Assessment Model for Distributed Wind Generation Hosting Capacity Considering Complex Spatial Correlations

Han Wu, *Member*, *IEEE*, Yue Yuan, *Member*, *IEEE*, Junpeng Zhu, *Member*, *IEEE*, and Yundai Xu

***Abstract*—To facilitate the large-scale integration of distributed wind generation (DWG), the uncertainty of DWG outputs needs to be quantified, and the maximum DWG hosting capacity (DWGHC) of distribution systems must be assessed. However, the structure of the high-dimensional nonlinear dependencies and the abnormal marginal distributions observed in geographically dispersed DWG outputs lead to the increase of the complexity of the uncertainty analysis. To address this issue, this paper proposes a novel assessment model for DWGHC that considers the spatial correlations between distributed generation (DG) outputs. In our method, an advanced dependence modeling approach called vine copula is applied to capture the high-dimensional correlation between geographically dispersed DWG outputs and generate a sufficient number of correlated scenarios. To avoid an overly conservative hosting capacity in some extreme scenarios, a novel chance-constrained assessment model for DWGHC is developed to determine the optimal sizes and locations of DWG for a given DWG curtailment probability. To handle the computational challenges associated with large-scale scenarios, a bilinear variant of Benders decomposition (BD) is employed to solve the chance-constrained problem. The effectiveness of the proposed method is demonstrated using a typical 38-bus distribution system in eastern China.**

***Index Terms*—Correlation, Benders decomposition (BD), distributed wind generation (DWG), hosting capacity, vine copula.**

I. Introduction

FUTURE distribution networks will witness an increasing integration of renewable-energy-based distributed generation (DG) due to its low cost and low emission. Compared with the conventional dispatchable generation, renewable-energy-based DG output, especially distributed wind generation (DWG), is more intermittent and stochastic owing to time-varying weather conditions. When a high level of DG penetration is reached, the uncertain DG output will result in severe pressure (e.g., voltage increases and line overloads) on the distribution system. To successfully and economically integrate renewable-energy-based DG into a distribution network, a hosting capacity assessment process is required to quantify the uncertainty of local renewable energy sources and determine the amount of DG that can be installed in the distribution network without network enhancement.

Recently, many capacity assessment methods that include uncertain renewable energies have been proposed. In [1], a probabilistic approach is proposed to assess the maximum DG penetration levels in low-voltage networks. In [2], an analytical method is proposed to calculate the DG hosting capacity of the feeder considering the hourly variation in the load demand and DG outputs. In [3], a probabilistic framework considering the uncertainties associated with the type, amount, location, and hourly variation of DGs is presented. In [4], the uncertainty of DG forecasting and active management methods are considered. Additional literature reviews of existing hosting capacity assessment technologies are provided in [5].

Although many studies on the uncertainties associated with renewable energies have been conducted, they solely focus on the randomness of the DG outputs. The spatial variability and dependencies among DG outputs are not considered. As is well-known, the outputs of neighboring DG outputs usually have strong spatial correlations owing to similar meteorological conditions in the distribution system area; these correlations amplify the uncertainty in the distribution network [6]. Reference [7] demonstrates that such such correlation has a strong impact on PV hosting capacity. Furthermore, the complex geographical environment of the distribution system area leads to various abnormal marginal distributions and structures of the nonlinear dependencies among different DG outputs [8], [9]. Ignoring the spatial correlation or modeling the correlation incorrectly may result in a severe underestimation of the power flow volatility and erroneous capacity evaluation results [10], [11]. For an accurate assessment of the capacity, a flexible multivarite distribution model is required to address the complex dependencies among geographically dispersed DGs.

Copula functions are powerful tools for modeling abnormal joint distributions with nonlinear dependencies. According to Sklar's theorem, joint distribution functions can be

Manuscript received: January 12, 2021; revised: April 28, 2021; accepted: August 30, 2021. Date of CrossCheck: August 30, 2021. Date of online publication: November 12, 2021.

This work was supported by the National Key Research and Development Program of China (No. 2016YFB0900100) and High-level Talents Introduction & Research Start-up Fund Program of Nanjing Institute of Technology (No. YKJ202134).

H. Wu (corresponding author) is with Nanjing Institute of Technology, Nanjing 211167, China (e-mail: wuhanichina@vip.qq.com).

Y. Yuan, J. Zhu, and Y. Xu are with the College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China (e-mail: yyuan@hhu.edu.cn; jzhu@hhu.edu.cn; hanxyd@126.com).

DOI: 10.35833/MPCE.2020.000889

constructed from arbitrary marginal distributions and appropriate copula functions [12]. However, due to the parameter restriction, the existing copula functions are mostly restricted to bivariate cases [13]. Thus, an advanced high-dimensional modeling tool called vine copula is proposed. In vine copula, a cascade pair-copula construction (PCC) is used to capture the structure of the high-dimensional dependencies across a large number of variables [14]. Such a PCC structure substantially increases the flexibility of the model and allows the modeling of complex dependencies. Recently, vine copula has been proven to be successful in modeling large-scale photovoltaic (PV) generation [15] and wind generation forecasting errors [13]. However, there are no studies on the assessment of the hosting capacity that apply the vine copula method to model the spatial correlations between DG outputs.

Another problem arises when assessing the DG capacity with the consideration of nonlinear dependencies. The capacity assessment is an optimization problem in nature. To date, two predominant methods, i.e., robust optimization (RO) [4], [16] and scenario-based stochastic programming (SBSP) [17]-[19], have been applied to address the uncertainty in the DG hosting capacity assessment problem. In the RO method, the uncertainty is represented by norm-based uncertainty sets, which cannot account for the complex dependencies among variables. In SBSP, the uncertainty is treated as stochastic variables with a given probability density function (PDF). Unlike RO, SBSP can fully consider the stochastic nature of renewable energy and derive an optimal result of DG installation schemes. Nevertheless, some infrequent extreme DG output scenarios are also considered in SBSP, which leads to conservative and costly DG allocation results [20]. To prevent an overly conservative assessment of the capacity caused by extreme scenarios, the chance-constrained programming (CCP) approach is proposed to consider reasonable DG curtailment. By adjusting the DG curtailment probabilities (constraint violation probabilities), the distribution network operator (DNO) can control the trade-off between the DG hosting capacity and the renewable energy curtailment level [21]. However, due to the implicit form of chance constraints, the CCP with general multivariate distribution cannot be efficiently solved by off-the-shelf optimization tools [22]-[24].

To address these two problems while considering the complex dependencies among DWG outputs, we first employ the vine copula in the assessment model of DWG hosting capacity and then propose a novel bilinear chance-constrained DWG capacity model. Finally, a customized bilinear Benders decomposition (BD) method is applied and developed to reduce the computational cost. The main contributions of this paper can be summarized as follows.

1) To obtain comprehensive knowledge of the joint distribution of DG outputs, an advanced vine copula method is applied to construct the multivariate distribution of adjacent wind speeds in the distribution system area. Unlike the traditional Gaussian copula approach, vine copula can deal with various tail dependencies.

2) To consider the spatial correlations between the DG outputs when assessing the hosting capacity, a scenario-based chance-constrained DG capacity assessment problem is formulated. Unlike the previous capacity assessment model in the literature, the DNO can achieve the desired trade-offs between the hosting capacity and the curtailment probability by adjusting the constraint violation probabilities. In addition, the proposed chance-constrained capacity assessment model can be efficiently solved using off-the-shelf commercial optimization platforms by linearizing the chance constraints with its bilinear variant. Furthermore, to achieve a better computational performance, a customized bilinear BD method is applied and developed to solve the proposed model.

3) A case study with real datasets is presented to verify the effectiveness of the proposed method. A comparison with the Gaussian copula and canonical vine (C-vine) copula is also carried out to show the flexibility of the vine copula model. In addition, the traditional Big-*M* method is applied to solve the chance-constrained capacity assessment model for comparison.

The remainder of this paper is organized as follows. In Section II, the vine copula model and scenario generation process are presented to capture the complex spatial correlations between DG outputs. In Section III, the formulation of the novel bilinear chance-constrained DG hosting capacity assessment model is explained. Section IV introduces the bilinear BD algorithm. In Section V, a typical 38-bus distribution system located in eastern China is used to demonstrate the effectiveness of the proposed method and algorithm. Finally, this paper is concluded in Section VI.

II. Modeling High-dimensional Distributions by C-vine Copula

A. Copula Theory

For d random variables $x_1, x_2, \ldots, x_d$ having a joint distribution $F(x_1, x_2, \ldots, x_d)$ and the associated joint density function $f(x_1, x_2, \ldots, x_d)$, Sklar's theorem states that $F(x_1, x_2, \ldots, x_d)$ can be decomposed into its marginal distributions $F_1(x_1), F_2(x_2), \ldots, F_d(x_d)$ and a copula function $C(\cdot)$ [25].

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \tag{1}$$

Using the chain rule, $f(x_1, x_2, \ldots, x_d)$ is decomposed as:

$$f(x_1, x_2, \ldots, x_d) = c(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \prod_{m=1}^{d} f(x_m) \tag{2}$$

where $c(\cdot)$ denotes the density of the copula.

