# Data-Driven Chance-Constrained Planning for Distributed Generation: A Partial Sampling Approach

Shiyi Jiang; Jianqiang Cheng, *Member, IEEE*; Kai Pan, *Member, IEEE*;
Feng Qiu, *Senior Member, IEEE*; and Boshi Yang.

***Abstract*—The planning of distributed energy resources has been challenged by the significant uncertainties and complexities of distribution systems. To ensure system reliability, one often employs chance-constrained programs to seek a highly likely feasible solution while minimizing certain costs. The traditional sample average approximation (SAA) is commonly used to represent uncertainties and reformulate a chance-constrained program into a deterministic optimization problem. However, the SAA introduces additional binary variables to indicate whether a scenario sample is satisfied and thus brings great computational complexity to the already challenging distributed energy resource planning problems. In this paper, we introduce a new paradigm, i.e., the partial sample average approximation (PSAA) using real data, to improve computational tractability. The innovation is that we sample only a part of the random parameters and introduce only continuous variables corresponding to the samples in the reformulation, which is a mixed-integer convex quadratic program. Our extensive experiments on the IEEE 33-Bus and 123-Bus systems show that the PSAA approach performs better than the SAA because the former provides better solutions in a shorter time in in-sample tests and provides better guaranteed probability for system reliability in out-of-sample tests. All the data used in the experiments are real data acquired from Pecan Street Inc. and ERCOT. More importantly, our proposed chance-constrained model and PSAA approach are general enough and can be applied to solve other valuable problems in power system planning and operations. Thus, this paper fits one of the journal scopes: *Distribution System Planning in Power System Planning and Implementation*.**

***Index Terms*—Planning, distributed energy resources, renewable distributed generation, energy storage, stochastic programming, chance-constrained programming, data-driven**

## I. Nomenclature

### A. Indices and Sets

$n/k/r$ Index of buses/ renewable distributed generation (RDG) units/ energy storage (ES) units.

$\pi/\kappa/\tau$ Index of different data samples.

$\mathcal{N}/\mathcal{E}/\mathcal{N}_n$ Set of all the buses/ all the power distribution lines/ the buses connected to a given Bus $n \in \mathcal{N}$, i.e., $\mathcal{N}_n := \{m \mid (n,m) \in \mathcal{E}\}$.

$\mathcal{B}_0/\mathcal{B}_1/\mathcal{B}_2$ Set of buses that are connected to Bus 0/ installed with dispatchable distributed generation (DDG) units/ installed with reactive power sources.

$\mathcal{X}$ Set of standard capacities of RDG units, i.e., $\mathcal{X} := \{\bar{x}_1, \ldots, \bar{x}_L\}$.

$[K]$ $\{1, 2, \ldots, K\}$, for any $K \in \mathbb{Z}_+$.

### B. Parameters

$c^0_{kn}/d^0_{rn}$ Setup cost of placing the $k^{th}$ RDG unit/ the $r^{th}$ ES unit at Bus $n$.

$c^1_k/c^2_k$ Size-based investment/ maintenance cost of the $k^{th}$ RDG unit.

$d^1_r/d^2_r$ Size-based investment/ maintenance cost of the $r^{th}$ ES unit.

$c^t_p/c^t_q$ Electricity price of purchasing active/reactive power from the main grid via Bus 0.

$c^f_n/c^e_n$ Fuel/emission price for the DDG units at Bus $n$.

$\omega$ Emission factor of the DDG units (kg/kWh).

$K/R/T/N$ Total number of available RDG units/ available ES units/ all the time intervals in the planning horizon/ all the buses.

$\bar{K}$ Maximum number of RDG units to be installed.

$LC_{mn}$ Capacity of a power distribution line $(m,n) \in \mathcal{E}$.

$(\bar{p}^t_n, \underline{p}^t_n)$ Active power output bounds of DDG unit $n$ in period $t$.

$(\bar{q}^t_n, \underline{q}^t_n)$ Reactive power output bounds of reactive power source $n$ in period $t$.

$\mathfrak{R}_{mn}/\mathfrak{X}_{mn}$ Electrical resistance/reactance of line $(m,n)$.

$\delta_n/\tau_n$ Binary indicator of whether a DDG unit/ a reactive power source is at Bus $n$.

$\underline{v}/\bar{v}$ Upper/lower bound of voltage magnitude at a bus.

$\bar{y}_r/\underline{y}_r$ Maximum/minimum capacity of the $r^{th}$ ES unit.

$e_1/e_2$ ES charging/discharging unit cost.

$\eta/\gamma/b^0_r$ Violation probability/ ES efficiency/ Initial power level of the $r^{th}$ ES unit.

$\Pi_1/\Pi_2/\Pi_3$ Total number of data samples of different types.

### C. Random Variables

$d^t_{pn}/d^t_{qn}$ Active/reactive load at Bus $n$ in period $t$.

$s^t_k$ Active power output efficiency of the $k^{th}$ RDG unit in period $t$.

$\boldsymbol{\xi}^t$ Vector of uncertainty in compact form in period $t$, i.e., $[d^t_{p1}, \ldots, d^t_{pN}, d^t_{q1}, \ldots, d^t_{qN}, s^t_1, \ldots, s^t_k]^\top$.

### D. Decision Variables

$z_{kn}/u_{kl}/w_{rn}$ Binary indicator of whether the $k^{th}$ RDG unit is located at Bus $n$/ of whether the capacity of the $k^{th}$ RDG unit is the $l^{th}$ element in $\mathcal{X}$/ of whether the $r^{th}$ ES unit is located at Bus $n$.

$x_k/y_r$ Size of the $k^{th}$ RDG unit/ Capacity of the $r^{th}$ ES unit.

$f^t_r/g^t_r$ Active power that is charged/ discharged at the $r^{th}$ ES unit in period $t$.

Shiyi Jiang and Kai Pan are with the Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong. E-mails: shiyi.jiang@connect.polyu.hk and kai.pan@polyu.edu.hk. Jianqiang Cheng is with the Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ. E-mail: jqcheng@email.arizona.edu. Boshi Yang is with the School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC. E-mail:boshiy@clemson.edu. Feng Qiu is with the Energy Systems Division, Argonne National Laboratory, Lemont, IL. E-mail: fqiu@anl.gov.

$p_0^t/q_0^t$ Active/ reactive power purchased from the main grid via Bus 0 in period $t$.
$P_{mn}^t/Q_{mn}^t$ Active/ reactive power flow from Bus $m$ to $n$ in period $t$.
$V_n^t/|V_n^t|^2$ Complex voltage at Bus $n$ in period $t$/ its magnitude.
$I_{mn}^t/|I_{mn}^t|^2$ Complex current from Bus $m$ to $n$ in period $t$/ its magnitude.
$p_n^t/q_n^t$ Active/ reactive power output of the DDG unit/ reactive power source at Bus $n$ in period $t$.
$b_r^t$ Active power of the $r^{th}$ ES unit in period $t$.
$LS_{1mn}^t/LS_{2mn}^t$ Load-shedding variables.
$\boldsymbol{z}$ $[z_{kn}, \forall\, k \in [K], n \in \mathcal{N}]^\top$.
$\boldsymbol{x},\ \boldsymbol{u}$ $[x_1, \ldots, x_K]^\top$, $[u_{kl}, \forall\, k \in [K], l \in [L]]^\top$
$\boldsymbol{y}, \boldsymbol{w}$ $[y_1, \ldots, y_R]^\top$, $[w_{rn}, \forall\, r \in [R], n \in \mathcal{N}]^\top$.

## II. Introduction

With technological development and governments' support, renewable energy has drawn significant attention and investment. A widely used strategy to exploit renewable energy is to integrate renewable distributed generation (RDG) units into existing power distribution grids. Correct installation of RDG units can help power distribution grids provide customers with affordable and reliable energy, while improper placement may result in many problems, e.g., system instability and power losses [1], due to the intermittency of renewable energy. Such intermittency from RDG units may lead to cascading problems such as an imbalance of electricity supply and demand and system blackouts [2]. Therefore, proper siting (i.e., location) and sizing (i.e., capacity) decisions of the RDG units are of great significance to ensure the benefits of renewable energy and maintain reliable operations of the power distribution grids.

Besides RDG units, energy storage (ES) has also been considered for use in power distribution grids. It is because ES units can provide a buffer against an imbalance of supply and demand, thereby reducing operating costs and increasing a power distribution grid's probability of meeting demand. Such benefits may offset the installation and operating costs of ES units and even lead to profits. For example, [3] adopts ES to help support wind energy applications and [4] uses ES to increase the penetration of more general renewable generation by smoothing out the effects of intermittency. The positive results from these studies indicate the necessity to integrate ES units into a power distribution grid with RDG units. Thus, it is important to determine the optimal siting and sizing decisions of both the RDG and ES units.

For ease of exposition, we refer to the problem of siting and sizing both RDG and ES units in a power distribution grid as *the planning problem* thereafter. Such a problem largely relies on accurate power flow analysis. Two mathematical models are often considered to calculate the optimal power flow in power grids: the direct current optimal power flow (DCOPF) and the alternating current optimal power flow (ACOPF) models. The DCOPF model is composed of linear constraints and thus is easier to solve, while it oversimplifies the physical features. In contrast, the ACOPF model is relatively accurate by considering active and reactive power-generation limits, demand limits, bus voltage limits, and network flow limits [5], while it is nonlinear and nonconvex. The DCOPF model provides a linear approximation of the ACOPF model, with systematic errors though. Such inaccuracy is acceptable for large-scale power networks but is often unacceptable for local distribution grids. Therefore, this paper uses the ACOPF model to accurately simulate a power distribution grid.

Different variants of the planning problem have recently drawn much attention from academia and industry. The paper [6] provides a review of related studies on the RDG planning in the power distribution network, and [7] reviews related studies on more general generation expansion planning. Specifically, [8] plans the locations and capacity sizes of RDG units based on simplified load flow calculation in a multiobjective optimization model that is further solved by a genetic algorithm, while ES units are not considered. Similarly, without ES units included, [9] considers the siting and sizing of RDG units in a power distribution grid in a two-stage robust optimization model. The paper [10] makes the siting and sizing decisions for the ES units in a power transmission grid via a three-stage mixed-integer linear program, while the DCOPF model is adopted. A recent work can be found in [11], which considers only the sizing of RDG units in a two-stage distributionally robust optimization model and makes the siting decisions via sensitivity analyses. The above existing efforts demonstrate the significance of siting and sizing RDG and ES units.

However, the large-scale installation of RDG units adds significant uncertainties to a power distribution grid due to the intermittency of renewable energy, requiring methodological innovation to deal with the already challenging operations of a power distribution grid. To that end, many stochastic programming models dealing with uncertainties have been developed to support effective distribution grid operations. For example, [12] proposes a two-stage stochastic programming model for the optimal planning of distributed energy systems under demand and supply uncertainties. The paper [13] proposes a two-stage stochastic programming-based optimal power flow model for the operation of distribution networks with uncertainties from wind power.

In this paper, we adopt chance-constrained programming to model the planning and operational decisions under uncertainties, including renewable generation and load uncertainties, to ensure the feasibility of the distribution system under a high probability. Specifically, we note that, due to such uncertainties, a power distribution grid may face various reliability issues. For instance, power outages often happen when the power supply is insufficient, such as facing a natural disaster like a typhoon. A load bus that is far away from the upper stream grid may not receive the power injected from the upper stream grid or local distributed generators because of distribution line capacity limit and line loss. The distribution line contingency may also happen. Therefore, to help relieve such pressure and ensure system reliability, we build a two-stage chance-constrained (TCC) model (where the siting and sizing decisions are in the first stage and operational decisions are in the second stage) to solve *the planning problem*. More importantly, considering various reliability issues, we adopt

a chance constraint to ensure *all* the operational constraints in the second stage to be satisfied simultaneously, rather than ensuring load satisfaction only. As such, our proposed TCC model becomes extremely difficult to solve, requiring further innovation in solution approaches.

We note that the chance-constrained model is a risk-averse decision-making tool that can help grid operators actively control the probability of unfavorable outcomes (e.g., system blackouts). The chance-constrained model is widely used for power system operations. For instance, [14] solves the chance-constrained ACOPF problems to ensure that operational constraints are satisfied with the desired probability. The paper [15] further provides convex approximations via second-order conic programming, and [16] provides asymptotically tight conic approximations for the chance-constrained ACOPF problems. Similarly, [17] studies the chance-constrained unit commitment formulations and [18] investigates the chance-constrained day-ahead scheduling. Several studies use TCC models, which are much more complex than single-stage chance-constrained models. For instance, [19] is among the first to propose a TCC model for the unit commitment problem, while it considers a single chance constraint to ensure load satisfaction only. The paper [20] considers a similar TCC model for the unit commitment problem while presenting a bilinear mixed-integer reformulation solved by Benders decomposition following the study in [21]. The paper [22] formulates the chance constraints based on the definition of conditional value at risk in a TCC model for the unit commitment problem and reformulates these constraints using sampling-based approaches. The paper [23] specifically considers wind uncertainty in the chance-constrained unit commitment model. However, to the best of our knowledge, few studies apply the TCC models in the planning of RDG and ES units together in a distribution system considering the ACOPF model. More importantly, different from the existing studies above, this paper considers all the operational constraints simultaneously in a difficult joint chance constraint.

