



Statistical Machine Learning Model for Uncertainty Planning of Distributed Renewable Energy Sources in Distribution Networks

Xueqian Fu^{1*}, Xianping Wu¹ and Nian Liu²

¹College of Information and Electrical Engineering, China Agricultural University, Beijing, China, ²State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Beijing, China

OPEN ACCESS

Edited by:

Bin Zhou,
Hunan University, China

Reviewed by:

Yang Liu,
South China University of Technology,
China

Qiang Yang,
Zhejiang University, China

*Correspondence:

Xueqian Fu
fuxueqian@cau.edu.cn

Specialty section:

This article was submitted to
Process and Energy Systems
Engineering,
a section of the journal
Frontiers in Energy Research

Received: 04 November 2021

Accepted: 18 November 2021

Published: 13 December 2021

Citation:

Fu X, Wu X and Liu N (2021) Statistical
Machine Learning Model for
Uncertainty Planning of Distributed
Renewable Energy Sources in
Distribution Networks.
Front. Energy Res. 9:809254.
doi: 10.3389/fenrg.2021.809254

New energy power systems with high-permeability photovoltaic and wind power are high-dimensional dynamic large-scale systems with nonlinear, uncertain and complex operating characteristics. The uncertainty of new energies creates challenges in detailed analyses of operating conditions and the efficient planning of distribution networks. Probabilistic power flows (PPFs) are effective tools for uncertainty analyses of distribution networks, and they can be applied in stochastic programming, risk assessment and other fields. We propose different forms of PPFs, which are origin moments rather than means and variances, based on point estimation. We design a stochastic programming model suitable for new energy planning in practice, and the PPF results can be used to improve energy stochastic programming methods by considering the principle of maximum entropy (POME) and quadratic fourth-order moment (QFM) estimation. The origin moments of PPFs are transformed into central moments as inputs of QFM based on probability theory. QFM can efficiently estimate the constraint probability levels of stochastic optimal planning models, and the proposed method is verified based on an IEEE 33-node distribution network.

Keywords: statistical machine learning, stochastic optimal planning, distributed renewable energy sources, probabilistic power flow, uncertainty

INTRODUCTION

In the context of smart grids and low-carbon power, the broad application of intermittent renewable distributed generation (DG) has led to uncertainty in distribution network planning (Injeti and Thunuguntla, 2020). Hydrogen production from renewable energy has become a hot spot of new energy application because of its low energy consumption (Zhang et al., 2021). The impact of new energy uncertainty on the operation optimization of distribution networks cannot be ignored, and high requirements for planning and design have been proposed. The stochastic optimal planning of distribution networks has become one of the key problems in the development of smart grids. The impacts of various uncertain processes on distribution network planning results can be formulated as part of a complex, nonlinear mixed-integer programming problem. The traditional planning method is based on typical daily deterministic scenarios, and the influence of uncertain factors is often not considered. The main methods that include uncertain programming in power systems are robust programming and stochastic programming.

Robust power system planning commonly combines the worst-case scenario approach and interval optimization theory (Liu et al., 2016). Roldán *et al.* presented an adaptive robust optimization theory in which correlated uncertainty and the worst-case operating cost are fully considered (Roldán et al., 2019). Considering the uncertainty of DG and the corresponding demand, Chatthaworn and Chaitusaney presented a robust model for transmission network expansion planning in which all possible scenarios are analyzed (Chatthaworn and Chaitusaney, 2015). Zhang *et al.* presented a sequential quadratic interval programming model for reactive power optimization based on the interval uncertainty of DG (Zhang et al., 2019). To efficiently solve reactive power planning problems, the applied planning model often approximated as a linear model (Zhang et al., 2018). Zhang *et al.* presented a scenario method for reactive power optimization, and the uncertainty of DG was obtained from distribution functions (Zhang et al., 2016). Compared with deterministic programming, robust programming can ensure the security of power grid systems under extreme conditions. However, the fluctuations in DG uncertainty considerably influence robust programming. When the fluctuations in uncertain factors are small, the superiority of robust programming may not be apparent. When the fluctuations in uncertain factors exceed the corresponding threshold values, robust programming may not yield sufficient planning results. In addition, for safety, robust programming may be too extreme in many scenarios.

With respect to stochastic programming, probability distributions are often used in uncertain programming modeling. Gan *et al.* presented a wind farm planning method based on a linearized bilevel model that was solved using a linear solver (Gan et al., 2016). The mixed-integer linear stochastic model has become a popular stochastic planning model for distribution networks, and various scenarios should be considered to capture DG uncertainty (Jooshaki et al., 2020). Park *et al.* presented a two-stage stochastic mixed-integer programming model that considers DG uncertainty (Park et al., 2015). Haghighat and Zeng presented a two-stage stochastic programming model in which correlations among uncertainties are reflected by a Gaussian copula (Haghighat and Zeng, 2018). In regard to new energy power systems, the operation scenario of power systems is often determined according to the DG outputs. Chen *et al.* suggested that the operation scenario should be established according to the relevant state variables (Chen et al., 2020). The uncertainty of DG is considered in stochastic programming, so decision makers can assess the relationship between risk and benefits, which can help balance the economy and security of distribution networks. However, the probability distribution of DG uncertainty in stochastic programming models is limited by the amount of historical data available, and this limitation can bias the empirical distribution. In addition, nonlinear stochastic programming methods often have poor solution efficiency. Thus, establishing an efficient planning model is important and is the research motivation of this paper.

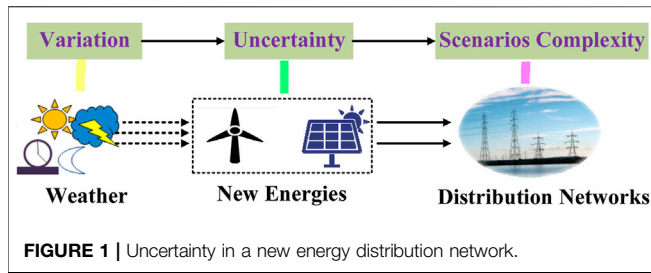
Scholars conducted research in the stochastic planning and inquiry field. The decision variables of the stochastic planning

scenario are related to DG levels, reactive power devices, PEV charging stations, etc. The decision variables include the optimal site, size, quantity or level. To decrease wind power (WP) curtailment, a novel reactive power allocation scheme was presented in (Niu et al., 2017) based on autonomous voltage security regions. To solve the problem of reactive power imbalance associated with DG, a robust reactive power potential estimation method was presented in (Li et al., 2019) based on two-stage robust optimization. To improve the efficiency of stochastic programming under the premise of fully considering refined scenarios, a new stochastic programming theory was presented in (Fu et al., 2020) based on statistical machine learning. To maximize the profit of PEV charging stations, a bilevel optimization model was formulated in (Zhao et al., 2020), and the optimal sites and station sizes were calculated. It can be concluded that the consideration of equivalence in programming models is a key to solving stochastic programming problems.

In uncertainty analyses of power systems, probabilistic digital information is important. Many works use probabilistic power flow (PPF) calculations to evaluate the influence of uncertainty on the power grid, and the final results include the mean and variance (Tang et al., 2016; Ren et al., 2017; Lin et al., 2020). In this paper, a novel concept is proposed for the stochastic planning of power grids. In stochastic programming models, the central moment is used instead of the mean and variance.