Sklar's theorem allows arbitrary marginal distributions when constructing joint distributions as long as they increase strictly and continuously. Therefore, several methods such as parametric distribution fitting and nonparametric methods, e.g., the empirical distribution and kernel density estimation (KDE), can be applied to model the marginal distribution. Here, the KDE method with a Gaussian kernel is applied to obtain a smooth distribution curve.

Two families of copula functions, i.e., Archimedean copulas and elliptical copulas, are widely used to model the joint

distribution. These two copula families have different strengths in describing dependence structures. Archimedean copulas (e. g., Clayton, Frank, and Gumbel copulas) have superior performance in modeling the tail and asymmetric dependencies, but they cannot model a dependence structure with more than two dimensions. Elliptical copulas, e.g., the Gaussian copula and *t*-copula, can be easily extended to multivariate formulations. However, the Gaussian copula cannot fit the tail dependence, and the *t*-copula fails to model the asymmetric tail dependence structure. Thus, these two existing copula families are limited in capturing complex multivariate dependencies.

The PCC is proposed in [25] to overcome the limitations of the existing copula families when fitting multivariate distributions. By combining different bivariate copulas as blocks to construct a high-dimensional dependence structure, PCC can make full use of existing copulas in high-dimensional distribution modeling. This substantially increases the flexibility in multivariate distribution fitting, especially in modeling the structures having tail dependencies with different characteristics, and has performed much better than multivariate elliptical copulas.

B. Pair-copula and C-vine Structures

The main idea of PCC is to convert a joint density function into a product of bivariate density functions using recursive conditioning. According to the chain rule, a *d*-dimensional joint density function can be recursively factorized as:

$$f(x_1,x_2,\ldots,x_d)=f_1(x_1)f_{2|1}(x_2|x_1)f_{3|2,1}(x_3|x_2,x_1)\cdot\ldots\cdot f_{d|d-1,d-2,\ldots,1}(x_d|x_{d-1},x_{d-2},\ldots,x_1) \tag{3}$$

Each univariate conditional probability distribution in (3) can be decomposed into the appropriate (conditional) pairwise copula times a conditional marginal density as follows [26]:

$$f(x|\boldsymbol{v})=c_{xv_m|\boldsymbol{v}_{-m}}\{F_{x|\boldsymbol{v}_{-m}}(x|\boldsymbol{v}_{-m}),F_{v_m|\boldsymbol{v}_{-m}}(v_m|\boldsymbol{v}_{-m})\}f(x|\boldsymbol{v}_{-m}) \tag{4}$$

where $c_{xv_m|\boldsymbol{v}_{-m}}$ denotes an appropriate pair-copula density for the pair of variables $F_{x|\boldsymbol{v}_{-m}}(x|\boldsymbol{v}_{-m})$ and $F_{v_m|\boldsymbol{v}_{-m}}(v_m|\boldsymbol{v}_{-m})$; $\boldsymbol{v}$ is a *d*-dimensional vector; v_m is an arbitrary component of $\boldsymbol{v}$; and $\boldsymbol{v}_{-m}$ denotes $\boldsymbol{v}$ without v_m.

By using (4) recursively, all conditional densities can be expressed as the products of bivariate copula densities and the corresponding marginal densities. Note that the decomposition of the joint density in (4) is not unique due to the arbitrarily chosen sequence from vector $\boldsymbol{v}$. In fact, there are $2d!$ ways to decompose a *d*-dimensional joint probability function. To help organize all possible PCCs, a specific graphical model called regular vine is proposed in [25]. In regular vine, the dependence structures are represented as a sequence of nested trees with nodes and edges. Each edge corresponds to a bivariate copula block. Although the regular vine structure is flexible when modeling the dependence structure, the enormous number of possible vine tree sequences makes it impossible to calculate all of these vines and choose the best one in practice [27]. Thus, a special case of the regular vine called C-vine is more frequently used when fitting the spatial correlations [28]-[30].

In the C-vine copula, each tree T_n has a unique node (denoted as the root node) with $d-n$ edges, which gives $d!/2$ different C-vines on the *d* nodes. This significantly reduces the computational cost when selecting the tree structure. Figure 1 shows one possible C-vine tree structure with five variables.

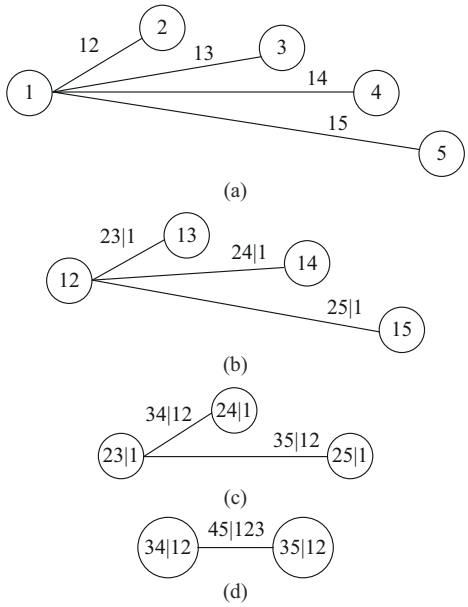

Fig. 1. C-vine tree structure with five variables. (a) T_1. (b) T_2. (c) T_3. (d) T_4.

The *d*-dimensional density corresponding to C-vine is given by:

$$f(x_1,x_2,\ldots,x_d)=\prod_{k=1}^{d}f_k(x_k)\cdot\prod_{m=1}^{d-1}\prod_{n=1}^{d-m}c_{m,m+n|1,2,\ldots,m-1}(F(x_m|x_1,\ldots,x_{m-1}),F(x_{m+n}|x_1,\ldots,x_{m-1})) \tag{5}$$

where the index *m* represents the trees; and the index *n* denotes the edges in each tree. Furthermore, we use the *h*-function notation introduced in [25] to denote the conditional distribution functions:

$$h(U_n,U_m,\Theta)=F(U_n|U_m)=\frac{\partial C_{u_n,u_m}(U_n,U_m,\Theta_{n,m})}{\partial U_m} \tag{6}$$

where Θ is the set of parameters for the bivariate copula; and U_n and U_m are two uniformly distributed variables.

C. Summary of C-vine Fitting and Simulation

To describe a multivariate dependence structure with the C-vine copula, the best C-vine tree structure and the associated optimal copula parameter should be identified. To obtain the best fit of a C-vine tree structure, a useful approach is to use Kendall's τ as a criterion to find the root node in each tree. Then, the optimal parameters of each edge are estimated using the maximum likelihood estimation and Akaike's information criterion (AIC).

After constructing the C-vine copula model, *N* correlated wind speed samples are generated on the basis of the probability integral transformation method and the inverse condi-

tional distribution given by the *h*-function. The procedure of the sampling algorithm can be summarized as follows. First, generate a *d*-dimensional independent and identically distributed (i.i.d.) random vector $[u_1, u_2, \ldots, u_d]$, which follows a uniform distribution on [0, 1]. Then, the n^{th} variable is sampled, and the conditional distribution functions needed for sampling the $(n+1)^{\text{th}}$ variable are computed by repeatedly using the *h*-function defined in (6). A detailed description of the sampling method is given in Appendix A Algorithm A1.

III. Mathematical Formulation of Chance-constrained Assessment Model for DWGHC

A. Stochastic Assessment Model for DG Hosting Capacity

The DG hosting capacity assessment problem is an optimization problem. To consider the dependencies among DG outputs, SBSP is employed to build a scenario-based DG capacity assessment model (SDGCAM). In this model, the dependent DG outputs are represented by numerous scenarios sampled from vine copula and there is no DG curtailment in any circumstances. The proposed SDGCAM is presented as follows.

1) Objective Function

The aim of assessing the DG hosting capacity is to determine the maximum DG capacity that can be installed in the distribution network without violating operation constraints. Thus, the objective function (7) sums up all of the DG capacities at any candidate bus:

$$\max \sum_{i \in \Psi_{DG}} c_i^{DG} \tag{7}$$

where Ψ_{DG} is the set of all candidate nodes connected to DG; and c_i^{DG} is the DG capacity installed at bus *i*.

2) DG Installation Constraints

A DNO should know the maximum DG installation capacity limited by the local environment before assessing the capacity through a careful investigation. Such an investigation will help improve the usability of the assessment results. However, if an investigation cannot be performed in advance of time, all buses in the distribution network could be set as candidate buses to obtain the extreme hosting capacity. The DG installation location and geographical constraints associated with the local environment are expressed as:

$$0 \le c_i^{DG} \le C_i^{DG,\max} \quad \forall i \in \Psi_{DG} \tag{8}$$

where $C_i^{DG,\max}$ is the maximum installation capacity at bus *i*.

