^i - x_{sc}^j\| \quad (31)$$

where I_{ES} is the ES index; N_{sc} is the number of generated scenarios; x_{sc}^{Ac} is the actual wind power sequence; and x_{sc}^i is the i^{th} scenario.

2) Time-series Characteristics

In this paper, the time-series characteristics are evaluated from 2 aspects: the fluctuation characteristics and the autocorrelation function (ACF) [35].

The fluctuation is defined as the first-order difference sequence of wind power output, as shown in (32). The quantile-quantile (Q-Q) diagram and the cumulative probability curve are introduced to compare the fluctuation characteristics of the generated scenarios and the actual wind power se-

quence.

$$x_t^{fluc} = \nabla^1 x_t \quad (32)$$

where x_t and x_t^{fluc} are the wind power output sequence and the corresponding fluctuation sequence, respectively; and ∇^1 denotes the first-order difference computation.

The ACF index is the correlation between wind power sequences x_t and x_{t+k} . It intuitively reflects the time-series characteristics of wind power output. The expression is [35]:

$$I_{ACF}(k) = \frac{E((x_t - \mu_t^y)(x_{t+k} - \mu_{t+k}^y))}{\sigma_t^y \sigma_{t+k}^y} \quad (33)$$

where I_{ACF} is the ACF index; μ_t^y and μ_{t+k}^y are the mean values of x_t and x_{t+k} , respectively; σ_t^y and σ_{t+k}^y are the standard deviations; and k is the delay time. When the delay time $k=0$, $I_{ACF}(0)=1$.

The scenarios with ACF values closer to the actual wind power sequence have better effectiveness.

3) Spatial-temporal Correlation

In [36], the cross-correlation function (CCF) is introduced to evaluate the spatial-temporal correlation between different wind farms. The expression is:

$$I_{CCF}(k) = \frac{E((x_t^{s1} - \mu_t^{y,s1})(x_{t+k}^{s2} - \mu_{t+k}^{y,s2}))}{\sigma_t^{y,s1} \sigma_{t+k}^{y,s2}} \quad (34)$$

where I_{CCF} is the CCF index; x_t^{s1} and x_{t+k}^{s2} are the output sequences of 2 wind farms; $\mu_t^{y,s1}$ and $\mu_{t+k}^{y,s2}$ are the mean values of x_t^{s1} and x_{t+k}^{s2} , respectively; and $\sigma_t^{y,s1}$ and $\sigma_{t+k}^{y,s2}$ are the standard deviations of x_t^{s1} and x_{t+k}^{s2} , respectively.

When the delay time $k=0$, the CCF is the commonly-used Pearson correlation coefficient, which reflects the overall correlation of the 2 wind farms. With the change of delay time, the CCF reflects the correlation with time-series characteristics.

The scenarios with CCF values closer to the actual power sequence have better effectiveness.

VII. CASE STUDY

A. Description of Case Study

In this subsection, the historical data from 8 wind farms in Northwest China is used for analysis. The sampling time is 3 months (from January 1st to March 31st), and the sampling frequency is 96 points per day.

For the power output of the 8 wind farms, the average Kendall correlation coefficient of each wind farm to other wind farms is shown in Table I. As shown in Table I, the correlation of the wind farms is relatively strong.

TABLE I
AVERAGE KENDALL CORRELATION COEFFICIENT OF EACH WIND FARM TO OTHER WIND FARMS

Wind farm	Average correlation	Wind farm	Average correlation
1	0.78	5	0.74
2	0.71	6	0.80
3	0.74	7	0.78
4	0.69	8	0.68

To better verify the superiority of the proposed model in this paper, the simulation work is divided into three parts. In Section VI-B and VI-C, the comparisons between different marginal distribution models and different Copula models are conducted. In Section VI-D, the effectiveness of the generated scenarios is evaluated from 3 aspects.

In addition, to verify the efficacy of the proposed method, based on the data from the 2012 Global Energy Forecasting Competition (GEFCOM 2012) [37], the scenarios are generated and evaluated as in Section VI-D. The simulation results are shown in Supplementary Material Part A.

