



Distributionally Robust Generation Expansion Planning With Unimodality and Risk Constraints

Pourahmadi, Farzaneh; Kazempour, Jalal

Published in:
IEEE Transactions on Power Systems

Link to article, DOI:
[10.1109/TPWRS.2021.3057265](https://doi.org/10.1109/TPWRS.2021.3057265)

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Pourahmadi, F., & Kazempour, J. (2021). Distributionally Robust Generation Expansion Planning With Unimodality and Risk Constraints. *IEEE Transactions on Power Systems*, 36(5), 4281-4295.
<https://doi.org/10.1109/TPWRS.2021.3057265>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Distributionally Robust Generation Expansion Planning With Unimodality and Risk Constraints

Farzaneh Pourahmadi, *Member, IEEE* and Jalal Kazempour, *Senior Member, IEEE*

Abstract—As more renewables are integrated into the power system, capacity expansion planners need more advanced long-term decision-making tools to properly model short-term stochastic production uncertainty and to explore its effects on expansion decisions. We develop a distributionally robust generation expansion planning model, accounting for a family of potential probability distributions of wind forecast error uncertainty. Aiming to include more realistic distributions, we construct more informed moment-based ambiguity sets by adding structural information of unimodality. We include operational-stage unit commitment constraints and model the risk of operational limit violations in two distinct forms: chance and conditional value-at-risk (CVaR) constraints. In both forms, the resulting expansion planning model is a mixed-integer second-order cone program. Using a thorough out-of-sample numerical analysis, we conclude: (i) the distributionally robust chance-constrained generation expansion planning model exhibits a better out-of-sample performance only if sufficiently accurate information about the first- and the second-order moments as well as the mode location of potential distributions is available; (ii) conversely, if such accurate information is unavailable, the distributionally robust CVaR-constrained generation expansion planning model outperforms; (iii) these two models have a similar performance when unimodality information is excluded.

Index Terms—Distributionally robust optimization, chance constraints, CVaR constraints, generation expansion planning, unimodality information.

I. INTRODUCTION

The increasing penetration of renewable energy sources with variable and uncertain production necessitates the improvement of existing long-term decision-making tools, enabling expansion planners to model short-term uncertainties more accurately and to explore their effects on generation expansion decisions [1]. One main challenge in long-term expansion studies is the estimation of the probability distribution of renewable power generation uncertainty. For decision-making under uncertainty, different stochastic approaches have been proposed, among which distributionally robust optimization (DRO) is potentially more efficient in terms of out-of-sample performance and computational tractability [2]–[4]. The advantages of DRO compared to scenario-based stochastic programming and robust optimization are discussed in [5].

Distributionally robust generation expansion planning accounts for a *family* of potential probability distributions of the underlying short-term uncertainty, collected in the so-called an *ambiguity set*. It then determines the optimal expansion decisions by minimizing the total *expected* expansion and operational cost of the system with respect to the *worst-case*

distribution picked from the ambiguity set. To construct an ambiguity set, two paradigms exist in the literature, namely metric- and moment-based approaches. In the former, the ambiguity set encompasses all distributions that are close to an empirical distribution, measured by a probability distance metric [6]–[9]. In the latter, the ambiguity set contains all distributions satisfying identical moment constraints [2], [5]. Ambiguity sets based on moment information, e.g., mean and covariance, provide superior tractability properties [8]. Therefore, the moment-based DRO approach has been used more often in the existing literature of capacity expansion planning [10]–[14].

Moment-based ambiguity set contains all distributions with identical moments but with potentially various shapes and different structures, which may cause overly conservative solutions. In practice, in addition to moment information, we may have access to further information about structural properties of potential distributions of renewable power generation achieved from historical data. Adding more information makes the ambiguity set contain more realistic distributions, resulting in a satisfactory trade off between the capacity expansion cost and the cost incurred by the lack of supply security in long run. This paper explores how important is to incorporate structural information of distributions describing the short-term uncertainty into a DRO-based expansion planning model in a computationally tractable way.

Another concern about DRO-based capacity expansion planning models is how to control the conservativeness level of the solution. An appropriate decision-making tool is expected to provide an expansion planner with different degrees of freedom to manage risks of operational limit violations. This expectation raises two research questions: (i) In addition to shrinking the ambiguity set by adding extra information, how can the expansion planner adjust the conservativeness level of the expansion solution? (ii) Under which circumstances does the expansion planner need to make more conservative decisions in order to reduce the operational risk?

To address the aforementioned challenges, we develop a centralized¹ generation capacity expansion planning model based on DRO. This model constructs one ambiguity set for each operational time period throughout the planning horizon,

¹By centralized, we refer to a model with a central expansion planner who decides on the technology, capacity and location of all new generating units to be installed throughout the underlying power system. The outcomes of this centralized model are not necessarily identical to those in a market-based model, wherein competing producers whose objective is to maximize their individual profit make investment decisions to expand their own generation portfolio. However, policy-makers and market regulators need such a centralized model in order to identify the optimal capacity expansion decisions for the whole system, and thereby design proper investment incentives in electricity markets. From this perspective, the centralized model can be used as an ideal benchmark that provides a lower bound for the total expansion and operational cost of the system in long run.