Nevertheless, the TCC model is intractable in general, specifically when the random parameters follow an underlying continuous (yet unknown) probability distribution and are of high dimensions, posing severe computational challenges [24]. The existing studies primarily adopt two approximation approaches to solve a chance-constrained model: a convex (or tractable) approximation approach [15] and a sampling-based approach [19]. The former is not applicable to a TCC model because the second-stage recourse decision is a function of the first-stage decision and random parameters and it is impossible to algebraically characterize this function. Thus, this paper uses the latter. Specifically, we propose two sampling techniques to reformulate our TCC model: the standard sample average approximation (SAA) method and a new partial sample average approximation (PSAA) method. Both approximations lead to mixed-integer convex quadratic programs. However, the SAA introduces many additional binary variables corresponding to the samples, creating computational complexity to the already challenging distributed energy resource planning problem. In contrast, the PSAA samples only a part of the random parameters, and we use a non-parametric estimation method to approximate the probability distribution of the remainder, leading to an efficient data-driven approach. We only need to introduce additional continuous variables corresponding to the samples to reformulate the model, reducing the computational complexity. Thus, the reformulation can be scaled up to solve large-scale instances.

The main contributions of the paper are as follows.

- We develop a novel TCC model for the distributed energy resource planning problem that considers both the placement and capacity of RDG and ES units under uncertainty, combined with the ACOPF model, in a distribution grid with multiple periods. We consider all the operational constraints in the second stage to be satisfied simultaneously in a joint chance constraint.
- We are the first to develop the PSAA approach using historical data to solve the above TCC model for a significant industry problem. We extend the PSAA idea that solves single-stage chance-constrained models.
- Our extensive experiments on the IEEE 33-Bus and 123-Bus systems using real data show that the PSAA approach performs better than the standard SAA approach because the former provides better solutions in a shorter time in in-sample tests and provides better guaranteed probability for system reliability in out-of-sample tests. The effectiveness of ES units in reducing total costs and improving system balance is also demonstrated.

The remainder of this paper is organized as follows. A two-stage model and its TCC counterpart for the planning problem are presented in Section III. The TCC model is then approximated using the SAA and PSAA methods in Section IV. We provide computational results and explanations in Section V. Section VI concludes this paper.

## III. Mathematical Model

In this section, we present a two-stage stochastic programming model and its TCC counterpart for the planning problem.

### A. A Two-Stage Model

We focus on a typical distribution grid topology: the radial network. Such a network has a tree structure and connects to a transmission network via a single bus (Bus 0). In our distribution network, the power supply comes from four sources: the transmission network, the traditional dispatchable distributed generation (DDG) units and reactive sources, the RDG units, and the storage units. The last three sources are located in some buses of the distribution grid. The operating costs for the supply sources include the payment to the transmission network, the cost of power generation from the DDG units, and the cost of charging/discharging the ES units. Here we investigate the optimal siting and sizing of $\bar{K}$ RDG units and $R$ ES units in a distribution grid with buses $\mathcal{N}$, to minimize the total cost across the planning horizon. The cost includes deterministic investment/maintenance costs and stochastic operating costs (due to the uncertainties in load and renewable power generation). In the following, we formulate the planning problem as a two-stage optimization

model and describe the corresponding first-stage and second-stage objectives and constraints.

1) **First-stage model**: The first-stage objective minimizes the total cost of building, maintaining, and operating the RDG and ES units, with the model formulated as follows.

$$\min_{\boldsymbol{\Omega}^1} \; C_1\left(\boldsymbol{\Omega}^1\right) + \mathbb{E}\left[Q\left(\boldsymbol{\Omega}^1, \boldsymbol{\xi}\right)\right] \tag{1a}$$

$$\text{s.t.} \sum_{n\in\mathcal{N}} z_{kn} \le 1, \; \forall\, k \in [K], \tag{1b}$$

$$\sum_{k=1}^{K}\sum_{n\in\mathcal{N}} z_{kn} \le \bar{K}, \tag{1c}$$

$$x_k = \sum_{l=1}^{L} u_{kl}\bar{x}_l, \; \forall\, k \in [K], \tag{1d}$$

$$\sum_{l=1}^{L} u_{kl} = \sum_{n\in\mathcal{N}} z_{kn}, \; \forall\, k \in [K], \tag{1e}$$

$$\sum_{n\in\mathcal{N}} w_{rn} \le 1, \; \forall\, r \in [R], \tag{1f}$$

$$\sum_{n\in\mathcal{N}} w_{rn}\underline{y}_r \le y_r \le \sum_{n\in\mathcal{N}} w_{rn}\bar{y}_r, \; \forall\, r \in [R], \tag{1g}$$

where $\boldsymbol{\Omega}^1 := [\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{w}, \boldsymbol{y}]^\top$ is the vector of first-stage variables. The first part in the objective function (1a)

$$C_1\left(\boldsymbol{\Omega}^1\right) := \sum_{k=1}^{K}\left(\sum_{n\in\mathcal{N}} c^0_{kn} z_{kn} + \left(c^1_k + T c^2_k\right) x_k\right) + \sum_{r=1}^{R}\left(\sum_{n\in\mathcal{N}} d^0_{rn} w_{rn} + \left(d^1_r + T d^2_r\right) y_r\right),$$

represents the total deterministic cost, including the setup costs and the size-based investment/maintenance costs of the RDG and ES units. The second part, $\mathbb{E}[Q(\boldsymbol{\Omega}^1, \boldsymbol{\xi})]$, represents the expected minimum operating costs over all $T$ periods, which is defined explicitly in (2). Regarding the constraints, (1b) means that a RDG unit, if installed, has to be in a certain bus, and (1c) restricts the maximum number of installed RDG units. Various regions and institutions have different regulations on the capacity of distributed generation, and thus the capacity of RDG units vary [25]. When a RDG unit is installed in a bus, (1d) and (1e) guarantee that the capacity of the RDG unit is one of the $L$ pre-defined values in $\mathcal{X}$. Similar to (1b), (1f) means that an ES unit can be installed at only one bus, and if so, its capacity lower and upper bounds are provided by (1g).

2) **Second-stage model**: Given a first-stage decision $\boldsymbol{\Omega}^1$ and a realization $\boldsymbol{\xi}$ of the uncertain load and renewable generation, the second-stage objective minimizes the distribution grid's operating costs $Q(\boldsymbol{\Omega}^1, \boldsymbol{\xi})$, where $\boldsymbol{\xi} := [\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^T]^\top$, while respecting a set of physical constraints such as the ACOPF constraints. The operating costs include the cost of purchasing active/reactive energy via Bus 0, the cost of fuel used and emissions created in generating active power in the DDG units, and the cost of charging and discharging the stored energy. The operating costs also include the load-shedding variables $LS^t_{1mn}$, $LS^t_{2mn}$, and a penalty factor $p$ to account for any unsatisfied load. Let $\boldsymbol{\Omega}^2$ be the vector of all second-stage variables and let

$$C_2\left(\boldsymbol{\Omega}^2\right) := \sum_{t=1}^{T}\left(c^t_p p^t_0 + c^t_q q^t_0 + \sum_{n\in\mathcal{B}_1} c^f_n p^t_n + \sum_{n\in\mathcal{B}_1} c^e_n \omega p^t_n + \sum_{r=1}^{R} e_1 f^t_r + \sum_{r=1}^{R} e_2 g^t_r + \sum_{(m,n)\in\mathcal{E}} p\left(LS^t_{1mn} + LS^t_{2mn}\right)\right)$$

be the total operating costs in the second stage. The second-stage problem, whose optimal value is denoted by $Q(\boldsymbol{\Omega}^1, \boldsymbol{\xi})$, can be formulated as follows. (For ease of exposition, all constraints with a superscript $t$ hold for all $t \in [T]$.)

$$\min_{\boldsymbol{\Omega}^2} \; C_2\left(\boldsymbol{\Omega}^2\right) \tag{2a}$$

$$\text{s.t.} \; \underline{\mathrm{p}}^t_n \le p^t_n \le \bar{p}^t_n, \; \forall\, n \in \mathcal{B}_1, \tag{2b}$$

$$\underline{\mathrm{q}}^t_n \le q^t_n \le \bar{q}^t_n, \; \forall\, n \in \mathcal{B}_2, \tag{2c}$$

$$\underline{v} \le \left|V^t_n\right|^2 \le \bar{v}, \; \forall\, n \in \mathcal{N}\setminus\{0\}, \tag{2d}$$

$$p^t_0 = \sum_{n\in\mathcal{B}_0} P^t_{0n}, \quad q^t_0 = \sum_{n\in\mathcal{B}_0} Q^t_{0n}, \tag{2e}$$

$$P^t_{mn} - \mathfrak{R}_{mn}\left|I^t_{mn}\right|^2 + LS^t_{1mn} = d^t_{pn} - \sum_{k=1}^{K} z_{kn} s^t_k x_k - \delta_n p^t_n + \sum_{l\in\mathcal{N}_n} P^t_{nl} + \sum_{r=1}^{R} w_{rn}\left(f^t_r - g^t_r\right), \; \forall\,(m,n)\in\mathcal{E}, \tag{2f}$$

$$Q^t_{mn} - \mathfrak{X}_{mn}\left|I^t_{mn}\right|^2 + LS^t_{2mn} = d^t_{qn} - \tau_n q^t_n + \sum_{l\in\mathcal{N}_n} Q^t_{nl}, \; \forall\,(m,n)\in\mathcal{E}, \tag{2g}$$

$$b^0_r = 0, \; \forall\, r \in [R], \tag{2h}$$

$$0 \le b^t_r \le y_r, \; \forall\, r \in [R], \tag{2i}$$

$$b^t_r - b^{t-1}_r = \gamma f^t_r - g^t_r/\gamma, \; \forall\, r \in [R], \tag{2j}$$

$$\left|V^t_m\right|^2 - \left|V^t_n\right|^2 = 2\mathfrak{R}_{mn}P^t_{mn} + 2\mathfrak{X}_{mn}Q^t_{mn} - \left(\mathfrak{R}^2_{mn} + \mathfrak{X}^2_{mn}\right)\left|I^t_{mn}\right|^2, \; \forall\,(m,n)\in\mathcal{E}, \tag{2k}$$

$$\left\|\left[2P^t_{mn}, 2Q^t_{mn}, \left|V^t_m\right|^2 - \left|I^t_{mn}\right|^2\right]\right\|_2 \le \left|V^t_m\right|^2 + \left|I^t_{mn}\right|^2 \quad \forall\,(m,n)\in\mathcal{E}, \tag{2l}$$

$$\left(P^t_{mn}\right)^2 + \left(Q^t_{mn}\right)^2 \le \left(LC_{mn}\right)^2, \; \forall\,(m,n)\in\mathcal{E}, \tag{2m}$$

$$p^t_0 \ge 0, \; q^t_0 \ge 0, \tag{2n}$$

$$LS^t_{1mn} \ge 0, \; LS^t_{2mn} \ge 0, \; \forall\,(m,n)\in\mathcal{E}. \tag{2o}$$

Here, $\boldsymbol{\Omega}^2$ consists of $p^t_0, q^t_0$ for $t \in [T]$, $p^t_n, q^t_n, V^t_n, P^t_{0n}, Q^t_{0n}$ for $n \in \mathcal{N}$, $t \in [T]$, $f^t_r, g^t_r, b^t_r$ for $r \in [R], t \in [T]$, $P^t_{nl}, Q^t_{nl}$ for $n$ such that $(m,n) \in \mathcal{E}$ for some $m$, $l \in \mathcal{N}_n$, $t \in [T]$, and $P^t_{mn}, Q^t_{mn}, I^t_{mn}, LS^t_{1mn}, LS^t_{2mn}$ for $(m,n) \in \mathcal{E}, t \in [T]$.

We explain all the constraints in the model (2) as follows. The power generated by the DDG units and the reactive sources is bounded by (2b) and (2c), respectively. Constraint (2d) sets the bounds on the voltage of each bus. Constraint (2e) represents the active and reactive balance equations at Bus 0. Constraints (2f) and (2g) are active and reactive power balance equations from Kirchhoff's current law. The following Fig. 1 illustrates the active power flow balance for each distribution line $(m,n) \in \mathcal{E}$. The balance of power in storage is initialized by (2h) and bounded by (2i). ES balance between two consecutive periods is shown in (2j), considering ES

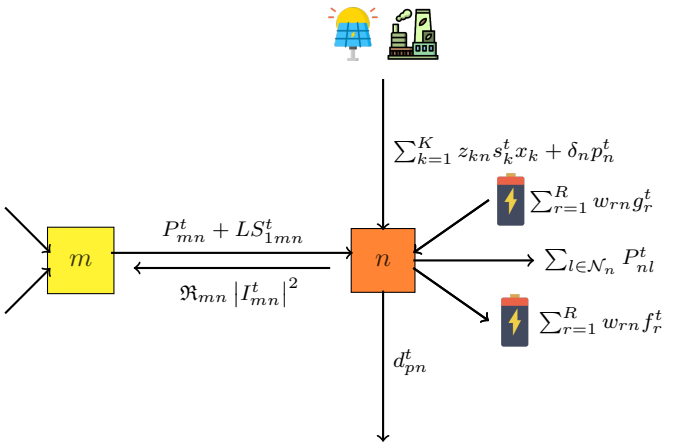


Fig. 1. Power Flow Blance

charging/discharging efficiency. Constraint (2k) represents the voltage drop on each line. Constraint (2l) is the branch power-flow constraint, and the capacity of each distribution line is limited by (2m). Nonnegativity constraints are listed in (2n) and (2o).