B. Comparison Between Different Marginal Distribution Models

Four marginal distribution models are selected for comparison: ① the ARIMA-GARCH- t model; ② the student- t model [8]; ③ the ECDF model [10]; ④ the KDE model [12].

The historical output data of 1 wind farm is taken to carry out the simulation. The data of the first 85 days are used to fit the model parameters, while the data of the last 5 days are used as test data. The reliability and the sharpness indexes are used for evaluation.

Figure 4 shows the reliability of the four models. A group of 5 confidence levels is selected: {55%, 65%, 75%, 85%, 95%}. At each confidence level, the reliability values of the 4 models are plotted in the corresponding direction, and the distance from the center point intuitively shows the reliability values. The simulation results based on the data of the 3rd day and all 5 days are shown in Fig. 4(a) and Fig. 4(b), respectively.

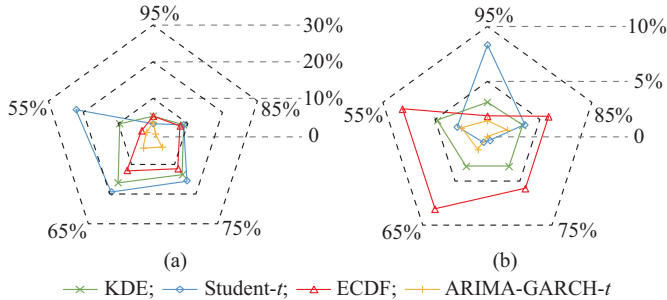


Fig. 4. Comparison of reliability index of different marginal distribution models at different confidence levels. (a) Comparison based on data of the third day. (b) Comparison based on data of all 5 days.

Figure 5 shows the sharpness of the four models. The simulation results at the 75% and 95% confidence levels are shown in Fig. 5(a) and Fig. 5(b), respectively.

As shown in Figs. 4 and 5, both the reliability values and the sharpness values of the ARIMA-GARCH- t model are smaller than those of the other 3 models. When the reliability values are small, the probability distribution fitted by the model is close to the statistical frequency distribution of the actual wind power sequence. When the sharpness values are small, the width of the probability distribution intervals is narrow with low redundancy. The simulation results show that the ARIMA-GARCH- t model achieves high accuracy as the marginal distribution model of wind power data. Therefore, it can provide more reliable input data for the Copula model.

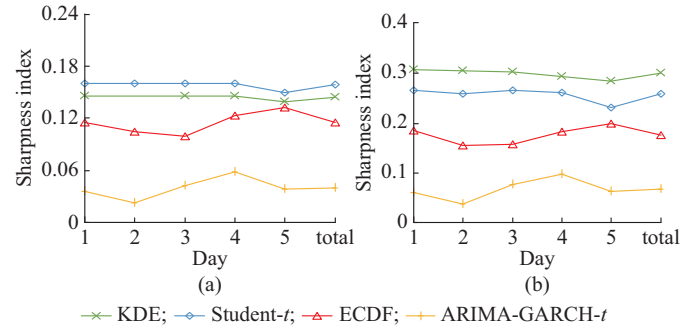


Fig. 5. Comparison of sharpness index of different marginal distribution models. (a) Comparison under 75% confidence level. (b) Comparison under 95% confidence level.

Through theoretical analysis, for the student- t , KDE, and ECDF models, the basic principle is to simulate the long-term frequency distribution of historical samples and take it as the PDF of the test data. However, due to the time-varying characteristics, the short-term probability distribution of wind power might be significantly different from the long-term probability distribution, which may lead to reliability defects in the other 3 models.

Compared with the other 3 models, the ARIMA-GARCH- t model has two advantages. First, it can track the time-varying process of the probability distribution of wind power. On this basis, it can provide an accurate PDF for wind power at each moment, as shown in Fig. 1(b), which contributes to reducing the reliability values. Second, the model can reflect the time-series characteristics of wind power output, as shown in Fig. 1(a). A simple example of the characteristics is that when the wind power is large at this moment, it is less likely that the wind power is small at the next moment. The width of the fitted probability distribution intervals is effectively reduced. As a consequence, the sharpness values become small.