Farzaneh Pourahmadi is with the Department of Mathematical Sciences, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark (e-mail: farzaneh.pourahmadi@gmail.com).

Jalal Kazempour is with the Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark (e-mail: seykaz@elektro.dtu.dk).

while considering the fact that the most potential distributions of wind power production have *unimodality* properties [15], [16]. By doing so, we incorporate the additional information of unimodality structure into moment-based ambiguity sets used in the proposed generation expansion planning model. It reduces the conservativeness of expansion decisions, and subsequently leads to a reduced capacity expansion cost while maintaining the system reliability in the desired level. The proposed expansion model includes the operational-stage unit commitment constraints, and models the risk of operational limit violations in two distinct forms: chance and conditional value-at-risk (CVaR) constraints. These operational-stage *probabilistic* constraints contain random variables with unknown probability distribution. A chance-constrained model limits the *violation probability* of probabilistic constraints only. On the contrary, a CVaR-constrained model restricts not only the violation probability but also the *violation magnitude*, which might be of interest for the expansion planner. As a result, the conservativeness can be tuned based on (i) the type of probabilistic constraints, and (ii) the amount of information included in ambiguity sets. We use tractable reformulations and some approximation techniques to recast the resulting DRO problems as mixed-integer second-order cone (SOC) programs. The performance of the proposed models is eventually explored using out-of-sample simulations.

To the best of our knowledge, there is no work in the existing literature proposing a distributionally robust risk-constrained generation expansion planning model given unimodality information. An extensive literature review about uncertainty modeling in expansion planning problems including DRO is available in [10] and [17]. The existing works on expansion planning problems, e.g., [10]–[14] and [18], neglect structured distributions in ambiguity sets. Additionally, the violation magnitude risk of operational limits caused by short-term uncertainties is not considered in the existing literature on distributionally robust expansion planning models [10]–[14].

The main findings of our thorough out-of-sample simulations are as follows: In the case the expansion planner does *not* access to accurate information about the first two moments, i.e., mean and covariance, and the mode location of probability distributions describing wind power generation uncertainty, it is more appealing to enforce operational-stage constraints in the form of CVaR constraints. In other words, when information drawn from training empirical dataset is different than that of unseen test dataset, using CVaR constraints improves the out-of-sample performance. In contrast, as the probability distance between training and test datasets decreases, the chance-constrained approach exhibits a comparatively stronger performance in terms of out-of-sample cost. Moreover, our numerical analysis highlights the importance of incorporating structured distributions with unimodality feature into the proposed generation capacity expansion model. Finally, we show that when structural information of unimodality is neglected, the chance- and CVaR-constrained approaches have similar performance.

The remainder of this paper is structured as follows. Section II provides some preliminaries about the proposed capacity expansion planning model, explains the unimodality structure,

presents the structured ambiguity sets, and differentiates the chance constraints from the CVaR constraints. Section III presents the proposed distributionally robust generation expansion planning model with unimodality and risk constraints. Section IV describes the solution methodology. Section V provides numerical results for a case study based on the IEEE 118-node test system. Section VI concludes the paper. Appendix A presents a nomenclature. Appendix B explains a tight relaxation technique used. Finally, Appendix C provides the final tractable model obtained by the implementation of several approximations and reformulations.

II. MODELING FRAMEWORK

A. Preliminaries

We focus on the expansion of generation assets, and therefore it is assumed that the network topology throughout the planning horizon is fixed. Similar to [10], the proposed model follows a static capacity expansion planning model with a single target year. See [19] for the comparison of static versus multi-year dynamic capacity expansion model. We consider both long- and short-term uncertainties. The long-term uncertainty of load growth and wind capacity expansion is described by a finite set of scenarios. For each long-term scenario, the short-term uncertainty of wind power production is modeled by an ambiguity set. We assume that wind power production is the only source of short-term uncertainty. Moreover, we incorporate unit commitment constraints into the operational stage of the proposed model to represent flexibility requirements, wherein the commitment status of conventional units is dependent on the long-term uncertainty, but independent of the short-term uncertainty. We represent the power flow across the network through a linearized lossless DC model. The chronological variability of net load across the target year is considered. In order to make the problem tractable, we assign a number of representative days, each comprising 24 operating hours to represent the target year. Accordingly, for each long-term scenario n , representative day d , and operating hour t , we construct a separate ambiguity set denoted by Π_{ndt} . For notational clarity, we drop indexes, and represent it by Π .

In this paper, \mathbb{R} refers to the set of real numbers, whereas \mathbb{R}_+ represents the set of positive real numbers. We consider Z number of wind power units, whose short-term production forecast errors are collected in vector $\xi_{ndt} \in \mathbb{R}^Z$. Again, for notational clarity, we drop indexes, and represent forecast error vector by ξ .

