DATA-DRIVEN CHANCE-CONSTRAINED PLANNING FOR DISTRIBUTED GENERATION: A PARTIAL SAMPLING APPROACH

SHIYI JIANG

MPhil

The Hong Kong Polytechnic University

2023

The Hong Kong Polytechnic University
Department of Logistics and Maritime Studies

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

Shiyi Jiang

A thesis submitted in partial fulfillment of the requirements for
the degree of Master of Philosophy
May 2023

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

Signature: _____

Name of Student: Shiyi Jiang

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 thesis, 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.

Acknowledgments

First, I am extremely grateful to my chief supervisor, Dr. Kai Pan, who has kindly provided me with guidance and support. It is my honor to benefit from his supervision and suggestions. I also want to give my heartfelt thanks to Dr. Jianqiang Cheng and Dr. Boshi Yang. They have helped me a lot in my research. Besides, I am full of gratitude to all teachers in our department. Of course, I am full of gratitude to my parents and girlfriend. Without their support and encouragement, I couldn't have done all my work. I hereby give you my genuine blessing.

Table of Contents

List of Figures

List of Tables

Chapter 1

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 [21], 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 [19]. 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, [2] adopts ES to help support wind energy applications and [5] 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 [22], 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 thesis 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 [14] provides a review of related studies on the RDG planning in the power distribution network, and [30] reviews related studies on more general generation expansion planning. Specifically, [11] 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, [27] considers the siting and sizing of RDG units in a power distribution grid in a two-stage robust optimization model. The paper [34] 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 [16], 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, [53] proposes a two-stage stochas-

tic programming model for the optimal planning of distributed energy systems under demand and supply uncertainties. The paper [45] proposes a two-stage stochastic programming-based optimal power flow model for the operation of distribution networks with uncertainties from wind power.

In this thesis, 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 are 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, [41] solves the chance-constrained ACOPF problems to ensure that operational constraints are satisfied with the desired probability. The paper [25] further provides convex approximations via second-order conic programming, and [17] provides asymptotically tight conic approximations for the chance-constrained ACOPF problems. Similarly, [31] studies the chance-constrained unit commitment formulations, and [47] 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, [46] 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 [52] 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 [50]. The paper [37] 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 [48] 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 thesis 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 [29]. The existing studies primarily adopt two approximation approaches to solve a chance-constrained model: a convex (or tractable) approximation approach [25] and a sampling-based approach [46]. 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 thesis 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 thesis 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 thesis is organized as follows. A two-stage model and its TCC counterpart for the planning problem are presented in Chapter 2. The TCC model is then approximated using the SAA and PSAA methods in Chapter 3. We provide computational results and explanations in Chapter 4. Chapter 5 concludes this thesis.

Chapter 2

Mathematical Model

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

2.1 Physical Model

In the planning problem, we consider the siting and sizing of RDG and ES units in the radial distribution system. This kind of network has a tree structure and connects to a transmission network via a single bus (Bus 0). We use $G = (\mathcal{N}, \mathcal{E})$, which is a connected graph, to represent this power network. Each node in $\mathcal{N}$ represents a bus and each link in $\mathcal{E}$ represents a line. Then, we can model the distribution system in two stages.

Assume that DDG units and reactive sources have been placed, in the first stage, we plan to place the RDG and ES units. We let z_{kn} be a binary variable indicating whether the kth candidate RDG unit is located at bus n and w_{rn} be a binary variable indicating whether the rth candidate ES unit is located at Bus n. Besides, x_k and y_r are continuous variables indicating the capacities of kth candidate RDG unit and rth candidate ES unit, respectively. Now, the aim of the first stage is to minimize the setup costs and investment/maintenance costs of RDG and ES units. At the same time, we consider the following constraints in the first stage,

$$\sum_{n \in \mathcal{N}} z_{kn} \leq 1, \ \forall\, k \in [K], \tag{2.1}$$

$$\sum_{k=1}^{K} \sum_{n \in \mathcal{N}} z_{kn} \leq \bar{K}, \tag{2.2}$$

$$x_k = \sum_{l=1}^{L} u_{kl} \bar{x}_l, \ \forall\, k \in [K], \tag{2.3}$$

$$\sum_{l=1}^{L} u_{kl} = \sum_{n \in \mathcal{N}} z_{kn}, \ \forall\, k \in [K], \tag{2.4}$$

$$\sum_{n \in \mathcal{N}} w_{rn} \leq 1, \ \forall\, r \in [R], \tag{2.5}$$

$$\sum_{n \in \mathcal{N}} w_{rn} \underline{y}_r \leq y_r \leq \sum_{n \in \mathcal{N}} w_{rn} \bar{y}_r, \ \forall\, r \in [R]. \tag{2.6}$$

Constraints (2.1) show that a given kth candidate RDG unit, if installed, should be in one of the buses in $\mathcal{N}$. Constraint (2.2) enforces that the total number of RDG units installed should not exceed the limit $\bar{K}$.

In constraints (2.3) and (2.4), u_{kl} is a binary variable indicating whether the capacity of the kth candidate RDG unit is the lth element in $\mathcal{X}$, $\bar{x}_l$. Thus, constraints (2.3) and (2.4) ensure that if a given kth candidate RDG unit is installed, then the capacity of this RDG unit should be one of the pre-defined values in $\mathcal{X}$. For any given kth candidate RDG unit, the summation $\sum_{l=1}^{L} u_{kl}$ is equivalent to the $\sum_{n \in \mathcal{N}} z_{kn}$, which may take only the value 0 or 1. Such equivalence is not related to the locations of candidate RDG units. Constraints (2.3) and (2.4) are motivated by the practice where various regions and institutions have different regulations on the capacity of distributed generation, and thus the capacity of RDG units vary [3].

Similar to (2.1), constraints (2.5) show that a given rth candidate ES unit, if installed, should be in one of the buses in $\mathcal{N}$. Constraints (2.6) show that if a given rth candidate ES unit is installed (i.e., $w_{rn} = 1$ for a bus n), then its capacity should be between the lower bound $\underline{y}_r$ and the upper bound $\bar{y}_r$. Note that multiple new assets (including RDG and ES units) may be installed eventually.

In the second stage, we minimize the operating costs and consider the ACOPF constraints such that the system can operate normally. The bus injection model and the branch flow model are two standard models for power flow analysis. The bus injection model focuses on nodal variables such as voltages, current and power injections and does not directly deal with power flows on individual branches. The branch flow model focuses on currents and powers on the

branches, which have been used mainly for modeling distribution systems. In this thesis, we focus on the radial network (because most distribution systems are radial) and use the branch flow model to analyze the power flow.

For each $(m,n) \in \mathcal{E}$, we let I_{mn} be the complex current form buses m to n, $z_{mn} = \mathfrak{R}_{mn} + \mathbf{i}\mathfrak{X}_{mn}$ be the complex impedance on the line, and $S_{mn} = P_{mn} + \mathbf{i}Q_{mn}$ be the sending-end complex power from buses m to n. For each node $n \in \mathcal{N}$, we let V_n be the complex voltage on bus n, $s_n = p_n + \mathbf{i}q_n$ be the net complex power injection, which is generation minus load on bus n. Then, we build the power flow model by the following steps.

First, these variables should satisfy Ohm's law:

$$V_m - V_n = z_{mn} I_{mn}, \ \forall (m,n) \in \mathcal{E}, \tag{2.7}$$

the definition of branch power flow:

$$S_{mn} = V_m I_{mn}^*, \ \forall (m,n) \in \mathcal{E}, \tag{2.8}$$

and power balance at each bus:

$$\sum_{n:m\to n} S_{mn} - \sum_{n:n\to m} \left(S_{nm} - z_{nm}\|I_{nm}\|^2\right) = s_m, \ \forall m \in \mathcal{N}. \tag{2.9}$$

These three equations are branch flow equations. Second, we impose the output constraints on power generation:

$$\underline{\mathrm{p}}_n \le p_n \le \bar{p}_n, \ \underline{\mathrm{q}}_n \le q_n \le \bar{q}_n, \ \forall\, n \in \mathcal{N}, \tag{2.10}$$

and the voltage magnitudes must be maintained in tight ranges:

$$\underline{v} \le |V_n|^2 \le \bar{v}, \ \forall\, n \in \mathcal{N} \setminus \{0\}. \tag{2.11}$$

Finally, we impose flow limits on branch currents:

$$\|I_{mn}\| \le LC_{mn}, \ \forall (m,n) \in \mathcal{E}. \tag{2.12}$$

Now, (2.7)–(2.12) are constraints that we should consider in the planning problem.

Before we move to the two-stage model, we reformulate constraints (2.7)–(2.9) into convex constraints. We first substitute (2.8) into (2.7) and obtain $V_n = V_m - z_{mn}S^*_{mn}/V^*_m$. Taking the magnitude squared, we obtain $\|V_n\|^2 = \|V_m\|^2 + \|z_{mn}\|^2\|I_{mn}\|^2 - (z_{mn}S^*_{mn} + z^*_{mn}S_{mn})$. Considering real variables in (2.8) and (2.9), we have the following relaxed branch flow equations:

$$\sum_{n:m\to n} P_{mn} - \sum_{n:n\to m} \left(P_{nm} - z_{nm}\|I_{nm}\|^2\right) = p_m,\ \forall m \in \mathcal{N}, \tag{2.13}$$

$$\sum_{n:m\to n} Q_{mn} - \sum_{n:n\to m} \left(Q_{nm} - z_{nm}\|I_{nm}\|^2\right) = q_m,\ \forall m \in \mathcal{N}, \tag{2.14}$$

$$\|V_n\|^2 = \|V_m\|^2 - 2\left(\mathfrak{R}_{mn}P_{mn} + \mathfrak{X}_{mn}Q_{mn}\right) + \left(\mathfrak{R}^2_{mn} + \mathfrak{X}^2_{mn}\right)\|I_{mn}\|^2,\ \forall (m,n) \in \mathcal{E}, \tag{2.15}$$

$$\|I_{mn}\|^2 = \frac{P^2_{mn} + Q^2_{mn}}{\|V_m\|^2},\ \forall (m,n) \in \mathcal{E}. \tag{2.16}$$

In the end, we relax the (2.16) to inequalities:

$$\|I_{mn}\|^2 \geq \frac{P^2_{mn} + Q^2_{mn}}{\|V_m\|^2},\ \forall (m,n) \in \mathcal{E}. \tag{2.17}$$

[15] show that all approximations and relaxations are actually exact when we consider the radial network. Thus, we use constraints (2.13)–(2.15), (2.17), and (2.10)–(2.12) to model the physical distribution system.

2.2 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 ES units. The last three sources are located in some buses of the distribution grid. While the DDG units and reactive sources have already been placed (i.e., given system input data), the RDG and ES units are to be installed (i.e., system decision variables). 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 $\bar{K}$ candidate RDG units and R candidate 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.

$$\min_{\boldsymbol{\Omega}^1} \; C_1\left(\boldsymbol{\Omega}^1\right) + \mathbb{E}\left[Q\left(\boldsymbol{\Omega}^1, \boldsymbol{\xi}\right)\right] \tag{2.18a}$$

$$\text{s.t.} \sum_{n\in\mathcal{N}} z_{kn} \le 1, \; \forall\, k \in [K], \tag{2.18b}$$

$$\sum_{k=1}^{K}\sum_{n\in\mathcal{N}} z_{kn} \le \bar{K}, \tag{2.18c}$$

$$x_k = \sum_{l=1}^{L} u_{kl}\bar{x}_l, \; \forall\, k \in [K], \tag{2.18d}$$

$$\sum_{l=1}^{L} u_{kl} = \sum_{n\in\mathcal{N}} z_{kn}, \; \forall\, k \in [K], \tag{2.18e}$$

$$\sum_{n\in\mathcal{N}} w_{rn} \le 1, \; \forall\, r \in [R], \tag{2.18f}$$

$$\sum_{n\in\mathcal{N}} w_{rn}\underline{y}_r \le y_r \le \sum_{n\in\mathcal{N}} w_{rn}\bar{y}_r, \; \forall\, r \in [R], \tag{2.18g}$$

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 (2.18a)

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

represents the total deterministic cost, including the setup costs and the size-based investment/maintenance costs of the RDG and ES units. Specifically, $C_1\left(\boldsymbol{\Omega}^1\right)$ includes two parts:

(i) the setup costs of the RDG and ES units, $\sum_{k=1}^{K}\left(\sum_{n\in\mathcal{N}} c_{kn}^{0} z_{kn}\right) + \sum_{r=1}^{R}\left(\sum_{n\in\mathcal{N}} d_{rn}^{0} w_{rn}\right)$, and (ii) the investment/maintenance costs of the RDG and ES units, $\sum_{k=1}^{K}\left(\left(c_k^1 + T c_k^2\right) x_k\right) + \sum_{r=1}^{R}\left(\left(d_r^1 + T d_r^2\right) y_r\right)$. In part (i), z_{kn} is a binary variable indicating whether the kth candidate RDG unit is located at bus n and w_{rn} is a binary variable indicating whether the rth candidate ES unit is located at Bus n. In part (ii), x_k and y_r are continuous variables indicating the capacities of kth candidate RDG unit and rth candidate ES unit, respectively.