The innovation of this paper lies in: 1) PPF is applied to uncertain programming for the first time, which verifies the application value of PPF theory. Instead of the mean and variance, we use central moments as the results of the PPF calculation, and they are used to estimate the opportunity constraint levels in stochastic programming problems. 2) The principle of maximum entropy (POME) is applied twice to uncertain planning of distributed renewable energy sources in distribution networks. The first involves estimating the probability characteristics of temperature, which can influence photovoltaic (PV) power generation, and the second involves estimating the probability level of constraint satisfaction in stochastic programming. 3) We propose a novel uncertain planning model with the mean value as the objective function and confidence levels as constraints. The practical significance of the planning model is clear and can be expanded for real problems.

The remainder of this paper is organized as follows. First, a statistical model for uncertainty planning of distributed renewable energy sources is proposed based on statistical machine learning in *Stochastic Programming Model* Section. Second, uncertainty calculation method is given for expanding PPF results into a stochastic programming model with the proposed method in *Uncertainty Calculation Method* Section. Specifically, we propose methods for weather uncertainty calculation, PPF calculation and uncertainty constraint calculation. Finally, three cases are simulated in MATLAB to verify the effectiveness of the proposed statistical model and solution method in *Simulation* Section. Specifically, we successively verify the effectiveness of uncertainty constraints, PPF and uncertainty programming.



STOCHASTIC PROGRAMMING MODEL

Stochastic programming is an important part of uncertain programming, and it regards random variables as uncertain parameters. Combined with probability theory, the new energy system can be planned based on stochastic programming. We first introduce the stochastic programming problem and then introduce the stochastic programming model used in this paper.

Problem Description

As shown in **Figure 1**, the uncertain components of new energy distributions network are important to consider. The energy sources of PV and WP are natural, but these sources are highly unpredictable. When new renewable energy capacity is added to the grid, the uncertainty of the distribution network will increase, and the operation scenario become increasingly complex, thus creating challenges related to the efficient planning and use of new energies. In stochastic programming, decision variables can only take certain values, thereby limiting the potential operating scenarios.

Programming Model

We propose a PPF-based stochastic programming model for new energies in distribution networks as follows:

$$f_{\text{obj}}(\text{num}_{\text{PV}}, p_{\text{PV}}, \text{num}_{\text{WP}}, p_{\text{WP}}) = \min E(p_{\text{loss}}), \quad (1)$$

subject to

$$0 \leq Y_{\text{PV}} \leq Y_{\text{PV}}^-, \quad (2)$$

$$0 \leq Y_{\text{WP}} \leq Y_{\text{WP}}^-, \quad (3)$$

$$2 \leq \text{num}_{\text{PV}} < \text{num}_{\text{sys}}, \quad (4)$$

$$2 \leq \text{num}_{\text{WP}} < \text{num}_{\text{sys}}, \quad (5)$$

$$\Pr(V_i > V_-) > \alpha, \quad (6)$$

where $f_{\text{obj}}(\bullet)$ is the objective function of the planning scheme, p_{loss} is the total active power loss in distribution and can be changed to other economic indicators based on the relevant decision-making requirements, $E(\bullet)$ is the mean function, Y_{PV} is a decision variable and represents the rated capacity of PV power, Y_{PV}^- is the upper limit of the PV planning capacity, Y_{WP} is another decision variable that represents the rated capacity of the WP, Y_{WP}^- is the upper limit of the WP planning capacity, num is the number of buses with new energy grid connections, num_{sys} is the number of distribution network buses, V_i is the i th bus voltage amplitude, V_- is the allowable lower limit of v_i , $\Pr(\bullet)$ is a probability function, and α is the confidence probability.

The above formula cannot clearly explain how the decision variables affect the objective and constraint functions. Therefore, we will explain these relations through the following formula.

$$E(p_{\text{loss}}) = \text{PPF}(\text{num}_{\text{PV}}, p_{\text{PV}}, \text{num}_{\text{WP}}, p_{\text{WP}}), \quad (7)$$

where $\text{PPF}(\bullet)$ is the PPF estimation method, which is a point estimation method (PEM) in this paper. Additionally, p_{PV} is the active PV power, and p_{WP} is the active WP power.

The power generated by new energy must be a function of the planned capacity, as shown below (Rohani and Nour, 2014).

$$p_{\text{PV}} = Y_{\text{PV}} f_{\text{PV}} \left(\frac{G}{G_{\text{STC}}} \right) \left[1 + \alpha_p (T_c - T_{c,\text{STC}}) \right] \quad (8)$$

$$T_c = \frac{T_a + (T_{c,\text{NOCT}} - T_{a,\text{NOCT}}) \frac{G}{G_{\text{NOCT}}} \left[1 - \frac{\eta_{\text{mp,STC}} (1 - \alpha_p T_{c,\text{STC}})}{\tau \alpha_{\text{ab}}} \right]}{1 + (T_{c,\text{NOCT}} - T_{a,\text{NOCT}}) \frac{G}{G_{\text{T,NOCT}}} \frac{\alpha_p \eta_{\text{mp,STC}}}{\tau \alpha_{\text{ab}}}} \quad (9)$$

$$p_{\text{WP}} = \begin{cases} 0 & v < v_{\text{ci}} \text{ OR } v \geq v_{\text{co}} \\ \frac{v^3 - v_{\text{ci}}^3}{v_r^3 - v_{\text{ci}}^3} Y_{\text{WP}} & v_{\text{ci}} \leq v < v_r \\ Y_{\text{WP}} & v_r \leq v < v_{\text{co}} \end{cases} \quad (10)$$

where f_{PV} is the PV derating factor, G is the amount of current solar radiation, STC is a subscript that represents standard test conditions, α_p is the temperature coefficient, T_c is the PV cell temperature, NOCT is a subscript that represents the nominal operating cell temperature, T_a is the ambient temperature, η_{mp} is the maximum power efficiency, τ is the solar transmittance of a PV cell, α_{ab} is the solar absorptance of a PV cell, v is the current wind speed, v_r is the nominal wind speed, v_{ci} is the cut-in wind speed, and v_{co} is the cut-out wind speed.

By substituting **Equations 8–10** into **Eq. 7** and treating the capacities and locations of new energies as decision variables, we can obtain

$$E(p_{\text{loss}}) = \text{PPF}(\text{num}_{\text{PV}}, Y_{\text{PV}}, \text{num}_{\text{WP}}, Y_{\text{WP}}), \quad (11)$$

Now, let us summarize the stochastic programming model for new energies in distribution networks. Notably, (2), (3), (4), and (5) are the upper and lower limits of decision variables, and these constraints can be implemented by limiting the feasible region when solving the stochastic programming model for new energies in distribution networks.

Remark 1: The key to the solution is to efficiently and accurately solve **Eqs 6, 11**. As in the estimation in **Eq. 1** based on **Eqs 6, 11** can also be estimated based on a PPF. The solution processes for (6) and (11) are discussed in detail in the next section.