B. Unimodality Structure

We explain unimodal distributions known as an important class of distributions commonly encountered in practice. A distribution admitting a density function is unimodal with mode ν , if the density function is non-increasing along any rays starting from ν . In the following, we introduce the generalized definition of unimodality structure stated in [20]–[23], that also covers the distributions without density function. In this definition, distribution \mathbb{P} on \mathbb{R}^Z is α -unimodal with mode $\mathbf{0}$, if for any fixed $\alpha \in \mathbb{R}_+$, $u^\alpha \mathbb{P}(B/u)$ is non-decreasing in $u > 0$ for any Borel set $B \in \mathcal{B}^Z$ [20]–[23]. Note that

\mathcal{B} represents the Borel σ -algebra containing all Borel sets on a topological space. From the definition above, it can be concluded if \mathbb{P} is a continuous α -unimodal distribution with density function $f(\tau)$ and mode $\mathbf{0}$, for every fixed $\tau \neq \mathbf{0}$, $u^{Z-\alpha}f(\tau u)$ is non-increasing in $u \in (0, \infty)$, meaning that density function $f(\tau u)$ does not grow faster than $u^{\alpha-Z}$ on $(0, \infty)$ wherein α controls the rate of increase. As a result, parameter α specifies the shape of the unimodal distribution. In the case α is equal to the number of random variables, i.e., $\alpha = Z$, then $f(\tau u)$ or $f(\tau)$ is non-increasing on $(0, \infty)$ for all τ . In other words, the density $f(\tau u)$ is non-increasing along any rays starting from the mode located at the origin [22], [23]. Relying on our observations from the historical data of wind power production that we will use later in our case study, we suppose that the fitted density function of wind power production is non-increasing along any rays emanating from the mode (e.g., see Fig. 2 in Section V). Accordingly, considering the definition of α -unimodality, we set parameter α equal to Z in this paper.

It is also stated in [16] and [23] that a random vector ξ is α -unimodal if and only if there exists a random vector $\mathcal{X} \in \mathbb{R}^Z$ such that $\xi = \mathcal{U}^{1/\alpha} \mathcal{X}$ where \mathcal{U} is uniform in $(0,1)$ and independent of ξ . Therefore, it can be deduced that as $\alpha \rightarrow \infty$, the unimodality requirement disappears. Note that without loss of generality, when the true mode value is ν , we can define a new random vector, i.e., ξ minus ν with mode $\mathbf{0}$, and then apply the definition, accordingly.

C. Structured Ambiguity Set with Unimodality Information

We describe the moment-based ambiguity set used in this paper to model the uncertainty of wind power forecast error. Considering the fact that empirical distributions of wind power production and thereby distributions of their forecast error often follow a unimodal structure, the moment-based ambiguity set including all probability distributions with identical moments and unimodality requirement is written as

$$\Pi = \{\mathbb{P} \in \Omega^\alpha(\mathbb{R}^Z) : \mathbb{E}_{\mathbb{P}}[\xi] = \mu, \mathbb{E}_{\mathbb{P}}[\xi\xi^\top] = \Sigma, M_{O\mathbb{P}}[\xi] = \nu\}, \quad (1)$$

where Ω^α denotes plausible α -unimodal distributions on \mathbb{R}^Z . Vector $\mu \in \mathbb{R}^Z$ and matrix $\Sigma \in \mathbb{R}^{Z \times Z}$ represent the first- and second-order moments, i.e., mean and covariance, of wind forecast error ξ obtained from historical data. Further, $(\cdot)^\top$ is the transpose operator and $\mathbb{E}_{\mathbb{P}}[\cdot]$ represents the expectation operator under probability distribution \mathbb{P} . Without loss of generality, it is assumed that the mean of forecast error of each wind power unit is zero, i.e., $\mu = \mathbf{0}$. Operator $M_{O\mathbb{P}}[\cdot]$ determines the mode location of unimodal distribution under probability distribution \mathbb{P} . Parameter α determines the degree of unimodality and $\nu \in \mathbb{R}^Z$ specifies the mode location. We assume that the values for moments and the mode location estimated from historical data are exact. The last term in (1) represents the structural information of unimodality, which leads to excluding non-unimodal distributions from the ambiguity set. We hypothesize adding this extra information to the moment-based ambiguity set brings about a lower total capacity expansion and operational cost. To show that, in the following, we aim to build a model to stress the importance of taking into consideration such structural information.

D. Chance Constraint Versus CVaR Constraint

In this paper, we use two risk measures for restricting the violation of probabilistic constraints in the operational stage. Let us consider a linear probabilistic constraint as

$$a(\mathbf{y})^\top \xi \leq b(\mathbf{y}), \quad (2)$$

where \mathbf{y} is a decision vector and ξ is a random vector. Note that (2) is an infinite-dimensional constraint. One way to overcome such an infinite-dimensional nature while managing constraint violation is to enforce (2) as a distributionally robust chance constraint, i.e.,

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{a(\mathbf{y})^\top \xi \leq b(\mathbf{y})\} \geq 1 - \epsilon, \quad (3)$$

where ϵ is a given non-negative parameter whose value is lying between zero and one. The term $1 - \epsilon$ denotes the *confidence level* of satisfying constraint (3) for the worst-case distribution \mathbb{P} within ambiguity set Π . Constraint (3), however, does not provide any guarantee on the severity of violations. This may motivate expansion planners to employ another risk measure, the so-called distributionally robust CVaR constraints, i.e.,