C. Comparison Between Different Copula Models

In this part, 4 high-dimensional Copula models are selected for comparison: ① the TRVMC model in this paper; ② the static C-vine Copula model [19]; ③ the static D-vine Copula model [20]; ④ the high-dimensional Gaussian Copula model [18].

Taking the AIC and the BIC as the evaluation indexes, the simulation results are as follows.

As shown in Table II, the TRVMC model has the smallest AIC and BIC values. The simulation results show that, on one hand, the TRVMC model has the best accuracy in fitting the dependence structure of multiple wind farms. On the other hand, the complexity of the TRVMC is not much higher than the other 3 models.

Through theoretical analysis, the TRVMC model is superior to the other 3 Copula models from 3 aspects.

1) The model does not need to make assumptions on the correlation between multiple wind farms. The flexible dependence structure enables the TRVMC model to capture the spatial-temporal correlation of multiple wind farms effectively.

TABLE II
COMPARISON OF AIC AND BIC OF COPULA MODELS

Copula model	AIC	BIC
TRVMC	-40397.3	-38023.7
C-vine	-27548.0	-25174.4
D-vine	-29533.4	-27159.8
Gaussian	-24375.6	-22002.0

2) The TMC model is used as the pair Copulas, which has higher accuracy than the classic bivariate Copula models in describing the complex joint distribution of every 2 wind farms.

3) The dynamic calculation method of the model parameters is introduced into the TRVMC model.

Therefore, the model can track the time-varying process of the correlation between the wind farms. The above factors greatly enhance the applicability of the TRVMC to wind farms in different regions and different periods.

D. Evaluation of Output Scenarios of Multiple Wind Farms

In this part, 4 scenario generation models are selected for comparison: ① model 1, the proposed model in this paper; ② model 2, the independent ARIMA-GARCH- t model; ③ model 3, the static D-vine Copula model [20]; ④ model 4, the independent KDE model [12]. In models 2 and 4, the correlations of multiple wind farms are ignored.

The typical characteristics of the 4 models are compared as in Table III.

TABLE III
COMPARISON OF TYPICAL CHARACTERISTICS OF SCENARIO GENERATION MODELS

Model	Time-series characteristic	Spatial-temporal correlation
1	✓	✓
2	✓	
3		✓
4		

Take the wind power data on March 28th as the test data. A total of 100 scenarios are generated for evaluation. First, the output scenarios of each wind farm are generated. Then, the joint-output scenarios of 8 wind farms are obtained through superposition calculation.

The evaluation work consists of the following 3 parts.

1) Overall Evaluation of Scenarios

In Fig. 6, the generated scenarios of the joint-output of 8 wind farms are shown. The effectiveness of the scenarios is evaluated by the SE index, as shown in Fig. 7.

As shown in Fig. 7, for each single wind farm, the scenarios generated by model 1 and model 2 have smaller SE values than those of model 3 and model 4. In the case of the joint-output of 8 wind farms, the scenarios generated by model 1 have the smallest SE index value. The simulation results verify the effectiveness and superiority of the proposed model in this paper.