The second part in the objective function (2.18a), $\mathbb{E}[Q(\boldsymbol{\Omega}^1, \boldsymbol{\xi})]$, represents the expected minimum operating costs over all T periods, which is defined explicitly in (2.19). Regarding the constraints (2.18b) - (2.18g), they link the variables z_{kn}, w_{rn}, x_k, y_r, and u_{kl}, by which the setup costs are linked with the investment/maintenance cost. Specifically, constraints (2.18b) show that a given kth candidate RDG unit, if installed, should be in one of the buses in $\mathcal{N}$. Constraint (2.18c) enforces that the total number of RDG units installed should not exceed the limit $\bar{K}$.

In constraints (2.18d) and (2.18e), u_{kl} is a binary variable indicating whether the capacity of the kth candidate RDG unit is the lth element in $\mathcal{X}$, $\bar{x}_l$. Thus, constraints (2.18d) and (2.18e) ensure that if a given kth candidate RDG unit is installed, then the capacity of this RDG unit should be one of the pre-defined values in $\mathcal{X}$. For any given kth candidate RDG unit, the summation $\sum_{l=1}^{L} u_{kl}$ is equivalent to the $\sum_{n\in\mathcal{N}} z_{kn}$, which may take only the value 0 or 1. Such equivalence is not related to the locations of candidate RDG units. Constraints (2.18d) and (2.18e) are motivated by the practice where various regions and institutions have different regulations on the capacity of distributed generation, and thus the capacity of RDG units vary [3].

Similar to (2.18b), constraints (2.18f) show that a given rth candidate ES unit, if installed, should be in one of the buses in $\mathcal{N}$. Constraints (2.18g) show that if a given rth candidate ES unit is installed (i.e., $w_{rn} = 1$ for a bus n), then its capacity should be between the lower bound $\underline{y}_r$ and the upper bound $\bar{y}_r$. Note that multiple new assets (including RDG and ES units) may be installed eventually.

From the above constraints, we can observe that z_{kn} impacts x_k and w_{rn} impacts y_r, i.e., the binary variables indicating the location of new assets have an impact on the investment/maintenance cost.

2) **Second-stage model**: Given a first-stage decision $\mathbf{\Omega}^1$ and a realization $\boldsymbol{\xi}$ of the uncertain load and renewable generation, the second-stage objective minimizes the distribution grid's operating costs $Q(\mathbf{\Omega}^1, \boldsymbol{\xi})$, 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.

As for constraints, we mainly use ACOPF constraints to model the distribution network. We begin by representing the distribution network's physical layout, including its components such as buses, branches, transformers, loads, distributed generators, RDG units, and ES units. This representation can be done using a graph-based approach, where buses represent the network nodes, and branches represent the network edges. We then use the AC power flow equations to describe the relationship between the voltage magnitudes, voltage angles, and power injections at each bus in the distribution network. These equations ensure that power flow is balanced across the network and that Kirchhoff's laws are satisfied. Next, we consider other distribution system constraints. These constraints can include voltage magnitude limits, current limits on branches, power factor limits, and other operational constraints. These constraints ensure that the distribution system operates within its physical limits and complies with operational requirements. The output of each distributed generator is bounded below and above. In the end, we consider the constraints of ES units. We explain the specific constraints later.

Let $\mathbf{\Omega}^2$ be the vector of all second-stage variables and let

$$C_2\left(\mathbf{\Omega}^2\right) := \sum_{t=1}^{T} \left(c_p^t p_0^t + c_q^t q_0^t + \sum_{n\in\mathcal{B}_1} c_n^f p_n^t + \sum_{n\in\mathcal{B}_1} c_n^e \omega p_n^t + \sum_{r=1}^{R} e_1 f_r^t + \sum_{r=1}^{R} e_2 g_r^t + \sum_{(m,n)\in\mathcal{E}} p\left(LS^t_{1mn} + LS^t_{2mn}\right) \right)$$

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

with a superscript t hold for all $t \in [T]$.)

$$\min_{\mathbf{\Omega}^2} \quad C_2\left(\mathbf{\Omega}^2\right) \tag{2.19a}$$

$$\text{s.t.} \quad \underline{\mathrm{p}}_n^t \le p_n^t \le \bar{p}_n^t,\ \forall\, n \in \mathcal{B}_1, \tag{2.19b}$$

$$\underline{\mathrm{q}}_n^t \le q_n^t \le \bar{q}_n^t,\ \forall\, n \in \mathcal{B}_2, \tag{2.19c}$$

$$\underline{v} \le \left|V_n^t\right|^2 \le \bar{v},\ \forall\, n \in \mathcal{N} \setminus \{0\}, \tag{2.19d}$$

$$p_0^t = \sum_{n \in \mathcal{B}_0} P_{0n}^t, \quad q_0^t = \sum_{n \in \mathcal{B}_0} Q_{0n}^t, \tag{2.19e}$$

$$\begin{aligned} P_{mn}^t - \mathfrak{R}_{mn}\left|I_{mn}^t\right|^2 + LS_{1mn}^t = d_{pn}^t - \sum_{k=1}^{K} z_{kn} s_k^t x_k \\ - \delta_n p_n^t + \sum_{l \in \mathcal{N}_n} P_{nl}^t + \sum_{r=1}^{R} w_{rn}\left(f_r^t - g_r^t\right), \\ \forall\, (m,n) \in \mathcal{E}, \end{aligned} \tag{2.19f}$$

$$\begin{aligned} Q_{mn}^t - \mathfrak{X}_{mn}\left|I_{mn}^t\right|^2 + LS_{2mn}^t = d_{qn}^t - \tau_n q_n^t \\ + \sum_{l \in \mathcal{N}_n} Q_{nl}^t,\ \forall\, (m,n) \in \mathcal{E}, \end{aligned} \tag{2.19g}$$

$$b_r^0 = 0,\ \forall\, r \in [R], \tag{2.19h}$$

$$0 \le b_r^t \le y_r,\ \forall\, r \in [R], \tag{2.19i}$$

$$b_r^t - b_r^{t-1} = \gamma f_r^t - g_r^t/\gamma,\ \forall\, r \in [R], \tag{2.19j}$$

$$\begin{aligned} \left|V_m^t\right|^2 - \left|V_n^t\right|^2 = 2\mathfrak{R}_{mn} P_{mn}^t + 2\mathfrak{X}_{mn} Q_{mn}^t \\ - \left(\mathfrak{R}_{mn}^2 + \mathfrak{X}_{mn}^2\right)\left|I_{mn}^t\right|^2,\ \forall\, (m,n) \in \mathcal{E}, \end{aligned} \tag{2.19k}$$

$$\begin{aligned} \left\|\left[2P_{mn}^t, 2Q_{mn}^t, \left|V_m^t\right|^2 - \left|I_{mn}^t\right|^2\right]\right\|_2 \le \left|V_m^t\right|^2 + \left|I_{mn}^t\right|^2 \\ \forall\, (m,n) \in \mathcal{E}, \end{aligned} \tag{2.19l}$$

$$\left(P_{mn}^t\right)^2 + \left(Q_{mn}^t\right)^2 \le (LC_{mn})^2,\ \forall\, (m,n) \in \mathcal{E}, \tag{2.19m}$$

$$p_0^t \ge 0,\ q_0^t \ge 0, \tag{2.19n}$$

$$LS_{1mn}^t \ge 0,\ LS_{2mn}^t \ge 0,\ \forall\, (m,n) \in \mathcal{E}. \tag{2.19o}$$

Here, $\mathbf{\Omega}^2$ consists of p_0^t, q_0^t for $t \in [T]$, $p_n^t, q_n^t, V_n^t, P_{0n}^t, Q_{0n}^t$ for $n \in \mathcal{N}$, $t \in [T]$, f_r^t, g_r^t, b_r^t for $r \in [R], t \in [T]$, P_{nl}^t, Q_{nl}^t for n such that $(m,n) \in \mathcal{E}$ for some m, $l \in \mathcal{N}_n$, $t \in [T]$, and $P_{mn}^t, Q_{mn}^t, I_{mn}^t, LS_{1mn}^t, LS_{2mn}^t$ for $(m,n) \in \mathcal{E}, t \in [T]$.

We explain all the constraints in the model (2.19) as follows. The power generated by the DDG

units and the reactive sources is bounded by (2.19b) and (2.19c), respectively. Constraint (2.19d) sets the bounds on the voltage of each bus. Constraint (2.19e) represents the active and reactive balance equations at Bus 0. Constraints (2.19f) and (2.19g) are active and reactive power balance equations from Kirchhoff's current law. The following Fig. 2.1 illustrates the active power flow balance for each distribution line $(m, n) \in \mathcal{E}$. The balance of power in storage

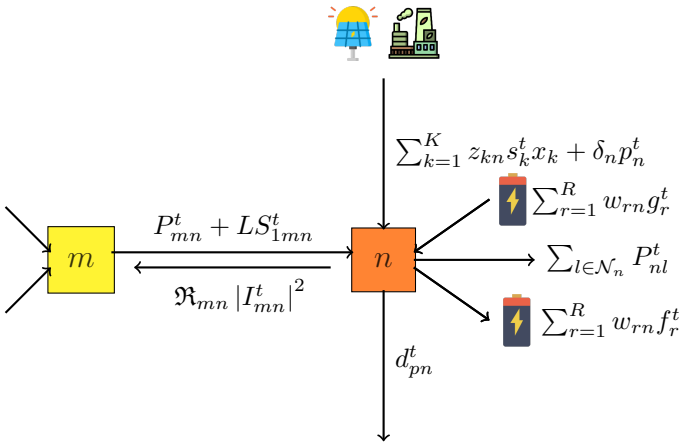

Figure 2.1: Power Flow Blance.

is initialized by (2.19h) and bounded by (2.19i). ES balance between two consecutive periods is shown in (2.19j), considering ES charging/discharging efficiency. Constraint (2.19k) represents the voltage drop on each line. Constraint (2.19l) is the branch power-flow constraint, and the capacity of each distribution line is limited by (2.19m). Nonnegativity constraints are listed in (2.19n) and (2.19o).

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 (2.19f) – (2.19g) and (2.19k) – (2.19m), where (2.19l) includes a set of a second-order conic (SOC) constraints. These constraints represent the ACOPF constraints via convex relaxation following the study in [15]. Specifically, [15] 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, [15] shows that the obtained convex relaxation is exact when the distribution network is radial. As most practical distribution networks are radial grids [15], we also consider a radial network in this thesis. Therefore, the constraints (2.19f) – (2.19g) and (2.19k) – (2.19m) form an exact reformulation

of the ACOPF constraints. As such, we obtain a second-order conic programming (SOCP) formulation in (2.19), which enables large-scale applications due to the computational efficiency of SOCP formulations.

Indeed, there are several linearized power flow models that have been developed to approximate the behavior of distribution systems and facilitate faster analysis and optimization. Some of these models include:

Linearized Distribution Flow (LinDistFlow) [4]: LinDistFlow is a simplified ACOPF model that approximates the nonlinear power flow equations of distribution systems by linearizing the relationships between voltage drop and line power flows with respect to power injections. It considers power injections at buses, resistive and reactive line losses, and voltage drops across the network. However, it assumes that the network's impedance matrix remains constant and neglects the impact of voltage magnitude and reactive power variations on line impedance. This linearization simplifies the calculations and allows for faster analysis compared to solving the full nonlinear equations.