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^\epsilon[a(\mathbf{y})^\top \xi] \leq b(\mathbf{y}), \quad (4)$$

where the right tail of $a(\mathbf{y})^\top \xi$ is examined for the worst-case distribution \mathbb{P} within ambiguity set Π . Note that the conditional expectation of $a(\mathbf{y})^\top \xi$ is measured by $\text{CVaR}_{\mathbb{P}}^\epsilon[a(\mathbf{y})^\top \xi]$, which is defined as [23]–[26]

$$\text{CVaR}_{\mathbb{P}}^\epsilon[a(\mathbf{y})^\top \xi] = \min_{\theta \in \mathbb{R}} \left\{ \theta + \frac{1}{\epsilon} \mathbb{E}_{\mathbb{P}}[a(\mathbf{y})^\top \xi - \theta]^+ \right\}. \quad (5)$$

It is well-known that distributionally robust chance constraints can be approximated using distributionally robust CVaR constraints in a conservative manner [8], [27]. The rationale behind such a conservative approximation is that distributionally robust CVaR constraints guarantee the satisfaction of distributionally robust chance constraints. In [27], it is proven that in the case the ambiguity set includes the exact values for the first two moment information without any other information, the approximation is exact, i.e.,

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^\epsilon[a(\mathbf{y})^\top \xi] \leq b(\mathbf{y}) \Leftrightarrow \min_{\mathbb{P} \in \Pi} \mathbb{P}\{a(\mathbf{y})^\top \xi \leq b(\mathbf{y})\} \geq 1 - \epsilon. \quad (6)$$

The above equivalence means if the last term in (1) is neglected, distributionally robust chance and CVaR constraints can be enforced, interchangeably. Because of that, in [27] and [28], distributionally robust CVaR constraints are used to reformulate distributionally robust chance constraints. However, as we consider structural information in the ambiguity set, the equivalence (6) does not necessarily hold. As a result, we examine both types of risk measures to restrict violations of operational-stage probabilistic constraints.

Note that in the rest of the paper, we consider lower-case letters for variables, upper-case letters for parameters, bold lower-case letters for variable vectors, and bold upper-case letters for parameter matrices. All symbols are defined within the text. In addition, a nomenclature is provided in Appendix A.

III. MODEL

The proposed distributionally robust generation expansion planning model writes as

$$\min_{\mathbf{z}, \mathbf{x}, \mathbf{v}} \mathbf{K}^\top \mathbf{z} + \sum_{n, d, t} \iota_n \omega_d \left(\mathbf{S}^\top \mathbf{v}_{ndt} + \max_{\mathbb{P} \in \Pi} \min_{\mathbf{p}} \mathbb{E}_{\mathbb{P}}[\mathbf{C}^\top \mathbf{p}_{ndt}(\boldsymbol{\xi})] \right) \quad (7a)$$

subject to:

$$x_{indt} \leq z_i, \quad \forall i \in I^{G^C}, n, d, t \quad (7b)$$

$$-x_{ind(t-1)} + x_{indt} - x_{indr} \leq 0, \quad \forall r \in \{t, \dots, U_i^{\text{up}} + t - 1\}, \forall i, n, d, t \quad (7c)$$

$$x_{ind(t-1)} - x_{indt} + x_{indr} \leq 1, \quad \forall r \in \{t, \dots, U_i^{\text{dn}} + t - 1\}, \forall i, n, d, t \quad (7d)$$

$$-x_{ind(t-1)} + x_{indt} - v_{indt} \leq 0, \quad \forall i, n, d, t \quad (7e)$$

$$\mathbf{z} \in \{0, 1\} \quad (7f)$$

$$\mathbf{x}_{ndt}, \mathbf{v}_{ndt} \in \{0, 1\}, \quad \forall n, d, t. \quad (7g)$$

$$\mathbf{1}^\top \mathbf{p}_{ndt}(\boldsymbol{\xi}) + \mathbf{1}^\top (\mathbf{w}_{ndt} + \boldsymbol{\xi}) = \mathbf{1}^\top \mathbf{l}_{ndt}, \quad \mathbb{P}\text{-a.s. } \forall n, d, t \quad (7h)$$

$$\text{Probabilistic constraints.} \quad (7i)$$

Objective function (7a) minimizes the total cost, including the annualized expansion cost which is a deterministic cost, as well as the expected annualized operational cost under long-term uncertainty and the worst-case probability distribution of short-term uncertainty. Parameter vector $\mathbf{K} \in \mathbb{R}_+^{G^C}$ gives the annualized capital cost of candidate conventional generating units $i \in I^{G^C} = \{1, \dots, G^C\}$ to be built. Binary variable vector $\mathbf{z} = \{z_1, \dots, z_i, \dots, z_{G^C}\}^\top$ indicates which candidate units are selected to be built. As operational costs, parameter vectors $\mathbf{C} \in \mathbb{R}_+^G$ and $\mathbf{S} \in \mathbb{R}_+^G$, respectively, represent production and start-up costs of all existing and candidate conventional units $i \in I^G = \{1, \dots, G\}$. Note that $I^{G^C} \subset I^G$. The operational cost of wind power units is assumed to be zero. Operational variable vectors are $\mathbf{p}_{ndt}(\boldsymbol{\xi}) \in \mathbb{R}_+^G$, $\mathbf{v}_{ndt} \in \{0, 1\}^G$ and $\mathbf{x}_{ndt} \in \{0, 1\}^G$. In details, continuous variable $p_{indt}(\boldsymbol{\xi}) \in \mathbf{p}_{ndt}(\boldsymbol{\xi})$ as well as binary variables $v_{indt} \in \mathbf{v}_{ndt}$ and $x_{indt} \in \mathbf{x}_{ndt}$, respectively, refer to the production level, start-up status and on/off commitment status of conventional generating unit i under long-term scenario n , representative day d and hour t . For each representative day d , a weight is assigned by $\omega_d \in \mathbb{N}$, representing the number of days in the target year grouped by a clustering method, such that $\sum_d \omega_d = 365$. Note that \mathbb{N} is the set of natural numbers. In addition, for each long-term scenario n , a probability is given by $\iota_n \in \mathbb{R}_+$, such that $\sum_n \iota_n = 1$.