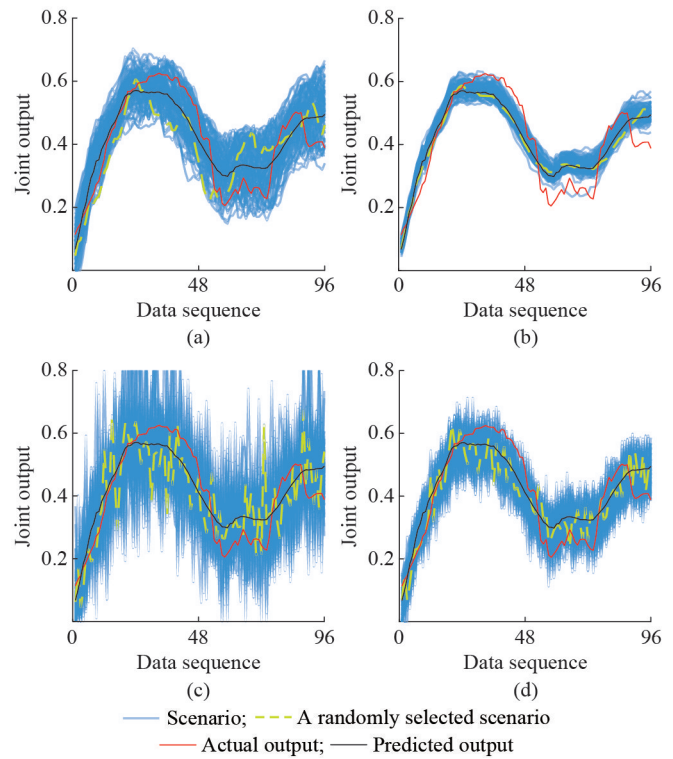


Fig. 6. Generated scenarios of joint-output of 8 wind farms. (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.

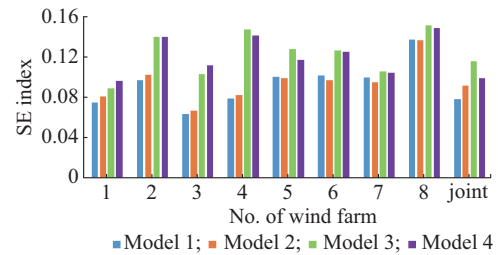


Fig. 7. SE index comparison of each wind farm.

Through theoretical analysis, model 2 is the marginal distribution of model 1. Therefore, the SE values of the two models are similar when generating scenarios for a single wind farm. For the same reason, the SE values of model 3 and model 4 are also similar. The scenarios of 2 random wind farms are provided in Supplementary Material Part B.

When generating scenarios for the joint-output of 8 wind farms, the fitting results of the 4 models are quite different. The comparison can be divided into 2 aspects.

On one hand, compared with model 1 and model 2, model 3 and model 4 ignore the time-series characteristics, i.e., the auto-correlation of wind power in the temporal dimension. As a result, the scenarios generated by model 3 and model 4 fluctuate more frequently and sharply. Theoretically, over a short period of time, the wind power output at the next moment is correlated with that in the previous period. For example, when the current wind power is large, the wind power is unlikely to be quite small at the next moment. If the temporal-dimensional correlation of wind power output is ignored, the adjacent data points in the generated

scenarios are more independent. When generating the output scenario at the next moment, the trend of the previous sequence is ignored, and the randomness of the calculation result is stronger. As a result, the generated scenarios fluctuate more frequently with greater amplitude. In some cases, the fluctuation range might be much larger than the actual wind power output sequence, as shown in Fig. 6(c).

On the other hand, compared with model 1 and model 3, model 2 and model 4 ignore the spatial-temporal correlation between multiple wind farms. As a result, the scenarios of different wind farms are more independent. After superposition calculation, the fluctuation range of the joint-output scenarios is largely reduced. Theoretically, since the wind farms are located in the same region, the environmental factors are similar. Therefore, the output of the wind farms is correlated. And the changing process of the output sequences tends to follow similar trends. If the spatial-temporal correlation is ignored, in the generated scenarios, when the output of one wind farm is large, the output of the other wind farms may be small. In the superposition calculation, the peak output of one wind farm may be added with the valley output of the other wind farms. As a result, the fluctuation range of the joint-output scenarios is largely reduced. In some cases, the scenarios might not be able to envelop the actual joint-output sequence, as shown in Fig. 6(b).

2) Evaluation of Time-series Characteristics

In this part, the scenarios of the joint-output of 8 wind farms are used for simulation.