Note that the load-shedding variables $LS_{1mn}^t$ and $LS_{2mn}^t$ are defined over each distribution line $(m,n) \in \mathcal{E}$. Once the model is solved, one can also compute the load-shedding values at each bus easily. We also note that the branch flow model is applied here to formulate the ACOPF constraints in (2f) – (2g) and (2k) – (2m), where (2l) includes a set of a second-order conic (SOC) constraints. These constraints represent the ACOPF constraints via convex relaxation following the study in [26]. Specifically, [26] removes the voltage and current angles while introducing squared voltage and current magnitudes, and relaxes the nonconvex quadratic constraints with convex SOC constraints. More importantly, [26] shows that the obtained convex relaxation is exact when the distribution network is radial. As most practical distribution networks are radial grids [26], we also consider a radial network in this paper. Therefore, the constraints (2f) – (2g) and (2k) – (2m) form an exact reformulation of the ACOPF constraints. As such, we obtain a second-order conic programming (SOCP) formulation in (2), which enables large-scale applications due to the computational efficiency of SOCP formulations. We note that there are also other types of approximations to formulate the ACOPF constraints, such as linearized distribution flow (LinDistFlow) [27], where voltage drop and line power flows are approximately linearly related to power injections.

We summarize the two-stage stochastic programming model of the planning problem as

$$\begin{aligned} \min_{\mathbf{\Omega}^1, \mathbf{\Omega}^2} \quad & C_1\left(\mathbf{\Omega}^1\right) + \mathbb{E}\left[C_2\left(\mathbf{\Omega}^2\right)\right] \\ \text{s.t.} \quad & (1b) - (1g), (2b) - (2o). \end{aligned} \tag{3}$$

Note that $\mathbf{\Omega}^2$ and constraints (2b) – (2o) are dependent on $\boldsymbol{\xi}$. As the first-stage problem (1) is an integer program and the second-stage problem (2) is an SOCP, the entire two-stage problem (3) is a mixed-integer SOCP.

### B. A TCC Model

Chance constraints perfectly fit power system operators' decision-making requirements because the operators often need controlled reliability with minimized operating costs. In our model, the distribution grid reliability is measured as the probability of load satisfaction. In the second-stage model, the load is satisfied over all time periods for $(\mathbf{\Omega}^1, \boldsymbol{\xi}, \mathbf{\Omega}^2)$ if and only if all the constraints in (2) are satisfied, such that

$$LS_{1mn}^t = LS_{2mn}^t = 0, \ \forall (m,n) \in \mathcal{E}, \ t \in [T]. \tag{4}$$

Thus, to maintain a high probability of load satisfaction, we add the following joint chance constraint

$$\mathbb{P}\left(\mathbf{\Omega}^2 \text{ in (2) satisfies (4)} \,|\, \mathbf{\Omega}^1\right) \geq 1 - \eta \tag{5}$$

to the two-stage model (3) to strengthen the problem, where $\mathbb{P}$ is a probability function. Different from some existing studies on two-stage chance-constrained programming, our chance constraint (5) consider all the constraints (2b) – (2o) in the second-stage problem (2) as the nominal constraints. For instance, in [19], the authors consider a single demand satisfaction constraint in their chance constraints, see constraints (20) – (22) in [19]. In contrast, our chance constraint (5) requires that (4) is satisfied by any $\mathbf{\Omega}^2$ that is feasible to all the constraints in the second-stage model (2). That is, we ensure the entire system (rather than the load satisification constraints only) to be feasible under a high probability, which can practically incorporate many possible reliability issues such as the distribution line capacity limit and loss. As a result, the TCC model can be formulated as:

$$\begin{aligned} \min_{\mathbf{\Omega}^1, \mathbf{\Omega}^2} \quad & C_1\left(\mathbf{\Omega}^1\right) + \mathbb{E}\left[C_2\left(\mathbf{\Omega}^2\right)\right] \\ \text{s.t.} \quad & (1b) - (1g), (2b) - (2o), (5). \end{aligned} \tag{6}$$

However, the above problem (6) is difficult to solve. Specifically, constraint (5) is nonconvex and thus problem (6) becomes intractable. Next, we introduce two approximation methods to address this challenge.

## IV. Solution Approaches

In this section, we describe two approximation methods for solving the problem (6): the SAA and PSAA methods.

### A. The SAA Formulation

The SAA is a classic sampling technique that is widely used in chance-constrained problems. It approximates the expectation of random variables using their sample means. The probability of an event $E$ can be reformulated as an expectation, as follows:

$$\mathbb{P}(E) = \mathbb{E}\left[\mathbb{I}(E)\right],$$

where $\mathbb{I}(\cdot)$ is an indicator function that takes a value of 1 when the event happens and 0 otherwise. Let $\Pi_1$ be the total number of samples of the random vector $\boldsymbol{\xi}$ and $[\Pi_1]$ be the set of all the samples, and let $\boldsymbol{\xi}_\pi$ be a specific sample for any $\pi \in [\Pi_1]$.

The SAA approximates $\mathbb{P}\left\{\boldsymbol{\Omega}^2 \text{ in (2) satisfies (4)} \,|\, \boldsymbol{\Omega}^1\right\}$ in (5) with

$$\frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1}\mathbb{I}\left(\boldsymbol{\Omega}_\pi^2 \text{ satisfies (4)} \,|\, \boldsymbol{\Omega}^1\right),$$

where $\boldsymbol{\Omega}_\pi^2$ is a copy of the second-stage variables $\boldsymbol{\Omega}^2$ corresponding to $\boldsymbol{\xi}_\pi$ for each $\pi \in [\Pi_1]$. Let $(2b)_\pi$ – $(2o)_\pi$ be constraints (2b) – (2o) with $\boldsymbol{\xi}$ replaced by $\boldsymbol{\xi}_\pi$ and $\boldsymbol{\Omega}^2$ replaced by $\boldsymbol{\Omega}_\pi^2$. The SAA approximation of (6) thus becomes:

$$\begin{aligned}\min_{\boldsymbol{\Omega}^1,\, \boldsymbol{\Omega}_\pi^2, \forall \pi\in[\Pi_1]} \; & C_1\left(\boldsymbol{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\boldsymbol{\Omega}_\pi^2\right)\\ \text{s.t. } & \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1}\mathbb{I}\left(\boldsymbol{\Omega}_\pi^2 \text{ satisfies (4)} \,|\, \boldsymbol{\Omega}^1\right) \geq 1-\eta. \qquad (7)\\ & \text{(1b)} - \text{(1g)},\ (2b)_\pi - (2o)_\pi,\ \forall\, \pi \in [\Pi_1].\end{aligned}$$

We further introduce a binary variable $\theta_\pi \in \{0,1\}$ for each sample $\pi \in [\Pi_1]$. When $\theta_\pi = 0$, it indicates that $\mathbb{I}(\boldsymbol{\Omega}_\pi^2 \text{ satisfies (4)} \,|\, \boldsymbol{\Omega}^1) = 1$, i.e., all the constraints in the second-stage model (2) are satisfied; when $\theta_\pi = 1$, it indicates $\mathbb{I}(\boldsymbol{\Omega}_\pi^2 \text{ satisfies (4)} \,|\, \boldsymbol{\Omega}^1) = 0$, i.e., all the constraints in the second-stage model (2) are not satisfied. Therefore, we can reformulate (7) as the following mixed-integer quadratic program:

$$\begin{aligned}\min_{\boldsymbol{\Omega}^1,\, \boldsymbol{\Omega}_\pi^2, \pi\in[\Pi_1]} \quad & C_1\left(\boldsymbol{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\boldsymbol{\Omega}_\pi^2\right) && (8a)\\ \text{s.t.} \quad & \text{(1b)} - \text{(1g)}, && (8b)\\ & (2b)_\pi - (2o)_\pi,\ \forall\, \pi \in [\Pi_1], && (8c)\\ & LS_{1\pi mn}^t \leq \theta_\pi M_\pi,\ LS_{2\pi mn}^t \leq \theta_\pi M_\pi,\\ & \qquad \forall\,(m,n)\in\mathcal{E}, \forall\, \pi\in[\Pi_1], && (8d)\\ & \sum_{\pi=1}^{\Pi_1}\theta_\pi \leq \Pi_1\eta, && (8e)\\ & \theta_\pi \in \{0,1\},\ \forall\ \pi\in[\Pi_1], && (8f)\end{aligned}$$

where $M_\pi$ is a sufficiently large number for any $\pi \in [\Pi_1]$. Specifically, with constraints (8d) and (8e), we ensure that all but a few number (i.e., $\eta\Pi_1$) of samples in $[\Pi_1]$ satisfy the constraints in the second-stage model (2). That is, with the probability of $1-\eta$, the constraints in the second-stage model (2) are satisfied. Moreover, [28] shows that the objective value difference between model (8) (i.e., model (7)) and model (6) converges to zero with probability one when $\Pi_1$ goes to infinity. In addition, as in the second-stage model (2), all constraints in (8) with a superscript $t$ hold for all $t \in [T]$.

All of the constraints in (8) are convex except for $(2f)_\pi, \pi \in [\Pi_1]$. After substituting $x_k$ with (1d), we achieve bilinear terms $z_{kn}u_{kl}$ for $k \in [K], n \in \mathcal{N}, l \in [L]$. The bilinear terms can be linearized using McCormick inequalities. Specifically, McCormick inequalities are commonly used to linearize a bilinear term, say $w = xy$ with $x^L \leq x \leq x^U$ and $y^L \leq y \leq y^U$, in (mixed-integer) nonlinear programming [29]. The general form of McCormick inequalities for $w = xy$ can be written as:

$$\begin{aligned}& w \geq x^L y + xy^L - x^L y^L,\ w \geq x^U y + xy^U - x^U y^U,\\ & w \leq x^U y + xy^L - x^U y^L,\ w \leq xy^U + x^L y - x^L y^U.\end{aligned}$$

When $x$ and $y$ are both continuous variables, the above four McCormick inequalities provide convex and concave envelopes of the bilinear term $xy$. When at least one of $x$ and $y$ is binary, $w = xy$ can be implied by the above four McCormick inequalities, resulting in an equivalent mixed-integer linear reformulation of the bilinear expression.

Thus, for all $k \in [K], n \in \mathcal{N}, l \in [L]$, we replace $z_{kn}u_{kl}$ with $\beta_{knl}$, and add the following constraints

$$\begin{aligned}& \beta_{knl} \geq 0,\ \beta_{knl} \geq z_{kn} + u_{kl} - 1,\\ & \beta_{knl} \leq u_{kl},\ \text{and } \beta_{knl} \leq z_{kn},\end{aligned}$$

to enforce $\beta_{knl} = z_{kn}u_{kl}$, where both $z_{kn}$ and $u_{kl}$ are binary. Using the same technique, for all $r \in [R], n \in \mathcal{N}, t \in [T], \pi \in [\Pi_1]$, we denote $\hat{f}_{\pi rn}^t := w_{rn}f_{\pi r}^t$, where $w_{rn}$ is binary, and add the following constraints:

$$\begin{aligned}& \hat{f}_{\pi rn}^t \geq w_{rn}\underline{f}_{\pi r}^t,\ \hat{f}_{\pi rn}^t \geq f_{\pi r}^t + w_{rn}\overline{f}_{\pi r}^t - \overline{f}_{\pi r}^t,\\ & \hat{f}_{\pi rn}^t \leq f_{\pi r}^t + w_{rn}\underline{f}_{\pi r}^t - \underline{f}_{\pi r}^t,\ \text{and } \hat{f}_{\pi rn}^t \leq w_{rn}\overline{f}_{\pi r}^t,\end{aligned}$$

where $\overline{f}_{\pi r}^t$ and $\underline{f}_{\pi r}^t$ are the lower and upper bounds of $f_{\pi r}^t$. The bilinear term $w_{rn}g_{\pi r}^t$ can be managed similarly. As a result, the SAA formulation (8) is transformed into a mixed-integer convex quadratic program.

### *B. The PSAA Formulation*

The SAA is relatively accurate when there are sufficient samples. However, more samples lead to more binary auxiliary variables (i.e., $\theta_\pi$), greatly increasing the computational burden. Thus, we use partial sampling to reduce the computational difficulty and improve the solution quality. We extend the preliminary studies on partial sampling in [30] to approximate our proposed TCC model, which is more complicated than the single-stage chance-constrained model considered in [30], as evidenced in [31]. This leads to an extended PSAA model, referred to as *the PSAA model* for simplicity. It samples a part of the random parameters and estimates the probability distribution of the remainder. We first present the basic PSAA idea and then detail our PSAA model.