3) DG Output Constraints

Wind generation is a common form of renewable energy generation in a distribution network; thus, DWG is employed in this paper to represent the intermittent and stochastic renewable-energy-based DG in a distributed network. However, the uncertainty and dependence of the output of wind generation originate from the wind speeds. Thus, (9) is employed to convert the wind speeds into DG outputs [31]. For simplicity, all the DGs are assumed to operate in the maximum power point tracking mode to make full use of renewable energy. Note that for other renewable energy resources (such as PVs), we can employ (10) for the expected output in scenario *s*.

$$\eta_s^{DG} = \begin{cases} 0 & 0 \le V_s^{DG} \le V^{ci}, V_s^{DG} > V^{co} \\ \dfrac{V_s^{DG} - V^{ci}}{V^r - V^{ci}} & V^{ci} < V_s^{DG} < V^r \\ 1 & V^r \le V_s^{DG} \le V^{co} \end{cases} \tag{9}$$

$$\begin{cases} P_{i,s}^{DG} = \eta_s^{DG} c_i^{DG} \\ Q_{i,s}^{DG} = \tan\varphi_i P_{i,s}^{DG} \end{cases} \quad \forall i \in \Psi_{DG}, s \in S \tag{10}$$

where *S* is the set of all scenarios; η_s^{DG} is the efficiency factor of the DG output at bus *i* in scenario *s*; V_s^{DG} is the wind speed in scenario *s*; V^{ci}, V^{co}, and V^r are the cut-in speed, cut-out speed, and related output speed of the wind turbine, respectively; $P_{i,s}^{DG}$ and $Q_{i,s}^{DG}$ are the active and reactive DG outputs at bus *i* in scenario *s*, respectively; and φ_i is the power factor angle of the DG output at bus *i* in scenario *s*.

4) Power Balance Constraints

Power balance is a basic operation constraint in the distribution system. The active and reactive power balances at each bus are represented as:

$$\begin{cases} \sum_{ij \in \Phi_b, j \in \xi(i)} p_{ij,s} = P_{i,s}^{\text{sub}} + P_{i,s}^{DG} - P_{i,s}^{L} \\ \sum_{ij \in \Phi_b, j \in \xi(i)} q_{ij,s} = Q_{i,s}^{\text{sub}} + Q_{i,s}^{DG} - Q_{i,s}^{L} \end{cases} \quad \forall i \in \Psi_n, s \in S \tag{11}$$

where $\xi(i)$ denotes the bus connected to bus *i*; $p_{ij,s}$ and $q_{ij,s}$ are the active and reactive power flows in branch *ij*, respectively; $P_{i,s}^{L}$ and $Q_{i,s}^{L}$ are the active and reactive load demands at bus *i* in scenario *s*, respectively; $P_{i,s}^{\text{sub}}$ and $Q_{i,s}^{\text{sub}}$ are the active and reactive power outputs of the substation at bus *i* in scenario *s*, respectively; Φ_b is the set of all branches; and Ψ_n is the set of all buses.

5) Bus Voltage Constraints

The bus voltage limits play a vital role when assessing the hosting capacity. Here, the Distflow formulation is employed to describe the power flow constraints and the relation between each bus. Equations (12) and (13) provide the bounds on the bus voltage, and (14) gives the Distflow formulation of the power flow.

$$U_{i,s} = V_{i,s}^2 \quad \forall i \in \Psi_n, s \in S \tag{12}$$

$$\underline{V}_i^2 \le U_{i,s} \le \overline{V}_i^2 \quad \forall i \in \Psi_n, s \in S \tag{13}$$

$$U_{i,s} - U_{j,s} = 2(r_{ij} p_{ij,s} + x_{ij} q_{ij,s}) - (x_{ij}^2 + x_{ij}^2) \frac{p_{ij,s}^2 + q_{ij,s}^2}{U_{i,s}} \quad \forall ij \in \Phi_b, s \in S \tag{14}$$

where $V_{i,s}$ and $U_{i,s}$ are the bus voltage and its square in scenario *s*, respectively; $\underline{V}_i$ and $\overline{V}_i$ are the lower and upper values of V_i, respectively; and r_{ij} and x_{ij} are the resistance and reactance of each branch, respectively.

Equation (14) is nonconvex owing to the quadratic term of line power. It is intractable for an on-the-shelf commercial solver for the computation of a large distribution system. Nevertheless, the quadratic term is far smaller than the linear terms, and ignoring the quadratic term does not significantly impact the allocation results [19]. Therefore, ignoring this term leads to the following linear power flow constraint:

$$U_{i,s} - U_{j,s} = 2(r_{ij} p_{ij,s} + x_{ij} q_{ij,s}) \quad \forall ij \in \Phi_b, s \in S \tag{15}$$

6) Branch Thermal Constraints

The thermal constraint is also vital in the hosting capacity assessment. Traditionally, the branch thermal constraint is expressed in a quadratic form as:

$$\sqrt{p_{ij,s}^2+q_{ij,s}^2} \le S_{ij}^{\max} \quad \forall ij \in \Phi_b, s \in S \tag{16}$$

where $S_{ij}^{\max}$ is the maximum apparent power of line ij.

However, the quadratic form in (16) cannot be efficiently handled by the commercial solver. Thus, a circular constraint linearization method that employs several square constraints to approximate the quadratic constraint is adopted to linearize (16) [32]. The linearized branch thermal constraint is given in (17), and a schematic diagram of the circular constraint linearization method is shown in Fig. 2, where the blue circular denotes the branch thermal constraints in (16); and the overlapping area of the two rectangles denotes the relaxed (16) after linearization.

$$\begin{cases} -S_{ij}^{\max} \le p_{ij,s} \le S_{ij}^{\max} \\ -S_{ij}^{\max} \le q_{ij,s} \le S_{ij}^{\max} \\ -\sqrt{2}\,S_{ij}^{\max} \le p_{ij,s}+q_{ij,s} \le \sqrt{2}\,S_{ij}^{\max} \\ -\sqrt{2}\,S_{ij}^{\max} \le p_{ij,s}-q_{ij,s} \le \sqrt{2}\,S_{ij}^{\max} \end{cases} \quad \forall ij \in \Phi_b, s \in S \tag{17}$$

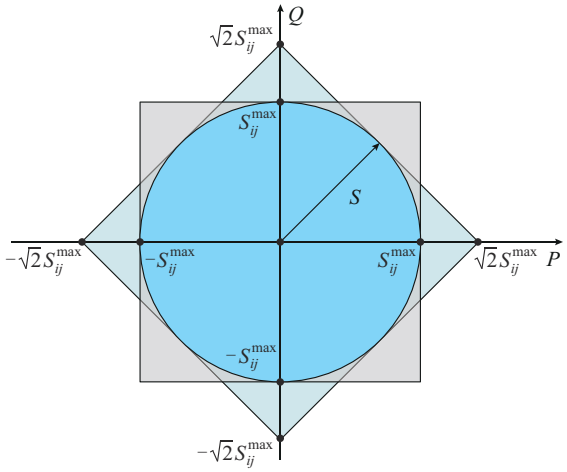

Fig. 2. Linearization of circular constraint.

The SDGCAM in (7)-(13), (15), and (17) is formulated as a linear programming (LP) problem, which has good performance for computing the hosting capacity in a large number of scenarios.

B. Modeling DG Curtailment

Although the SDGCAM can represent the uncertainty and dependencies between DG outputs, the assumption of full accommodation in all scenarios will lead to conservative and impractical DG integration. In fact, a DNO can control the output of DG in some extreme scenarios and curtail it if necessary, with the help of advanced distribution automation and active distribution management schemes. Thus, the security constraints in extreme scenarios can be ignored for economic hosting capacity. To balance the cost and curtailment risks, CCP is employed on the basis of the SDGCAM, and a chance-constrained DG capacity assessment model (CDGCAM) is built by replacing the constraints in (11) with constraint (18).

$$\begin{cases} \Pr\left(\sum_{ij\in\Phi_b} p_{ij,s} = P_{i,s}^{\text{sub}} + P_{i,s}^{DG} - P_{i,s}^{L}\right) \ge 1-\delta \\ \Pr\left(\sum_{ij\in\Phi_b} q_{ij,s} = Q_{i,s}^{\text{sub}} + Q_{i,s}^{DG} - Q_{i,s}^{L}\right) \ge 1-\delta \end{cases} \quad \forall ij \in \Phi_b, s \in S \tag{18}$$

Equation (18) implies that the probability of violating the power balance constraints in (11) is less than a predefined risk level $(1-\delta)\%$. A DNO can obtain a trade-off between the curtailment risk and the level of DG penetration by adjusting the risk level.

C. Bilinear Reformulation of Chance Constraints

Due to the non-convex distribution, the chance-constraint cannot be handled by the commercial solvers. Thus, the chance constraints are usually reformulated as a set of linear constraints using the Big-M formulation [33]:

$$\begin{cases} -Mw_s \le \sum_{ij\in\Phi_b} p_{ij,s} - P_{i,s}^{\text{sub}} - P_{i,s}^{DG} + P_{i,s}^{L} \le Mw_s \\ -Mw_s \le \sum_{ij\in\Phi_b} q_{ij,s} - Q_{i,s}^{\text{sub}} - Q_{i,s}^{DG} + Q_{i,s}^{L} \le Mw_s \end{cases} \quad \forall i \in \Psi_n, s \in S \tag{19}$$

where M is a sufficiently large number.