Linear Optimal Power Flow for Distribution (LOPF-D) [49]: LOPF-D is another linear optimization model used in distribution systems to determine the optimal operation of distributed energy resources, such as renewable generation, energy storage, and controllable loads. This model also linearizes the power flow equations and approximates the distribution network's behavior, considering the linear relationship between power injections, line losses, and voltage variations. It formulates the problem as a linear optimization program and provides reasonably accurate solutions when compared to ACOPF, while DCOPF tends to produce results with larger errors.

Linearized OPF Formulation for Unbalanced Distribution Systems: This novel linearized OPF formulation, as explored in [43], specifically focuses on unbalanced distribution systems. It approximates the ratios of voltages across phases in the feeder and generalizes the model to use arbitrary complex numbers to approximate voltage ratios at different locations in the system. This approach enables linearization of the unbalanced power flow equations, making it feasible to analyze and optimize unbalanced distribution networks.

These linearized power flow models provide efficient and computationally tractable alternatives

to the full nonlinear ACOPF model for distribution system analysis and optimization. They strike a balance between accuracy and computational complexity, allowing for faster solution times while still capturing key aspects of the distribution network's behavior. However, it's important to note that these linearized models are based on approximations, and their accuracy may vary depending on the specific characteristics and operating conditions of the distribution system. As such, their suitability should be evaluated based on the specific application and requirements of the analysis or optimization task at hand. For more comprehensive details on related models, the reader can refer to [28].

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

$$\begin{aligned} \min_{\mathbf{\Omega}^1, \mathbf{\Omega}^2} \quad & C_1\left(\mathbf{\Omega}^1\right) + \mathbb{E}\left[C_2\left(\mathbf{\Omega}^2\right)\right] \\ \text{s.t.} \quad & (2.18\text{b}) - (2.18\text{g}), (2.19\text{b}) - (2.19\text{o}). \end{aligned} \tag{2.20}$$

Note that $\mathbf{\Omega}^2$ and constraints (2b) – (2o) are dependent on $\boldsymbol{\xi}$. As the first-stage problem (2.18) is an integer program and the second-stage problem (2.19) is an SOCP, the entire two-stage problem (2.20) is a mixed-integer SOCP.

2.3 A TCC Model

In Section 2.2, we build the two-stage stochastic programming model (2.20) for the planning problem. However, the two-stage stochastic programming model is hard to capture all features of the planning problem. It is better to consider the two-stage chance-constrained programming model and it has the following benefits compared with the two-stage stochastic programming model:

- Chance constraints provide a better way to handle uncertainty. The planning problem involves dealing with various sources of uncertainty, such as load demand and renewable energy generation. While two-stage stochastic programming provides a probabilistic representation of uncertainty, it does not explicitly account for the risk associated with constraint violations. On the other hand, a chance-constrained model allows decision-makers to set probabilistic constraints on the system's operation, to ensure that the risk

of violating all constraints remains below acceptable levels. This explicit treatment of risk is essential in the planning problem, where power flow balance constraints are crucial.

- The chance-constrained model enhances the robustness and reliability of decisions. By setting chance constraints on one or more system constraints, decision-makers can explicitly manage and mitigate the risk of infeasible or unreliable decisions. This ensures that the proposed plan has a high probability of meeting operational constraints even under uncertain conditions. In contrast, a two-stage stochastic programming model may not capture all possible scenarios adequately, potentially leading to suboptimal or unreliable plans.

- Chance-constrained models facilitate trade-off analysis between conflicting objectives, such as cost, reliability, and demand. Decision-makers can adjust the level of risk tolerance by modifying the probability thresholds in the chance constraints, allowing them to explore the trade-off between minimizing costs and maintaining system reliability. This capability enables decision-makers to make informed decisions by explicitly considering the risk-reward trade-offs associated with different planning alternatives.

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 (2.19) are satisfied, such that

$$LS^t_{1mn} = LS^t_{2mn} = 0, \ \forall\, (m,n) \in \mathcal{E}, \ t \in [T]. \tag{2.21}$$

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

$$\mathbb{P}\left(\mathbf{\Omega}^2 \text{ in (2.19) satisfies (2.21)} \,|\, \mathbf{\Omega}^1\right) \geq 1 - \eta \tag{2.22}$$

to the two-stage model (2.20) 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 (2.22) consider all the constraints (2.19b) – (2.19o) in the second-stage problem (2.19) as the nominal constraints. For instance, in [46], the authors consider a single demand satisfaction constraint in their chance constraints, see constraints (20) – (22) in [46]. In contrast,

our chance constraint (2.22) requires that (2.21) is satisfied by any $\mathbf{\Omega}^2$ that is feasible to all the constraints in the second-stage model (2.19). That is, we ensure the entire system (rather than the load satisfaction constraints only) is 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:

$$\begin{aligned} \min_{\mathbf{\Omega}^1, \mathbf{\Omega}^2} \quad & C_1\left(\mathbf{\Omega}^1\right) + \mathbb{E}\left[C_2\left(\mathbf{\Omega}^2\right)\right] \\ \text{s.t.} \quad & (2.18\text{b}) - (2.18\text{g}), (2.19\text{b}) - (2.19\text{o}), (2.22). \end{aligned} \tag{2.23}$$

However, the above problem (2.23) is difficult to solve. Specifically, constraint (2.22) is non-convex and thus problem (2.23) becomes intractable. Next, we introduce two approximation methods to address this challenge.

2.3.1 Extensions of TCC Model

Before we move to the solution approaches section, we want to present some different ways to build the TCC model. One possible way is to put chance constraints on soft constraints.

Introducing chance constraints on soft constraints provides a more flexible approach in addressing the planning problems for distribution systems. In such cases, there are two categories of constraints to be satisfied: hard constraints, which must be fully met with no violation, and soft constraints, where a certain degree of violation is acceptable within predefined limits.

The hard constraints typically encompass power flow constraints, siting and sizing limits of DG, RDG, and ES units, as well as operating limits for these units. These constraints are critical for the stable and reliable operation of the distribution system and should be strictly adhered to.

On the other hand, the soft constraints include voltage profile constraints and capacity constraints. Voltage magnitude limits for each bus are essential to maintain a safe operating condition, and these limits are dependent on the specific system conditions. Capacity constraints of power lines are also considered as soft constraints.

Incorporating chance constraints on these soft constraints allows us to model the uncertainties

and tolerate a certain level of violation while still ensuring a high probability of satisfaction. For instance, we can introduce chance constraints for the voltage magnitude at each bus, ensuring that $\mathbb{P}(\boldsymbol{\Omega}^2 \text{ satisfies (2.19d)} \mid \boldsymbol{\Omega}^1) \geq 1-\eta$, where $\boldsymbol{\Omega}^1$ represents the uncertain parameters associated with the system.

Similarly, chance constraints can be applied to the capacity constraints of power lines, ensuring that $\mathbb{P}(\boldsymbol{\Omega}^2 \text{ satisfies (2.19m)} \mid \boldsymbol{\Omega}^1) \geq 1-\eta$. This approach allows for a probabilistic evaluation of the feasibility of the system, considering the uncertainties in the distribution system.

By incorporating chance constraints on soft constraints, we gain the advantage of obtaining a more flexible solution space. This approach provides the ability to handle uncertainties and accommodate minor violations in the voltage profile and line capacity limits while maintaining a high probability of satisfying these constraints. Ultimately, it leads to more robust and efficient planning solutions for distribution systems, striking a balance between reliability and cost-effectiveness.

Chapter 3

Solution Approaches

In this chapter, we describe two approximation methods for solving the problem (2.23): the SAA and PSAA methods.

3.1 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 Π_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 $\mathbb{P}\left\{\boldsymbol{\Omega}^2 \text{ in (2.19) satisfies (2.21)} \,|\, \boldsymbol{\Omega}^1\right\}$ in (2.22) with

$$\frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1}\mathbb{I}\left(\boldsymbol{\Omega}_\pi^2 \text{ satisfies (2.21)} \,|\, \boldsymbol{\Omega}^1\right),$$

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 $(2.19\text{b})_\pi$ – $(2.19\text{o})_\pi$ be constraints (2.19b) – (2.19o) with $\boldsymbol{\xi}$ replaced by $\boldsymbol{\xi}_\pi$ and $\boldsymbol{\Omega}^2$ replaced by

$\mathbf{\Omega}_\pi^2$. The SAA approximation of (2.23) thus becomes:

$$\begin{aligned}\min_{\mathbf{\Omega}^1,\, \mathbf{\Omega}_\pi^2, \forall \pi \in [\Pi_1]} \; & C_1\left(\mathbf{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\mathbf{\Omega}_\pi^2\right) \\ \text{s.t. } & \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} \mathbb{I}\left(\mathbf{\Omega}_\pi^2 \text{ satisfies (2.21)} \,|\, \mathbf{\Omega}^1\right) \geq 1-\eta. \\ & \text{(2.18b)} - \text{(2.18g)},\ \text{(2.19b)}_\pi - \text{(2.19o)}_\pi,\ \forall\, \pi \in [\Pi_1].\end{aligned} \tag{3.1}$$

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 $\mathbb{I}(\mathbf{\Omega}_\pi^2$ satisfies (2.21) $|\, \mathbf{\Omega}^1) = 1$, i.e., all the constraints in the second-stage model (2.19) are satisfied; when $\theta_\pi = 1$, it indicates $\mathbb{I}(\mathbf{\Omega}_\pi^2$ satisfies (2.21) $|\, \mathbf{\Omega}^1) = 0$, i.e., all the constraints in the second-stage model (2.19) are not satisfied. Therefore, we can reformulate (3.1) as the following mixed-integer quadratic program:

$$\min_{\mathbf{\Omega}^1,\, \mathbf{\Omega}_\pi^2, \forall \pi \in [\Pi_1]} \quad C_1\left(\mathbf{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\mathbf{\Omega}_\pi^2\right) \tag{3.2a}$$

$$\text{s.t.} \quad \text{(2.18b)} - \text{(2.18g)}, \tag{3.2b}$$

$$\text{(2.19b)}_\pi - \text{(2.19o)}_\pi,\ \forall\, \pi \in [\Pi_1], \tag{3.2c}$$

$$LS_{1\pi mn}^t \leq \theta_\pi M_\pi,\ LS_{2\pi mn}^t \leq \theta_\pi M_\pi, \quad \forall\, (m,n) \in \mathcal{E}, \forall\, \pi \in [\Pi_1], \tag{3.2d}$$

$$\sum_{\pi=1}^{\Pi_1} \theta_\pi \leq \Pi_1 \eta, \tag{3.2e}$$

$$\theta_\pi \in \{0,1\},\ \forall\ \pi \in [\Pi_1], \tag{3.2f}$$

where M_π is a sufficiently large number for any $\pi \in [\Pi_1]$. Specifically, with constraints (4.1d) and (4.1e), 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 (2.19). That is, with the probability of $1-\eta$, the constraints in the second-stage model (2.19) are satisfied. Moreover, [32] shows that the objective value difference between model (3.2) (i.e., model (3.1)) and model (2.23) converges to zero with probability one when Π_1 goes to infinity. In addition, as in the second-stage model (2.19), all constraints in (3.2) with a superscript t hold for all $t \in [T]$.

All of the constraints in (3.2) are convex except for $\text{(2.19f)}_\pi, \pi \in [\Pi_1]$. Specifically, constraints

(4.1b) are from the first-stage model, and all of them are linear constraints. Constraints (4.1d) - (4.1f) are also linear constraints. Constraints (4.1c), i.e., $(2.19b)_\pi - (2.19o)_\pi$, $\forall \pi \in [\Pi_1]$, are from the second-stage model, and they are either linear or second-order conic (SOC) constraints when the first-stage decision variables are given. However, as problem (3.2) needs to optimize both the first-stage and second-stage decision variables, constraints $(2.19f)_\pi, \pi \in [\Pi_1]$ include bilinear terms ($z_{kn}u_{kl}$ and $w_{rn}f^t_{\pi r}$), by which these constraints are nonconvex. Specifically, after substituting x_k with (2.18d), 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.