We consider a general chance constraint

$$\mathbb{P}\{g(\boldsymbol{x},\boldsymbol{\xi}) \geq 0\} \geq 1-\eta, \qquad (9)$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ and $\boldsymbol{\xi}_1$ is independent of $\boldsymbol{\xi}_2$. Clearly, $\mathbb{P}\{g(\boldsymbol{x},\boldsymbol{\xi}) \geq 0\} = \mathbb{E}[\mathbb{I}(g(\boldsymbol{x},\boldsymbol{\xi}) \geq 0)] = \mathbb{E}_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2}[\mathbb{I}(g(\boldsymbol{x},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \geq 0)] = \mathbb{E}_{\boldsymbol{\xi}_1}\mathbb{E}_{\boldsymbol{\xi}_2}[\mathbb{I}(g(\boldsymbol{x},\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) \geq 0)]$, where the third equation is because of the independence between $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. The PSAA idea then reformulates one of the above two expectations (i.e., $\mathbb{E}_{\boldsymbol{\xi}_1}$ and $\mathbb{E}_{\boldsymbol{\xi}_2}$) by its sample mean. For instance, if we replace the inner expectation $\mathbb{E}_{\boldsymbol{\xi}_2}$ by

a sample mean of $N$ independent samples of $\boldsymbol{\xi}_2$ (denoted by $\hat{\boldsymbol{\xi}}_2^1, \ldots, \hat{\boldsymbol{\xi}}_2^N$), then the PSAA formulation of (9) is as follows:

$$\begin{aligned}&\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}_{\boldsymbol{\xi}_1}[\mathbb{I}(g(\boldsymbol{x},\boldsymbol{\xi}_1,\hat{\boldsymbol{\xi}}_2^k))]\\ =&\frac{1}{N}\sum_{k=1}^{N}\mathbb{P}\{g(\boldsymbol{x},\boldsymbol{\xi}_1,\hat{\boldsymbol{\xi}}_2^k)\geq 0\}\geq 1-\eta,\end{aligned}$$

which is further equivalent to the following

$$\mathbb{P}\{g(\boldsymbol{x},\boldsymbol{\xi}_1,\hat{\boldsymbol{\xi}}_2^k)\geq 0\}\geq y_k, \forall k\in[N], \tag{10}$$

$$\frac{\sum_{k=1}^{N} y_k}{N}\geq 1-\eta,\ y_k\geq 0, \forall k\in[N]. \tag{11}$$

In constrast to constraints (8d) - (8f) (SAA formulation), constraints (10)-(11) (PSAA formulation) introduce only continuous variable $y_k$, while $N$ new chance constraints are added. In our proposed PSAA model, We will show that (10) has a convex approximation for *the planning problem* in this paper, which contributes to the existing literature.

To apply the above PSAA idea, we need to have the sampled random parameters independent of the unsampled ones. We first convert the random vector $\boldsymbol{\xi}$ into an uncorrelated random vector $\boldsymbol{\xi}'$ using an affine transformation, thereby approximating the independence requirement. Specifically, let $\Sigma$ be the covariance matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\mu}$ be the mean vector of $\boldsymbol{\xi}$. Suppose that $\Sigma = U\Lambda U^\top$ is an eigenvalue decomposition of $\Sigma$, where $U$ is an orthogonal matrix and $\Lambda$ is a diagonal matrix with the eigenvalues of $\Sigma$ on the diagonal. Without loss of generality, we assume that $\Lambda_{11}$ is the largest eigenvalue of $\Sigma$. Let

$$\boldsymbol{\xi}' = \Lambda^{-\frac{1}{2}}U^\top(\boldsymbol{\xi}-\boldsymbol{\mu}),\ \text{ or equivalently, }\ \boldsymbol{\xi} = U\Lambda^{\frac{1}{2}}\boldsymbol{\xi}'+\mu.$$

It is straightforward to see that $\boldsymbol{\xi}'$ is an uncorrelated random vector with a mean of 0.

We partition $\boldsymbol{\xi}'$ as $(\xi_1', \boldsymbol{\xi}_2')$, where $\xi_1'$ is the first component of $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}_2'$ is the vector of the other components. Note that when $\boldsymbol{\xi}$ is sampled, $\xi_1'$ is the first principal component. The PSAA then considers $\Pi_2$ Monte Carlo samples $\boldsymbol{\xi}_{2\kappa}', \kappa\in[\Pi_2]$ of $\boldsymbol{\xi}_2'$ and approximates the probability of an event $E$ as

$$\mathbb{P}_{(\xi_1',\boldsymbol{\xi}_2')}(E)\approx\frac{1}{\Pi_2}\sum_{\kappa=1}^{\Pi_2}\mathbb{P}_{\xi_1'}\left(E\,|\,\boldsymbol{\xi}_{2\kappa}'\right).$$

In our PSAA model, we retain the objective function of (7) and constraints (1b)$-$(1g) and $(2b)_\pi-(2o)_\pi$. We also develop a different approximation of the chance constraint using the PSAA. To be compatible with the PSAA framework, given the first-stage variables $\boldsymbol{\Omega}^1$, instead of requiring (2b) – (2o) and (4) to be satisfied with a high probability by a specified $\boldsymbol{\Omega}^2$, we relax the chance constraint (5) to require the consistency of (2b) – (2o) and (4). That is, we consider the following chance constraint:

$$\mathbb{P}\left(\exists \text{ a solution satisfying (2b)} - \text{(2o), (4)}\,|\,\boldsymbol{\Omega}^1\right)\geq 1-\eta. \tag{12}$$

Let $(2b)_\kappa-(2o)_\kappa$ and $(4)_\kappa$ be a copy of (2b) – (2o) and (4), with $\boldsymbol{\xi}$ replaced by $(\xi_1', \boldsymbol{\xi}_{2\kappa}')$ for any $\kappa\in[\Pi_2]$. The PSAA approximates (12) with

$$\frac{1}{\Pi_2}\sum_{\kappa=1}^{\Pi_2}\mathbb{P}_{\xi_1'}\left(\xi_1'\in A\left(\boldsymbol{\Omega}^1,\kappa\right)\right)\geq 1-\eta,$$

where $A(\boldsymbol{\Omega}^1,\kappa) := \{\xi_1' \,|\, (2b)_\kappa\text{-}(2o)_\kappa, (4)_\kappa \text{ are consistent with } \boldsymbol{\Omega}^1\}$. For a given $\boldsymbol{\Omega}^1$ and $\kappa$, $A(\boldsymbol{\Omega}^1,\kappa)$ is a convex set of $\xi_1'$. That is, if $\hat{\xi}_1'<\bar{\xi}_1'$ are both in $A(\boldsymbol{\Omega}^1,\kappa)$, then $\xi_1'\in A(\boldsymbol{\Omega}^1,\kappa)$ for all $\xi_1'\in[\hat{\xi}_1',\bar{\xi}_1']$. Let $\Psi(\cdot)$ be the cumulative distribution function (CDF) of $\xi_1'$. Then,

$$\begin{aligned}&\mathbb{P}_{\xi_1'}\left(\xi_1'\in A\left(\boldsymbol{\Omega}^1,\kappa\right)\right)\\ &\quad=\sup_{Z_1,Z_2}\left\{\Psi(Z_2)-\Psi(Z_1)\,|\,Z_1,Z_2\in A\left(\boldsymbol{\Omega}^1,\kappa\right)\right\}.\end{aligned}$$

Therefore, our PSAA model is as follows.

$$\begin{aligned}\min\ & C_1\left(\boldsymbol{\Omega}^1\right)+\frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1}C_2\left(\boldsymbol{\Omega}_\pi^2\right)\\ \text{s.t. }& \text{(1b)}-\text{(1g)}, && \text{(13a)}\\ & (2b)_\pi-(2o)_\pi,\ \forall\,\pi\in[\Pi_1], && \text{(13b)}\\ & (\boldsymbol{\Omega}^1,\boldsymbol{\Omega}_{1\kappa}^2) \text{ satisfies } (2b)_\kappa-(2o)_\kappa, (4)_\kappa \text{ with } (Z_{1\kappa},\boldsymbol{\xi}_{2\kappa}'),\ \forall\,\kappa\in[\Pi_2], && \text{(13c)}\\ & (\boldsymbol{\Omega}^1,\boldsymbol{\Omega}_{2\kappa}^2) \text{ satisfies } (2b)_\kappa-(2o)_\kappa, (4)_\kappa \text{ with } (Z_{2\kappa},\boldsymbol{\xi}_{2\kappa}'),\ \forall\,\kappa\in[\Pi_2], && \text{(13d)}\\ & \Psi(Z_{2\kappa})-\Psi(Z_{1\kappa})\geq\eta_\kappa,\ \forall\,\kappa\in[\Pi_2], && \text{(13e)}\\ & \sum_{\kappa=1}^{\Pi_2}\eta_\kappa\geq\Pi_2(1-\eta), && \text{(13f)}\\ & \eta_\kappa\geq 0,\ \forall\,\kappa\in[\Pi_2]. && \text{(13g)}\end{aligned}$$

Here, the decision variables are $\boldsymbol{\Omega}^1$, $\boldsymbol{\Omega}_\pi^2$ for any $\pi\in[\Pi_1]$, and $\boldsymbol{\Omega}_{1\kappa}^2, \boldsymbol{\Omega}_{2\kappa}^2, Z_{1\kappa}, Z_{2\kappa}, \eta_\kappa$ for any $\kappa\in[\Pi_2]$.

For the model to be practical, we need to estimate the CDF $\Psi(\cdot)$ of $\xi_1'$. Two types of methods are primarily used for estimating distributions: parametric and nonparametric estimation methods. Parametric estimation methods assume that the sample data conform to a parametrized family of probability distributions, and the sample data are used to find the best-fitting parameters. In contrast, nonparametric estimation methods do not depend on any prior assumption of the distribution family, and they fit the distribution according to the characteristics and properties of the data. Here we make no assumptions on the distribution of $\xi_1'$ and estimate $\Psi(\cdot)$ using the kernel density estimation, a commonly used nonparametric method proposed by Rosenblatt [32] and Parzen [33].

Let $\Pi_3$ be the total number of samples of $\xi_1'$, and $\xi_{1\tau}'$ be a specific sample for any $\tau\in[\Pi_3]$. The kernel density estimation of the probability density function $\psi$ of $\xi_1'$ can be written as

$$\psi(\xi_1')\approx\frac{1}{\Pi_3 h}\sum_{\tau=1}^{\Pi_3}\phi\left(\frac{\xi_1'-\xi_{1\tau}'}{h}\right),$$

where $h$ is a user-specified bandwidth parameter and $\phi$ is a kernel function. Among the popular choices, we choose the standard normal density function as the kernel function for our estimation. Thus, the CDF of $\xi_1'$ can be estimated by

$$\Psi(\xi_1')\approx\frac{1}{\Pi_3}\sum_{\tau=1}^{\Pi_3}\Phi\left(\frac{\xi_1'-\xi_{1\tau}'}{h}\right), \tag{14}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

We further approximate $\Phi(\cdot)$ in (14) using the following piecewise linear function:

$$\Phi(x) \approx \begin{cases} \min\limits_{\epsilon\in[\Delta_1]} \{a_\epsilon x + \alpha_\epsilon\} & \text{if } x \geq 0.5 \\ \max\limits_{\zeta\in[\Delta_2]} \{a_\zeta x + \alpha_\zeta\} & \text{if } x < 0.5, \end{cases} \tag{15}$$

where $\Delta_1$ and $\Delta_2$ are the number of pieces used to approximate the upper half and the lower half of $\Phi(\cdot)$, respectively. An example of such an approximation is depicted in the following Fig. 2.

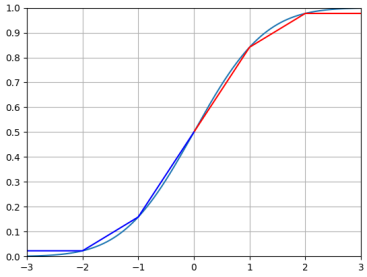

Fig. 2. Piecewise linear approximation of the CDF of the standard normal distribution $\Phi(\cdot)$ ($\Delta_1 = \Delta_2 = 3$)

When the probability level $1-\eta$ is close to 1, $\Phi(\frac{Z_{2\kappa}-\xi'_{1\tau}}{h})$ is usually greater than 0.5 and thus concave, while $\Phi(\frac{Z_{1\kappa}-\xi'_{1\tau}}{h})$ is usually less than 0.5 and thus convex [34]. Therefore, with the approximation in (15), we can remove the min and max operators in (15) and approximate constraint (13e) by the following constraints:

$$\rho^\kappa_{1\tau} \geq a_\zeta \left(\frac{Z_{1\kappa} - \xi'_{1\tau}}{h}\right) + \alpha_\zeta, \ \forall \zeta \in [\Delta_2], \tau \in [\Pi_3], \kappa \in [\Pi_2],$$

$$\rho^\kappa_{2\tau} \leq a_\epsilon \left(\frac{Z_{2\kappa} - \xi'_{1\tau}}{h}\right) + \alpha_\epsilon, \ \forall \epsilon \in [\Delta_1], \tau \in [\Pi_3], \kappa \in [\Pi_2],$$

$$\frac{1}{\Pi_3} \sum_{\tau=1}^{\Pi_3} (\rho^\kappa_{2\tau} - \rho^\kappa_{1\tau}) \geq \eta_\kappa, \ \forall \kappa \in [\Pi_2].$$

Finally, the bilinear terms in (13b) – (13d) can be linearized using McCormick inequalities, as in the SAA model. As a result, the PSAA model (13) is simplified to a mixed-integer convex quadratic program.