However, the Big-M formulation incurs a large computational burden, which cannot be applied to large-scale planning problems. Therefore, we follow the idea in [34] and apply a new bilinear reformulation of the chance constraints in this paper. This bilinear reformulation has exhibited better performance than the Big-M formulation for the unit commitment problem [35] but has not been applied to the hosting capacity assessment. The bilinear formulation of the constraints in (18) is as follows:

$$\begin{cases} \left(\sum_{ij\in\Phi_b} p_{ij,s} - P_{i,s}^{\text{sub}} - P_{i,s}^{DG} + P_{i,s}^{L}\right)(1-w_s) = 0 \\ \left(\sum_{ij\in\Phi_b} q_{ij,s} - Q_{i,s}^{\text{sub}} - Q_{i,s}^{DG} + Q_{i,s}^{L}\right)(1-w_s) = 0 \end{cases} \quad \forall i \in \Psi_n, s \in S \tag{20}$$

$$\sum_{s=1}^{S} \pi_s w_s \le \delta \quad s \in S \tag{21}$$

where w_s is a binary indicator of whether the associated constraint is satisfied in scenario s or not; and π_s is the probability of occurrence of scenario s. In this paper, we assume that each scenario has the same probability of occurrence, i. e., $\pi_s = 1/N$ $(N=|S|)$. The validity of the reformulation (20) is clear; by setting the scenario indicator variable w_s as 1, a nonresponsive scenario will be removed from the entire formulation. Equation (21) is equivalent to the chance constraints in (18), where the predefined risk level is converted to the total number of nonresponsive scenarios. In addition, the bilinear terms $p_{ij,s}w_s$, $P_i^{DG}w_s$, $q_{ij,s}w_s$, and $Q_i^{DG}w_s$ in (20) can be linearized using the McCormick linearization technique [36], which theoretically provides the tightest convex hull.

$$\begin{cases} \sum_{ij\in\Phi_b} p_{ij,s} - \sum_{ij\in\Phi_b} \tilde{p}_{ij,s} + \tilde{P}^{\text{sub}}_{i,s} - P^{\text{sub}}_{i,s} + \tilde{P}^{DG}_{i,s} - P^{DG}_{i,s} + P^{L}_{i,s}(1-w_s) = 0 \\ \sum_{ij\in\Phi_b} q_{ij,s} - \sum_{ij\in\Phi_b} \tilde{q}_{ij,s} + \tilde{Q}^{\text{sub}}_{i,s} - Q^{\text{sub}}_{i,s} + \tilde{Q}^{DG}_{i,s} - Q^{DG}_{i,s} + Q^{L}_{i,s}(1-w_s) = 0 \end{cases} \quad \forall i \in \Psi_n, s \in S \tag{22}$$

$$\begin{cases} -S^{\max}_{ij} w_s \le \tilde{p}_{ij,s} \le S^{\max}_{ij} w_s \\ S^{\max}_{ij}(w_s - 1) + p_{ij,s} \le \tilde{p}_{ij,s} \le p_{ij,s} - S^{\max}_{ij}(w_s - 1) \\ 0 \le \tilde{P}^{DG}_{i,s} \le \eta^{DG}_s C^{DG,\max}_i w_s \\ P^{DG}_{i,s} + \eta^{DG}_s C^{DG,\max}_i (1-w_s) \le \tilde{P}^{DG}_{i,s} \le P^{DG}_{i,s} \end{cases} \quad \forall i \in \Psi_n, ij \in \Phi_b, s \in S \tag{23}$$

$$\begin{cases} -S^{\max}_{ij} w_s \le \tilde{q}_{ij,s} \le S^{\max}_{ij} w_s \\ S^{\max}_{ij}(w_s - 1) + q_{ij,s} \le \tilde{q}_{ij,s} \le q_{ij,s} - S^{\max}_{ij}(w_s - 1) \\ 0 \le \tilde{Q}^{DG}_{i,s} \le \eta^{DG}_s C^{DG,\max}_i w_s \\ Q^{DG}_{i,s} + \eta^{DG}_s C^{DG,\max}_i (1-w_s) \le \tilde{Q}^{DG}_{i,s} \le Q^{DG}_{i,s} \end{cases} \quad \forall i \in \Psi_n, ij \in \Phi_b, s \in S \tag{24}$$

where $\tilde{p}_{ij,s} = p_{ij,s} w_s$, $\tilde{q}_{ij,s} = q_{ij,s} w_s$, $\tilde{P}^{\text{sub}}_{i,s} = P^{\text{sub}}_{i,s} w_s$, $\tilde{P}^{DG}_{i,s} = P^{DG}_{i,s} w_s$, $\tilde{Q}^{\text{sub}}_{i,s} = Q^{\text{sub}}_{i,s} w_s$, and $\tilde{Q}^{DG}_{i,s} = Q^{DG}_{i,s} w_s$. Through the McCormick linearization technique, (20) is rewritten as (22). The convex McCormick envelopes of the bilinear terms are given in (23) and (24). The CDGCAM in (7)-(13), (15)-(17), and (20)-(23) is thus converted to a large-scale deterministic mixed-integer linear programming (MILP) problem.

IV. Bilinear BD of CDGCAM

A large number of scenarios are required to obtain an accurate approximation of the multivariate distribution. However, the associated large-scale MILP problem cannot be efficiently solved using off-the-shelf MILP solvers. To deal with the computational burden resulting from large-scale scenarios, a specific approach of BD called bilinear BD algorithm is applied to solve the large-scale CDGCAM [36]. In the bilinear BD algorithm, the CDGCAM is reformulated in a master-subproblem structure, where the master problem determines the capacity of each DG and the subproblem checks the feasibility corresponding to all selected scenarios and feeds back Benders cuts to the master problem. Then, the master problem and subproblem are solved iteratively.

A. Master Problem

In the master problem, the aim is to maximize the total DG hosting capacity with respect to constraints (8)-(10), (21) and the Benders cut in (25).

$$\left[\hat{\Phi}^{(k)}_s + \sum_{i\in\Psi_{DG}} \mu^{(k)}_{i,s}(c^{DG}_i - \hat{c}^{DG,(k)}_i)\right](1-w_s) \le 0 \tag{25}$$

where $\hat{\Phi}^{(k)}_s$ is the optimal objective of the feasibility check subproblem in scenario s at the k^{th} iteration; $\hat{c}^{DG,(k)}_i$ is the fixed value of the first-stage variables c^{DG}_i at the k^{th} iteration; and $\mu^{(k)}_{i,s}$ is the dual variables provided by the subproblem at the k^{th} iteration.

Since the objective function (7) only contains the capacity of each DG, we only consider the Benders feasibility cut in the master problem. Unlike the conventional BD method, which simply generates Benders cuts from the dual variable in the subproblem, we follow the method in [34] and modify our Benders cut into a bilinear form. The binary integer variable w_s indicates the inclusion or omission of feedback from the feasibility check subproblem in scenario s. If w_s is set to be 1, the Benders cut in scenario s will be ignored. The k^{th} iteration of the bilinear master problem is described as:

$$\begin{cases} \max \sum_{i\in\Psi_{DG}} c^{DG}_i \\ \text{s.t. (8)-(10), (21), (25)} \end{cases} \tag{26}$$

Similar to the former transformations, the bilinear terms of constraints in (25) can be linearized using McCormick linearization method as (27)-(29). Then, (25) is converted into linear inequalities, and the whole formulation is converted into an MILP problem.

$$\hat{\Phi}^{(k)}_s + \sum_{i\in\Psi_{DG}} \mu^{(k)}_{i,s}(c^{DG}_i - \hat{c}^{DG,(k)}_i) - w_s\hat{\Phi}^{(k)}_s - \sum_{i\in\Psi_{DG}} \mu^{(k)}_{i,s}(\tilde{c}^{DG}_i - \hat{c}^{DG,(k)}_i w_s) \le 0 \tag{27}$$

$$0 \le \tilde{c}^{DG}_i \le C^{DG,\max}_i w_s \tag{28}$$

$$c^{DG}_i - C^{DG,\max}_i (1-w_s) \le \tilde{c}^{DG}_i \le c^{DG}_i \tag{29}$$

Equations (27)-(29) are the McCormick convex envelopes of the bilinear terms $\tilde{c}^{DG}_i = c^{DG}_i \mu^{(k)}_{i,s}$.

B. Subproblem

As the name suggests, the feasibility check is utilized to check whether the installed DG can be accommodated in each scenario, with respect to the constraints (12), (13), (15), (17), (30), (31).

$$\begin{cases} \sum_{ij\in\Phi_b} p_{ij,s} - P^{\text{sub}}_{i,s} - P^{DG}_{i,s} + P^{L}_{i,s} + \varepsilon^{+}_{i,s} - \varepsilon^{-}_{i,s} = 0 \\ \sum_{ij\in\Phi_b} q_{ij,s} - Q^{\text{sub}}_{i,s} - Q^{DG}_{i,s} + Q^{L}_{i,s} + \zeta^{+}_{i,s} - \zeta^{-}_{i,s} = 0 \end{cases} \quad \forall ij \in \Phi_b, s \in S \tag{30}$$

$$c^{DG}_i = \hat{c}^{DG,(k)}_i \mu^{(k)}_{i,s} \quad \forall i \in \Psi_{DG}, s \in S \tag{31}$$

In scenario s at the k^{th} iteration, the feasibility check subproblem is modeled as:

$$\begin{cases} \min \Phi^{(k)}_s = \sum_{i\in\Psi_n} (\varepsilon^{+}_{i,s} + \varepsilon^{-}_{i,s} + \zeta^{+}_{i,s} + \zeta^{-}_{i,s}) \\ \text{s.t. (12), (13), (15), (17), (30), (31)} \end{cases} \tag{32}$$

By introducing nonnegative slack variables $\varepsilon^{+}_{i,s}$, $\varepsilon^{-}_{i,s}$, $\zeta^{+}_{i,s}$, and $\zeta^{-}_{i,s}$ into the power balance constraint (11), the feasibility of the problem is ensured. The objective of the problem is to minimize the sum of all slack variables. Specifically, $\Phi^{(k)}_s$ denotes whether the DG allocation result at the k^{th} iteration is acceptable in scenario s. If $\Phi^{(k)}_s$ is larger than a preset threshold (denoted as 0 in this paper), the DG allocation result cannot match the operation constraints, and the curtailment of DG output is triggered. In this case, the linearized bilinear feasibility cut in (27)-(29) is generated and added to the master problem to guide the master problem to find another proper DG allocation result in the next iteration. Equation (31) denotes the DG capacity used in the subproblem that comes from the master problem at the k^{th} iteration.