Fig. 1 shows the chronological sequence of decision-making and uncertainty realization. Accordingly, capacity expansion decisions \mathbf{z} are made before the realization of long-term uncertainty, and therefore these decisions are not indexed by long-term scenario n , and are not a function of short-term uncertainty $\boldsymbol{\xi}$. In addition, these capacity expansion decisions are made for the whole planning horizon, and therefore they are not indexed by representative day d and hour t . The operational commitment and start-up variables \mathbf{x}_{ndt} and \mathbf{v}_{ndt} are indexed by long-term scenario n , because their optimal values are determined after the realization of the long-term

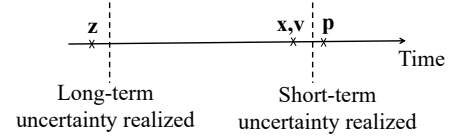


Fig. 1. Chronological sequence of decision-making and uncertainty realization

TABLE I
THE RELATION OF VARIABLES TO LONG- AND SHORT-TERM UNCERTAINTIES

Variable	Indexed by long-term scenario n	Function of short-term uncertainty $\boldsymbol{\xi}$
Expansion (\mathbf{z})	No	No
Commitment (\mathbf{x})	Yes	No
Start-up (\mathbf{v})	Yes	No
Production (\mathbf{p})	Yes	Yes

uncertainty. In addition, \mathbf{x}_{ndt} and \mathbf{v}_{ndt} are indexed by d and t , since they are operational-stage variables. However, as we make these decisions before the short-term uncertainty realization, they are not considered as a function of short-term uncertainty $\boldsymbol{\xi}$. By doing so, we implicitly assume that conventional generating unit i cannot be started up or shut down in the real-time operation when the short-term uncertainty is realized. Finally, the production variable $\mathbf{p}_{ndt}(\boldsymbol{\xi})$ is indexed by long-term scenario n , and considered as a function of short-term uncertainty $\boldsymbol{\xi}$. This implies that if conventional generating unit i is on (i.e., $x_{indt} = 1$), its production level can be modified after the realization of short-term uncertainty in order to offset the supply-demand power imbalance of the system. Table I summarizes the relation of variables to long- and short-term uncertainties.

Constraint (7b) enforces that a candidate conventional generating unit can be committed only if it is selected to be built. Constraints (7c) and (7d) limit the minimum up and down time of unit i to U_i^{up} and U_i^{dn} , respectively. Constraint (7e) determines the start-up status of unit i . Constraints (7f) and (7g) impose the integrality conditions. The power balance is enforced by (7h) under any realization of short-term uncertainty with a probability of almost surely one. Note that $\mathbf{1}$ is a vector of ones with an appropriate dimension. According to (7h), the production of conventional units, i.e., $\mathbf{p}_{ndt}(\boldsymbol{\xi})$, and the production of wind farms, i.e., $\mathbf{w}_{ndt} + \boldsymbol{\xi}$, supply inelastic loads. Note that parameter vector $\mathbf{w}_{ndt} \in \mathbb{R}_+^Z$ represents the mean of wind power production forecast, while $\boldsymbol{\xi}$ represents the forecast error with respect to \mathbf{w}_{ndt} . Furthermore, parameter vector $\mathbf{l}_{ndt} \in \mathbb{R}_+^L$ gives the load level of inelastic demands. Finally, (7i) includes the rest of operational-stage probabilistic constraints which are dependent on $\boldsymbol{\xi}$. Unlike power balance conditions (7h) which is also $\boldsymbol{\xi}$ -dependent, we allow probabilistic constraints (7i) to be violated to some extent in order to control conservativeness. These constraints will be presented in form of either distributionally robust chance or CVaR constraints in the next two subsections.