## V. Numerical Results

We conduct two sets of experiments on the IEEE 33-Bus system and the IEEE 123-Bus system using real data acquired from Pecan Street Inc. and ERCOT. We first compare the effectiveness of the SAA and PSAA models and then investigate the potential benefits of installing ES units. We use in-sample and out-of-sample tests to validate the quality of the obtained solutions to the planning problem. All numerical tests are executed on the high performance computing (HPC) cluster of Ieria [35] with 27 computing nodes. We allocate four CPUs to every instance, and every CPU is allocated 4 GB of memory. CPLEX 22.1.0 with its default setting is used to solve all optimization models. For ease of exposition, we use the following flowchart in Fig. 3 to summarize the sequential steps that we follow to perform the numerical experiments in this section.

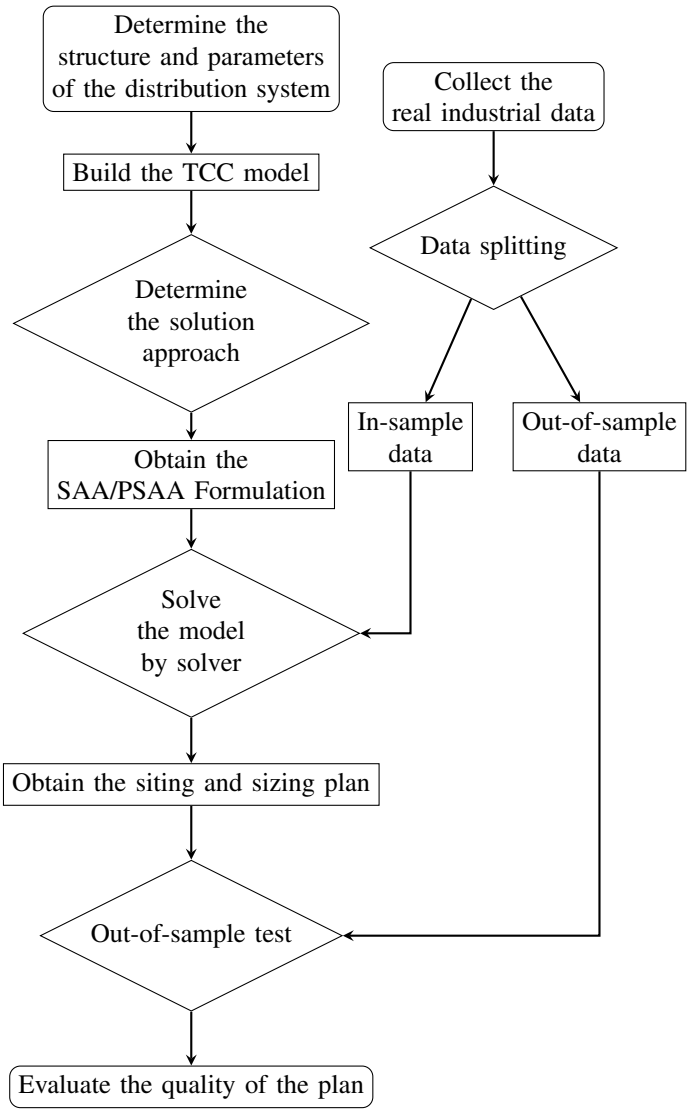


Fig. 3. The Procedure of Numerical Experiments

### A. IEEE 33-Bus System

We first consider the modified IEEE 33-Bus radial distribution network examined in [11] (see Fig. 4). In the network, Bus 0 is connected to the major transmission network, from which we can purchase active and reactive power via Bus 0 if needed. Buses 1–32 are connected to Bus 0, directly or indirectly. Two DDG units are located at Buses 15 and 29, and three reactive power sources are located at Buses 11, 13, and 32. With fixed location of the DDG units and reactive power sources, we then focus on the location and capacity planning of RDG and ES units.

*1) Data:* We consider two sources of uncertainties: the weekly active/reactive load at each bus, i.e., $d^t_{pn}/d^t_{qn}$, and the renewable generation efficiency of each RDG unit, i.e., $s^t_k \in [0, 1]$. The active-load data are obtained from Pecan Street Inc., and the wind generation data from ERCOT. The reactive-load

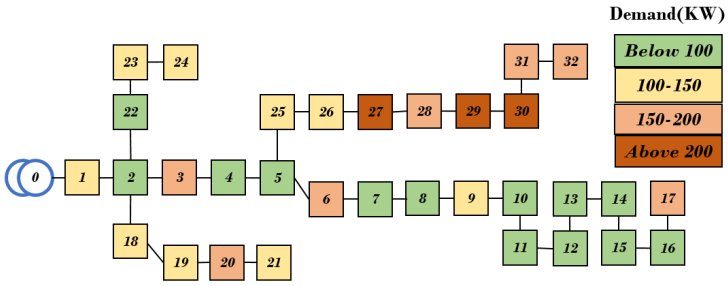


Fig. 4. IEEE 33-Bus Distribution Network

data are randomly generated based on the bounds of the total reactive-power output. Specifically, the reactive-power load is uniformly generated in the interval $[-0.01, 0.019]$. All data are for a range of 4 years, leading to 208 $(= 52 \text{ weeks} \times 4)$ data samples for each specific random variable. To perform more practical tests, we randomly generate more data samples to better demonstrate our proposed models' effectiveness, via in-sample and out-of-sample tests. To that end, we first calculate the mean value and covariance matrix using the given data samples, and then generate 3792 data samples by following the multivariate log-normal distribution, which has been widely adopted in academia and industry in similar scenarios [36, 37, 38]. Thus, we have 4000 data samples in total.

*2) Parameters:* All of the parameters used in our experiments are slightly modified based on the parameters[1] used in [11]. For instance, the electricity price of purchasing active/reactive power from the main grid is mainly from ERCOT. The detailed modification is as follows. In the first stage, the setup costs of RDG units $c^0_{kn}$ are uniformly generated in the interval $[0.95 \times 2000, 1.05 \times 2000]$. The size-based investment costs $c^1_k$ and maintenance costs $c^2_k$ of RDG units are uniformly generated in the intervals $[0.9 \times 238, 1.1 \times 238]$ and $[0.9 \times 4, 1.1 \times 4]$, respectively. The setup costs of ES units $d^0_{rn}$ are uniformly generated in the interval $[0.9 \times 200, 1.1 \times 200]$. The size-based investment costs $d^1_r$ and maintenance costs $d^2_r$ of ES units are both uniformly generated in the interval $[0.9 \times 2, 1.1 \times 2]$. The active power purchase prices $c^t_p$ are uniformly generated in the interval $[0.9 \times 130, 1.1 \times 130]$. The reactive power purchase prices $c^t_q$ are uniformly generated in the interval $[0.9 \times 4, 1.1 \times 4]$. The emission costs for the DDG units at Buses 15 and 29 are $c^f_{15} = c^f_{29} = 630$. The emission factor $\omega$ of the DDG units is 3 kg/MWh. A maximum of $\bar{K} = 3$ out of $K = 4$ candidates of RDG units are to be installed in this distribution network. The maximum number of ES units to be installed is $R = 3$. The active-power output bounds $(\underline{p}^t_n, \bar{p}^t_n)$ of both DDG units are $(0.5, 4.5)$. The reactive-power output bounds $(\underline{q}^t_n, \bar{q}^t_n)$ are $(-0.1, 0.2)$, $(-0.15, 0.25)$, and $(-0.1, 0.2)$ for the three reactive-power sources at Buses 11, 13, and 32, respectively. We further consider four types of RDG units ($\bar{x}_1 = 4$ MW, $\bar{x}_2 = 5$ MW, $\bar{x}_3 = 6$ MW, and $\bar{x}_4 = 7$ MW). The maximum capacity $\bar{y}_r$ of an ES unit is 3 MW, and the minimum capacity $\underline{y}_r$ is 0. The initial power level $b^0_r$ of an ES unit is set to 0. When an ES unit is charged, the unit cost $e_1$ is 0.1, whereas the discharging cost $e_2$ is 0.1. The energy loss factor $\gamma$ is set to 0.9.

### B. IEEE 123-Bus System

We then consider the commonly used IEEE 123-Bus radial distribution network [39, 40, 41] (see Fig. 5). In the network, Bus 149 is connected to the major transmission network. Eight DDG units are located at Buses 8, 25, 44, 57, 67, 87, 97 and 108, and twelve reactive power sources are located at Buses 7, 14, 15, 25, 47, 54, 62, 68, 80, 91, 98, and 109.

*1) Data:* All the data are obtained following the same process used for the IEEE 33-Bus system. The only difference is the dimension of uncertainty. Here we consider the renewable generation efficiency of each RDG unit $s^t_k$ to be uncertain and use an estimated value for the weekly active/reactive load at each bus, i.e., $d^t_{pn}/d^t_{qn}$. In particular, as the IEEE 123-Bus system is of large scale, both the SAA and PSAA formulation become difficult to solve when the test system is large. Thus, to better show the performance of these two approaches, we consider the system loads are given and the renewable generation is uncertain, by which the computational difficulty is relatively reduced.

*2) Parameters:* We continue to use the parameters designed for the IEEE 33-Bus system except that some parameters are modified as follows. First, the electrical resistance ($\mathfrak{R}_{mn}$) and reactance ($\mathfrak{X}_{mn}$) of each line $(m, n) \in \mathcal{E}$ and the upper/lower bound of voltage magnitude (i.e., $\underline{v}/\bar{v}$) at each bus are obtained from the IEEE PES Test Feeders[2]. Second, we consider a maximum of $\bar{K} = 6$ out of $K = 8$ candidates of RDG units are to be installed in this distribution network. We consider three types of RDG units ($\bar{x}_1 = 8$ MW, $\bar{x}_2 = 10$ MW, and $\bar{x}_3 = 12$ MW). The maximum capacity $\bar{y}_r$ of an ES unit is 6 MW. The maximum number of ES units to be installed is $R = 6$. The reactive-power output bounds $(\underline{q}^t_n, \bar{q}^t_n)$ are (-0.15, 0.25).

### C. Decomposition Framework

To reduce the computational difficulty of solving both the SAA and PSAA formulations, we adopt the Benders decomposition algorithm [42] to improve the computational efficiency. Specifically, we first linearize the SOCP constraints (2l) and (2m) as [11] does by using the polyhedral $\epsilon$-approximation in

[1]See https://www.dropbox.com/s/psqv9yr3atg46bk. Accessed: Jul. 2022.

[2]See https://cmte.ieee.org/pes-testfeeders/resources/. Accessed: Jul. 2022.

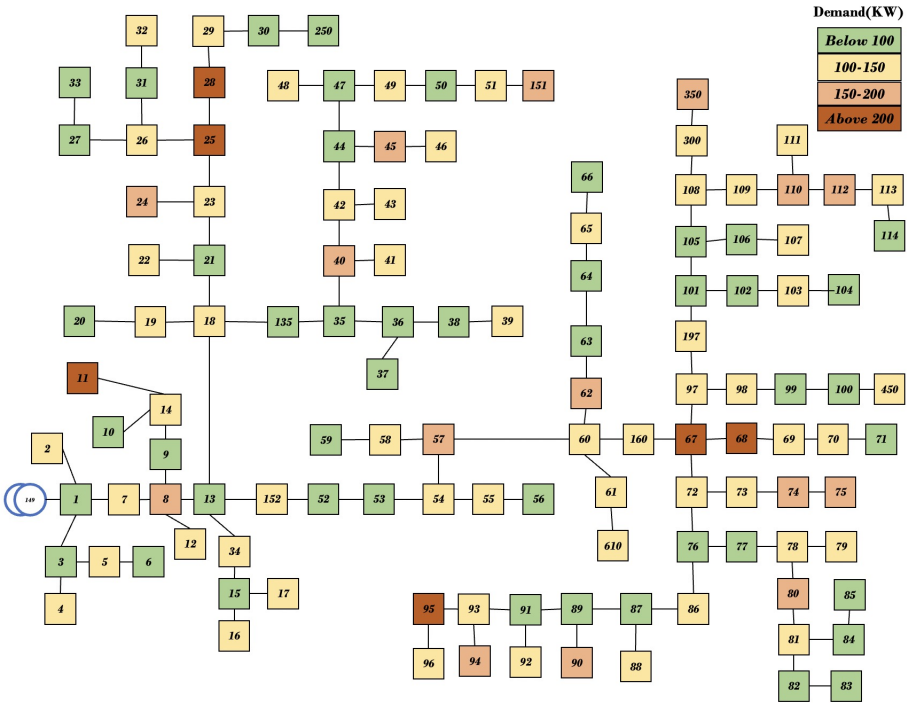


Fig. 5. IEEE 123-Bus Distribution Network

[43]. With such an approximation, both the SAA formulation (8) and PSAA formulation (13) are transformed into mixed-integer linear programming (MILP) formulations, which can be used practically in large-scale settings.