UNCERTAINTY CALCULATION METHOD

First, we introduce simulated weather data, which are the inputs of **Eqs 8–10**. Second, the PPF method is introduced, and the results are origin moments. Finally, the method used to solve **Eq. 6** is introduced based on central moments and particle swarm

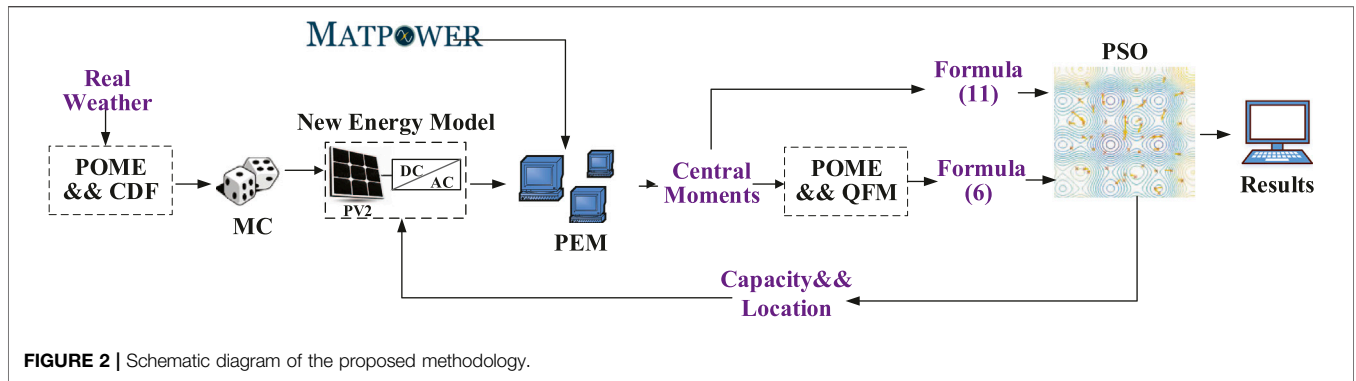


FIGURE 2 | Schematic diagram of the proposed methodology.

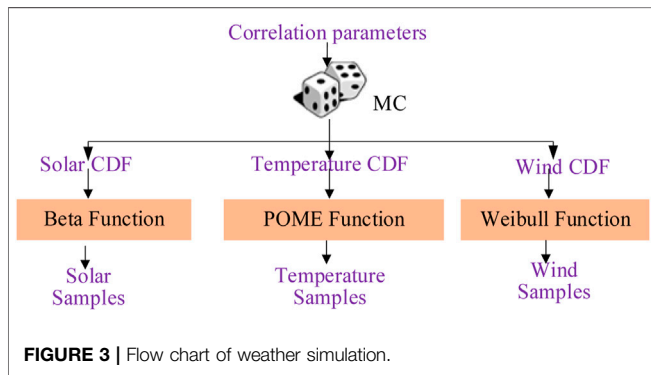


FIGURE 3 | Flow chart of weather simulation.

optimization (PSO). The proposed methodology is briefly shown in **Figure 2**.

Uncertainty Weather

The weather simulation is based on a copula function, marginal probability distributions and logic, as shown in **Figure 3**. First, we introduce the marginal probability distributions for different weather types. Second, the copula function is introduced.

The marginal probability distribution of G can be expressed as (Karaki et al., 1999):

$$pdf(G) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{G}{G_{max}}\right)^{\alpha-1} \left(1 - \frac{G}{G_{max}}\right)^{\beta-1} \quad (12)$$

where $pdf(\bullet)$ is a probability density function (PDF), α and β are shape parameters, Γ is the gamma function, and G_{max} is the maximum solar radiation.

The marginal probability distribution of v can be expressed as (Karaki et al., 1999):

$$pdf(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^k\right] \quad (13)$$

where k is a shape parameter and c is a scale parameter.

The marginal probability distribution of T_a can be expressed using POME in (Shore and Johnson, 1980):

$$\max H(T_a) = - \int pdf(T_a) \ln pdf(T_a) dT_a, \quad (14)$$

subject to

$$\int pdf(T_a) dT_a = 1, \quad (15)$$

$$\int g_k(T_a) pdf(T_a) dT_a = \mu_k, k = 0, \dots, m, \quad (16)$$

$$pdf(T_a) = e^{-\sum_{k=0}^m \lambda_k g_k(T_a)}, \quad (17)$$

where $H(\bullet)$ is the entropy function, μ_k is the k th given expectation, $g_k(\bullet)$ is the k th known function, and λ_k is the k th Lagrange parameter.

λ_k in **Eq. 18** can be estimated using the program in (Mohammad-Djafari et al., 1992).

$$\int g_k(T_a) e^{-\sum_{k=0}^m \lambda_k g_k(T_a)} dT_a = \mu_k, k = 0, \dots, m \quad (18)$$

A Gaussian copula function is used to model the correlation among G , v and T_a . We estimate rho for the matrix of linear correlation parameters using real weather data.

$$rho = \text{copulafit}(cdf(G), cdf(v), cdf(T_a)), \quad (19)$$

where rho is a coefficient estimated for a matrix of linear correlation parameters based on a Gaussian copula, $\text{copulafit}(\bullet)$ is a fitting function, and $cdf(\bullet)$ is a cumulative distribution function (CDF).

The Monte Carlo (MC) method is used to generate random weather simulation samples from the Gaussian copula with rho .

$$[cdf(G^s), cdf(v^s), cdf(T_a^s)] = \text{copularnd}(rho), \quad (20)$$

where s is a superscript that indicates that a variable is simulated and $\text{copularnd}(\bullet)$ represents the generation of random vectors using a MC approach. According to the simulated CDFs, we can obtain the simulated weather samples.

Remark 2: We input the weather variables into the model so that the research results can be applied to power grid planning under different weather conditions. There are classical probability functions for solar radiation and wind speed. With respect to the simulation of temperature, POME is used to estimate the unknown probability distribution.

PPF Calculation

PEM uses samples to obtain estimates. We can determine the statistical moments of the population using the PEM. The statistical moments of power flow solutions can be obtained from weighted samples at different positions. MATPOWER in (Zimmerman et al., 2011) is used for deterministic power flow (DPF) calculations. A two-point estimation (2PEM) approach is applied, and two values are determined on both sides of the mean value of each uncertain variable.

$$x_{i,k} = X_{\text{mean},i} + sp_{i,k} \times X_{\text{standard},i}, \quad (21)$$

$$sp_{i,k} = sk_i/2 + (-1)^{(3-k)} \times \sqrt{m + (sk_i/2)^2}, \quad (22)$$

where x_i is the i th uncertain variable (i.e., p_{PV} or p_{WP}), $i = 1, 2, \dots, m$, m is the number of uncertain variables, $X_{\text{mean},i}$ is the mean of x_i , $X_{\text{standard},i}$ is the standard deviation of x_i , $k = 1$ or 2 (where 2 reflects the values on both sides of $X_{\text{mean},i}$), sk_i is the sample skewness of x_i , and $sp_{i,k}$ is a location-specific measurement.

We introduce the following data into the DPF calculation using MATPOWER.

$$P_{i,k} = [X_{\text{mean},1}, X_{\text{mean},2}, \dots, X_{i,k}, \dots, X_{\text{mean},m}], \quad (23)$$

$$u_{i,k} = dpf(P_{i,k}), \quad (24)$$

where $dpf(\bullet)$ is the DPF calculated using MATPOWER and u is the obtained voltage, power flow, or other variable.

Next, we give the origin moment formula for u :

$$M_{\text{origin}}^j(u) = \sum_i^m \sum_k^2 w_{i,k} \times u_{i,k}^j, \quad (25)$$

$$w_{i,k} = \frac{1}{m} \times (-1)^k \frac{sp_{i,3-k}}{\delta_i}, \quad (26)$$

$$\delta_i = 2 \times \sqrt{m + (sk_i/2)^2} \quad (27)$$

where $M_{\text{origin}}(\bullet)$ is the function for the origin moment of order j and $w_{i,k}$ represents the weights for a given location set.