A. Distributionally Robust Chance Constraints

As the first alternative, probabilistic constraints (7i) can be written in form of distributionally robust chance constraints as

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{indt}(\boldsymbol{\xi}) \leq P_i^{\max} x_{indt}\} \geq 1 - \epsilon, \quad \forall i, n, d, t \quad (8a)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{indt}(\boldsymbol{\xi}) \geq P_i^{\min} x_{indt}\} \geq 1 - \epsilon, \quad \forall i, n, d, t \quad (8b)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{indt}(\boldsymbol{\xi}) - p_{ind(t-1)}(\boldsymbol{\xi}) \leq R_i^{\text{up}} x_{ind(t-1)} + R_i^{\text{st}}(1 - x_{ind(t-1)})\} \geq 1 - \epsilon, \quad \forall i, n, d, t \quad (8c)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{ind(t-1)}(\boldsymbol{\xi}) - p_{indt}(\boldsymbol{\xi}) \leq R_i^{\text{dn}} x_{indt} + R_i^{\text{st}}(1 - x_{indt})\} \geq 1 - \epsilon, \quad \forall i, n, d, t \quad (8d)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{\mathbf{M}_j^{\text{G}} \mathbf{p}_{ndt}(\boldsymbol{\xi}) + \mathbf{M}_j^{\text{W}}(\mathbf{w}_{ndt} + \boldsymbol{\xi}) - \mathbf{M}_j^{\text{D}} \mathbf{1}_{ndt} \leq F_j^{\max}\} \geq 1 - \epsilon, \quad \forall j, n, d, t \quad (8e)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{\mathbf{M}_j^{\text{G}} \mathbf{p}_{ndt}(\boldsymbol{\xi}) + \mathbf{M}_j^{\text{W}}(\mathbf{w}_{ndt} + \boldsymbol{\xi}) - \mathbf{M}_j^{\text{D}} \mathbf{1}_{ndt} \geq -F_j^{\max}\} \geq 1 - \epsilon, \quad \forall j, n, d, t. \quad (8f)$$

Without loss of generality, we consider an identical ϵ for all distributionally robust chance constraints (8). The upper bound P_i^{\max} and lower bound P_i^{\min} for production level of conventional unit i are enforced in (8a) and (8b), respectively. In addition, the ramping constraints are enforced by (8c) and (8d), wherein R_i^{up} and R_i^{dn} give the ramp-up and ramp-down capabilities while R_i^{st} represents the start-up and shut-down ramp rate capability. Finally, the power flow in transmission line $j \in J$ is restricted by F_j^{\max} in (8e) and (8f), wherein power transfer distribution factor matrices $\mathbf{M}^{\text{G}} \in \mathbb{R}^{J \times G}$, $\mathbf{M}^{\text{W}} \in \mathbb{R}^{J \times Z}$ and $\mathbf{M}^{\text{D}} \in \mathbb{R}^{J \times L}$ map the power of conventional generating units, wind farms, and loads on the power flow of line j , respectively. Note that subscript j in $\mathbf{M}_j^{(\cdot)}$ picks a row corresponding to line j from matrix $\mathbf{M}^{(\cdot)}$.

B. Distributionally Robust CVaR Constraints

As the second alternative, constraints (7i) are enforced in form of distributionally robust CVaR constraints as

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{indt}(\boldsymbol{\xi})] \leq P_i^{\max} x_{indt}, \quad \forall i, n, d, t \quad (9a)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[-p_{indt}(\boldsymbol{\xi})] \leq -P_i^{\min} x_{indt}, \quad \forall i, n, d, t \quad (9b)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{indt}(\boldsymbol{\xi}) - p_{ind(t-1)}(\boldsymbol{\xi})] \leq R_i^{\text{up}} x_{ind(t-1)} + R_i^{\text{st}}(1 - x_{ind(t-1)}), \quad \forall i, n, d, t \quad (9c)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{ind(t-1)}(\boldsymbol{\xi}) - p_{indt}(\boldsymbol{\xi})] \leq R_i^{\text{dn}} x_{indt} + R_i^{\text{st}}(1 - x_{indt}), \quad \forall i, n, d, t \quad (9d)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[\mathbf{M}_j^{\text{G}} \mathbf{p}_{ndt}(\boldsymbol{\xi}) + \mathbf{M}_j^{\text{W}}(\mathbf{w}_{ndt} + \boldsymbol{\xi}) - \mathbf{M}_j^{\text{D}} \mathbf{1}_{ndt}] \leq F_j^{\max}, \quad \forall j, n, d, t \quad (9e)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[-\mathbf{M}_j^{\text{G}} \mathbf{p}_{ndt}(\boldsymbol{\xi}) - \mathbf{M}_j^{\text{W}}(\mathbf{w}_{ndt} + \boldsymbol{\xi}) + \mathbf{M}_j^{\text{D}} \mathbf{1}_{ndt}] \leq -F_j^{\max}, \quad \forall j, n, d, t. \quad (9f)$$

IV. SOLUTION STRATEGY

Both proposed models, i.e., distributionally robust chance- and CVaR-constrained models, are computationally intractable owing to the relatively high number of binary variables as well as the infinite dimension of the problem. Similar to [10], we reduce the number of binary variables by deploying a tight relaxation of unit commitment constraints [29]. This technique