McCormick inequalities are commonly used to linearize a bilinear term, say $w = xy$ with $x^L \le x \le x^U$ and $y^L \le y \le y^U$, in (mixed-integer) nonlinear programming [26]. The general form of McCormick inequalities for $w = xy$ can be written as:

$$w \ge x^L y + xy^L - x^L y^L,\ w \ge x^U y + xy^U - x^U y^U,$$
$$w \le x^U y + xy^L - x^U y^L,\ w \le xy^U + x^L y - x^L y^U.$$

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 β_{knl}, and add the following constraints

$$\beta_{knl} \ge 0,\ \beta_{knl} \ge z_{kn} + u_{kl} - 1,$$
$$\beta_{knl} \le u_{kl},\ \text{and}\ \beta_{knl} \le z_{kn},$$

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 $\hat{f}^t_{\pi rn} := w_{rn}f^t_{\pi r}$, where w_{rn} is binary, and add the following constraints:

$$\hat{f}^t_{\pi rn} \ge w_{rn}\underline{f}^t_{\pi r},\ \hat{f}^t_{\pi rn} \ge f^t_{\pi r} + w_{rn}\overline{f}^t_{\pi r} - \overline{f}^t_{\pi r},$$
$$\hat{f}^t_{\pi rn} \le f^t_{\pi r} + w_{rn}\underline{f}^t_{\pi r} - \underline{f}^t_{\pi r},\ \text{and}\ \hat{f}^t_{\pi rn} \le w_{rn}\overline{f}^t_{\pi r},$$

where $\overline{f}^t_{\pi r}$ and $\underline{f}^t_{\pi r}$ are the lower and upper bounds of $f^t_{\pi r}$. The bilinear term $w_{rn}g^t_{\pi r}$ can be managed similarly. As a result, the SAA formulation (3.2) is transformed into a mixed-integer convex quadratic program.

3.2 The PSAA Formulation

The SAA is relatively accurate when there are sufficient samples. However, more samples lead to more binary auxiliary variables (i.e., θ_π), 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 [12] to approximate our proposed TCC model, which is more complicated than the single-stage chance-constrained model considered in [12], as evidenced in [23]. 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

$$\mathbb{P}\{g(\boldsymbol{x},\boldsymbol{\xi}) \geq 0\} \geq 1-\eta, \tag{3.3}$$

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 (3.3) 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

$$\mathbb{P}\{g(\boldsymbol{x}, \boldsymbol{\xi}_1, \hat{\boldsymbol{\xi}}_2^k) \geq 0\} \geq y_k, \forall k \in [N], \tag{3.4}$$

$$\frac{\sum_{k=1}^{N} y_k}{N} \geq 1 - \eta,\ y_k \geq 0, \forall k \in [N]. \tag{3.5}$$

In contrast to constraints (4.1d) - (4.1f) (SAA formulation), constraints (3.4)-(3.5) (PSAA formulation) introduce only continuous variable y_k, while N new chance constraints are added. In our proposed PSAA model, We will show that (3.4) has a convex approximation for *the planning problem* in this thesis, 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 Σ 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 Σ, where U is an orthogonal matrix and Λ is a diagonal matrix with the eigenvalues of Σ on the diagonal. Without loss of generality, we assume that Λ_{11} is the largest eigenvalue of Σ. Let

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

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 ξ_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, ξ_1' is the first principal component. The PSAA then considers Π_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 (3.1) and constraints (2.18b) – (2.18g) and $(2.19b)_\pi - (2.19o)_\pi$. 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 (2.19b) – (2.19o) and (2.21) to be satisfied with a high probability by a specified $\boldsymbol{\Omega}^2$, we relax the chance constraint (2.22) to require the consistency of (2.19b) –

(2.19o) and (2.21). That is, we consider the following chance constraint:

$$\mathbb{P}\left(\exists \text{ a solution satisfying (2.19b)} - \text{(2.19o)}, \text{(2.21)} \,|\, \mathbf{\Omega}^1\right) \geq 1 - \eta. \tag{3.6}$$

Let $(2.19b)_\kappa - (2.19o)_\kappa$ and $(2.21)_\kappa$ be a copy of (2.19b) – (2.19o) and (2.21), with $\boldsymbol{\xi}$ replaced by $(\xi_1', \boldsymbol{\xi}_{2\kappa}')$ for any $\kappa \in [\Pi_2]$. The PSAA approximates (3.6) with

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

where $A(\mathbf{\Omega}^1, \kappa) := \{\xi_1' \,|\, (2.19b)_\kappa\text{-}(2.19o)_\kappa, (2.21)_\kappa \text{ are consistent with } \mathbf{\Omega}^1\}$. For a given $\mathbf{\Omega}^1$ and κ, $A(\mathbf{\Omega}^1, \kappa)$ is a convex set of ξ_1'. That is, if $\hat{\xi}_1' < \bar{\xi}_1'$ are both in $A(\mathbf{\Omega}^1, \kappa)$, then $\xi_1' \in A(\mathbf{\Omega}^1, \kappa)$ for all $\xi_1' \in [\hat{\xi}_1', \bar{\xi}_1']$. Let $\Psi(\cdot)$ be the cumulative distribution function (CDF) of ξ_1'. Then,

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

Therefore, our PSAA model is as follows.

$$\begin{aligned}
\min\ & C_1\left(\mathbf{\Omega}^1\right) + \frac{1}{\Pi_1}\sum_{\pi=1}^{\Pi_1} C_2\left(\mathbf{\Omega}_\pi^2\right) \\
\text{s.t. }\ & \text{(2.18b)} - \text{(2.18g)}, && \text{(3.7a)}\\
& (2.19b)_\pi - (2.19o)_\pi,\ \forall\, \pi \in [\Pi_1], && \text{(3.7b)}\\
& (\mathbf{\Omega}^1, \mathbf{\Omega}_{1\kappa}^2) \text{ satisfies } (2.19b)_\kappa - (2.19o)_\kappa, (2.21)_\kappa \\
& \qquad \text{with } (Z_{1\kappa}, \boldsymbol{\xi}_{2\kappa}'),\ \forall\, \kappa \in [\Pi_2], && \text{(3.7c)}\\
& (\mathbf{\Omega}^1, \mathbf{\Omega}_{2\kappa}^2) \text{ satisfies } (2.19b)_\kappa - (2.19o)_\kappa, (2.21)_\kappa \\
& \qquad \text{with } (Z_{2\kappa}, \boldsymbol{\xi}_{2\kappa}'),\ \forall\, \kappa \in [\Pi_2], && \text{(3.7d)}\\
& \Psi(Z_{2\kappa}) - \Psi(Z_{1\kappa}) \geq \eta_\kappa,\ \forall\, \kappa \in [\Pi_2], && \text{(3.7e)}\\
& \sum_{\kappa=1}^{\Pi_2} \eta_\kappa \geq \Pi_2(1-\eta), && \text{(3.7f)}\\
& \eta_\kappa \geq 0,\ \forall\, \kappa \in [\Pi_2]. && \text{(3.7g)}
\end{aligned}$$

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

$\kappa \in [\Pi_2]$.

For the model to be practical, we need to estimate the CDF $\Psi(\cdot)$ of ξ_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 ξ_1' and estimate $\Psi(\cdot)$ using the kernel density estimation, a commonly used nonparametric method proposed by Rosenblatt [42] and Parzen [36].

Let Π_3 be the total number of samples of ξ_1', and $\xi_{1\tau}'$ be a specific sample for any $\tau \in [\Pi_3]$. The kernel density estimation of the probability density function ψ of ξ_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 ϕ 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 ξ_1' can be estimated by

$$\Psi(\xi_1') \approx \frac{1}{\Pi_3} \sum_{\tau=1}^{\Pi_3} \Phi\left(\frac{\xi_1' - \xi_{1\tau}'}{h}\right), \tag{3.8}$$

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

We further approximate $\Phi(\cdot)$ in (3.8) using the following piecewise linear function:

$$\Phi(x) \approx \begin{cases} \min\limits_{\epsilon \in [\Delta_1]} \{a_\epsilon x + \alpha_\epsilon\} & \text{if } x \geq 0.5 \\ \max\limits_{\zeta \in [\Delta_2]} \{a_\zeta x + \alpha_\zeta\} & \text{if } x < 0.5, \end{cases} \tag{3.9}$$

where Δ_1 and Δ_2 are the numbers 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. 3.1.

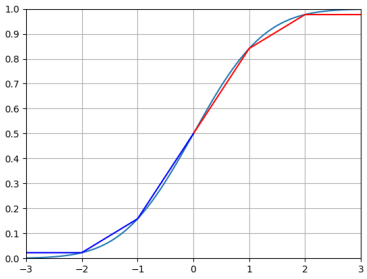

Figure 3.1: Piecewise linear approximation of the CDF of the standard normal distribution $\Phi(\cdot)$ ($\Delta_1 = \Delta_2 = 3$)

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[20]. Therefore, with the approximation in (3.9), we can remove the min and max operators in (3.9) and approximate constraint (3.7e) 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 (3.7b) – (3.7d) can be linearized using McCormick inequalities, as in the SAA model. As a result, the PSAA model (3.7) is simplified to a mixed-integer convex quadratic program.

Chapter 4

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 [1] 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. 4.1 to summarize the sequential steps we follow to perform the numerical experiments in this chapter.

4.1 IEEE 33-Bus System

We first consider the modified IEEE 33-Bus radial distribution network examined in [16] (see Fig. 4.2). 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. The reactive power sources are of the hybrid (capacitive and inductive) compensator type, and they can both generate and absorb reactive

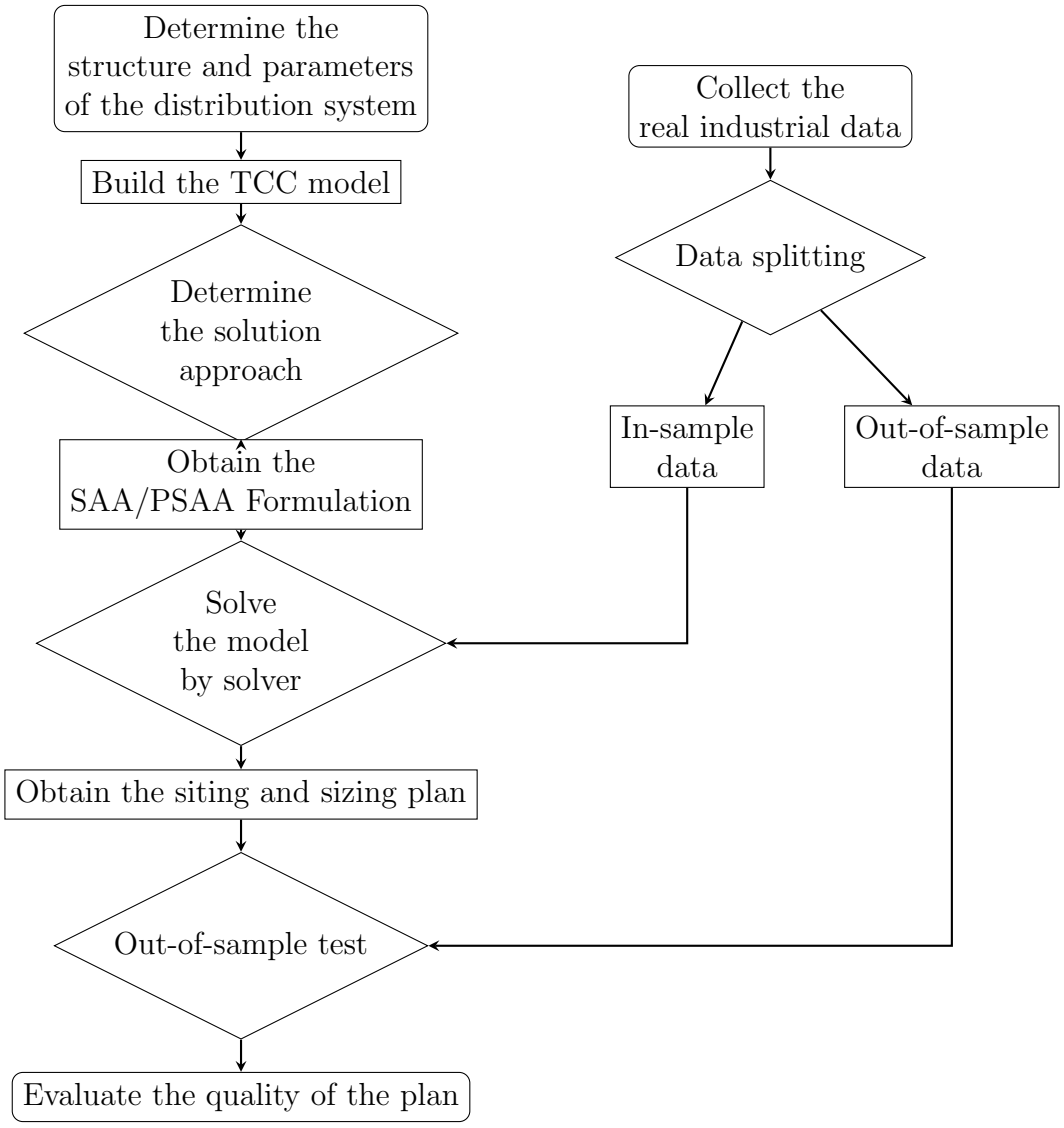

Figure 4.1: The Procedure of Numerical Experiments

power to stabilize the voltage. Nevertheless, our model can also consider other types. With the fixed location of the DDG units and reactive power sources, we then focus on the location and capacity planning of candidate RDG and ES units to be installed.