From Eq. 25, we can obtain the following equations:

$$E(p_{\text{loss}}) = M_{\text{origin}}^1(p_{\text{loss}}), \quad (28)$$

$$V_{\text{om},j} = M_{\text{origin}}^j(V), \quad j = 1, 2, 3, 4, \quad (29)$$

where p_{loss} is one type of u , V is the PQ node voltage set related to u , and $V_{\text{om},j}$ is the origin moment of order j .

At this point, we have obtained the objective function formula, that is, (28), using the PPF calculation based on 2PEM. The next problem to address is how to transform PPF results into constrained probabilities. The quadratic fourth-order moment (QFM) estimation method is introduced for this probability transformation by converting 29 into (6). However, the inputs of QFM are central moments. Hence, we convert the origin moments to central moments. It is well known that

$$\mu_j(V) = E\{[V - E(V)]^j\}, \quad (30)$$

where $\mu_j(\bullet)$ is the function for a central moment of order j .

This equation can be used to obtain

$$\mu_1(V) = 0, \quad (31)$$

$$\mu_2(V) = V_{\text{om},2} - V_{\text{om},1}^2, \quad (32)$$

$$\mu_3(V) = V_{\text{om},3} - 3V_{\text{om},1}V_{\text{om},2} + 2V_{\text{om},1}^3, \quad (33)$$

$$\mu_4(V) = V_{\text{om},4} - 4V_{\text{om},3}V_{\text{om},1} + 6V_{\text{om},2}V_{\text{om},1}^2 - 3V_{\text{om},1}^4, \quad (34)$$

Remark 3: PEM can be used to quickly and accurately obtain moment information for the output random variables, but Gram-Charlier series fitting is needed to obtain the PDFs of the output random variables. The proposed method combines PEM and QFM in stochastic programming to encompass the advantages of PEM (quick and accurate) and avoid the corresponding disadvantages (series expansion).

Uncertainty Constraint

A method for solving (6) with the QRM and central moments is introduced below. QRM is a theory for estimating failure probability, and we estimate the qualified voltage probability based on the principle that the sum of failure probability and qualified probability is equal to one.

$$P_f = 1 - \alpha, \quad (35)$$

where P_f is the failure probability, which is an unqualified concept.

The limit state function for the voltage amplitude is expressed as:

$$Z_i = g(V) = V_i - V_-, \quad (36)$$

By expanding $g(V_i)$ to a Taylor's series at V_i^* , we obtain the following equation:

$$Z_i = g(V_i^*) + (V_i - V_i^*)^T \nabla g(V_i^*) + \frac{1}{2}(V_i - V_i^*)^T \nabla^2 g(V_i^*) (V_i - V_i^*) \quad (37)$$

where $\nabla g(\bullet)$ is a gradient function and $\nabla^2 g(\bullet)$ is a Hessian matrix function. The moments of Z are

$$E(Z_i) = g(V_{\text{om},1}) + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 g(V_{\text{om},1})}{\partial V_i^2} \mu_2(V_i), \quad (38)$$

$$\begin{aligned} \mu_2(Z_i) = & \sum_{i=1}^n \left[\frac{\partial g(V_{\text{om},1})}{\partial V_i} \right]^2 \mu_2(V_i) + \sum_{i=1}^n \frac{\partial g(V_{\text{om},1})}{\partial V_i} \frac{\partial^2 g(V_{\text{om},1})}{\partial V_i^2} \mu_3(V_i) \\ & + \frac{1}{4} \sum_{i=1}^n \left[\frac{\partial^2 g(V_{\text{om},1})}{\partial V_i^2} \right]^2 [\mu_4(V_i) - 3\mu_2^2(V_i)] \\ & + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left[\frac{\partial^2 g(V_{\text{om},1})}{\partial V_i \partial V_j} \right]^2 \mu_2(V_i) \mu_2(V_j), \end{aligned} \quad (39)$$

$$\begin{aligned} \mu_3(Z_i) = & \sum_{i=1}^n \left[\frac{\partial g(V_{\text{om},1})}{\partial V_i} \right]^3 \mu_3(V_i) \\ & + \frac{3}{2} \sum_{i=1}^n \left[\frac{\partial g(V_{\text{om},1})}{\partial V_i} \right]^2 \frac{\partial^2 g(V_{\text{om},1})}{\partial V_i^2} (\mu_4(V_i) - 3\mu_2^2(V_i)) + 3 \sum_{i=1}^n \\ & \times \sum_{j=1}^n \frac{\partial g(V_{\text{om},1})}{\partial V_i} \frac{\partial g(V_{\text{om},1})}{\partial V_j} \frac{\partial^2 g(V_{\text{om},1})}{\partial V_i \partial V_j} \mu_2(V_i) \mu_2(V_j), \end{aligned} \quad (40)$$

$$\mu_4(Z_i) = \sum_{i=1}^n \left[\frac{\partial g(V_{om,1})}{\partial V_i} \right]^4 (\mu_4(V_i) - 3\mu_{X_{2,i}}^2) + 3 \sum_{i=1}^n \sum_{j=1}^n \left[\frac{\partial g(V_{om,1})}{\partial V_i} \right]^2 \left[\frac{\partial g(V_{om,1})}{\partial V_j} \right]^2 \mu_2(V_i) \mu_2(V_j), \quad (41)$$

We change the moments of the voltage amplitude into the moments of voltage amplitude constraints *via* (38)-(41).

The following formulas no longer limit the number of buses, and a portion of Z is removed. We transform moments to other digital features.

$$Z_{\text{standard}} = \sqrt{E(Z)}, \quad (42)$$

$$csZ = \frac{\mu_3(Z)}{Z_{\text{standard}}^3}, \quad (43)$$

$$ckZ = \frac{\mu_4(Z)}{Z_{\text{standard}}^4}, \quad (44)$$

where csz is the skewness coefficient and cks is the kurtosis coefficient.

We transform Z to a standard normal vector Y , and the moments of Y are calculated with

$$\mu_j(Z) = \sigma^j(Z)E(Y^j), \quad (45)$$

where Y is a standard normal vector and $\sigma(\bullet)$ is the standard deviation function.

From Eq. 42, we can obtain the moments of Y , as follows:

$$\mu_3(Z) = csZ, \quad (46)$$

$$\mu_4(Z) = ckZ, \quad (47)$$

After the second step in the POME approach with four moments, we obtain

$$\int_{-\infty}^{\infty} y^j pdf(y) dy = \mu_j(Y), \quad (48)$$

$$pdf(y) = \exp\left(-\sum_{j=0}^m \lambda_j y^j\right) \quad (49)$$

where (48) and (17) have the same form but different physical meanings. By inserting (46), (47), and (49) into Eq. 48, λ_k can be estimated.

P_f can be obtained using $pdf(y)$:

$$P_f = P(z \leq 0) = \int_{-\infty}^{-\frac{E(Z)}{Z_{\text{standard}}}} pdf(y) dy, \quad (50)$$

By inserting Eq. 50 into Eqs 6, 35 can be estimated. The voltage constraints for each node in the distribution network can be solved according to the above formula. Now, both 6) and 11) can be adjusted for a set of specific solutions. PSO in (Fu et al., 2020) is used to find the minimum to 11) while satisfying 6) by searching the solutions of Eqs 2-5.

Remark 4: The voltage probability information is transformed into voltage constraints *via* the QFM approach, which is a new way to apply PPFs to obtain constraint solutions in stochastic programming.