For each MILP formulation, we then decompose the problem into two parts: a master problem and a set of subproblems. The master problem includes all the integer variables and associated constraints, and the subproblems contain the remaining continuous variables and associated constraints. As such, we iteratively solve the master problem and subproblems until convergence. At each iteration, the master problem is solved to optimality and its optimal solution is then used to construct the subproblems. Feasible and optimality cuts are generated after solving the subproblems and added back to the master problem.

Note that we have $\Pi_1$ subproblems for the SAA formulation, where each sample $\pi \in [\Pi_1]$ corresponds to one subproblem, as shown in (8c), and these subproblems can be solved in parallel. For the PSAA formulation, we have $\Pi_1 + 1$ subproblems, where each sample $\pi \in [\Pi_1]$ corresponds to one subproblem, as shown in (13b), and constraints (13c) – (13g) are included in one subproblem.

### *D. SAA vs. PSAA*

Here we analyze the performance of the SAA and PSAA models using the data and parameters mentioned above. We thus ignore ES units and consider $T = 1$. Specifically, terms related to ES units, including $w_{rn}$, $y_r$, $f_r^t$, $g_r^t$, and $b_r^t$, are temporarily removed from the models.

We first divide the 4000 data samples into two sets for the experiment: a training data set and a testing data set. The former is used to obtain our planning decision in the first stage, and the latter is used to test the effectiveness of the obtained decision. To make full use of the real data and better simulate real-world decision-making, we ensure that the training data are selected from the first 208 samples, as it is very straightforward to feed the available historical real data into an optimization model to support decision-making. Each data sample is used as a scenario in the SAA and PSAA models. We solve the SAA and PSAA models using the training data and obtain two optimal sizing/siting plans. We then compare the performance of the plans using the testing data. Specifically, we calculate the first-stage cost and the average second-stage cost of all test samples for each plan. To verify our approximation formulations and demonstrate their ability to ensure that the demand can be satisfied with a high probability, we also calculate the actual feasible probability of the test samples. This probability is defined as the percentage of the test samples for which constraints (2b) – (2o) and (4) can be simultaneously satisfied.

The experiments are conducted using different settings for (i) the size of the training data (i.e., $\Pi_1 = \Pi_2$), (ii) the maximum running time (i.e., time limit) of the solver (denoted by $\vartheta$ in hours), and (iii) the desired feasible probability of the chance constraint (i.e., $1-\eta$), as shown in Tables I – IV. The training data size $\Pi_1 = \Pi_2 \in \{60, 100, 140\}$ for the IEEE 33-Bus system and $\Pi_1 = \Pi_2 \in \{30, 45, 60\}$ for the IEEE 123-Bus system. The maximum running time $\vartheta \in \{4, 8, 12\}$ for the IEEE 33-Bus system and $\vartheta \in \{10, 13, 16\}$ for the IEEE 123-Bus system. The desired feasible probability $1-\eta \in \{0.8, 0.9\}$ for both systems. These settings lead to $18 = 3 \times 3 \times 2$ combinations in total for each system. For some instances, when both models are too large to be solved to optimality

within the given time limits, we take the incumbent solutions returned by the solver as the optimal solutions. We also record the relative optimality gap, $(z_p - z_d)/z_p$, where $z_p$ is the primal objective bound (i.e., the incumbent objective value) and $z_d$ is the dual objective bound (i.e., the lower bound for minimization problems). Intuitively, a smaller gap indicates better-quality incumbent solution.

We illustrate the performance of both models in Tables I – IV. The columns in the SAA/PSAA section represent, from left to right, the first-stage cost, the average second-stage cost of testing samples, the average total cost, the relative optimality gap, and the actual feasible probability.

Compared with the SAA model, the PSAA model leads to lower first-stage costs, lower second-stage costs, and thus lower total costs in all cases. The lower costs indicate that the siting and sizing decisions provided by the PSAA approach help more effectively satisfy the same required demands than those provided by the SAA approach. Thus, compared to the SAA solutions, the PSAA solutions require fewer RDG units to be installed and/or the installed RDG units can be of smaller capacity. Moreover, the effective plans made by the PSAA approach lead to lower operating costs for the distribution grid than the plans made by the SAA approach. A direct comparison of the total costs is shown in Fig. 6, where the horizontal axis represents the training-data size, the vertical axis represents the total cost, and different colors represent different running times. The total costs of the SAA solutions are labeled with triangular symbols, and the total costs of the PSAA solutions are marked with circular symbols.

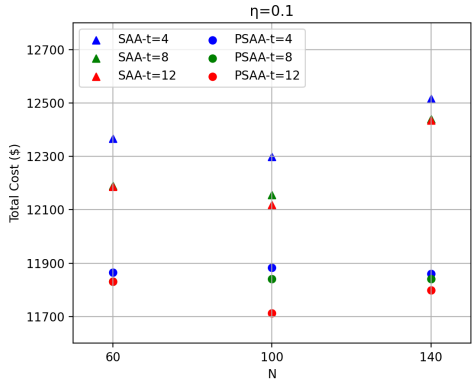


Fig. 6. SAA vs. PSAA solutions in the total cost ($\eta = 0.1$)

Within the same time limits, we observe that the relative optimality gap of the PSAA solution is always less than that of the SAA solution in all cases. Specifically, for the cases where the PSAA model can solve the instance to optimality (i.e., the optimality gap is 0) within the time limits, we report the corresponding computational time in hours used by the PSAA model in the column "Gap" and label it by $\star$. The SAA model cannot solve any instance to optimality within the time limits. The result clearly indicates that the PSAA model is more computationally efficient than the SAA model. The difference is due to how the chance constraint is dealt with in the models. The SAA model introduces $\Pi_1$ binary variables (i.e., $\theta_\pi$), whereas the PSAA model introduces $\Pi_2$ continuous variables (i.e., $\eta_\kappa$). Although there are more continuous variables and more constraints in the PSAA model, the binary variables in the SAA model are more difficult to manage. A direct comparison of the gap is shown in Fig. 7, where the vertical axis represents the relative optimality gap.

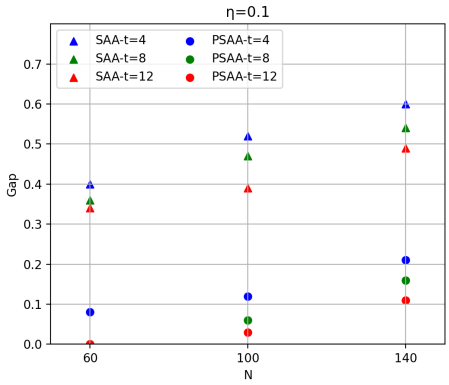


Fig. 7. SAA vs. PSAA solutions in the optimality gap ($\eta = 0.1$)

To further illustrate the computational improvement of the PSAA approach compared with the SAA approach, we summarize the optimality gap improvement (i.e., reduction) from the SAA approach to the PSAA approach in Fig. 8 and Fig. 9. In the figures, the horizontal axis represents the setting of $\eta$ and $\Pi_1$. For instance, "$0.1-60$" means that $\eta = 0.1$ and $\Pi_1 = 60$. The vertical axis represents the optimality gap improvement in percentage, as given by:

$$\frac{|\text{the gap by PSAA} - \text{the gap by SAA}|}{\text{the gap by SAA}} \times 100\%.$$

From Fig. 8 and Fig. 9, we find that the improvement is mostly above $50\%$ and even reaches $100\%$ when the PSAA solves an instance to optimality.

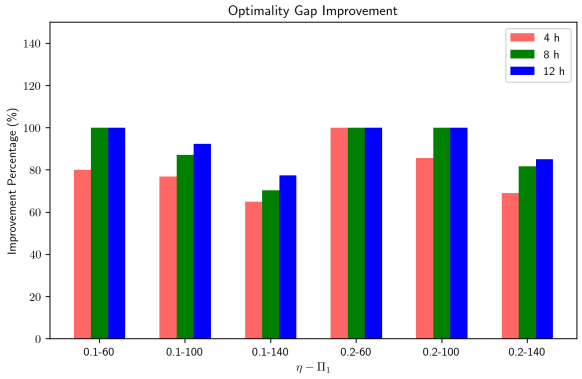


Fig. 8. Optimality Gap Improvement (IEEE 33-Bus system)

In each of the tested cases, the actual feasible probability of the PSAA solution is clearly higher than that of the SAA solution and is almost equal to the desired probability. This indicates that the PSAA performs better than the SAA as an approximation method for the chance constraint. In fact, the

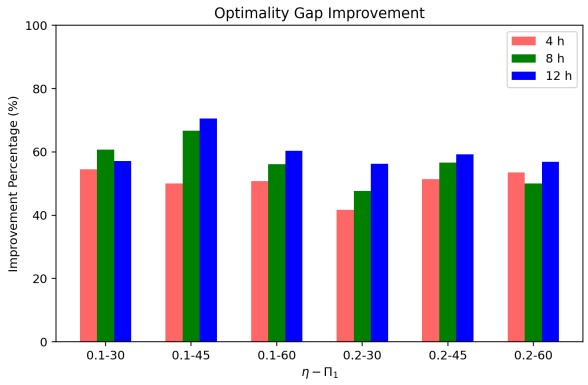


Fig. 9. Optimality Gap Improvement (IEEE 123-Bus system)

actual feasible probability of the PSAA solution is less than 1% different from the desired solution, which in practice will give the grid decision-makers more control of the confidence level. A direct comparison is shown in Fig. 10, where the vertical axis represents the actual feasible probability.

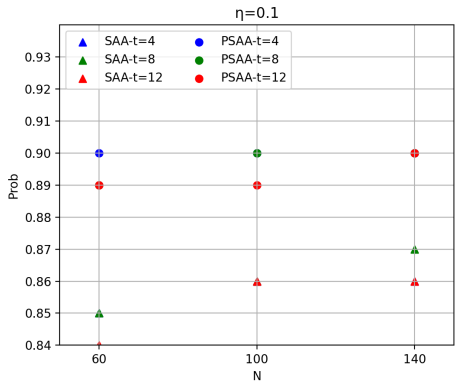


Fig. 10. SAA vs. PSAA solutions in the actual feasible probability ($\eta = 0.1$)

To conclude this section, we illustrate the performance difference between the two models with a specific example. When $\Pi_1 = \Pi_2 = 140$, $\vartheta = 8$, and $\eta = 0.1$ in the IEEE 33-Bus system, the SAA solution sites (sizes) the RDG units at Buses 3 (6 MW), 9 (6 MW), and 12 (4 MW), and the PSAA solution sits (sizes) the RDG units at Buses 3 (5 MW), 8 (4 MW), and 26 (6 MW). Thus, compared to the SAA solution, the PSAA solution results in a lower first-stage cost, due to the smaller total capacities of the installed RDG units. In addition, the output $p_n^t$ of the DDG unit at Bus 29 in the PSAA solution is significantly less than that in the SAA solution for most testing data, and the load-shedding penalties of the PSAA solution are also much less than those of the SAA solution. The two factors above account for most of the difference between the methods in the second-stage costs. We know that Buses 27, 29, and 30 have higher loads, and that the DDG units have been installed at Buses 15 and 29. Thus, it is reasonable to place an RDG unit at Bus 26 to reduce the output pressure on the DDG unit at Bus 29. In addition, in a distribution network structure, it is likely that the power purchased from Bus 0 and generated by the RDG unit at Bus 3 (6 MW) or Bus 3 (5 MW) is mainly used to fulfill the loads at Buses 1-4 and 18-24. Given the desired feasible probability of 0.9 (i.e., the power system operator expects to satisfy the load with a confidence level of 0.9), the PSAA solution shows that a 5 MW RDG unit is sufficient to satisfy the load, and thus a 6 MW unit (given by the SAA solution) may be excessive.

### *E. Storage vs. No Storage*

Here we compare the experimental results to show the effects of ES installation. From previous experiments, we observe that the PSAA model produces higher-quality solutions than the SAA model. Thus, all subsequent experiments are conducted using the PSAA model.

In the experiments, the maximum running time $\vartheta = 12$ hours for the IEEE 33-Bus system and $\vartheta = 16$ hours for the IEEE 123-Bus system, and the planning horizon $T = 3$ weeks. Note that our proposed TCC model (6) is general enough to consider a longer-term setting because one can always set $T$ to be years or seasons. Here our experiments consider a representative snapshot of the long-term future by setting $T$ to be a relatively small number. Correspondingly, the cost parameters in the first stage of model (6), including the setup costs and the size-based investment/maintenance costs of the RDG and ES units, have also been levelized over the specific $T$ time periods (weeks).