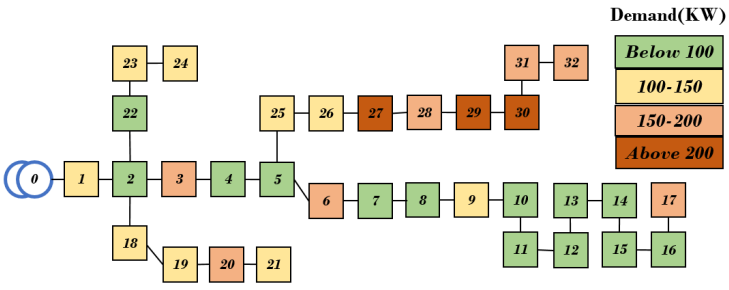

Figure 4.2: IEEE 33-Bus Distribution Network

4.1.1 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 candidate 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 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 $[-0.01, 0.019]$. Here $[-0.01, 0.019]$ is an interval, which specifies a range of possible values of reactive-power load that we can possibly generate. By using such a random data generation, we obtain different values of reactive-power loads at different buses in different periods. All data are for a range of 4 years, leading to 208 ($= 52$ 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 [35, 39, 38]. Thus, we have 4000 data samples in total.

4.1.2 Data Processing for PSAA

In the PSAA approach, we assume that random parameters $\boldsymbol{\xi}$ can be divided into independent two parts ξ_1 and $\boldsymbol{\xi}_2$. Here, we transform all data samples corresponding to $\boldsymbol{\xi}$ to uncorrelated data samples, thereby approximating the independence requirement. We first collect all historical data samples of $\boldsymbol{\xi}$ and then calculate its mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Next, we do an eigenvalue decomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$. Now, we can represent $\boldsymbol{\xi}$ as a linear combination of random vector $\boldsymbol{\xi}'$,

$$\boldsymbol{\xi} = \mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi}' + \boldsymbol{\mu}.$$

Through this equation, we can transform all data samples of $\boldsymbol{\xi}$ into those corresponding to $\boldsymbol{\xi}'$. More importantly, the mean vector of $\boldsymbol{\xi}'$ is $\mathbf{0}$ and the covariance matrix is $\mathbf{I}$. Thus, all data samples corresponding to $\boldsymbol{\xi}'$ are uncorrelated and we use them in experiments. However, this transformation is not strict and we cannot measure the gap between the irrelevance and

independence of an arbitrary random vector. In further research, we will consider more specific random vectors satisfying special distributions, for example, normal distribution. Then, we may quantify the gap between irrelevance and independence.

4.1.3 Parameters

All of the parameters used in our experiments are slightly modified based on the parameters[1] used in [16]. 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 candidate 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 candidate 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 candidate 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 candidate 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 ω of the DDG units is 3 kg/MWh. A maximum of $\bar{K} = 3$ out of $K = 4$ candidate 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 a candidate 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 γ is set to 0.9.

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

4.2 IEEE 123-Bus System

We then consider the commonly used IEEE 123-Bus radial distribution network [6, 13, 51] (see Fig. 4.3). 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.

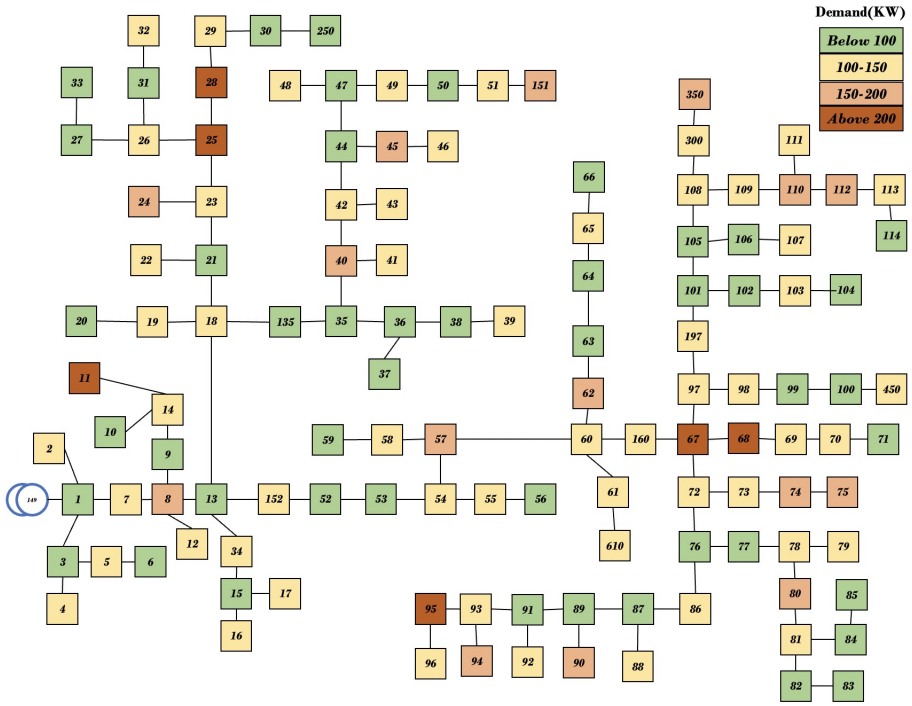

Figure 4.3: IEEE 123-Bus Distribution Network

4.2.1 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 candidate RDG unit s_k^t uncertain and use an estimated value for the weekly active/reactive load at each bus, i.e., d_{pn}^t/d_{qn}^t. 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.

4.2.2 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 $K = 8$ candidate RDG units 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 candidate ES units to be installed is $R = 6$. The reactive-power output bounds $(\underline{q}_n^t, \bar{q}_n^t)$ are (-0.15, 0.25).

4.3 Decomposition Framework

To reduce the computational difficulty of solving both the SAA and PSAA formulations, we adopt the Benders decomposition algorithm [8] to improve the computational efficiency. Specifically, we first linearize the SOCP constraints (2.19l) and (2.19m) as [16] does by using the polyhedral ϵ-approximation in [7]. With such an approximation, both the SAA formulation (3.2) and PSAA formulation (3.7) 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 Π_1 subproblems for the SAA formulation, where each sample $\pi \in [\Pi_1]$ corresponds to one subproblem, as shown in (4.1c), and these subproblems can be solved in parallel. For the PSAA formulation, we have $\Pi_1 + 1$ subproblems, where each sample $\pi \in [\Pi_1]$

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

corresponds to one subproblem, as shown in (3.7b), and constraints (3.7c) – (3.7g) are included in one subproblem.

4.3.1 Further Refinement

In this thesis, we employ Benders decomposition as a means to enhance computational efficiency. Specifically, we utilize the Benders decomposition package integrated within the Cplex solver, which represents a viable approach. Additionally, we explore the possibility of leveraging other advanced algorithms in the context of two-stage stochastic programming and SAA. There are mainly two types of methods: primary method and dual method. The primary method works with subproblems assigned to time stages, such as Benders decomposition and L-shaped decomposition. For example, [9] combine an accelerated sample average approximation approach with an accelerated Benders' decomposition algorithm to solve the two-stage model with binary variables in the first stage and continuous variables in the second stage.

Another method is the dual method, which works with subproblems assigned to scenarios. By using SAA or PSAA, we can approximate the two-stage stochastic programming into the mixed-integer problem with Π_1 scenarios (see Problem (3.2) and (3.7)). To solve problems with this kind of structure, we can use dual decomposition and Lagrangean relaxation [44, 10, 33, 40]. We take Problem (3.2) as an example. We first introduce first-stage variables $\mathbf{\Omega}_\pi^1$ for each scenario $\pi \in [\Pi_1]$ and add non-anticipativity constraints $\mathbf{\Omega}_1^1 = \cdots = \mathbf{\Omega}_{\Pi_1}^1$. We then introduce $\theta_{\pi_2}^{\pi_1}$ for any $\pi_1 \in [\Pi_1]$ and $\pi_2 \in [\Pi_1]$ and add non-anticipativity constraints $\theta_1^{\pi_1} = \cdots = \theta_{\Pi_1}^{\pi_1}$ for any $\pi_1 \in [\Pi_1]$. For each $\pi \in [\Pi_1]$, we implement the non-anticipativity constraints by equations $\sum_{\pi=1}^{\Pi_1} K^\pi \mathbf{\Omega}_\pi^1 = 0$ and $\sum_{\pi=1}^{\Pi_1} K^\pi \theta_\pi^{\pi_1} = 0,\ \forall \pi_1 \in [\Pi_1]$, where $K^1 = 1 - \Pi_1$ and $K^\pi = 1,\ \forall \pi = 2, \ldots, \Pi_1$. Now, Problem (3.2) can be reformulated as:

$$\min_{\mathbf{\Omega}_\pi^1,\, \mathbf{\Omega}_\pi^2, \forall \pi \in [\Pi_1]} \quad \frac{1}{\Pi_1} \sum_{\pi=1}^{\Pi_1} \left(C_1\left(\mathbf{\Omega}_\pi^1\right) + C_2\left(\mathbf{\Omega}_\pi^2\right)\right) \tag{4.1a}$$

$$\text{s.t.} \quad (2.18\text{b})_\pi - (2.18\text{g})_\pi,\ \forall\, \pi \in [\Pi_1], \tag{4.1b}$$

$$(2.19\text{b})_\pi - (2.19\text{o})_\pi,\ \forall\, \pi \in [\Pi_1], \tag{4.1c}$$

$$LS_{1\pi mn}^t \le \theta_i^\pi M_\pi,\ LS_{2\pi mn}^t \le \theta_i^\pi M_\pi,$$
$$\forall\, (m,n) \in \mathcal{E}, \forall\, i \in [\Pi_1], \forall\, \pi \in [\Pi_1], \tag{4.1d}$$

$$\sum_{i=1}^{\Pi_1} \theta_\pi^i \leq \Pi_1 \eta, \ \forall\, \pi \in [\Pi_1], \tag{4.1e}$$

$$\theta_\pi^i \in \{0,1\}, \forall\, i \in [\Pi_1], \ \forall\ \pi \in [\Pi_1], \tag{4.1f}$$

$$\sum_{\pi=1}^{\Pi_1} K^\pi \mathbf{\Omega}_\pi^1 = 0, \tag{4.1g}$$

$$\sum_{\pi=1}^{\Pi_1} K^\pi \theta_\pi^{\pi_1} = 0, \ \forall \pi_1 \in [\Pi_1]. \tag{4.1h}$$