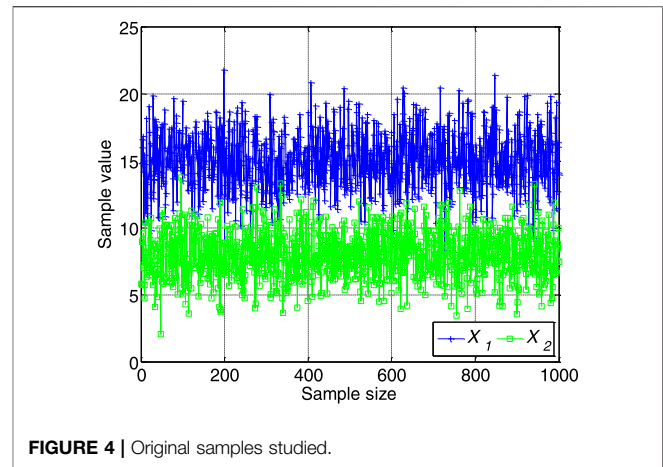


FIGURE 4 | Original samples studied.

SIMULATION

Verification is performed in a step-by-step manner in this study, and the specific examples are as follows. Case 1 verifies the feasibility of estimating confidence levels using central moments from the perspective of pure mathematics, and we show that the QFM approach is suitable for solving the constraint satisfaction probability problem. Case 2 verifies the correctness of the proposed method for estimating different forms of PPF results. Case 3 uses the power quality qualification as a constraint to verify the effectiveness of the proposed method in stochastic programming for distribution networks.

Case 1

The result of the stochastic programming constraint function is a confidence level. The index accompanying the confidence level is the failure probability, which is a common reliability index, and the sum of the two is equal to 1. In using different methods to estimate the variable disqualification probability (i.e., failure probability), we make the following assumptions.

- 1) \mathbf{X}_1 and \mathbf{X}_2 are two uncertain variables, both of which obey normal distributions. The reason why we assume a normal distribution is that there are many algorithms that are based on the normal distribution; thus, the effectiveness of the proposed method can be verified. It should be noted that the normal distribution does not have to be used based on the use conditions of the proposed method.
- 2) We assume that the limit state equation with independent variables \mathbf{X}_1 and \mathbf{X}_2 is

$$\mathbf{Z} = \mathbf{X}_1 - \mathbf{X}_2. \quad (51)$$

- 3) The original sample set is shown in Figure 4, and the number of samples is one thousand.

We compared the proposed method (QFM estimation method) with classic, mature and effective methods that have been previously published, including the Iman method in (Iman

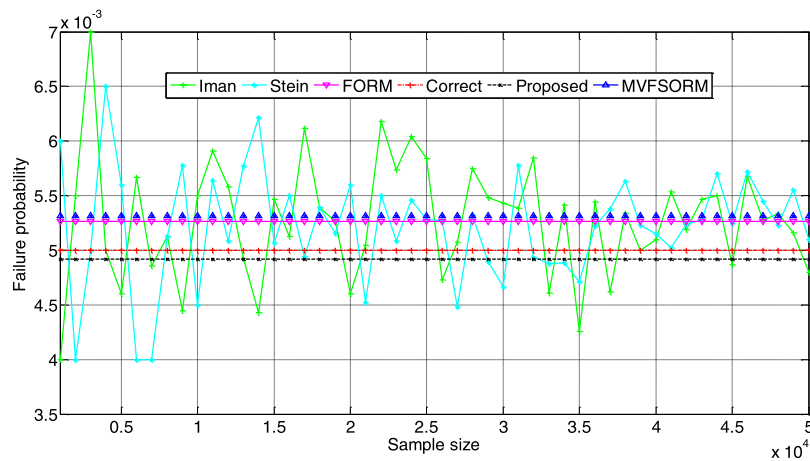


FIGURE 5 | Failure probability calculated with different algorithms.

TABLE 1 | Input probability information for the different algorithms.

Algorithm	Probability distribution	Probability information
Iman	Normal	mean, standard deviation, correlation coefficients
Stein	Normal	mean, standard deviation, correlation coefficients
FORM	known	mean, standard deviation, correlation coefficients
MVFSORM	unknown	mean, standard deviation
QFM (Proposed)	unknown	mean, standard deviation skewness, kurtosis

and Conover, 1982), the Stein method in (Stein, 1987), the first-order reliability method (FORM) in (Sankararaman et al., 2014), and the mean value first-order second moment (MVFSORM) method in (Du and Huang, 2008), as references to verify the proposed method. Considering the need for sampling, the Iman and Stein methods are applied with step size of 100–5,000. The simulation results are shown in **Figure 5**.

The results of different algorithms vary. Compared with those of other algorithms, the results of the proposed method are closest to the real values. The simulation results are related to the information that each algorithm considers and the algorithm design. The usage conditions and input probability information for the different algorithms are shown in **Table 1**.

It should be noted that the central moments can be calculated from the mean, standard deviation, skewness, and kurtosis. We can also obtain central moments using origin moments. The mean is an origin moment, and another origin moment (standard deviation, which is the arithmetic square root of the variance) is the second central moment.

The mean and standard deviation are the first two moments. Both MVFSORM and the proposed method do not use a probability distribution function, so they are not limited by a certain probability distribution and have high practical value.

More probability information is used when considering four moments than when considering two moments, so the proposed method based on four moments is more accurate than the MVFSORM method based on two moments. Compared with

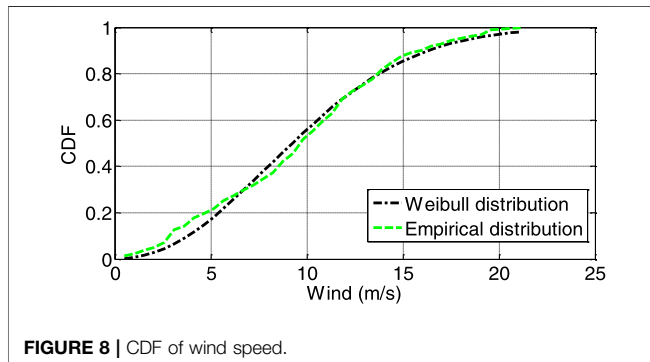
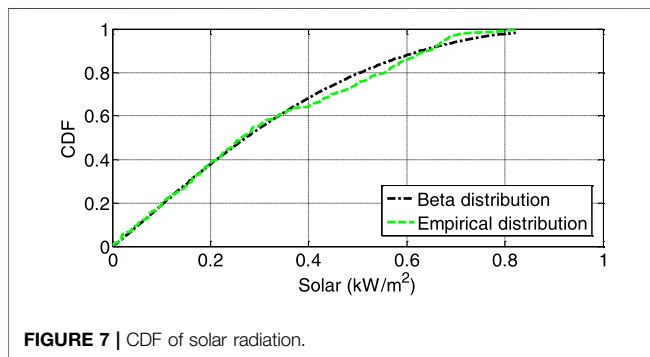
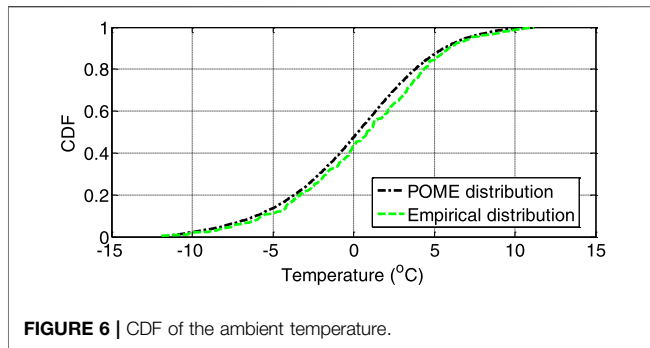
the Iman, Stein, and FORM methods, the proposed method does not need to solve the probability distribution function. In addition, the Iman and Stein methods are only suitable for the normal distribution, but the proposed method does not have this limitation. When the probability distribution is unknown, the proposed method can still determine the reliability or chance constraint level.