C. Solution Algorithm

The procedure of the proposed bilinear BD algorithm is shown in Algorithm 1. Furthermore, the flow chart of the algorithm is presented in Fig. 3.

Algorithm 1: procedure of proposed bilinear BD algorithm

Initialization: set the iteration counter $k=0$; set the lower bound $LB=-\infty$ and the upper bound $UB=+\infty$; and set the optimality tolerance σ.

Iteration:

Step 1: compute the bilinear master problem; derive its optimal value $\hat{V}_U^{(k)}$, the optimal place and size solution $\hat{c}_k^{DG,(k)}$, and the scenario indicator $\hat{w}_s^{(k)}$; update the upper bound $UB=\hat{V}_U^{(k)}$.

Step 2: for all scenarios, solve the feasibility check subproblem, and derive the optimal value $\Phi_s^{(k)}$. If $\Phi_s^{(k)}$ is larger than zero (original problem infeasible), then generate the linearized Benders feasibility cut in equations (27)-(29), and provide feedback to the master problem.

Step 3: construct an MILP to derive the *LB*. This problem is composed of the objective function in (7) and the constraints in (8)-(13), (17), and (20) with fixed $\hat{w}_s$ from the master problem. Then, solve this problem and obtain its optimal objective $\hat{V}_L^{(k)}$. Finally, update $LB=\max\{LB,\hat{V}_L^{(k)}\}$.

Stop Criteria: if $|UB-LB|/LB\le\sigma$, terminate with the DG capacity solution associated with the latest *LB*; otherwise, update the iteration $k=k+1$ and return to *Step 2*.

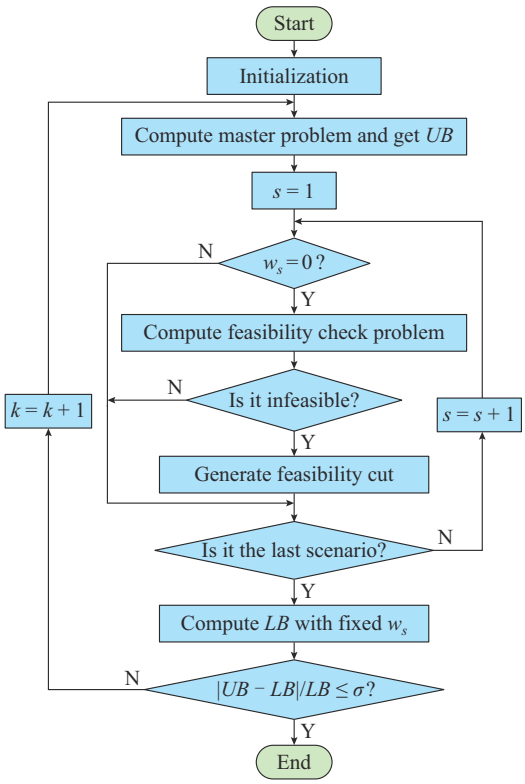

Fig. 3. Flow chart of bilinear BD algorithm.

V. Case Study

To verify the proposed method, we study a typical 38-bus distribution system located in eastern China. In the test distribution system, there are six neighboring wind speed observation stations. The wind speed dataset consists of measurements over one year with a quarter-hourly resolution (35040 periods). Figure 4 shows the locations of six wind speed observation stations W1-W6, where the maximum distance among the observation stations is 3.2 km (W1 and W5). The linear correlation matrix of the six wind speeds is illustrated by the heat map in Fig. 5, which shows strong correlations between the geographically dispersed wind speed observation stations in the distribution system. Figure 6 shows the marginal distributions of the observed wind speeds. Since there are no obstacles around W1, the marginal distribution of W1 is different from the others. The network structure of the 38-bus distribution system is shown in Fig. 7, where the candidate wind generation is integrated at buses 17, 18, 21, 34, 36, and 38. Note that the integrated bus of DG will impact the hosting capacity of the distribution network; however, we only present one selected interconnected bus combination plan in this paper to verify the effectiveness of the proposed method. The maximum DG capacity $C_i^{DG,\max}$ of each bus is set to be 10 MW. As we focus on the planning of small enough and geographically DWG, the wake effect of wind generation can be omitted. Detailed data of the test system are given in Appendix A Tables AI and AII.

Fig. 4. Locations of six wind speed observation stations.

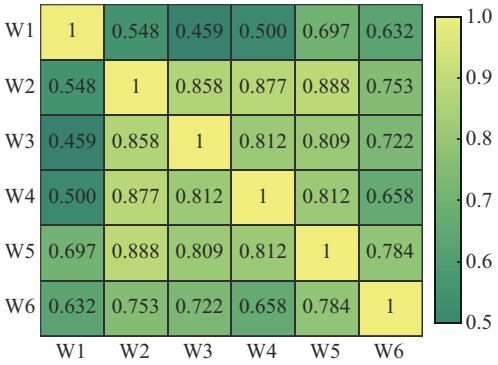

Fig. 5. Linear correlation matrix of wind speed.

For a better illustration of the proposed model, we assume that the cut-in speed, cut-out speed, and related output speed of all wind turbines are 3 m/s, 25 m/s, and 12 m/s, respectively. The voltage at the point of common coupling (PCC) node is set to be 1.05 p. u.. The R package vine copula [36] is employed to determine the optimal parameter of vine copula. The formulations of the DG hosting capacity assessment in the SDGCAM and CDGCAM are programmed in GAMS environment and solved using CPLEX 12.5. The optimality gap of CPLEX is set to be 10^{-4}, and the time limit is set to be 3600 s. In the BD, the optimality tolerance σ is set to be 0.02.

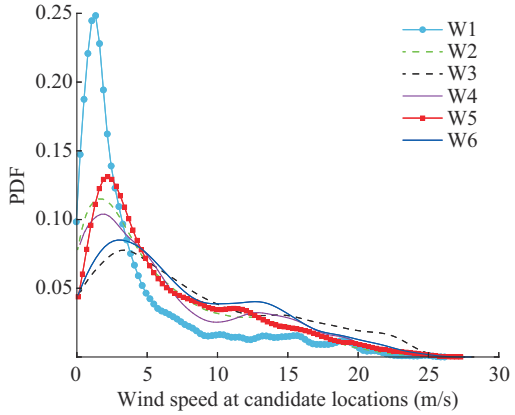

Fig. 6. Marginal distributions of observed data at different locations.

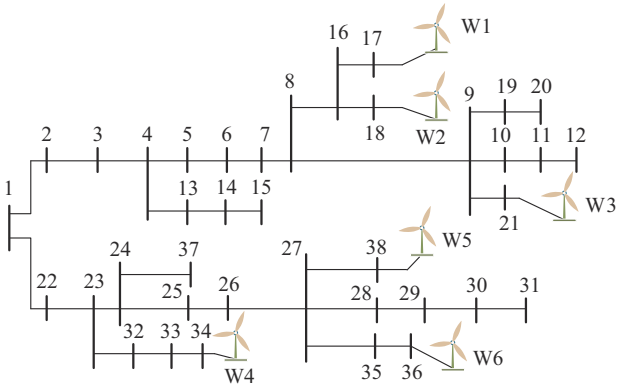

Fig. 7. Diagram of 38-bus distribution network in eastern China.

All experiments are implemented on an HP Z840 workstation with 10-core processors clocked at 2.4 GHz and 16 GB of RAM.

A. Modeling Correlated Wind Speed with Copula and C-vine Copula

In this paper, the families of Gaussian copula, *t*-copula, Frank copula, Gumbel copula, and Clayton copula are considered as candidate bivariate copulas. To show the differences in their abilities to describe complicated dependence structures, the results of a statistical model quality index called AIC and scatter plots of some variable pairs are shown in Figs. 8 and 9. The AIC is obtained by:

$$AIC=-2L(\boldsymbol{\theta}|\boldsymbol{u})+3\varsigma \tag{33}$$

where L is the log-likelihood; $\boldsymbol{\theta}$ is the parameter set; ς is the number of parameters estimated by the model; and $\boldsymbol{u}$ is the observation. The AIC provides a means for model selection by estimating the quality of each model. The model with the lowest AIC is regarded as the best-fit copula model. For a better illustration of the dependence structures in the scatter plots, the variables in Figs. 8 and 9 are transformed into a uniform domain with their marginal distributions.

Owing to different types of tail dependencies and symmetries of these copula functions, the fitting results are distinct from each other. For instance, the Gaussian copula and *t*-copula cannot capture the tail dependence. The Gumbel copula is sensitive to the lower tail dependence.