The fluctuation characteristics of the scenarios are evaluated by the Q-Q diagram shown in Fig. 8(a) and the cumulative probability curve as shown in Fig. 8(b). In the ideal case, the Q-Q curve is a 45-degree line. Besides, for better demonstration, only a part of the cumulative probability curve is shown. The ACF values are shown in Fig. 9.

As shown in Figs. 8 and 9, the fluctuation characteristics and the ACF of the scenarios generated by model 1 are the closest to those of the actual joint-output sequence. The simulation results prove that the proposed model can effectively describe the time-series characteristics of wind power output.

Through theoretical analysis, since model 3 and model 4 ignore the time-series characteristics, the adjacent data points in the generated scenarios are relatively independent. More specifically, when the wind power is large at this moment, the wind power might be rather small at the next moment. As a result, the overall fluctuation range of the scenarios largely exceeds that of the actual joint-output sequence, which is directly reflected in the generated scenario, as shown in Fig. 6. Therefore, the ACF values are always lower than the actual values.

Compared with model 1, model 2 ignores the correlation of multiple wind farms. Therefore, the fluctuation range of the joint-output scenarios generated by model 2 is greatly reduced after the superposition calculation. As the direct performances, in Fig. 6, the scenarios are unreasonably smooth. In Fig. 8(a) and Fig. 8(b), the Q-Q curve and the cumulative probability curve of model 2 are almost always higher than those of the actual joint-output sequence.

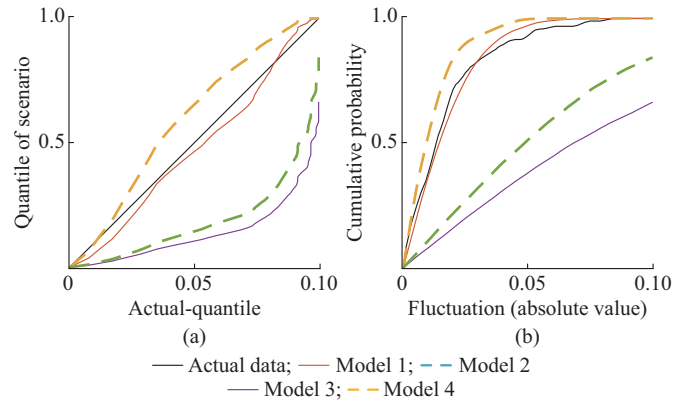


Fig. 8. Comparison of probability distribution of fluctuations. (a) Q-Q diagram of fluctuations. (b) Cumulative probability curve of fluctuations.

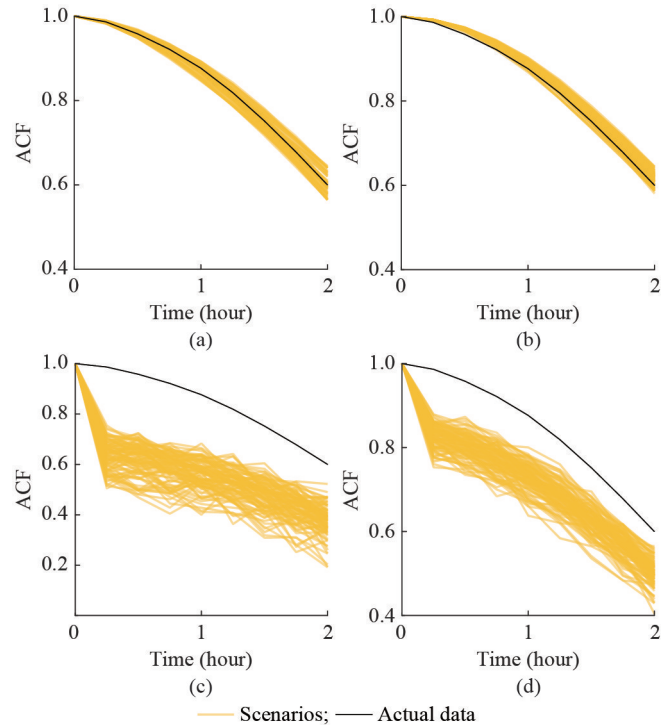


Fig. 9. ACF comparison of scenarios generated by different models. (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.