Discussion 1: This case shows how to transform digital features into unqualified probabilities, and qualified probabilities are used in the constraint conditions in stochastic programming. The typical results of PPF calculations are the mean and variance, which are first-order and second-order moments. However, the above results indicate that it is not appropriate to rely only on the mean and variance to estimate unqualified probabilities. Specifically, it is necessary to use additional central moments from the PPF results to estimate the probability of constraint satisfaction in stochastic programming.

Case 2

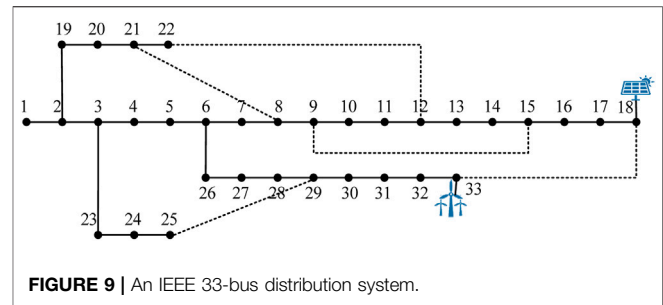
The main advantage of the proposed method is its computational efficiency. The computing time depends on both the efficiency of the algorithm and the hardware configuration of the computer. The simulations in cases 2 and 3 were performed on a ThinkBook with an Intel(R) Core (TM) i7-1065G7 CPU @ 1.50 GHz and 15.7 GB of available memory.

POME is used to describe the cumulative distribution function (CDF) of the ambient temperature, which affects



the PV cell temperature. We calculate the Lagrange multipliers of the maximum entropy probability density functions from the knowledge of the four-moment constraints (i.e., the first, second, third and fourth central sample moments of the values at a given ambient temperature). λ is a vector containing the resulting Lagrange parameters, and the λ vector in this case is $[2.4003, -0.0447, 0.0292, 0.0008, -0.0001]$. The solar radiation that reaches a PV array is simulated using a beta distribution. The maximum likelihood estimates of the parameters of the beta distribution for solar radiation are $\alpha = 1.1112$ and $\beta = 2.4690$. The maximum likelihood estimates for the parameters of the Weibull distribution for wind speed are as follows: scale parameter = 1.1112 and shape parameter = 2.4690.

As shown in **Figures 6–8**, the POME, beta and Weibull distributions are very similar to the empirical distributions, so



the simulated ambient temperature, solar radiation and wind speed values are generally accurate. We use a Gaussian copula to model the correlation among the three weather CDFs. The MC method is used for sampling, and the sample size for sets of weather variables is 10,000.

The IEEE 33-bus distribution system from (Baran and Wu, 1989) is used to demonstrate the proposed method. We assume that both the PV and WP rated capacities are 2,000 kW, and the access buses are shown in **Figure 9**.

Figures 10, 11 cannot be used to verify the appropriateness of the PEM theory but do verify the suitability of the point PEM formula and the corresponding programming method in this paper. The proposed PEM method provides results that can support subsequent simulation analyses. The MC simulation takes 266.829 s, and the PEM simulation takes 0.147 s; thus, the disadvantage of MC is the comparatively longer computing time.

&& stands for logical AND. It should be noted that the proposed method refers to PEM&&QFM. In essence, the proposed method transforms the digital characteristics of PEMs to other digital characteristics that are suitable for stochastic programming. The purpose of assessing the simulation results of MC&&QFM is to verify the accuracy of the QFM approach and the proposed method. As shown in **Figure 12**, the voltage qualification probability of MC&&QFM is almost the same as that of MC&&Statistics (i.e., direct statistical qualification probability for the voltage amplitude), and it can be used as a reference standard. The voltage qualification probability of PEM&&QFM is close to that of MC&&Statistics.

As shown in **Figure 13**, although the PEM result is very similar to the MC result, there are still some errors associated with the mean value and the second, third and fourth central moments of the voltage amplitude. The errors in the voltage amplitude moments lead to differences in the qualified probability of the voltage amplitude in the MC&&QFM and PEM&&QFM cases, as shown in **Figure 12**.

Discussion 2: We use the maximum entropy principle to transform the quadratic fourth-order moment of the voltage amplitude into probability density function (PDF) information and then solve for the voltage qualification probability. The most important contribution of this paper is the novel PPF results, which include the mean loss and the probability of voltage qualification. These results provide the basis for the stochastic planning of distribution networks

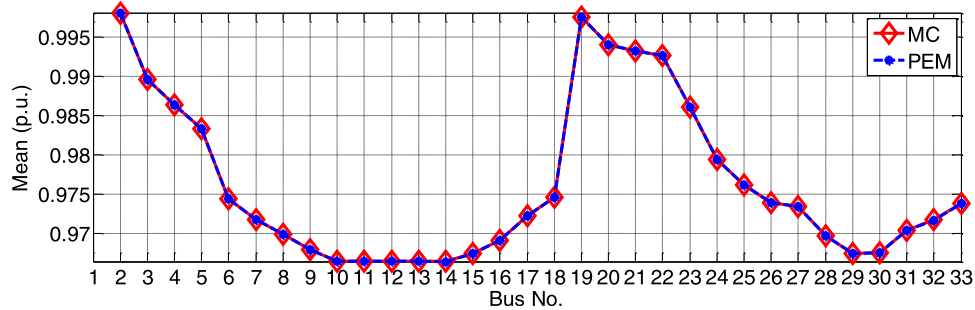


FIGURE 10 | Mean value of the elements from bus voltage profiles.

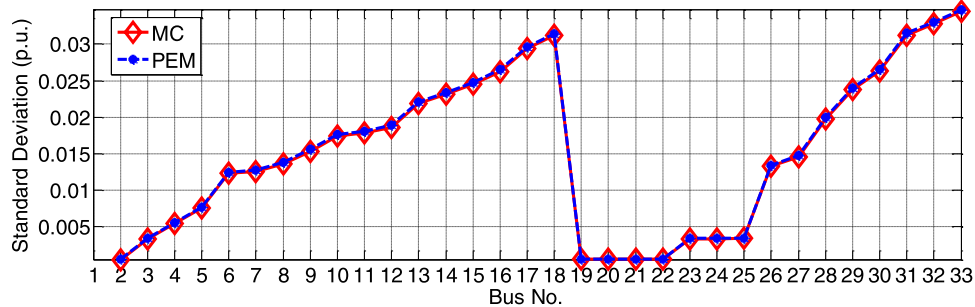


FIGURE 11 | Standard deviation of the elements from bus voltage profiles.