relaxes the integrality condition of operational constraints (7g) in form of $0 \leq x_{indt} \leq 1$ and $0 \leq v_{indt} \leq 1$. Additional constraints are added to the problem in order to tighten the relaxation. These additional inequalities are considered in the form of risk constraints which are given in Appendix B. Note that the discrete nature of expansion decisions z_i imposed in (7f) is still preserved. In order to further reduce the computational burden, we use a *linear decision rule* [30] to approximate the recourse action of conventional generating units. In this way, production $p_{indt}(\boldsymbol{\xi})$ of unit i under long-term scenario n , representative day d and hour t is approximated by $p_{indt} + \beta_{indt} \mathbf{1}^{\top} \boldsymbol{\xi}$, where variable $p_{indt} \in \mathbb{R}_+$ is the nominal (tentative) dispatch of unit i before the short-term uncertainty realization. In addition, variable $\beta_{indt} \in \mathbb{R}$, the so-called *participation factor*, is the linear response of unit i to short-term uncertainty. This variable takes a value lying between -1 and 1, and gives the contribution of unit i to coping with total renewable forecast error uncertainty of the system, i.e., $\mathbf{1}^{\top} \boldsymbol{\xi}$. In the following, we explain how to reformulate different parts of the proposed models in the previous section, leading to computationally tractable models.

A. Reformulation of Constraints (8)

For example, let us consider the distributionally robust chance constraint (8a) enforcing the capacity constraint of unit i . By applying the linear decision rule, it reformulates as

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{\beta_{indt} \mathbf{1}^{\top} \boldsymbol{\xi} \leq P_i^{\max} x_{indt} - p_{indt}\} \geq 1 - \epsilon, \quad \forall i, n, d, t. \quad (10a)$$

From now on, we drop indices n, d and t for notational clarity from $\{p_{indt}, x_{indt}, \beta_{indt}, \eta_{indt}\}$ and present them in form of $\{p_i, x_i, \beta_i, \eta_i\}$. According to [15], constraint (10a) can be exactly reformulated as

$$\sqrt{\frac{1 - \epsilon - \eta_i^{-\alpha}}{\epsilon}} \|\beta_i \mathbf{1}^{\top} \boldsymbol{\Phi}\| \leq \eta_i (P_i^{\max} x_i - p_i - \beta_i \mathbf{1}^{\top} \boldsymbol{\nu}) - \left(\frac{\alpha + 1}{\alpha}\right) (\beta_i \mathbf{1}^{\top}) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall i, \eta_i \geq \left(\frac{1}{1 - \epsilon}\right)^{1/\alpha}, \quad (10b)$$

where $\boldsymbol{\Phi} = \frac{\alpha+2}{\alpha}(\boldsymbol{\Sigma} - \boldsymbol{\mu}\boldsymbol{\mu}^{\top}) - \frac{1}{\alpha^2}(\boldsymbol{\mu} - \boldsymbol{\nu})(\boldsymbol{\mu} - \boldsymbol{\nu})^{\top}$. Note that (10b) includes an infinite number of SOC constraints, originating by the infinite number of η_i . This maintains the intractability of the problem. To achieve tractability, [15] suggests to consider a finite number of η_i , indicated by parameters η_{is} , whose values are systematically selected as explained in iterative Algorithm 1. By implementing it, a sequence of monotonically increasing values for objective function (7a) is obtained, which eventually converges to a fixed value. The convergence rate of Algorithm 1 will be numerically explored later in Section V. The proposed algorithm does not guarantee optimality to the original problem.

B. Reformulation of Constraints (9)

Similar to the reformulation of distributionally robust chance constraints (8), the reformulation of distributionally robust CVaR constraints (9) results in a mixed-integer SOC program. However, unlike Algorithm 1, the reformulation procedure is not iterative. For example, we provide here the reformulation of CVaR constraint (9a). The other CVaR

Algorithm 1 Iterative procedure for selection of η_{is}

Step 1. Input: A small tolerance ϑ to control convergence.

Step 2. Initialization: Set iteration index $k = 1$ and $\eta_{i0} = (\frac{1}{1-\epsilon})^{1/\alpha}$, $\forall i$.

Step 3. Optimization: Given values for η_{is} where $s = \{0, \dots, k-1\}$, replace original infinite-dimensional constraint (10b) by

$$\sqrt{\frac{1-\epsilon-\eta_{is}^{-\alpha}}{\epsilon}} \|\beta_i \mathbf{1}^\top \Phi\| \leq \eta_{is} (P_i^{\max} x_i - p_i - \beta_i \mathbf{1}^\top \nu) - \left(\frac{\alpha+1}{\alpha}\right) (\beta_i \mathbf{1}^\top) (\boldsymbol{\mu} - \nu), \quad \forall i, s, \quad (11)$$

which includes a finite number of SOC constraints. Replace other distributionally robust chance constraints (8b)-(8f) in the same manner. Solve the resulting mixed-integer SOC program, and save the optimal solutions \mathbf{p}^* , \mathbf{x}^* and $\boldsymbol{\beta}^*$, where superscript $(\cdot)^*$ stands for the optimal value. Finally, save $f^{(k)}$ indicating the optimal value of the objective function (7a) in iteration k .

Step 4. Parameter η Selection: Given optimal values \mathbf{p}^* , \mathbf{x}^* and $\boldsymbol{\beta}^*$ while treating η_i as a variable, find the worst value for η_i , indicated as $\eta_i^{(k)*}$, causing the largest violation of (11). Repeat this step for other SOC constraints related to distributionally robust chance constraints (8b)-(8f).