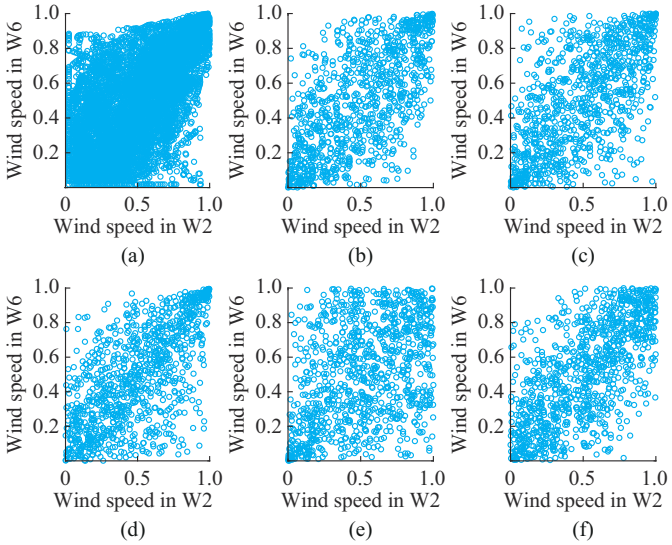

Fig. 8. Scatter diagrams of W2 and W6. (a) Actual data. (b) Gaussian copula, $AIC=-5609.38$. (c) *t*-copula, $AIC=-5718.23$. (d) Gumbel copula, $AIC=-6132.38$. (e) Clayton copula, $AIC=-5878.17$. (f) Frank copula, $AIC=-5602.55$.

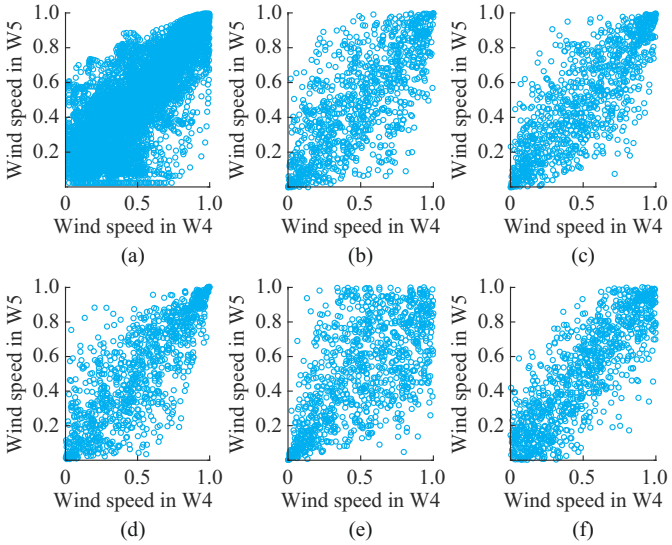

Fig. 9. Scatter diagrams of W4 and W5. (a) Actual data. (b) Gaussian copula, $AIC=-9245.88$. (c) *t*-copula, $AIC=-9399.91$. (d) Gumbel copula, $AIC=-9693.72$. (e) Clayton copula, $AIC=-8797.128$. (f) Frank copula, $AIC=-9789.12$.

On the contrary, the Clayton copula is heavily concentrated in the upper tail. The Frank copula has a wide and uniform degree of scattering. In Fig. 8, the blue scatters are mainly dispersed at the lower-left corner of Fig. 8(a), which shows a distinct lower tail dependence; therefore, the Gumbel copula has a better fit than the other copula functions. In Fig. 9, the blue scatters are distributed evenly along the diagonal of Fig. 9(a) and spread out at the lower-left corner. It is difficult to obtain a proper function just by the shape of the scatters, since it shows the characteristics of both of the Gumbel copula and Frank copula. The AIC provides a more convincing and precise result. The Frank copula has a lower AIC and fits the observations more appropriately. However, neither the Gumbel copula nor the Frank copula can exactly

fit all of the bivariate models. Actually, the dependence relation varies with the installation locations, and such a variable dependence structure is far beyond the fitness of any specific bivariate copula.

The bivariate case demonstrates the necessity of using a mix of different bivariate copula functions. Thus, we compare the performance of the C-vine copula and the popular multivariate Gaussian copula in modeling geographically dispersed wind speed distribution. To quantify the performance of each copula model, three indexes including AIC, Bayesian information criterion (BIC), and log-likelihood are adopted. As we know, a higher log-likelihood value indicates a better approximation. However, a higher log-likelihood is usually accompanied by overfitting. Like the AIC, the BIC is also a criterion for model selection, which attempts to choose a model with fewer parameters and a high likelihood value to avoid overfitting. The BIC is obtained by:

$$BIC = k\ln(n) - 2L(\boldsymbol{\theta}|\boldsymbol{u}) \tag{34}$$

where n is the number of observations. The model with the lowest BIC is preferred.

The results of the fitting tests of different copula models are summarized in Table I.

TABLE I
FITTING TEST OF DIFFERENT COPULA MODELS

Index	AIC	BIC	Log-likelihood
Gaussian copula	−72782.54	−72633.91	36412.27
C-vine copula	−77875.87	−77621.07	38973.94

It can be observed that the C-vine copula has smaller values of the AIC and BIC and larger values of the log-likelihood than those of the Gaussian copula, which means that the C-vine copula has better performance in fitting this data sample, while the Gaussian copula is relatively inaccurate.

B. DG Hosting Capacity of Typical Distribution System

1) Chance-unconstrained Case

In this subsection, detailed experiments are presented using SDGCAM to reveal the impact of different dependence models of renewable resources on the maximum hosting capacity. The computation results are listed in Table II. Four test datasets with different dependence relations are specified as follows.

1) *IND*: the wind speeds are assumed to be independent, and the marginal distributions of the wind speeds are the same as those of the actual data. The Monte Carlo sampling method is applied to generate 1000 scenarios.

2) *VINE*: the correlations between different wind speeds are modeled by C-vine copula, and 1000 samples are generated by Appendix A Algorithm A1.

3) *COPULA*: the correlations between wind speeds are modeled by Gaussian copula, and the probability integral transformation method is applied to generate 1000 samples.

4) *ACTUAL*: the actual data of the wind speeds (35040 periods) are used to directly compute the SDGCAM.

After sampling a sufficient number of scenarios from the cases, the SDGCAM is computed to determine how many DGs can be installed in the testing network without curtailment. The DG hosting capacities for the four cases are presented in Table II.

TABLE II
HOSTING CAPACITIES OF SDGCAM FOR DATASETS WITH DIFFERENT DEPENDENCIES

Dataset	Hosting capacity (MW)	Time (s)
IND	20.38	7.01
COPULA	16.77	9.61
VINE	16.56	5.80
ACTUAL	16.79	2180.63

It can be observed from Table II that the total hosting capacity will be highly overestimated (3.59 MW, 21.38% of the actual DG hosting capacity) if the spatial dependence is not taken into account. This overestimate originates from the ignorance of the positive spatial correlations between DG outputs. Such positive spatial correlations aggravate the fluctuation in the total wind generation and lead to more extreme scenarios and lower DG hosting capacity. The optimistic results given by the assumption of independence pose a significant threat to the operation of the distribution system. The results in Table II demonstrate that it is necessary to consider the spatial correlation between DG outputs in the DG hosting capacity scheme.

In addition, the hosting capacity at each bus is shown in Fig. 10 to reveal the impact of different dependence models.

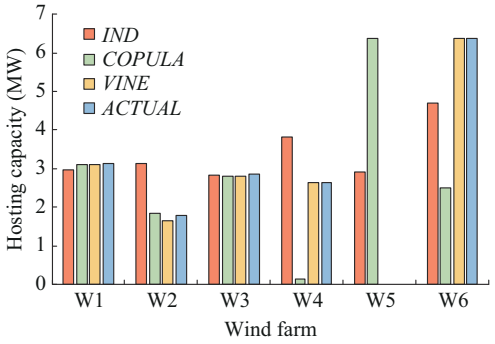

Fig. 10. Hosting capacity at each bus for *IND*, *COPULA*, *VINE*, and *ACTUAL* datasets.

As shown in Fig. 10, the C-vine copula is the best fit to the actual data, whereas the Gaussian copula leads to erroneous capacity evaluation results at W4, W5, and W6. In particular, the hosting capacity of W5 is zero in the *ACTUAL* dataset, while the hosting capacity is more than 3 MW for the *IND* and *COPULA* datasets. The reasons for these differences are understandable. In the *IND* dataset, the wind speeds at different locations can be quite different, which may cause complementary wind speeds (power) at different locations at the same time, which is the reason why the capacity of each wind farm is positive and close in the *IND* dataset. In the *COPULA* dataset, the complementary wind speeds are limited by the correlation between wind farms, which leads to different hosting capacities than those of the *IND* dataset. However, other main dependence structures

such as tail and asymmetric dependencies are not considered because only linear correlations are considered in the Gaussian copula model, and the hosting capacity is still far from that of the *ACTUAL* dataset. Neither an assumption of independence nor a copula-based model can correctly reflect the true wind speed correlations in this case. For the best fit of the various tail dependencies among wind speeds, the C-vine copula is the best choice. Note that although the *ACTUAL* dataset leads to the most actual result, computing this dataset is intractable in CDGCAM. Considering both the runtime and accuracy, the C-vine copula method is adopted as the primary modeling method in the following tests.