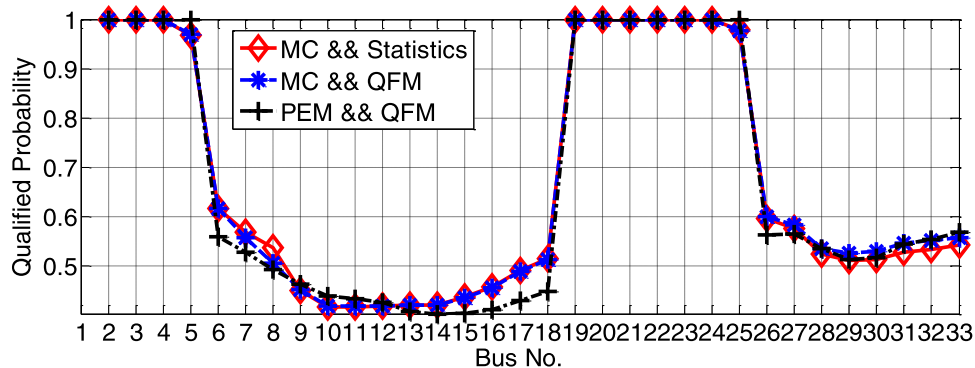


FIGURE 12 | Simulation results for the qualified probability of the voltage amplitude.

because traditional PPF results cannot be used directly in stochastic programming.

Case 3

What we want to verify in this case is the efficiency of achieving a feasible solution to the objective function and constraint functions in programming. To verify the effectiveness of the proposed method, we compare the accuracy of some specific solutions, as listed in **Table 2**. We assume that the rated

capacities of PV and WP are the same, and the access buses are shown in **Figure 9**.

It should be noted that one hundred percent accuracy is logically impossible for an efficient approximation solution algorithm. Stochastic optimal planning allows for certain error margins, and the error problem in power grid dispatching can be solved with this relative approach. It can be concluded that the proposed method supports stochastic programming in distribution networks. The simulation results using PSO are

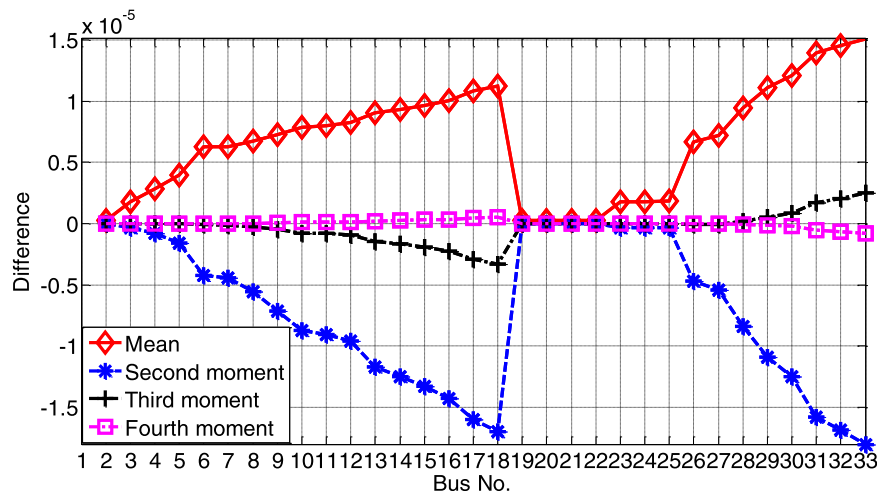


FIGURE 13 | Digital characteristic error of voltage variables.

TABLE 2 | Input simulation results for the objective and constraint functions.

Item		Solution 1	Solution 2	Solution 3	Solution 4	Solution 5
f_{obj}	Y_{PV} (kW)	10,000	10,500	11,000	11,500	12,000
	MC (MW)	0.501	0.548	0.598	0.650	0.704
	Proposed (MW)	0.503	0.551	0.601	0.653	0.708
α	MC	0.836	0.844	0.853	0.866	0.866
	Proposed	0.839	0.843	0.847	0.850	0.853

TABLE 3 | Optimum planning results.

Item	Value
$[Y_{PV}, Y_{WP}]$	$[10^4 \text{ kW}, 10^4 \text{ kW}]$
V_-	0.95 p.u.
α	0.82
Number of particles	20
Number of iterations	10
$[num_{PV}, num_{WP}]$	Buses 8 and 30
$[Y_{PV}, Y_{WP}]$	$[478 \text{ kVAr}, 935 \text{ kVAr}]$
f_{obj}	0.2425 MW
Solution time	4,989 s

listed in Table 3. Compared with the values in Table 2, the values in Table 3 are better.

Discussion 3: We consider mean loss in the objective function in stochastic programming, and the probability of a constraint meeting a certain confidence threshold is a programming constraint based on a quadratic fourth-order moment. Our goal is not to improve the programming algorithm but to improve the efficiency of estimating the probabilistic numerical characteristics of objective functions and constraint functions corresponding to feasible solutions.

The superiorities of the proposed method can be summarized as follows. 1) The results of case 1 show that

the proposed QFM method are more precise than other algorithms, and it is not limited by a certain probability distribution. 2) The results of case 2 show that the proposed PEM method spend much less time than MC, but PEM can achieve accuracy levels equivalent to MC. The proposed method can save more than 266 s 3) The results of case 3 show that PPF can be used in uncertainty planning of distributed renewable energy sources in distribution networks. It takes only 4,989 s for optimization solution, and the importance of simplifying the planning model is verified. Compared with the traditional programming methods, we improve the uncertainty planning model rather than the optimization solver.

CONCLUSION

The mean, variance and probability distribution function are commonly obtained from PPF calculations. However, the conventional PPF calculation results cannot be used to directly solve many actual operation and planning problems. The PPF results in this paper are central moments that are used to model objective and constraint functions in stochastic programming and provide a feasible way to address uncertainty problems in smart grid planning.

The essence of uncertain planning is to achieve optimal economic performance under safety constraints. It is reasonable to use different statistical machine learning theories such as PPF, PEM and QFM to obtain probability values for uncertainty planning of distributed renewable energy sources.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