To match the planning horizon, we integrate every 3 of the 4000 weekly samples into a 3-week-long sample (without repetition) to obtain 1333 new samples. For computational efficiency, the size of the training data is set to $\Pi_1 = \Pi_2 = 50$ for the IEEE 33-Bus system and $\Pi_1 = \Pi_2 = 30$ for the IEEE 123-Bus system. As the training set is relatively small compared with the number of random variables, we use the *k*-means clustering algorithm to improve the reliability of the training samples. In particular, we randomly choose 200 of the 1333 samples, and divide them into 50 groups by the *k*-means algorithm. We then use the centers of the 50 groups as our training samples. The remaining 1133 samples are used for testing.

We modify the active-power upper bounds of the DDG units (i.e., $\bar{p}_n^t$) and the standard capacities in $\mathcal{X}$ of the RDG units to match their designed load share in different cases. In particular, let $\varrho_1$ and $\varrho_2$ be two nonnegative parameters, and let $SUM$ be the expected total active load (estimated from the real data and invariant to $t$). For each of the two DDG units, the active-power upper bound $\bar{p}_n^t$ is adjusted to $\varrho_1 \times SUM/2$ for all $t \in [T]$. The maximum standard capacity of the RDG units $\bar{x}_4$ is adjusted from 7 MW to $\varrho_2 \times SUM/4$. The other three standard capacities in $\mathcal{X}$ are adjusted proportionally. For example, the minimum standard capacity $\bar{x}_1$ is adjusted from 4 MW to $\varrho_2 \times SUM/4 \times 4/7$. We conduct two sets of experiments with $\eta = 0.1$ and $\eta = 0.2$, respectively. For each set of experiments, we set $(\varrho_1, \varrho_2)$ to six different pairs of values. The results are shown in Tables V – VIII.

When $\varrho_1 + \varrho_2 = 1.0$, the solutions with ES units lead to higher second-stage costs but lower total costs than those without ES units. This indicates that ES installation is beneficial overall, despite leading to higher operational costs.

TABLE I
IEEE 33-BUS SYSTEM: SAA VS. PSAA ($\eta = 0.1$)

| $\Pi_1$ | $\vartheta\,(h)$ | SAA | | | | | PSAA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| 60 | 4 | 10674.5 | 1692.8 | 12367.3 | 0.40 | 0.85 | 10226.3 | 1639.2 | 11865.5 | 0.08 | 0.90 |
| | 8 | 10510.0 | 1679.4 | 12189.4 | 0.36 | 0.85 | 10184.6 | 1645.8 | 11830.4 | 7.33⋆ | 0.89 |
| | 12 | 10510.0 | 1679.4 | 12189.4 | 0.34 | 0.84 | 10184.6 | 1645.8 | 11830.4 | 7.33⋆ | 0.89 |
| 100 | 4 | 10586.3 | 1713.0 | 12299.3 | 0.52 | 0.86 | 10275.5 | 1607.6 | 11883.1 | 0.12 | 0.90 |
| | 8 | 10447.2 | 1708.5 | 12155.7 | 0.47 | 0.86 | 10224.5 | 1616.2 | 11840.7 | 0.06 | 0.90 |
| | 12 | 10397.4 | 1721.3 | 12118.7 | 0.39 | 0.86 | 10105.6 | 1606.5 | 11712.1 | 0.03 | 0.89 |
| 140 | 4 | 10800.2 | 1716.0 | 12516.2 | 0.60 | 0.86 | 10231.8 | 1629.0 | 11860.8 | 0.21 | 0.90 |
| | 8 | 10749.0 | 1691.2 | 12440.2 | 0.54 | 0.87 | 10209.0 | 1632.7 | 11841.7 | 0.16 | 0.90 |
| | 12 | 10723.9 | 1711.9 | 12435.8 | 0.49 | 0.86 | 10173.0 | 1625.1 | 11798.1 | 0.11 | 0.90 |

TABLE II
IEEE 33-BUS SYSTEM: SAA VS. PSAA ($\eta = 0.2$)

| $\Pi_1$ | $\vartheta\,(h)$ | SAA | | | | | PSAA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| 60 | 4 | 9947.9 | 1756.2 | 11704.1 | 0.21 | 0.78 | 9536.3 | 1723.3 | 11259.6 | 2.55⋆ | 0.80 |
| | 8 | 9947.9 | 1756.2 | 11704.1 | 0.12 | 0.78 | 9536.3 | 1723.3 | 11259.6 | 2.55⋆ | 0.80 |
| | 12 | 9664.2 | 1742.8 | 11407.0 | 0.08 | 0.77 | 9536.3 | 1723.3 | 11259.6 | 2.55⋆ | 0.80 |
| 100 | 4 | 9772.3 | 1722.6 | 11494.9 | 0.35 | 0.78 | 9652.7 | 1720.1 | 11372.8 | 0.05 | 0.80 |
| | 8 | 9651.9 | 1743.9 | 11395.8 | 0.27 | 0.78 | 9574.9 | 1652.2 | 11227.1 | 4.73⋆ | 0.80 |
| | 12 | 9633.8 | 1739.7 | 11373.5 | 0.21 | 0.77 | 9574.9 | 1652.2 | 11227.1 | 4.73⋆ | 0.80 |
| 140 | 4 | 10039.0 | 1740.1 | 11779.1 | 0.42 | 0.79 | 9705.0 | 1692.1 | 11397.1 | 0.13 | 0.80 |
| | 8 | 9974.3 | 1748.7 | 11723.0 | 0.33 | 0.79 | 9639.5 | 1683.0 | 11322.5 | 0.06 | 0.80 |
| | 12 | 9974.3 | 1748.7 | 11723.0 | 0.27 | 0.79 | 9592.1 | 1666.8 | 11258.9 | 0.04 | 0.80 |

TABLE III
IEEE 123-BUS SYSTEM: SAA VS. PSAA ($\eta = 0.1$)

| $\Pi_1$ | $\vartheta\,(h)$ | SAA | | | | | PSAA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| 30 | 10 | 26715.3 | 6422.7 | 33138.0 | 0.33 | 0.86 | 25937.3 | 6396.0 | 32333.3 | 0.15 | 0.91 |
| | 13 | 26127.5 | 6248.7 | 32376.2 | 0.28 | 0.86 | 25372.6 | 6210.7 | 31583.3 | 0.11 | 0.90 |
| | 16 | 25972.4 | 6255.3 | 32227.7 | 0.21 | 0.85 | 25047.2 | 6251.8 | 31299.0 | 0.09 | 0.90 |
| 45 | 10 | 27019.2 | 6392.3 | 33411.5 | 0.42 | 0.87 | 26252.8 | 6305.3 | 32558.1 | 0.21 | 0.90 |
| | 13 | 26407.7 | 6268.9 | 32676.6 | 0.39 | 0.86 | 26043.7 | 6365.3 | 32409.0 | 0.13 | 0.90 |
| | 16 | 26149.3 | 6274.3 | 32423.6 | 0.34 | 0.86 | 25392.6 | 6216.6 | 31609.2 | 0.10 | 0.89 |
| 60 | 10 | 26931.0 | 6561.9 | 33492.9 | 0.61 | 0.87 | 25892.4 | 6410.8 | 32303.2 | 0.30 | 0.91 |
| | 13 | 26772.4 | 6347.2 | 33119.6 | 0.57 | 0.87 | 25465.3 | 6304.3 | 31769.6 | 0.25 | 0.91 |
| | 16 | 26073.5 | 6403.6 | 32477.1 | 0.53 | 0.87 | 25428.0 | 6309.8 | 31737.8 | 0.21 | 0.90 |

TABLE IV
IEEE 123-BUS SYSTEM: SAA VS. PSAA ($\eta = 0.2$)

| $\Pi_1$ | $\vartheta\,(h)$ | SAA | | | | | PSAA | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| 30 | 10 | 25894.3 | 6407.1 | 32301.4 | 0.24 | 0.78 | 25283.4 | 6392.2 | 31675.6 | 0.14 | 0.80 |
| | 13 | 25607.6 | 6379.2 | 31986.8 | 0.21 | 0.78 | 25076.9 | 6359.4 | 31436.3 | 0.11 | 0.80 |
| | 16 | 25313.7 | 6392.4 | 31706.1 | 0.16 | 0.77 | 24764.0 | 6271.2 | 31035.2 | 0.07 | 0.79 |
| 45 | 10 | 26021.9 | 6307.6 | 32329.5 | 0.37 | 0.78 | 25506.3 | 6293.2 | 31799.5 | 0.18 | 0.80 |
| | 13 | 25528.3 | 6517.4 | 32045.7 | 0.30 | 0.78 | 24892.6 | 6268.1 | 31160.7 | 0.13 | 0.80 |
| | 16 | 24986.1 | 6492.2 | 31478.3 | 0.27 | 0.78 | 24508.3 | 6238.9 | 30747.2 | 0.11 | 0.80 |
| 60 | 10 | 26328.0 | 6398.2 | 32726.2 | 0.56 | 0.78 | 26017.3 | 6344.5 | 32361.8 | 0.26 | 0.81 |
| | 13 | 26148.7 | 6362 | 32510.7 | 0.48 | 0.78 | 25370.4 | 6308.8 | 31679.2 | 0.24 | 0.80 |
| | 16 | 25693.3 | 6417.4 | 32110.7 | 0.44 | 0.78 | 25091.2 | 6284.5 | 31375.7 | 0.19 | 0.80 |

TABLE V
IEEE 33-BUS SYSTEM: STORAGE VS. NO STORAGE ($\eta = 0.1$)

| $(\varrho_1, \varrho_2)$ | without energy storage | | | | | with energy storage | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| (0.5,0.5) | 10592.0 | 5243.3 | 15835.3 | 0.24 | 0.89 | 9952.0 | 5501.4 | 15453.4 | 0.26 | 0.90 |
| (0.4,0.6) | 10464.3 | 5372.8 | 15837.1 | 0.23 | 0.90 | 9873.8 | 5427.6 | 15301.4 | 0.29 | 0.91 |
| (0.6,0.4) | 10726.9 | 5194.2 | 15921.1 | 0.27 | 0.89 | 9908.2 | 5462.9 | 15371.1 | 0.28 | 0.90 |
| (0.3,0.3) | | | * | | | 10143.7 | 5576.1 | 15719.8 | 0.38 | 0.89 |
| (0.2,0.4) | | | * | | | 10471.5 | 5560.8 | 16032.3 | 0.39 | 0.89 |
| (0.4,0.2) | | | * | | | 10239.2 | 5602.5 | 15841.7 | 0.43 | 0.89 |

TABLE VI
IEEE 33-BUS SYSTEM: STORAGE VS. NO STORAGE ($\eta = 0.2$)

| $(\varrho_1, \varrho_2)$ | without energy storage | | | | | with energy storage | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| (0.5,0.5) | 10527.7 | 5293.7 | 15821.4 | 0.16 | 0.80 | 9903.2 | 5460.4 | 15363.6 | 0.22 | 0.82 |
| (0.4,0.6) | 10332.5 | 5268.3 | 15600.8 | 0.20 | 0.79 | 9764.0 | 5503.9 | 15267.9 | 0.26 | 0.80 |
| (0.6,0.4) | 10200.4 | 5372.9 | 15573.3 | 0.21 | 0.80 | 9672.4 | 5423.5 | 15095.9 | 0.23 | 0.81 |
| (0.3,0.3) | | | * | | | 10021.3 | 5690.3 | 15711.6 | 0.33 | 0.80 |
| (0.2,0.4) | | | * | | | 10206.5 | 5575.1 | 15781.6 | 0.40 | 0.80 |
| (0.4,0.2) | | | * | | | 10164.8 | 5625.0 | 15789.8 | 0.38 | 0.81 |

TABLE VII
IEEE 123-BUS SYSTEM: STORAGE VS. NO STORAGE ($\eta = 0.1$)

| $(\varrho_1, \varrho_2)$ | without energy storage | | | | | with energy storage | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| (0.5,0.5) | 28274.9 | 19707.3 | 47982.2 | 0.44 | 0.89 | 26922.5 | 20311.6 | 47234.1 | 0.46 | 0.90 |
| (0.4,0.6) | 27944.6 | 19638.2 | 47582.8 | 0.37 | 0.89 | 26327.3 | 20610.4 | 46937.7 | 0.46 | 0.91 |
| (0.6,0.4) | 27826.9 | 19648.2 | 47475.1 | 0.39 | 0.89 | 26281.7 | 20409.3 | 46691.0 | 0.49 | 0.90 |
| (0.3,0.3) | | | * | | | 28037.3 | 20416.3 | 48453.6 | 0.57 | 0.90 |
| (0.2,0.4) | | | * | | | 27614.0 | 20882.5 | 48496.5 | 0.54 | 0.89 |
| (0.4,0.2) | | | * | | | 27932.4 | 20741.3 | 48673.7 | 0.52 | 0.90 |

TABLE VIII
IEEE 123-BUS SYSTEM: STORAGE VS. NO STORAGE ($\eta = 0.2$)

| $(\varrho_1, \varrho_2)$ | without energy storage | | | | | with energy storage | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob | 1st ($) | 2nd ($) | Total cost ($) | Gap | Prob |
| (0.5,0.5) | 27970.4 | 19504.8 | 47475.2 | 0.38 | 0.80 | 26392.7 | 19903.2 | 46295.9 | 0.37 | 0.81 |
| (0.4,0.6) | 27822.6 | 19762.4 | 47585.0 | 0.29 | 0.80 | 27041.5 | 20113.4 | 47154.9 | 0.41 | 0.81 |
| (0.6,0.4) | 28203.0 | 19793.0 | 47996.0 | 0.33 | 0.80 | 26808.3 | 20513.6 | 47321.9 | 0.35 | 0.81 |
| (0.3,0.3) | 29409.4 | 19442.7 | 48852.1 | 0.52 | 0.78 | 27751.4 | 19862.4 | 47613.8 | 0.45 | 0.79 |
| (0.2,0.4) | 29143.6 | 19627.1 | 48770.7 | 0.44 | 0.79 | 27948.3 | 20172.3 | 48120.6 | 0.48 | 0.79 |
| (0.4,0.2) | 29527.1 | 20062.6 | 49589.7 | 0.49 | 0.79 | 27684.0 | 20194.0 | 47878.0 | 0.38 | 0.80 |

Regarding computational efficiency, the optimality gaps for the solutions with ES are no smaller than those without ES, but the difference is negligible. Thus, considering ES installation increases the problem complexity, but not significantly.