We then define λ and $\boldsymbol{\gamma}$ as the Lagrangean multipliers associated with the non-anticipativity constraints and relax these. The resulting Lagrangean relaxation is

$$LR(\lambda, \boldsymbol{\gamma}) = \min_{\mathbf{\Omega}_\pi^1, \mathbf{\Omega}_\pi^2, \forall \pi \in [\Pi_1]} \quad \frac{1}{\Pi_1} \sum_{\pi=1}^{\Pi_1} \left(C_1\left(\mathbf{\Omega}_\pi^1\right) + C_2\left(\mathbf{\Omega}_\pi^2\right) + \lambda K^\pi \mathbf{\Omega}_\pi^1 + \sum_{\pi_1=1}^{\Pi_1} \gamma_{\pi_1} K^\pi \theta_\pi^{\pi_1} \right) \tag{4.2a}$$

$$\text{s.t.} \quad (4.1\text{b}) - (4.1\text{f}). \tag{4.2b}$$

Note that Problem (4.2) is separable in scenarios. Then, we can find the best lower bound for the primary problem by solving the Lagrangean dual:

$$z_{LD} = LD = \max_{\lambda, \boldsymbol{\gamma}} LR(\lambda, \boldsymbol{\gamma}).$$

We let $\mathfrak{P}$ demote the list of current problems together with associated lower bound $z_{LD} = z_{LD}(P)$. Now we can design the following branch and bound algorithm:

- Step 1: Initialization: Set $\overline{z} = \infty$ and let $\mathfrak{P}$ consist of primary problem.
- Step 2: Termination: If $\mathfrak{P} = \emptyset$ then the solution is optimal.
- Step 3: Node selection: Select and delete a problem P from $\mathfrak{P}$, solve the dual problem that yields the bound $z_{LD} = z_{LD}(P)$. If P is infeasible, go to Step 2.
- Step 4: Bounding: If $z_{LD}(P) \geq \overline{z}$, go to Step 2. (i) The non-anticipativity constraints are satisfied. Let $\overline{z}$ be the corresponding objective value of the primary problem and delete from $\mathfrak{P}$ all problem P' with $z_{LD}(P') \geq \overline{z}$. Go to Step 2. (ii) The non-anticipativity constraints are not satisfied. Compute the average solution and round it by some heuristic to obtain $\bar{x}$. If $\bar{x}$ is feasible, the corresponding objective is denoted by $\overline{z}$ and delete from $\mathfrak{P}$ all problem P' with $z_{LD}(P') \geq \overline{z}$. Go to Step 5.

- Branching: Select a component x_i and add two new problems to $\mathfrak{P}$ by adding the constraints $x_i \leq \lfloor \bar{x}_i \rfloor$ and $x_i \geq \lfloor \bar{x}_i \rfloor + 1$, respectively (if x_i is an integer component) or $x_i \leq \bar{x}_i - \epsilon$ and $x_i \geq \bar{x}_i + \epsilon$, respectively.

[10] show that this branch and bound algorithm terminates in finitely many steps.

4.4 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 (2.19b) – (2.19o) and (2.21) 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 ϑ in hours), and (iii) the desired feasible probability of the chance constraint (i.e., $1 - \eta$), as shown in Tables 4.1 – 4.4. 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 a better-quality incumbent solution.

We illustrate the performance of both models in Tables 4.1 – 4.4. The columns in the SAA/P-SAA 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. 4.4, 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.

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

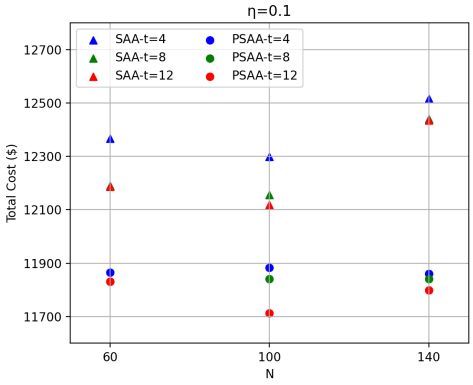

Figure 4.4: SAA vs. PSAA solutions in the total cost ($\eta = 0.1$)

constraint is dealt with in the models. The SAA model introduces Π_1 binary variables (i.e., θ_π), whereas the PSAA model introduces Π_2 continuous variables (i.e., η_κ). 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. 4.5, where the vertical axis represents the relative optimality gap.

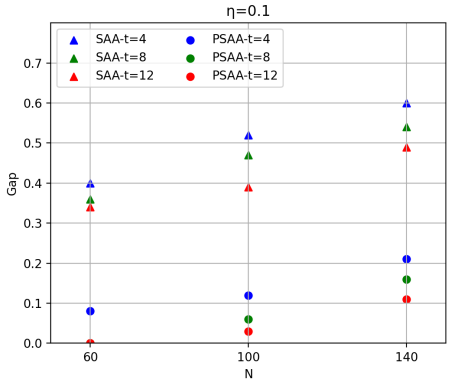

Figure 4.5: SAA vs. PSAA solutions in the optimality gap ($\eta = 0.1$)

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. 4.6 and Fig. 4.7. In the figures, the horizontal axis represents the setting of η and Π_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. 4.6 and Fig. 4.7, we find that the improvement is mostly above 50% and even reaches 100% when the PSAA solves an instance to optimality.

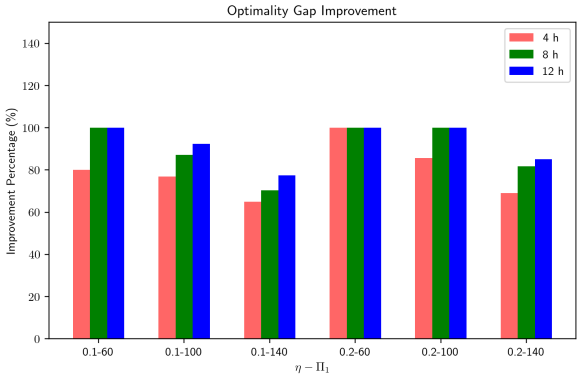

Figure 4.6: Optimality Gap Improvement (IEEE 33-Bus system)

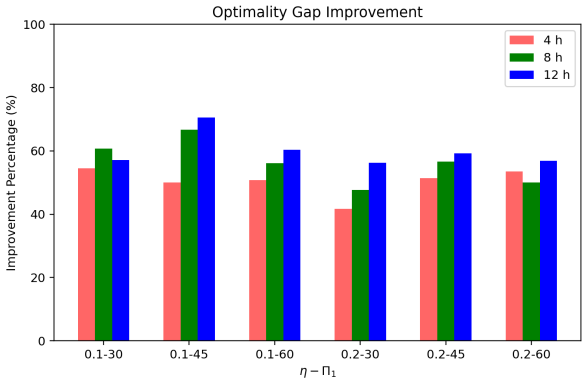

Figure 4.7: Optimality Gap Improvement (IEEE 123-Bus system)

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 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. 4.8, where the vertical axis represents the actual feasible probability.

To conclude this chapter, we illustrate the performance difference between the two models with some specific examples. When $\Pi_1 = \Pi_2 = 140$, $\vartheta = 8$, and $\eta = 0.1$ in the IEEE 33-Bus system, the SAA solution sites (sizes) the candidate RDG units at Buses 3 (6 MW), 9 (6 MW), and 12

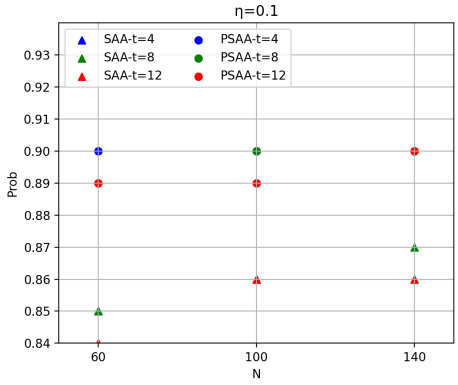

Figure 4.8: SAA vs. PSAA solutions in the actual feasible probability ($\eta = 0.1$)

(4 MW), and the PSAA solution sits (sizes) the candidate 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 a candidate 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 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.

When considering the IEEE 33-Bus system with $\Pi_1 = \Pi_2 = 140$, we examine the effect of computational time on the solutions while varying the parameter ϑ. We also maintain the desired feasible probability at $\eta = 0.1$ for the chance-constrained PSAA model. The specific results for each case are as follows:

For $\vartheta = 4$: In the SAA solution, candidate RDG units are sited (sized) at Buses 4 (6 MW), 8 (6 MW), and 12 (5 MW). In contrast, the PSAA solution places candidate RDG units at Buses

3 (5 MW), 9 (5 MW), and 26 (6 MW).

For $\vartheta = 8$: In the SAA solution, candidate RDG units are sited (sized) at Buses 3 (6 MW), 9 (6 MW), and 12 (4 MW). In the PSAA solution, candidate RDG units are sited (sized) at Buses 3 (5 MW), 8 (4 MW), and 26 (6 MW).

For $\vartheta = 12$: In the SAA solution, candidate RDG units are sited (sized) at Buses 3 (5 MW), 9 (6 MW), and 12 (4 MW). In the PSAA solution, candidate RDG units are sited (sized) at Buses 3 (5 MW), 8 (4 MW), and 26 (5 MW).

The results clearly indicate that as the computational time (represented by ϑ) increases, the PSAA approach consistently yields better siting and sizing decisions compared to the SAA approach. In each case, the PSAA solution provides more optimal locations and capacities for the candidate RDG units, potentially resulting in enhanced system performance and reduced operational costs.

By allowing for more computational time, the PSAA method can explore a larger solution space and consider a broader range of possibilities, leading to improved decision-making. The trade-off between computational resources and solution quality is evident, with the PSAA approach providing superior results at the cost of increased computation time. Ultimately, this demonstrates the significance of allocating sufficient computational resources to achieve more robust and effective RDG unit siting and sizing decisions in distribution networks.

When keeping $\vartheta = 8$ fixed and varying the number of samples $\Pi_1 = \Pi_2$, we examine how the number of samples affects the RDG unit siting and sizing decisions in both the SAA and PSAA models. We maintain the desired feasible probability at $\eta = 0.1$ for the PSAA model. The specific results for each case are as follows:

For $\Pi_1 = \Pi_2 = 60$: In the SAA solution, candidate RDG units are sited (sized) at Buses 4 (6 MW), 8 (6 MW), and 12 (5 MW). The PSAA solution, however, places candidate RDG units at Buses 5 (5 MW), 9 (6 MW), and 26 (6 MW).

For $\Pi_1 = \Pi_2 = 100$: In the SAA solution, candidate RDG units are sited (sized) at Buses 3 (6 MW), 9 (6 MW), and 11 (5 MW). In contrast, the PSAA solution sites (sizes) candidate RDG units at Buses 3 (5 MW), 8 (4 MW), and 26 (6 MW).

For $\Pi_1 = \Pi_2 = 140$: In the SAA solution, candidate RDG units are sited (sized) at Buses 3 (6 MW), 9 (6 MW), and 12 (4 MW). The PSAA solution sites (sizes) candidate RDG units at Buses 3 (5 MW), 8 (4 MW), and 26 (6 MW).

The results clearly demonstrate the impact of the number of samples on the solution quality in both SAA and PSAA models. As the number of samples increases, both methods yield better siting and sizing decisions. The PSAA approach, in particular, requires fewer samples to achieve superior solutions compared to the SAA approach. This indicates that the PSAA method is more efficient in utilizing the available samples to make informed decisions and find robust solutions.

By increasing the number of samples, the PSAA method can better capture the underlying stochastic nature of the problem, leading to improved solution accuracy and reliability. Additionally, the PSAA approach exhibits its ability to provide higher-quality solutions with a reduced computational burden, making it a favorable choice for optimizing RDG unit siting and sizing decisions in distribution networks. The findings emphasize the importance of employing an appropriate number of samples to ensure accurate and efficient solution outcomes in stochastic optimization problems.

Table 4.1: IEEE 33-Bus System: SAA vs. PSAA ($\eta = 0.1$)

Π_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

4.5 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.