REFERENCES

- Baran, M. E., and Wu, F. F. (1989). Network Reconfiguration in Distribution Systems for Loss Reduction and Load Balancing. *IEEE Trans. Power Deliv.* 4 (2), 1401–1407. doi:10.1109/61.25627
- Chatthaworn, R., and Chaitusaney, S. (2015). Improving Method of Robust Transmission Network Expansion Planning Considering Intermittent Renewable Energy Generation and Loads. *IET Generation, Transm. Distribution* 9 (13), 1621–1627. doi:10.1049/iet-gtd.2015.0363
- Chen, Z., Shi, G., Li, Y., and Fu, X. (2020). “Optimal Planning Method for a Multi-Energy Complementary System with New Energies Considering Energy Supply Reliability,” in 2020 Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 29–31 May 2020, 952–956. doi:10.1109/aeecs48850.2020.9121425
- Du, X., and Huang, B. (2008). Probabilistic Uncertainty Analysis by Mean-Value First Order Saddlepoint Approximation. *Reliability Eng. Syst. Saf.* 93 (2), 325–336. doi:10.1016/j.res.2006.10.021
- Fu, X., Guo, Q., and Sun, H. (2020). Statistical Machine Learning Model for Stochastic Optimal Planning of Distribution Networks Considering a Dynamic Correlation and Dimension Reduction. *IEEE Trans. Smart Grid* 11 (4), 2904–2917. doi:10.1109/tsg.2020.2974021
- Gan, L., Li, G., and Zhou, M. (2016). Coordinated Planning of Large-Scale Wind Farm Integration System and Transmission Network. *CSEE Power Eng. Syst.* 2 (1), 19–29. doi:10.17775/cseejpes.2016.00005
- Haghighat, H., and Zeng, B. (2018). Stochastic and Chance-Constrained Conic Distribution System Expansion Planning Using Bilinear Benders Decomposition. *IEEE Trans. Power Syst.* 33 (3), 2696–2705. doi:10.1109/tpwrs.2017.2751514
- Iman, R. L., and Conover, W. J. (1982). A Distribution-free Approach to Inducing Rank Correlation Among Input Variables. *Commun. Stat. - Simulation Comput.* 11 (3), 311–334. doi:10.1080/03610918208812265
- Injeti, S., and Thunuguntla, V. (2020). Optimal Integration of DGs into Radial Distribution Network in the Presence of Plug-In Electric Vehicles to Minimize Daily Active Power Losses and to Improve the Voltage Profile of the System Using Bioinspired Optimization Algorithms. *Prot. Control. Mod. Power Syst* 5 (1), 21–35. doi:10.1186/s41601-019-0149-x
- Jooshaki, M., Farzin, H., Abbaspour, A., Fotuhi-Firuzabad, M., and Lehtonen, M. (2020). A Model for Stochastic Planning of Distribution Network and Autonomous DG Units. *IEEE Trans. Ind. Inf.* 16 (6), 3685–3696. doi:10.1109/tii.2019.2936280
- Karaki, S. H., Chedid, R. B., and Ramadan, R. (1999). Probabilistic Performance Assessment of Autonomous Solar-Wind Energy Conversion Systems. *IEEE Trans. Energy Convers.* 14 (3), 766–772. doi:10.1109/60.790949
- Li, Z., Wang, J., Sun, H., Qiu, F., and Guo, Q. (2019). Robust Estimation of Reactive Power for an Active Distribution System. *IEEE Trans. Power Syst.* 34 (5), 3395–3407. doi:10.1109/tpwrs.2019.2902136
- Lin, C., Bie, Z., Pan, C., and Liu, S. (2020). Fast Cumulant Method for Probabilistic Power Flow Considering the Nonlinear Relationship of Wind Power Generation. *IEEE Trans. Power Syst.* 35 (4), 2537–2548. doi:10.1109/tpwrs.2019.2959404
- Liu, C., Lee, C., Chen, H., and Mehrotra, S. (2016). Stochastic Robust Mathematical Programming Model for Power System Optimization. *IEEE Trans. Power Syst.* 31 (1), 821–822. doi:10.1109/tpwrs.2015.2394320

AUTHOR CONTRIBUTIONS

XF organized case studies and wrote the manuscript. XW and NL contributed to the theoretical research of this paper. All authors contributed to manuscript, and approved the submitted version.

FUNDING

This study is supported by the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (Grant No. LAPS21016), the National Natural Science Foundation of China (Grant No. 52007193), and The 2115 Talent Development Program of China Agricultural University.

- Mohammad-Djafari, A. (1992). “A Matlab Program to Calculate the Maximum Entropy Distributions,” in *Maximum Entropy and Bayesian Methods*. Editors C.R. Smith, G.J. Erickson, and P.O. Neudorfer (New York, NY, USA: Springer), 221–233. doi:10.1007/978-94-017-2219-3_16
- Niu, T., Guo, Q., Jin, H., Sun, H., Zhang, B., and Liu, H. (2017). Dynamic Reactive Power Optimal Allocation to Decrease Wind Power Curtailment in a Large-Scale Wind Power Integration Area. *IET Renew. Power Generation* 11 (13), 1667–1678. doi:10.1049/iet-rpg.2017.0144
- Park, H., Baldick, R., and Morton, D. P. (2015). A Stochastic Transmission Planning Model with Dependent Load and Wind Forecasts. *IEEE Trans. Power Syst.* 30 (6), 3003–3011. doi:10.1109/tpwrs.2014.2385861
- Ren, Z., Wang, K., Li, W., Jin, L., and Dai, Y. (2017). Probabilistic Power Flow Analysis of Power Systems Incorporating Tidal Current Generation. *IEEE Trans. Sustain. Energ.* 8 (3), 1195–1203. doi:10.1109/tste.2017.2669139
- Rohani, G., and Nour, M. (2014). Techno-Economical Analysis of Stand-Alone Hybrid Renewable Power System for Ras Musherib in United Arab Emirates. *Energy* 64, 828–841. doi:10.1016/j.energy.2013.10.065
- Roldán, C., Mínguez, R., García-Bertrand, R., and Arroyo, J. M. (2019). Robust Transmission Network Expansion Planning under Correlated Uncertainty. *IEEE Trans. Power Syst.* 34 (3), 2071–2082. doi:10.1109/TPWRS.2018.2889032
- Sankararaman, S., Daigle, M. J., and Goebel, K. (2014). Uncertainty Quantification in Remaining Useful Life Prediction Using First-Order Reliability Methods. *IEEE Trans. Rel.* 63 (2), 603–619. doi:10.1109/tr.2014.2313801
- Shore, J., and Johnson, R. (1980). Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Trans. Inform. Theor.* 26 (1), 26–37. doi:10.1109/tit.1980.1056144
- Stein, M. (1987). Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics* 29 (2), 143–151. doi:10.1080/00401706.1987.10488205
- Tang, J., Ni, F., Ponci, F., and Monti, A. (2016). Dimension-Adaptive Sparse Grid Interpolation for Uncertainty Quantification in Modern Power Systems: Probabilistic Power Flow. *IEEE Trans. Power Syst.* 31 (2), 907–919. doi:10.1109/tpwrs.2015.2404841
- Zhang, C., Chen, H., Liang, Z., Guo, M., Hua, D., and Ngan, H. (2018). Reactive Power Optimization under Interval Uncertainty by the Linear Approximation Method and its Modified Method. *IEEE Trans. Smart Grid* 9 (5), 4587–4600. doi:10.1109/tsg.2017.2664816
- Zhang, C., Chen, H., and Ngan, H. (2016). Reactive Power Optimisation Considering Wind Farms Based on an Optimal Scenario Method. *IET Generation, Transm. Distribution* 10 (15), 3736–3744. doi:10.1049/iet-gtd.2016.0036
- Zhang, C., Chen, H., Shi, K., Liang, Z., Mo, W., and Hua, D. (2019). A Multi-Time Reactive Power Optimization under Interval Uncertainty of Renewable Power Generation by an Interval Sequential Quadratic Programming Method. *IEEE Trans. Sustain. Energ.* 10 (3), 1086–1097. doi:10.1109/tste.2018.2860901
- Zhang, K., Zhou, B., Or, S. W., Li, C., Chung, C. Y., and Voropai, N. I. (2021). Optimal Coordinated Control of Multi-Renewable-To-Hydrogen Production System for Hydrogen Fueling Stations. *IEEE Trans. Ind. Appl.*, 1. doi:10.1109/TIA.2021.3093841 Available at: <https://ieeexplore.ieee.org/document/9468931/authors#authors>.

- Zhao, Y., Guo, Y., Guo, Q., Zhang, H., and Sun, H. (2020). Deployment of the Electric Vehicle Charging Station Considering Existing Competitors. *IEEE Trans. Smart Grid* 11 (5), 4236–4248. doi:10.1109/tsg.2020.2991232
- Zimmerman, R. D., Murillo-Sanchez, C. E., and Thomas, R. J. (2011). MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Trans. Power Syst.* 26 (1), 12–19. doi:10.1109/tpwrs.2010.2051168

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Fu, Wu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.