When $\varrho_1 + \varrho_2 = 0.6$ (i.e., the DDG and RDG units may be insufficient to satisfy the load), ES units are more crucial. Without ES units, we cannot find a feasible solution within the time limit in any case. This indicates that the problems without ES units are likely to be infeasible. However, with ES units, feasible solutions are found within the time limit, and the actual feasible probabilities of the solutions are very close to the desired probabilities. This is because more active power can be purchased or generated in advance when there are ES units, and thus fulfill the load when the demand is high.

As an illustrative example, we consider the instance with $(\varrho_1, \varrho_2) = (0.6, 0.4)$ and $\eta = 0.2$ in the IEEE 33-Bus system. Without ES units, the solution sites (sizes) RDG units at Buses 4 (7 MW), 9 (6 MW), and 28 (7 MW). With ES units, the result shows that RDG units should be installed at Buses 3 (5 MW), 12 (3 MW), and 29 (7 MW), and that ES units should be installed at Buses 2 (2.3 MW), 12 (1.8 MW), and 27 (2.8 MW). The latter solution agrees with our intuition that storage units placed close to high-load buses play an important role in balancing the supply and demand in the power grid. We also observe that ES installation reduces the total required capacities of RDG units. This reduction lowers the first-stage costs so much that the total costs are reduced, even though the operating costs increase due to the operation of storage units.

Finally, all the above numerical results demonstrate that our proposed TCC model and PSAA approach can effectively deal with the RDG and ES planning problem under significant uncertainties. We note that, although we focus on such a planning problem in this paper, the proposed model and approach can also be applied to other practical problems under uncertainty in the industry. For instance, we can apply the PSAA approach to solve the chance-constrained unit commitment problems in [19] and chance-constrained optimal power flow problems in [14], thereby reducing computational challenge. In general, many practical problems that consider two-stage decision-making under uncertainty may be formulated as a TCC model and solved by the PSAA approach. In addition, our proposed PSAA approach is a data-driven approach because (i) we use historical data to represent the possible scenarios of uncertain parameters and accordingly characterize chance constraints in our model; (ii) we use historical data to estimate the cumulative distribution function of a single random parameter $\xi_1'$ by a non-parametric estimation technique; and (iii) once we obtain the first-stage solution of our model, we use real data to test the effectiveness of the obtained solution, simulating real practices. Such an approach can be applied to a wider range of practical problems.

## VI. CONCLUSIONS

Distribution grid operators face great challenges in deciding the locations and capacities of RDG and ES units due to significant uncertainties and complexities of distribution

systems (e.g., ACOPF). To support such a decision-making problem, we develop a novel TCC model to ensure system reliability, minimize costs, and improve renewable energy penetration. One key feature of our model is that the chance constraint ensures that all the operational constraints are satisfied simultaneously with a high probability, leading to system reliability. We use two sampling techniques to reformulate our developed model, leading to the standard SAA formulation and our proposed PSAA formulation. The novelty of the PSAA formulation is that it introduces only continuous variables corresponding to the samples (as compared to integer variables in the SAA formulation) and uses historical data to improve its performance. Our extensive experiments show that the PSAA formulation performs better than the SAA formulation. The PSAA provides better locations and capacities of the RDG and ES units in a shorter time with a lower total cost and achieves a better desired probability of ensuring system feasibility than the SAA. The PSAA also reduces the optimality gap by more than 50% as compared to the SAA. We finally demonstrate the significance of ES units in reducing total costs and improving the power system balance.

This research can be extended in various directions. First, as our proposed TCC model and PSAA approach is general enough, it would be interesting to apply the TCC model and PSAA approach to solve other practical problems in power system planning and operations. Second, the PSAA approach always finds a better solution in a shorter time than the SAA approach in our numerical experiments, but we do not have a theoretical proof for such results. A theoretical study would be appealing. Third, although we consider a radial distribution network in this paper, there can be other types of distribution networks, e.g., meshed grids [26] and multiphase grids [44]. One can apply various approximations (e.g., semidefinite programming [45]) to formulate the corresponding ACOPF constraints. Fourth, although we adopt Bender's decomposition algorithms to improve computational efficiency, more advanced algorithms can be developed. We leave them to future research.

## REFERENCES


[1] A. Keane *et al.*, "State-of-the-art techniques and challenges ahead for distributed generation planning and optimization," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1493–1502, 2013.

[2] P. S. Georgilakis and N. D. Hatziargyriou, "Optimal distributed generation placement in power distribution networks: Models, methods, and future research," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3420–3428, 2013.

[3] C. Abbey and G. Joos, "Supercapacitor energy storage for wind energy applications," *IEEE Trans. Ind Appl.*, vol. 43, no. 3, pp. 769–776, 2007.

[4] J. P. Barton and D. G. Infield, "Energy storage and its use with intermittent renewable energy," *IEEE Trans. Energy Convers.*, vol. 19, no. 2, pp. 441–448, 2004.

[5] X. Kuang, B. Ghaddar, J. Naoum-Sawaya, and L. F. Zuluaga, "Alternative LP and SOCP hierarchies for ACOPF problems," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2828–2836, 2016.

[6] A. Ehsan and Q. Yang, "Optimal integration and planning of renewable distributed generation in the power distribution networks: A review of analytical techniques," *Appl. Energy*, vol. 210, pp. 44–59, 2018.

[7] V. Oree, S. Z. S. Hassen, and P. J. Fleming, "Generation expansion planning optimisation with renewable energy integration: A review," *Renew. Sustain. Energy Rev.*, vol. 69, pp. 790–803, 2017.

[8] G. Celli, E. Ghiani, S. Mocci, and F. Pilo, "A multiobjective evolutionary algorithm for the sizing and siting of distributed generation," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 750–757, 2005.

[9] O. D. Melgar-Dominguez, M. Pourakbari-Kasmaei, and J. R. S. Mantovani, "Adaptive robust short-term planning of electrical distribution systems considering siting and sizing of renewable energy based dg units," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 158–169, 2018.

[10] H. Pandžić, Y. Wang, T. Qiu, Y. Dvorkin, and D. Kirschen, "Near-optimal method for siting and sizing of distributed storage in a transmission network," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2288–2300, 2014.

[11] A. M. Fathabad, J. Cheng, K. Pan, and F. Qiu, "Data-driven planning for renewable distributed generation integration," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4357–4368, 2020.

[12] Z. Zhou, J. Zhang, P. Liu, Z. Li, M. Georgiadis, and E. Pistikopoulos, "A two-stage stochastic programming model for the optimal design of distributed energy systems," *Appl. Energy*, vol. 103, pp. 135–144, 2013.

[13] Y. Tan, Y. Cao, C. Li, Y. Li, J. Zhou, and Y. Song, "A two-stage stochastic programming approach considering risk level for distribution networks operation with wind power," *IEEE Syst. J.*, vol. 10, no. 1, pp. 117–126, 2014.

[14] L. Roald and G. Andersson, "Chance-constrained AC optimal power flow: Reformulations and efficient algorithms," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2906–2918, 2017.

[15] M. Lubin, Y. Dvorkin, and L. Roald, "Chance constraints for improving the security of AC optimal power flow," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1908–1917, 2019.

[16] A. M. Fathabad, J. Cheng, K. Pan, and B. Yang, "Asymptotically tight conic approximations for chance-constrained AC optimal power flow," *Eur. J. Oper. Res.*, vol. 0, no. 0, 2022.

[17] U. A. Ozturk, M. Mazumdar, and B. A. Norman, "A solution to the stochastic unit commitment problem using chance constrained programming," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1589–1598, 2004.

[18] H. Wu, M. Shahidehpour, Z. Li, and W. Tian, "Chance-constrained day-ahead scheduling in stochastic power system operation," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1583–1591, 2014.

[19] Q. Wang, Y. Guan, and J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 206–215, 2012.

[20] Y. Zhang, J. Wang, B. Zeng, and Z. Hu, "Chance-constrained two-stage unit commitment under uncertain load and wind output using bilinear benders," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3637–3647, 2017.

[21] B. Zeng, Y. An, and L. Kuznia, "Chance constrained mixed integer program: Bilinear and linear formulations, and benders decomposition," *arXiv preprint arXiv:1403.7875*, 2014.

[22] D. Pozo and J. Contreras, "A chance-constrained unit commitment with an $n-k$ security criterion and significant wind generation," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2842–2851, 2012.

[23] Z. Wu, P. Zeng, X.-P. Zhang, and Q. Zhou, "A solution to the chance-constrained two-stage stochastic program for unit commitment with wind integration," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4185–4196, 2016.

[24] A. Nemirovski and A. Shapiro, "Convex approximations of CC programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, 2007.

[25] T. Ackermann, G. Andersson, and L. Söder, "Distributed generation: a definition," *Electr. Power Syst. Res.*, vol. 57, no. 3, pp. 195–204, 2001.

[26] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification—part I," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2554–2564, 2013.

[27] M. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Trans. Power Deliv.*, vol. 4, no. 1, pp. 735–743, 1989.

[28] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, "Sample average approximation method for chance constrained programming: theory and applications," *J. Optim. Theory Appl.*, vol. 142, no. 2, pp. 399–416, 2009.

[29] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I—convex underestimating problems," *Math. Program.*, vol. 10, no. 1, pp. 147–175, 1976.

[30] J. Cheng, C. Gicquel, and A. Lisser, "Partial sample average approximation method for chance constrained problems," *Optim. Lett.*, vol. 13, no. 4, pp. 657–672, 2019.

[31] X. Liu, S. Küçükyavuz, and J. Luedtke, "Decomposition algorithms for two-stage chance-constrained programs," *Math. Program.*, vol. 157, no. 1, pp. 219–243, 2016.

[32] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Stat.*, vol. 27, no. 3, pp. 832–837, 1956.

[33] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, 1962.

[34] R. Karimi, J. Cheng, and M. A. Lejeune, "A framework for solving chance-constrained linear matrix inequality programs," *INFORMS J. Comput.*, vol. 33, no. 3, pp. 1015–1036, 2021.
[35] "High performance computing cluster of leria," 2018, slurm/debian cluster of 27 nodes(700 logical CPU, 2 nvidia GPU tesla k20m, 1 nvidia P100 GPU), 120TB of beegfs scratch storage.
[36] H. Park, R. Baldick, and D. P. Morton, "A stochastic transmission planning model with dependent load and wind forecasts," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3003–3011, 2015.
[37] Z. Qin, W. Li, and X. Xiong, "Generation system reliability evaluation incorporating correlations of wind speeds with different distributions," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 551–558, 2012.
[38] K. Qian, C. Zhou, M. Allan, and Y. Yuan, "Modeling of load demand due to EV battery charging in distribution systems," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 802–810, 2010.
[39] M. Bazrafshan and N. Gatsis, "Comprehensive modeling of three-phase distribution systems via the bus admittance matrix," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2015–2029, 2017.
[40] N. Daratha, B. Das, and J. Sharma, "Coordination between oltc and svc for voltage regulation in unbalanced distribution system distributed generation," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 289–299, 2013.
[41] S. Zhang, H. Cheng, L. Zhang, M. Bazargan, and L. Yao, "Probabilistic evaluation of available load supply capability for distribution system," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3215–3225, 2013.
[42] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, no. 1, pp. 238–252, 1962.
[43] A. Ben-Tal and A. Nemirovski, "On polyhedral approximations of the second-order cone," *Math. Oper. Res.*, vol. 26, no. 2, pp. 193–205, 2001.
[44] L. Gan and S. H. Low, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *2014 Power Syst. Comput. Conf.* IEEE, 2014, pp. 1–9.
[45] S. H. Low, "Convex relaxation of optimal power flow—part I: Formulations and equivalence," *IEEE Trans. Control. Netw. Syst.*, vol. 1, no. 1, pp. 15–27, 2014.