3) Evaluation of Spatial-temporal Correlation

In this part, the generated scenarios of 2 wind farms are used for simulation. The effectiveness of the scenarios is evaluated by the CCF index. The simulation results are shown in Fig. 10.

As shown in Fig. 10, the CCF values of scenarios generated by model 1 are the closest to the actual values. The simulation results prove that the proposed model can effectively describe the spatial-temporal correlation of different wind farms.

Through theoretical analysis, since model 2 and model 4 ignore the correlation between the wind farms, the CCF values are almost always smaller than the actual values. Besides, the distribution of the CCF curves is relatively dispersed, which indicates that the correlation between the wind farms in different scenarios is quite different. This is not consistent with the actual situation.

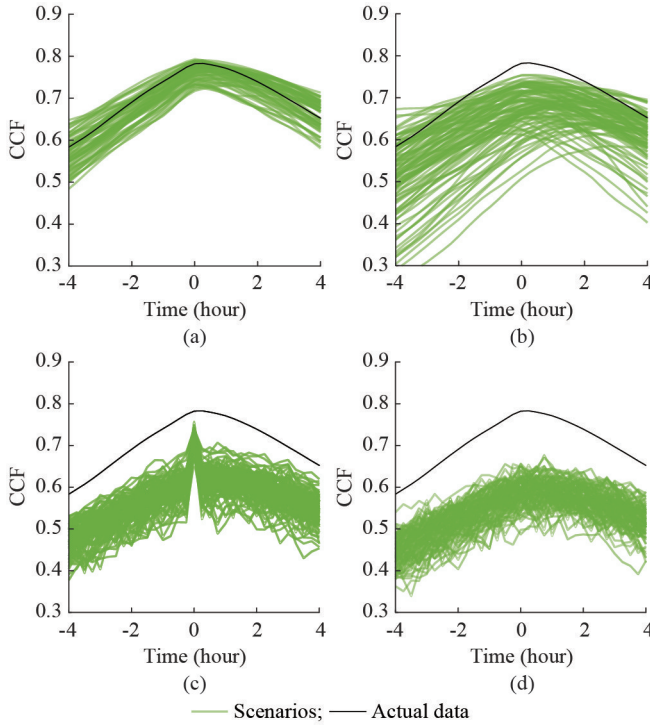


Fig. 10. CCF comparison of output scenarios of 2 wind farms. (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4.

Although model 1 and model 3 both fit the correlation of the multiple wind farms by the Copula model, model 1 further considers the time-series characteristics of wind power output and the time-varying characteristics of the correlations. As a result, the CCF values of all scenarios generated by model 1 are close to the actual values. On the contrary, model 3 only considers the overall correlation of the wind farms, more specifically, the Kendall correlation coefficient. Consequently, the CCF values of the scenarios generated by model 3 are close to the actual values only when the delay time is 0 and still smaller than the actual value.

VIII. CONCLUSION

In this paper, a scenario generation method for the output of multiple wind farms considering the time-series characteristics and spatial-temporal correlation is proposed. The main conclusions are as follows.

1) The ARIMA-GARCH- t model can accurately fit the marginal distribution of wind power output, i.e., the independent CDF. For 1-day wind power output data, the reliability index value is within 10%, and the sharpness index value is within 0.1. Therefore, it can provide reliable input data for the Copula model.

2) Compared with the Copula models in the existing research, the TRVMC model has higher fitting accuracy for the joint-distribution of the output of multiple wind farms, which has smaller AIC and BIC values.

3) The ARIMA-GARCH- t model and the TRVMC model are combined to generate the output scenarios of multiple wind farms. The generated scenarios have similar time-series characteristics and spatial-temporal correlation with the actual

al wind power sequences. Specifically, the scenarios have good SE index performance, and the fluctuation characteristics, the ACF, and the CCF are similar to the actual wind power sequence.