Table 4.2: IEEE 33-Bus System: SAA vs. PSAA ($\eta = 0.2$)

Π_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 4.3: IEEE 123-Bus System: SAA vs. PSAA ($\eta = 0.1$)

Π_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 4.4: IEEE 123-Bus System: SAA vs. PSAA ($\eta = 0.2$)

Π_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 4.5: IEEE 33-Bus System: Storage vs. No Storage ($\eta = 0.1$)

(ϱ_1, ϱ_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 4.6: IEEE 33-Bus System: Storage vs. No Storage ($\eta = 0.2$)

(ϱ_1, ϱ_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 4.7: IEEE 123-Bus System: Storage vs. No Storage ($\eta = 0.1$)

(ϱ_1, ϱ_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 4.8: IEEE 123-Bus System: Storage vs. No Storage ($\eta = 0.2$)

(ϱ_1, ϱ_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

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 (2.23) 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 (2.23), 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 ϱ_1 and ϱ_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 (ϱ_1, ϱ_2) to take six different pairs of values, i.e., $(0.5, 0.5)$, $(0.4, 0.6)$, $(0.6, 0.4)$, $(0.3, 0.3)$, $(0.2, 0.4)$, and $(0.4, 0.2)$. Note that (ϱ_1, ϱ_2) does not represent an interval. When $(\varrho_1, \varrho_2) = (0.4, 0.6)$, it means that we set $\varrho_1 = 0.4$ and $\varrho_2 = 0.6$. The results are shown in Tables 4.5 – 4.8.

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. 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) candidate RDG units at Buses 4 (7 MW), 9 (6 MW), and 28 (7 MW). With ES units, the result shows that candidate 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 thesis, 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 [46] and chance-constrained optimal power flow problems in [41], 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 ξ_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.

4.6 Chance Constraints vs. No Chance Constraints

In this section, we conduct a comparative analysis of two models: one with chance constraints and the other without. The stochastic programming model without chance constraints is denoted as (2.20), while the model with chance constraints is represented as (2.23). To facilitate the analysis, we omit the consideration of ES units, resulting in the temporary removal of terms related to ES units such as w_{rn}, y_r, f_r^t, g_r^t, and b_r^t from both models. The evaluation is performed for a single time period $T = 1$, and we employ the same data and parameters as used in Section 4.4.

To solve the two models, we utilize the SAA approach. The experimentation involves the IEEE 33-Bus system, and we vary the training data size $\Pi_1 = \Pi_2$ among the set $60, 100, 140$. To ensure tractability, the maximum running time for all instances is limited to 40 hours, thereby guaranteeing the attainment of optimal solutions. For the model with chance constraints, we specify a desired feasible probability of $1 - \eta$, where η takes values from the set $0.1, 0.2$. The results of all instances are presented in Table 4.9, which provides a comprehensive overview of the performance comparison between the models with and without chance constraints.

Table 4.9: IEEE 33-Bus System: Chance Constraints vs. No Chance Constraints

Π_1	No Chance Constraints				Chance Constraints ($\eta = 0.2$)				Chance Constraints ($\eta = 0.1$)			
	1st ($)	2nd ($)	Total cost ($)	Prob	1st ($)	2nd ($)	Total cost ($)	Prob	1st ($)	2nd ($)	Total cost ($)	Prob
60	7371.8	1704.6	9076.4	0.32	9664.2	1742.8	11407.0	0.77	10510.0	1679.4	12189.4	0.84
100	7112.4	1925.0	9037.4	0.42	9633.8	1739.7	11373.5	0.77	10397.4	1721.3	12118.7	0.86
140	7676.2	1855.6	9531.8	0.35	9974.3	1748.7	11723.0	0.79	10471.3	1704.6	12175.9	0.84

Indeed, the inclusion of chance constraints in the models results in a higher probability of satisfying the demand, signifying an improvement in out-of-sample performance. By incorporating chance constraints, the models provide a greater level of assurance in meeting the decision-maker's requirements. This increased level of reliability is particularly crucial in systems with

inherent uncertainties, as it helps mitigate potential risks associated with demand satisfaction. On the other hand, the model without chance constraints exhibits lower first-stage costs and lower total costs across all instances. This finding can be attributed to the absence of probabilistic constraints, allowing for more flexibility in the optimization process. Consequently, the optimal solutions in the absence of chance constraints may necessitate the installation of fewer RDG units or the use of RDG units with smaller capacities. However, the advantage of lower costs in the model without chance constraints comes at the expense of potentially compromising the system's ability to reliably meet demand and ensure stable operation. This drawback highlights the necessity of considering chance-constrained models, which strike a balance between cost-effectiveness and risk management. By incorporating chance constraints, these models provide more robust solutions that account for uncertainties and reduce the likelihood of demand satisfaction failures or operational instabilities. In conclusion, the incorporation of chance constraints in the optimization models enhances their out-of-sample performance, ensuring a higher probability of meeting demand requirements. While the model without chance constraints may yield lower costs, it may sacrifice reliability and system stability. Therefore, the adoption of chance-constrained models is essential to strike an optimal balance between cost efficiency and system robustness in the presence of uncertainty.

4.7 Sensitivity Analysis

In this section, we investigate the impact of changing two key parameters on the planning results for the distribution system. Specifically, we vary the maximum number of RDG units to be installed, denoted as $\bar{K}$, to examine its influence on the comparison between SAA and PSAA. We also vary the total number of available ES units, denoted as R, to assess its impact on the results with ES units. We fix all other parameters related to the distribution system, such as setup cost, size-based cost, electricity price, fuel/emission price, capacity of the power distribution line, active/reactive power output bound, and electrical resistance/reactance. Then, we vary the maximum number of RDG units to be installed, $\bar{K}$, and observe how this parameter affects the comparison between SAA and PSAA. We analyze the solutions obtained from both methods for different values of $\bar{K}$ to understand the trade-offs between the two approaches in

terms of cost, reliability, and feasibility. By examining the effects of these parameters, we gain valuable insights into the behavior of the optimization models under different system configurations and constraints. Understanding how changes in $\bar{K}$ and R impact the planning results will assist in making informed decisions regarding RDG unit deployment and the inclusion of ES units in the distribution network.

4.7.1 $\bar{K}$

In this experiment, we focus on the IEEE-33 Bus system, considering the impact of varying the maximum number of RDG units to be installed, denoted as $\bar{K}$. We ignore ES units and assume a single time period $T = 1$ with the same data and parameters as mentioned earlier. The training data size $\Pi_1 = \Pi_2$ is varied among the set $60, 100, 140$. The desired feasible probability is set at $1 - \eta = 0.9$, and the maximum running time is limited to 40 hours to ensure all instances obtain the optimal solution.

In Table 4.10, the performance of all instances is illustrated for the values of $\bar{K} \in 2, 3, 4$. Notably, the model becomes infeasible when $\bar{K} = 2$, resulting in no reported results for this case. This outcome indicates that reducing the number of RDG units to only two is insufficient to ensure normal operation of the system.

On the other hand, the results clearly demonstrate that $\bar{K} = 3$ emerges as the most favorable choice for the IEEE-33 Bus system. When $\bar{K} = 3$, the system achieves the best trade-off between cost, reliability, and feasibility. Interestingly, even when $\bar{K} = 4$, the optimal number of RDG units to be installed remains at three, indicating that adding more RDG units does not provide any significant improvement.

Based on the experimental findings, it is evident that setting $\bar{K} = 3$ is the most suitable option for the IEEE-33 Bus system. This decision ensures that the system can operate normally, without the risk of infeasibility that arises when reducing the number of RDG units. Moreover, increasing the number of RDG units beyond three does not yield any additional benefits in terms of cost or system performance. Therefore, the optimal configuration with three RDG units is the most efficient and reliable choice for this distribution system.

Table 4.10: IEEE 33-Bus System: $\bar{K} = 3$ or $\bar{K} = 4$

Π_1	SAA				PSAA			
	1st ($)	2nd ($)	Total cost ($)	Prob	1st ($)	2nd ($)	Total cost ($)	Prob
60	10510.0	1679.4	12189.4	0.84	10184.6	1645.8	11830.4	0.89
100	10397.4	1721.3	12118.7	0.86	10105.6	1606.5	11712.1	0.89
140	10471.3	1704.6	12175.9	0.84	10173.0	1625.1	11798.1	0.90

4.7.2 R

In this analysis, we consider the impact of varying the total number of available ES units, denoted as R, in the IEEE-33 Bus system. The other parameters and data remain the same as mentioned in Section 4.5. The desired feasible probability is set at $1 - \eta = 0.9$, and the maximum running time is limited to 40 hours to ensure all instances obtain the optimal solution. We illustrate the performance of all instances in Table 4.11 for the values of $R \in 2, 3, 4$. The results highlight the trade-offs involved when installing ES units in the power system.

When $\rho_1 + \rho_2 = 1$: In this case, the increase in the number of ES units may not lead to a significant reduction in the total cost. This is because the output in the system may already be sufficient, and it may not be necessary to increase the number of ES units. The additional setup costs and operating costs of ES units may offset the potential cost reduction achieved by reducing the capacity of RDG units. When $\rho_1 + \rho_2 = 0.6$: Here, increasing the number of ES units can result in a notable reduction in the total cost. The reason is that the output of the system is insufficient, and installing more ES units can substantially decrease the cost of RDG units. The costs of the additional ES units are relatively small compared to the reduced costs of RDG units, leading to a net cost reduction.

In addition to cost considerations, the number of ES units also affects the probability of satisfying the demand in the testing samples. Increasing the number of ES units can enhance the system's ability to meet demand with a higher probability. This improvement in reliability is an important factor to consider when making decisions about the number of ES units to install. In conclusion, the decision to install ES units in the power system involves various trade-offs. When the output in the system is sufficient (i.e., $\rho_1 + \rho_2 = 1$), increasing the number of ES units may not yield significant cost reductions. However, when the output is insufficient (i.e., $\rho_1 + \rho_2 = 0.6$), increasing the number of ES units can lead to cost savings and improve the

probability of satisfying the demand. Careful consideration of these factors is essential to make informed decisions regarding the optimal number of ES units in the distribution network.

Table 4.11: IEEE 33-Bus System: Change R

(ϱ_1, ϱ_2)	$R=2$				$R=3$				$R=4$			
	1st ($)	2nd ($)	Total cost ($)	Prob	1st ($)	2nd ($)	Total cost ($)	Prob	1st ($)	2nd ($)	Total cost ($)	Prob
(0.5,0.5)	10418.2	5492.0	15910.2	0.90	9952.0	5501.4	15453.4	0.90	9407.3	5641.1	15048.4	0.91
(0.4,0.6)	9741.6	5297.4	15039.0	0.89	9873.8	5427.6	15301.4	0.91	10231.4	5713.2	15944.6	0.91
(0.6,0.4)	9814.7	5620.0	15434.7	0.90	9908.2	5462.9	15371.1	0.90	9631.5	5666.0	15297.5	0.90
(0.3,0.3)	11371.7	5391.6	16763.3	0.89	10143.7	5576.1	15719.8	0.89	9613.6	5432.9	15046.5	0.89
(0.2,0.4)	10649.3	5409.2	16058.5	0.87	10471.5	5560.8	16032.3	0.89	9904.3	5409.2	15313.5	0.91
(0.4,0.2)	11027.1	5419.1	16446.2	0.88	10239.2	5602.5	15841.7	0.89	10015.0	5548.6	15563.6	0.91

Chapter 5

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 thesis, there can be other types of distribution networks, e.g., meshed grids [15] and multiphase grids [18]. One can apply various approximations (e.g., semidefinite programming [24]) 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] 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.

[2] Chad Abbey and Gza Joos. Supercapacitor energy storage for wind energy applications. *IEEE Trans. Ind Appl.*, 43(3):769–776, 2007.

[3] Thomas Ackermann, Göran Andersson, and Lennart Söder. Distributed generation: a definition. *Electr. Power Syst. Res.*, 57(3):195–204, 2001.

[4] Mesut Baran and Felix F Wu. Optimal sizing of capacitors placed on a radial distribution system. *IEEE Trans. Power Deliv.*, 4(1):735–743, 1989.

[5] John P Barton and David G Infield. Energy storage and its use with intermittent renewable energy. *IEEE Trans. Energy Convers.*, 19(2):441–448, 2004.