2) Chance-constrained Case

In this subsection, the impact of the acceptable wind curtailment probability is studied using the proposed CDGCAM model. Here, 1000 data samples are generated from independent marginal distributions, the C-vine copula, and the Gaussian copula. The wind generation capacity assessment results obtained via the CDGCAM with different curtailment probabilities are presented in Table III. Note that the CDGCAM model will be reduced to the SDGCAM model when the wind curtailment probability is set to be 0.

TABLE III
HOSTING CAPACITY RESULTS OF CDGCAM FOR DATASETS WITH DIFFERENT DEPENDENCIES AND DIFFERENT WIND CURTAILMENT PROBABILITIES

Curtailment probability (%)	Hosting capacity (MW)		
	IND	*COPULA*	*VINE*
1	20.76	17.36	16.72
5	21.44	18.56	17.34
10	21.76	19.39	17.72
15	22.39	20.70	18.40
20	23.98	21.69	19.37

From the results in Table III, it is obvious that the hosting capacity has a monotonically increasing relation with the curtailment probability. In fact, a higher curtailment probability leads to more discarded scenarios. Therefore, more optimistic results can be obtained for the hosting capacity. By controlling the curtailment probability, the DNO can achieve a satisfactory trade-off between the level of penetration and the renewable energy curtailment probability. Moreover, the modeling errors of the independence and Gaussian models are amplified in the higher curtailment probability case, which increases the operation risk of the DG planning scheme. Hence, it is important to select a proper dependence model to assess the DG hosting capacity.

For better understanding of the impact of different wind curtailment probabilities, we present the results for the bus voltage and apparent power of each line of the C-vine-copula-based CDGCAM in Figs. 11 and 12, respectively. Accommodating a high penetration of DG will lead to a reverse power flow, which increases the bus voltages at the DG nodes. As shown in Fig. 11, the voltages of buses 17, 18, and 21 reach the upper limit in the full accommodation case. When we relax the power balance constraint to a chance constraint, the bus voltage in some extreme scenarios can reach a higher level to accommodate more renewable energy. Compared with the bus voltage, the apparent power in Fig. 12 is less sensitive to the wind curtailment probability, which will not exceed the line capacity limit until the wind curtailment probability increases to 15%. This sensitivity analysis provides a hint for distribution system management, which means that the current distribution system can accommodate 16.56 MW of DWG without further network management and accommodate up to 18.40 MW of DWG with proper voltage management. Further, network expansion may be required to accommodate more than 18.40 MW of DWG.

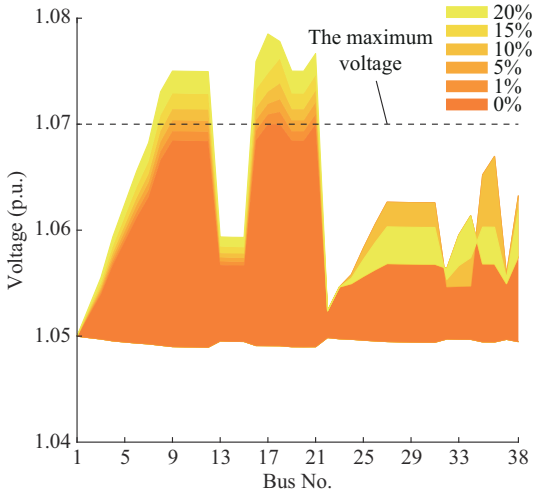

Fig. 11. Bus voltage with different wind curtailment probabilities.

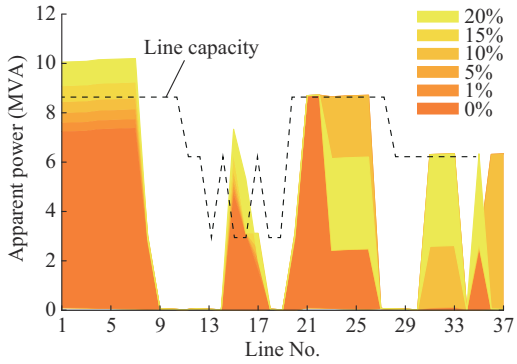

Fig. 12. Apparent power of each line with different wind curtailment probabilities.

3) Performance Comparisons

In this subsection, comparisons of the performance of the Big-*M* method, bilinear variant, and bilinear BD via 1000 C-vine-copula data samples are presented. If an optimal solution cannot be derived owing to the time limit (3600 s), the instance is labeled with an "F" The computational performance is summarized in Table IV.

It can be observed from Table IV that even a small increase in the wind curtailment probability can lead to a significant increase in the computational time. Moreover, the state-of-the-art CPLEX solver does not have the capacity to handle cases with a wind curtailment probability greater than 10% with the Big-*M* formulation. In contrast, the proposed bilinear variant of the chance-constrained formulation has a

better solution capacity and better computational performance. Compared with direct solution of the Big-*M* formulation and bilinear variant formulations, the bilinear BD method shows a drastically improved solution capacity and better performance and solves all cases within the time limit.

TABLE IV
COMPUTATIONAL TIME OF CDGCAM WITH DIFFERENT CURTAILMENT PROBABILITIES

Curtailment probability (%)	Computational time (s)		
	Big-*M*	Bivariant	BD
1	234.97	122.49	29.91
5	2821.93	1033.74	550.81
10	3374.47	1461.37	1356.79
15	F(N/A)	3567.64	1834.67
20	F(N/A)	F(N/A)	3257.20

VI. CONCLUSION

In this paper, we propose a novel approach for assessing the DG hosting capacity that considers the spatial correlations between uncertain DG outputs and the risk preference of DNO. In this approach, an advanced C-vine copula is applied to model the complex high-dimensional dependencies among geographically dispersed DGs. After generating a sufficient number of scenarios from the C-vine copula model, the dependencies of the geographically dispersed DG outputs are introduced to the chance-constrained assessment model for DWGHC, where the risk preference of DNO is considered by a predefined DG curtailment probability. Furthermore, a bilinear variant of the chance constraints and a customized BD algorithm are developed to solve the chance-constrained assessment model for DWGHC and to deal with the computational challenge resulting from the large number of random scenarios.

Numerical results for wind generation in a real distribution system in eastern China verify the effectiveness of the proposed method and algorithm. The results show that the maximum hosting capacity would be too optimistic without the consideration of the spatial correlations between wind generation outputs, which will lead to inaccurate and unrealistic solutions during the assessment of the capacity. Specifically, the hosting capacity of the test system under the assumption that the independent wind speed is 21% higher than that of the actual wind speed scenario. Thus, the correlations between DG outputs should be considered for authentic hosting capacity results. After further investigation of the dependence models, we find that the selection of the dependence model has a significant impact on the accuracy of the capacity assessment results. In the test system, the C-vine copula has better fitting performance than Gaussian copula, and thus provides a more accurate hosting result for each candidate bus. Overall, the correlations between renewable energy sources impact the total hosting capacity of the distribution network, and the dependence structure has an impact on the hosting capacity of each bus. The dependencies between DG resources should be carefully handled in the hosting capacity problem.

Another numerical test shows the impact of wind generation curtailment. On the basis of the proposed bilinear chance-constrained assessment model, we found that the curtailment probability has a significant impact on the hosting capacity; a higher curtailment probability leads to a higher hosting capacity. These results are consistent with previous experience in the operation of distribution networks. A further study found that hosting capacity models with different dependence structures have different sensitivities to wind curtailment. In our case study, the vine copula model is less sensitive than the Gaussian copula model.

Further study can use a more powerful dependence modeling tool to analyze the impacts of the correlations among different loads and renewable resources on the hosting capacity.

APPENDIX A

Algorithm A1: simulation algorithm for a canonical vine

Sample $u_1, u_2, \ldots, u_d$ independent uniform on [0, 1].
$x_1 = v_{1,1} = u_1$
for $n = 2, 3, \ldots, d$
  $v_{n,1} = u_n$
  for $k = n-1, n-2, \ldots, 1$
    $v_{n,1} = h^{-1}(v_{n,1}, v_{k,k}, \Theta_{k,n-k})$
  end for
  $x_n = v_{n,1}$
  if $n == d$ **then**
    stop
  end if
  for $j = 1, 2, \ldots, n-1$
    $v_{n,m+1} = h(v_{n,m}, v_{m,m}, \Theta_{m,n-m})$
  end for
end for

TABLE AI
LOAD DATA

Bus No.	Active load (kW)	Reactive load (kvar)	Bus No.	Active load (kW)	Reactive load (kvar)
2	83.640	22.018	21	28.988	7.631
3	65.290	21.942	22	68.320	22.961
4	87.390	27.957	23	77.210	24.700
5	122.471	38.880	24	183.630	58.295
6	102.760	34.484	25	208.080	69.828
7	163.232	50.462	26	106.510	32.927
8	81.600	21.629	27	49.612	13.150
9	102.695	33.145	28	24.500	7.907
10	30.600	10.329	29	36.720	12.395
11	61.265	18.435	30	32.640	9.822
12	37.678	11.170	31	112.200	33.262
13	132.636	41.179	32	116.530	36.179
14	22.112	6.695	33	97.850	29.625
15	51.230	17.190	34	22.177	7.441
16	61.240	18.605	35	28.980	8.804
17	81.600	21.310	36	20.400	5.328
18	122.403	32.038	37	22.179	5.805
19	28.969	9.167	38	20.480	6.481
20	20.410	6.360			