Moreover, the proposed scenario generation method in this paper can be further applied to decision-making problems such as dispatch planning and optimization for trading strategies. Further studies are planned and will be reported.

REFERENCES

- [1] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "A new fuzzy-based combined prediction interval for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 18-26, Jan. 2016.
- [2] A. Shukla and S. N. Singh, "Clustering based unit commitment with wind power uncertainty," *Energy Conversion and Management*, vol. 111, pp. 89-102, Mar. 2016.
- [3] P. Pinson, N. Siebert, and G. Kariniotakis, "Forecasting of regional wind generation by a dynamic fuzzy-neural networks based upscaling approach," in *Proceedings of European Wind Energy Conference*, Madrid, Spain, Jun. 2003, pp. 16-19.
- [4] M. G. Lobo and I. Sanchez, "Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1990-1997, Nov. 2012.
- [5] X. Peng, L. Xiong, J. Wen *et al.*, "A summary of the state of the art for short-term and ultra-short-term wind power prediction of regions," *Proceedings of the CSEE*, vol. 36, no. 23, pp. 6596-6596, Dec. 2016.
- [6] P. Meibom, R. Barth, B. Hasche *et al.*, "Stochastic optimization model to study the operational impacts of high wind penetrations in Ireland," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1367-1379, Aug. 2011.
- [7] H. Bludszweit, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983-991, Aug. 2008.
- [8] X. Ma, Y. Sun, and H. Fang, "Scenario generation of wind power based on statistical uncertainty and variability," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 894-904, Oct. 2013.
- [9] J. B. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, no. 1, pp. 3-11, Apr. 2006.
- [10] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1845-1856, Nov. 2010.
- [11] R. J. Bessa, V. Miranda, A. Botterud *et al.*, "Time adaptive conditional kernel density estimation for wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 660-669, Oct. 2012.
- [12] Y. Zhang, J. Wang, and X. Luo, "Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel," *Energy Conversion and Management*, vol. 96, pp. 440-451, May 2015.
- [13] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 234-240, Apr. 2008.
- [14] A. Tuohy, P. Meibom, E. Denny *et al.*, "Unit commitment for systems with significant wind penetration," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 592-601, May 2009.
- [15] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843-855, Mar. 2010.
- [16] D. D. Le, G. Gross, and A. Berizzi, "Probabilistic modeling of multi-site wind farm production for scenario-based applications," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 748-758, Jul. 2015.
- [17] P. Pierre, M. Henrik, N. H. Aa *et al.*, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51-62, Jan. 2010.
- [18] M. Yang, Y. Lin, S. Zhu *et al.*, "Multi-dimensional scenario forecast for generation of multiple wind farms," *Journal of Modern Power Systems and Clean Energy*, vol. 3, no. 3, pp. 361-370, May 2015.
- [19] W. Wu, K. Wang, B. Han *et al.*, "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1337-1345, Oct. 2015.
- [20] H. V. Haghi and S. Lotfifard, "Spatiotemporal modeling of wind generation for optimal energy storage sizing," *IEEE Transactions on Sus-*