[6] Mohammadhafez Bazrafshan and Nikolaos Gatsis. Comprehensive modeling of three-phase distribution systems via the bus admittance matrix. *IEEE Trans. Power Syst.*, 33(2):2015–2029, 2017.

[7] Aharon Ben-Tal and Arkadi Nemirovski. On polyhedral approximations of the second-order cone. *Math. Oper. Res.*, 26(2):193–205, 2001.

[8] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.

[9] Hadi Mohammadi Bidhandi and Jonathan Patrick. Accelerated sample average approximation method for two-stage stochastic programming with binary first-stage variables. *Applied Mathematical Modelling*, 41:582–595, 2017.

[10] Claus C Carøe and Rüdiger Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1-2):37–45, 1999.

[11] Gianni Celli, Emilio Ghiani, Susanna Mocci, and Fabrizio Pilo. A multiobjective evolutionary algorithm for the sizing and siting of distributed generation. *IEEE Trans. Power Syst.*, 20(2):750–757, 2005.

[12] Jianqiang Cheng, Céline Gicquel, and Abdel Lisser. Partial sample average approximation method for chance constrained problems. *Optim. Lett.*, 13(4):657–672, 2019.

[13] Novalio Daratha, Biswarup Das, and Jaydev Sharma. Coordination between OLTC and SVC for voltage regulation in unbalanced distribution system distributed generation. *IEEE Trans. Power Syst.*, 29(1):289–299, 2013.

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

[15] Masoud Farivar and Steven H Low. Branch flow model: Relaxations and convexification—part I. *IEEE Trans. Power Syst.*, 28(3):2554–2564, 2013.

[16] A. M. Fathabad, J. Cheng, K. Pan, and F. Qiu. Data-driven planning for renewable distributed generation integration. *IEEE Trans. Power Syst.*, 35(6):4357–4368, 2020.

[17] Abolhassan Mohammadi Fathabad, Jianqiang Cheng, Kai Pan, and Boshi Yang. Asymptotically tight conic approximations for chance-constrained AC optimal power flow. *Eur. J. Oper. Res.*, 305(2):738–753, 2023.

[18] Lingwen Gan and Steven H Low. Convex relaxations and linear approximation for optimal power flow in multiphase radial networks. In *2014 Power Syst. Comput. Conf.*, pages 1–9. IEEE, 2014.

[19] Pavlos S Georgilakis and Nikos D Hatziargyriou. Optimal distributed generation placement in power distribution networks: Models, methods, and future research. *IEEE Trans. Power Syst.*, 28(3):3420–3428, 2013.

[20] Roya Karimi, Jianqiang Cheng, and Miguel A Lejeune. A framework for solving chance-constrained linear matrix inequality programs. *INFORMS J. Comput.*, 33(3):1015–1036, 2021.

[21] Andrew Keane et al. State-of-the-art techniques and challenges ahead for distributed generation planning and optimization. *IEEE Trans. Power Syst.*, 28(2):1493–1502, 2013.

[22] Xiaolong Kuang, Bissan Ghaddar, Joe Naoum-Sawaya, and Luis F Zuluaga. Alternative LP and SOCP hierarchies for ACOPF problems. *IEEE Trans. Power Syst.*, 32(4):2828–2836, 2016.

[23] Xiao Liu, Simge Küçükyavuz, and James Luedtke. Decomposition algorithms for two-stage chance-constrained programs. *Math. Program.*, 157(1):219–243, 2016.

[24] Steven H Low. Convex relaxation of optimal power flow—part I: Formulations and equivalence. *IEEE Trans. Control. Netw. Syst.*, 1(1):15–27, 2014.

[25] Miles Lubin, Yury Dvorkin, and Line Roald. Chance constraints for improving the security of AC optimal power flow. *IEEE Trans. Power Syst.*, 34(3):1908–1917, 2019.

[26] Garth P McCormick. Computability of global solutions to factorable nonconvex programs: Part I—convex underestimating problems. *Math. Program.*, 10(1):147–175, 1976.

[27] Ozy Daniel Melgar-Dominguez, Mahdi Pourakbari-Kasmaei, and José Roberto Sanches Mantovani. Adaptive robust short-term planning of electrical distribution systems considering siting and sizing of renewable energy based dg units. *IEEE Trans. Sustain. Energy*, 10(1):158–169, 2018.

[28] Daniel K Molzahn, Ian A Hiskens, et al. A survey of relaxations and approximations of the power flow equations. *Foundations and Trends® in Electric Energy Systems*, 4(1-2):1–221, 2019.

[29] Arkadi Nemirovski and Alexander Shapiro. Convex approximations of CC programs. *SIAM J. Optim.*, 17(4):969–996, 2007.

[30] Vishwamitra Oree, Sayed Z Sayed Hassen, and Peter J Fleming. Generation expansion planning optimisation with renewable energy integration: A review. *Renew. Sustain. Energy Rev.*, 69:790–803, 2017.

[31] U Aytun Ozturk, Mainak Mazumdar, and Bryan A Norman. A solution to the stochastic unit commitment problem using chance constrained programming. *IEEE Trans. Power Syst.*, 19(3):1589–1598, 2004.

[32] Bernardo K Pagnoncelli, Shabbir Ahmed, and Alexander Shapiro. Sample average approximation method for chance constrained programming: theory and applications. *J. Optim. Theory Appl.*, 142(2):399–416, 2009.

[33] Daniel Pérez Palomar and Mung Chiang. A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications*, 24(8):1439–1451, 2006.

[34] Hrvoje Pandžić, Yishen Wang, Ting Qiu, Yury Dvorkin, and Daniel Kirschen. Near-optimal method for siting and sizing of distributed storage in a transmission network. *IEEE Trans. Power Syst.*, 30(5):2288–2300, 2014.

[35] Heejung Park, Ross Baldick, and David P Morton. A stochastic transmission planning model with dependent load and wind forecasts. *IEEE Trans. Power Syst.*, 30(6):3003–3011, 2015.

[36] Emanuel Parzen. On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33(3):1065–1076, 1962.

[37] David Pozo and Javier Contreras. A chance-constrained unit commitment with an $n-k$ security criterion and significant wind generation. *IEEE Trans. Power Syst.*, 28(3):2842–2851, 2012.

[38] Kejun Qian, Chengke Zhou, Malcolm Allan, and Yue Yuan. Modeling of load demand due to EV battery charging in distribution systems. *IEEE Trans. Power Syst.*, 26(2):802–810, 2010.

[39] Zhilong Qin, Wenyuan Li, and Xiaofu Xiong. Generation system reliability evaluation incorporating correlations of wind speeds with different distributions. *IEEE Trans. Power Syst.*, 28(1):551–558, 2012.

[40] Ragheb Rahmaniani, Shabbir Ahmed, Teodor Gabriel Crainic, Michel Gendreau, and Walter Rei. The benders dual decomposition method. *Operations Research*, 68(3):878–895, 2020.

[41] Line Roald and Göran Andersson. Chance-constrained AC optimal power flow: Reformulations and efficient algorithms. *IEEE Trans. Power Syst.*, 33(3):2906–2918, 2017.

[42] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, 27(3):832–837, 1956.

[43] Michael D Sankur, Roel Dobbe, Emma Stewart, Duncan S Callaway, and Daniel B Arnold. A linearized power flow model for optimization in unbalanced distribution systems. *arXiv preprint arXiv:1606.04492*, 2016.

[44] Peter Schütz, Asgeir Tomasgard, and Shabbir Ahmed. Supply chain design under uncertainty using sample average approximation and dual decomposition. *European journal of operational research*, 199(2):409–419, 2009.

[45] Yi Tan, Yijia Cao, Canbing Li, Yong Li, Jinju Zhou, and Yan Song. A two-stage stochastic programming approach considering risk level for distribution networks operation with wind power. *IEEE Syst. J.*, 10(1):117–126, 2014.

[46] 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.*, 27(1):206–215, 2012.

[47] Hongyu Wu, Mohammad Shahidehpour, Zuyi Li, and Wei Tian. Chance-constrained day-ahead scheduling in stochastic power system operation. *IEEE Trans. Power Syst.*, 29(4):1583–1591, 2014.

[48] Zhi Wu, Pingliang Zeng, Xiao-Ping Zhang, and Qinyong Zhou. A solution to the chance-constrained two-stage stochastic program for unit commitment with wind integration. *IEEE Trans. Power Syst.*, 31(6):4185–4196, 2016.

[49] Haoyu Yuan, Fangxing Li, Yanli Wei, and Jinxiang Zhu. Novel linearized power flow and linearized opf models for active distribution networks with application in distribution lmp. *IEEE Transactions on Smart Grid*, 9(1):438–448, 2016.

[50] Bo Zeng, Yu An, and Ludwig Kuznia. Chance constrained mixed integer program: Bilinear and linear formulations, and benders decomposition. *arXiv preprint arXiv:1403.7875*, 2014.

[51] Shenxi Zhang, Haozhong Cheng, Libo Zhang, Masoud Bazargan, and Liangzhong Yao. Probabilistic evaluation of available load supply capability for distribution system. *IEEE Trans. Power Syst.*, 28(3):3215–3225, 2013.

[52] Yao Zhang, Jianxue Wang, Bo Zeng, and Zechun Hu. Chance-constrained two-stage unit commitment under uncertain load and wind output using bilinear benders. *IEEE Trans. Power Syst.*, 32(5):3637–3647, 2017.

[53] Zhe Zhou, Jianyun Zhang, Pei Liu, Zheng Li, Michael Georgiadis, and Efstratios Pistikopoulos. A two-stage stochastic programming model for the optimal design of distributed energy systems. *Appl. Energy*, 103:135–144, 2013.

Nomenclature

A. Indices and Sets

n Index of buses.

k Index of renewable distributed generation (RDG) units.

r Index of energy storage (ES) units.

$\mathcal{N}$ Set of all the buses.

$\mathcal{E}$ Set of all the power distribution lines.

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

$\mathcal{B}_0$ Set of buses that are connected to Bus 0.

$\mathcal{B}_1$ Set of buses installed with dispatchable distributed generation (DDG) units.

$\mathcal{B}_2$ Set of buses 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} Setup cost of placing the k^{th} RDG unit at Bus n.

d^0_{rn} Setup cost of placing 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_r^1/d_r^2 Size-based investment/ maintenance cost of the r^{th} ES unit.

c_p^t/c_q^t Electricity price of purchasing active/reactive power from the main grid via Bus 0.

c_n^f/c_n^e Fuel/emission price for the DDG units at Bus n.

ω Emission factor of the DDG units (kg/kWh).

K Total number of available RDG units.

R Total number of available ES units.

T All the time intervals in the planning horizon.

N Total number of all the buses.

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

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

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

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

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

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

δ_n/τ_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.

η Violation probability.

γ ES efficiency.

b_r^0 Initial power level of the r^{th} ES unit.

Π_i 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} Binary indicator of whether the k^{th} RDG unit is located at Bus n.

u_{kl} Binary indicator of whether the capacity of the k^{th} RDG unit is the l^{th} element in $\mathcal{X}$.

w_{rn} Binary indicator of whether the r^{th} ES unit is located at Bus n.

x_k Size of the k^{th} RDG unit.

y_r 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.

p^t_0/q^t_0 Active/ reactive power purchased from the main grid via Bus 0 in period t.

P^t_{mn} Active power flow from Bus m to n in period t.

Q^t_{mn} Reactive power flow from Bus m to n in period t.

V^t_n Complex voltage at Bus n in period t.

I^t_{mn} Complex current from Bus m to n in period t.

p^t_n/q^t_n Active/ reactive power output of the DDG unit/ reactive power source at Bus n in period t.

b^t_r Active power of the r^{th} ES unit in period t.

LS^t_{imn} Load-shedding variables.

$\boldsymbol{z}$ $[z_{kn}, \forall\, k \in [K], n \in \mathcal{N}]^\top$.

$\boldsymbol{x},\ \boldsymbol{u}$ $[x_1, \ldots, x_K]^\top$, $[u_{kl}, \forall\, k \in [K], l \in [L]]^\top$

$\boldsymbol{y}, \boldsymbol{w}$ $[y_1, \ldots, y_R]^\top$, $[w_{rn}, \forall\, r \in [R], n \in \mathcal{N}]^\top$.