TABLE AII
LINE DATA

Line No.	Bus out	Bus in	R (Ω)	X (Ω)	Thermal limit (kVA)	Line No.	Bus out	Bus in	R (Ω)	X (Ω)	Thermal limit (kVA)
1	1	2	0.029	0.048	8711.960	20	9	21	0.060	0.029	3117.600
2	2	3	0.029	0.048	8711.960	21	1	22	0.029	0.048	8711.960
3	3	4	0.041	0.066	8711.960	22	22	23	0.028	0.045	8711.960
4	4	5	0.033	0.053	8711.960	23	23	24	0.014	0.023	8711.960
5	5	6	0.033	0.053	8711.960	24	24	25	0.030	0.049	8711.960
6	6	7	0.028	0.046	8711.960	25	25	26	0.028	0.046	8711.960
7	7	8	0.051	0.082	8711.960	26	26	27	0.026	0.041	8711.960
8	8	9	0.075	0.121	8711.960	27	27	28	0.028	0.046	8711.960
9	9	10	0.033	0.054	8711.960	28	28	29	0.028	0.046	8711.960
10	10	11	0.033	0.054	8711.960	29	29	30	0.028	0.045	8711.960
11	11	12	0.058	0.094	8711.960	30	30	31	0.012	0.019	6339.120
12	4	13	0.017	0.021	6339.120	31	23	32	0.029	0.037	6339.120
13	13	14	0.084	0.107	6339.120	32	32	33	0.053	0.067	6339.120
14	14	15	0.063	0.031	3117.600	33	33	34	0.031	0.039	6339.120
15	8	16	0.042	0.053	6339.120	34	24	37	0.023	0.030	6339.120
16	16	17	0.053	0.026	3117.600	35	27	38	0.027	0.034	6339.120
17	16	18	0.100	0.049	3117.600	36	27	35	0.044	0.055	6339.120
18	9	19	0.023	0.029	6339.120	37	35	36	0.029	0.037	6339.120
19	19	20	0.034	0.017	3117.600						

REFERENCES

[1] M. Kolenc, I. Papic, and B. Blazic, "Assessment of maximum distributed generation penetration levels in low voltage networks using a probabilistic approach," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 505-515, Jan. 2015.

[2] A. Dubey and S. Santoso, "On estimation and sensitivity analysis of distribution circuit's photovoltaic hosting capacity," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2779-2789, Jul. 2017.

[3] M. Abad, J. Ma, D. Zhang *et al*., "Probabilistic assessment of hosting capacity in radial distribution systems," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1935-1947, Oct. 2018.

[4] X. Chen, W. Wu, and B. Zhang, "Robust capacity assessment of distributed generation in unbalanced distribution networks incorporating ANM techniques," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 651-663, Apr. 2018.

[5] S. Ismael, S. Aleem, A. Abdelaziz *et al*., "State-of-the-art of hosting capacity in modern power systems with distributed generation," *Renewable Energy*, vol. 130, pp. 1002-1020, Jan. 2019.

[6] C. Yang, A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 1, pp. 104-112, Jan. 2015.

[7] H. Wu, Y. Yuan, J. Zhu *et al*., "Potential assessment of spatial correlation to improve maximum distributed PV hosting capacity of distribution networks," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 4, pp. 800-810, Jul. 2021.

[8] X. Ran and S. Miao, "Three-phase probabilistic load flow for power system with correlated wind, photovoltaic and load," *IET Generation, Transmission & Distribution*, vol. 10, no. 12, pp. 3093-3101, Jun. 2016.

[9] Z. Qin, W. Li, and X. Xiong, "Incorporating multiple correlations among wind speeds, photovoltaic powers and bus loads in composite system reliability evaluation," *Applied Energy*, vol. 110, pp. 285-294, Oct. 2013.

[10] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 40-49, Feb. 2009.

[11] Q. Tu, S. Miao, F. Yao *et al*., "Forecasting scenario generation for multiple wind farms considering time-series characteristics and spatial-temporal correlation," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 4, pp. 837-848, Jul. 2021.

[12] S. Sreekumar, K. Sharma, and R. Bhakar, "Gumbel copula based aggregated net load forecasting for modern power systems," *IET Generation, Transmission & Distribution*, vol. 12, no. 19, pp. 4348-4358, Aug. 2018.

[13] 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, Jan. 2018.

[14] T. Bedford and R. Cooke, "Vines–a new graphical model for dependent random variables," *Annals of Statistics*, vol. 30, no. 4, pp. 1031-1068, Aug. 2002.

[15] 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.

[16] S. Wang, S. Chen, L. Ge *et al*., "Distributed generation hosting capacity evaluation for distribution systems considering the robust optimal operation of OLTC and SVC," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1111-1123, Jul. 2016.

[17] K. Zou, A. Agalgaonkar, K. Muttaqi *et al*., "Distribution system planning with incorporating DG reactive capability and system uncertainties," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 112-123, Jan. 2012.

[18] V. Martins and C. Borges, "Active distribution network integrated planning incorporating distributed generation and load response uncertainties," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2164-2172, Nov. 2011.

[19] X. Chen, W. Wu, B. Zhang *et al*., "Data-driven DG capacity assessment method for active distribution networks," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3946-3957, Sept. 2017.

[20] L. Ochoa, C. Dent, and G. Harrison, "Distribution network capacity assessment: variable DG and active networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 87-95, Feb. 2010.

[21] H. Wu, L. Wu, Y. Yuan *et al*., "A PV hosting capacity assessment method with spatial correlation," in *Proceedings of 2019 IEEE Sustainable Power and Energy Conference*, Beijing, China, Nov. 2019, pp. 1102-1107.

[22] X. Jiang, H. Chen, and T. Xiang, "Assessing the effect of wind power peaking characteristics on the maximum penetration level of wind power," *IET Generation, Transmission & Distribution*, vol. 9, no. 16, pp. 2466-2473, Nov. 2015.

[23] H. Park, R. Baldick, and D. Morton, "A stochastic transmission planning model with dependent load and wind forecasts," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3003-3011, Nov. 2015.

[24] F. Qiu and J. Wang, "Chance-constrained transmission switching with guaranteed wind power utilization," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1270-1278, May 2015.

[25] K. Aas, C. Czado, A. Frigessi *et al*., "Pair-copula constructions of mul-

tiple dependence," *Insurance Mathematics & Economics*, vol. 44, no. 2, pp. 182-198, Apr. 2009.
[26] R. Nelsen, *An Introduction to Copulas*. New York: Springer, 2007.
[27] H. Joe, "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters," *Lecture Notes-Monograph Series*, vol. 28, pp. 120-141, 1996.
[28] J. Hernandez, S. Hammoudeh, D. K. Nguyen *et al*., "Global financial crisis and dependence risk analysis of sector portfolios: a vine copula approach," *Applied Economics*, vol. 49, no. 25, pp. 2409-2427, Nov. 2016.
[29] R. Becker, "Generation of time-coupled wind power infeed scenarios using pair-copula construction," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1298-1306, Jul. 2018.
[30] M. Sun, I. Konstantelos, and G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2382-2393, May 2017.
[31] Y. Cao, Y. Zhang, H. Zhang *et al*., "Probabilistic optimal PV capacity planning for wind farm expansion based on nasa data," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1291-1300, Jul. 2017.
[32] X. Chen, W. Wu, and B. Zhang, "Robust restoration method for active distribution networks," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 4005-4015, Sept. 2016.
[33] H. Haghighat and B. Zeng, "Stochastic and chance-constrained conic distribution system expansion planning using bilinear benders decomposition," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2696-2705, May 2018.
[34] B. Zeng, Y. An, and L. Kuznia. (2014, Mar.). Chance constrained mixed integer program: bilinear and linear formulations, and benders decomposition. [Online]. Available: https://arxiv.org/abs/1403.7875
[35] Y. Zhang, J. Wang, B. Zeng *et al*., "Chance-constrained two-stage unit commitment under uncertain load and wind power output using bilinear benders decomposition," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3637-3647, Sept. 2017.
[36] Vine copula: statistical inference of vine copulas. (2021, Oct.). [Online]. Available: https://CRAN.R-roject.org/package=VineCopula

Han Wu received the M. E. and Ph. D. degrees in electrical engineering from Hohai University, Nanjing, China, in 2015 and 2020, respectively. He is currently a Lecturer with Nanjing Institute of Technology, Nanjing, China. His research interests include distribution system planning, probabilistic power system analysis, and power system optimization.

Yue Yuan received the B. E. and M. Sc. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1987 and 1990, respectively, and the Ph. D. degree from Hiroshima University, Hiroshima, Japan, in 2002. He joined the Faculty of Xi'an Jiaotong University, in 1990. He has been with Hohai University, Nanjing, China, as a Professor, since 2003. His research interests include power system operation and optimization, renewable energy, and distributed generation.

Junpeng Zhu received the B.S. degree in applied mathematics and the Ph. D. degree in electrical engineering from Southeast University, Nanjing, China, in 2012 and 2017, respectively. He is currently a Lecturer with the College of Energy and Electrical Engineering, Hohai University, Nanjing, China. His research interests include distributed generation integration, optimization, and control of active distribution networks.

Yundai Xu received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2012. He is currently working toward the Ph. D. degree in electrical engineering from Hohai University, Nanjing, China. His research interests include optimal power flow and renewable energy forecasting.