- tainable Energy*, vol. 6, no. 1, pp. 113-121, Jan. 2015.
- [21] L. Kamal and Y. Z. Jafri, "Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan," *Solar Energy*, vol. 61, no. 1, pp. 23-32, Jul. 1997.
- [22] C. Park, Y. Sun, K. T. Yoon *et al.*, "Dickey-fuller test for an extended MA model," *Quantitative Bio-Science*, vol. 38, no. 1, pp. 1-21, May 2019.
- [23] L. Li, S. Miao, Q. Tu *et al.*, "Dynamic dependence modeling of wind power uncertainty considering heteroscedastic effect," *International Journal of Electrical Power and Energy Systems*, vol. 116, pp. 105556-105558, Mar. 2020.
- [24] H. Liu, S. Jing, and X. Qu, "Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines," *Energy Conversion and Management*, vol. 67, pp. 8-17, Mar. 2013.
- [25] J. Dissmann, E. C. Brechmann, C. Czado *et al.*, "Selecting and estimating regular vine copulae and application to financial returns," *Data Analysis*, vol. 59, pp. 52-69, Nov. 2013.
- [26] X. Li, *Copula Method and Its Application*, Beijing: Economy and Management Publishing House, 2014.
- [27] R. Chou, C. Wu, and N. Liu, "Forecasting time-varying covariance with a range-based dynamic conditional correlation model," *Review of Quantitative Finance and Accounting*, vol. 33, no. 4, pp. 327-345, Mar. 2009.
- [28] A. J. Patton, "Modelling asymmetric exchange rate dependence," *International Economic Review*, vol. 47, no. 2, pp. 527-556, Jun. 2006.
- [29] E. C. Brechmann, C. Czado, and K. Aas, "Truncated regular vines in high dimensions with application to financial data," *Canadian Journal of Statistics*, vol. 40, no. 1, pp. 68-85, Jan. 2012.
- [30] H. Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle*, New York: Springer, 1998.
- [31] Y. Zhang, "The development of Bayesian theory and its applications in business and bioinformatics," in *Proceedings of IOP Conference*, Beijing, China, Dec. 2017, pp. 28-31.
- [32] Z. Wang, W. Wang, C. Liu *et al.*, "Probabilistic forecast for multiple wind farms based on regular vine copulas," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 578-589, Apr. 2018.
- [33] M. Bogdan, "Modifying the Schwarz Bayesian information criterion to locate multiple interacting quantitative trait loci," *Genetics*, vol. 167, no. 2, pp. 989-999, Jun. 2004.
- [34] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, pp. 12-20, Aug. 2012.
- [35] D. Li, W. Yan, W. Li *et al.*, "A two-tier wind power time series model considering day-to-day weather transition and intraday wind power fluctuations," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 1-10, Dec. 2016.
- [36] Z. Wang, W. Wang, C. Liu *et al.*, "Forecasted scenarios of regional wind farms based on regular vine copulas," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 77-85, Jan. 2020.
- [37] IEEE Power and Energy Society and IEEE Working Group on Energy Forecasting. (2021, Mar.). Global energy forecasting competition 2012-wind forecasting. [Online]. Available: <https://www.kaggle.com/c/GEF2012-wind-forecasting/overview>
- Qingyu Tu** received his B.S. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2017, where he is currently pursuing the Ph.D degree. His research interests include probabilistic forecasting method of renewable energy and load.
- Shihong Miao** received his M.S. and Ph.D. degrees in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1996 and 2003, respectively. He is currently a Professor with the School of Electrical and Electronic Engineering, HUST. His research interests include power system protective relaying and control, novel distribution network and microgrid technology, and energy storage technology.
- Fuxing Yao** received his B.S. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2020, where he is currently pursuing the Ph.D degree. His research interests include the uncertainty analysis and modeling method of renewable energy.
- Yaowang Li** received his B.S. and Ph.D degrees in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2015 and 2020, respectively. Currently, he is a Postdoc Researcher with Department of Electrical Engineering, Tsinghua University, Beijing, China. His research interests include cloud energy storage, compressed air energy storage, power system optimization, and integrated energy system.
- Haoran Yin** received his B.S. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2019, where he is currently pursuing the M.S. degree. His research interests include application of deep-learning algorithms in the power system.
- Ji Han** received his B.S. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2016, where he is currently pursuing the Ph.D degree. His research interests include application of machine learning in power systems, renewable energy technology, and fault diagnosis of power systems.
- Di Zhang** received his B.S. degree in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2015, where he is currently pursuing the Ph.D degree. His research interests include operation and control of power system integrated with renewable energy.
- Weichen Yang** received his B.S. degree in electrical engineering from North China Electric Power University, Baoding, China, in 2017. He is currently working toward the Ph.D. degree at Huazhong University of Science and Technology (HUST), Wuhan, China. His research interests include high-voltage direct current technology, renewable energy technology, and microgrid technology.