Step 5. Termination: If $\frac{f^{(k)} - f^{(k-1)}}{f^{(k-1)}} \leq \vartheta$, stop and report the optimal generation expansion decisions with a level of accuracy ϑ . Otherwise, set $\eta_{ik} \leftarrow \eta_i^{(k)*}$, $k \leftarrow k+1$, and continue the algorithm in Step 3.

constraints are reformulated in the same manner. Using the linear decision rule, (9a) can be rewritten as

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[\beta_i \mathbf{1}^\top \boldsymbol{\xi}] \leq P_i^{\max} x_i - p_i, \quad \forall i. \quad (12a)$$

Considering the definition of CVaR operator in (5), the left side of (12a) is equal to

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[\beta_i \mathbf{1}^\top \boldsymbol{\xi}] = \min_{\theta_i \in \mathbb{R}} \left\{ \theta_i + \frac{1}{\epsilon} \max_{\mathbb{P} \in \Pi} \mathbb{E}_{\mathbb{P}}[\beta_i \mathbf{1}^\top \boldsymbol{\xi} - \theta_i]^+ \right\}, \quad (12b)$$

where $\theta_i \in \mathbb{R}$ is an auxiliary variable. Therefore, (12a) boils down to

$$\theta_i + \frac{1}{\epsilon} \max_{\mathbb{P} \in \Pi} \mathbb{E}_{\mathbb{P}}[\beta_i \mathbf{1}^\top \boldsymbol{\xi} - \theta_i]^+ \leq P_i^{\max} x_i - p_i \quad \forall i. \quad (12c)$$

According to [23] and [26], computing $\max_{\mathbb{P} \in \Pi} \mathbb{E}_{\mathbb{P}}[\beta_i \mathbf{1}^\top \boldsymbol{\xi} - \theta_i]^+$ requires to calculate the worst-case expectation of a non-linear function, which is computationally prohibitive to solve. Hence, [23] and [26] propose to approximate this non-linear function by two piece-wise linear supporting functions, each of which consists of two linear pieces. By computing the worst-case expectation of the approximated functions, (12c) can be reformulated as

$$\begin{aligned} \left\| \theta_i - \left(\frac{\alpha+1}{\alpha}\right) \beta_i \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (P_i^{\max} x_i - p_i) \\ &- \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1 \right] \theta_i - \left(\frac{\alpha+1}{\alpha}\right) \beta_i \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i \\ \left\| \theta_i - \left(\frac{\alpha+1}{\alpha}\right) \beta_i \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (P_i^{\max} x_i - p_i) \end{aligned} \quad (13a)$$

$$\begin{aligned} &- \left[\frac{(2\epsilon-1)(\alpha+1)-1}{\alpha} \right] \theta_i - \left(\frac{\alpha+1}{\alpha}\right) \beta_i \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i \quad (13b) \\ \left\| \left(\frac{\alpha+1}{\alpha}\right) \theta_i - \left(\frac{\alpha+1}{\alpha}\right) \beta_i \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (P_i^{\max} x_i - p_i) \\ &- \left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha} \right] \theta_i - \left(\frac{\alpha+1}{\alpha}\right) \beta_i \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i. \quad (13c) \end{aligned}$$

The extensive proof is provided in [23]. Note that the exact reformulation of (12a) is intractable [22], [23]. Tighter approximations are proposed in [23] by increasing the number of pieces of the two linear supporting functions. However, the implementation of such conservative approximations results in semi-definite constraints, which may also make the problem computationally intractable. Therefore, we do not use a tighter approximation in this paper, and as a result, the proposed reformulation does not guarantee optimality to the original problem.

C. Reformulation of Power Balance and Objective Function

We first reformulate the power balance constraint (7h) in a tractable way. Given the linear response of conventional generating units to cope with the uncertainty of wind power forecast error, we rewrite the power balance constraint (7h) as

$$\mathbf{1}^\top (\mathbf{p}_{ndt} + (\mathbf{1}^\top \boldsymbol{\xi}) \boldsymbol{\beta}_{ndt}) + \mathbf{1}^\top (\mathbf{w}_{ndt} + \boldsymbol{\xi}) = \mathbf{1}^\top \mathbf{1}_{ndt}, \quad \forall n, d, t. \quad (14)$$

By separating the nominal ξ -independent terms from the ξ -dependent terms, (14) decomposes to the following two equations:

$$\mathbf{1}^\top \mathbf{p}_{ndt} + \mathbf{1}^\top \mathbf{w}_{ndt} = \mathbf{1}^\top \mathbf{1}_{ndt}, \quad \forall n, d, t \quad (15a)$$

$$\mathbf{1}^\top \boldsymbol{\beta}_{ndt} = -1, \quad \forall n, d, t. \quad (15b)$$

Note that the forecast error ξ has been cancelled out from the two sides of (15b).

Next, we reformulate the objective function (7a). Again, by applying the linear decision rule and reformulating the probabilistic and power balance constraints, the objective function and constraints no longer include the forecast error $\boldsymbol{\xi}$. Recall that we assume $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}] = \boldsymbol{\mu} = \mathbf{0}$. As a result, the objective function (7a) boils down to a single-stage function as

$$\min_{\mathbf{z}, \mathbf{p}, \mathbf{x}, \mathbf{v}, \boldsymbol{\beta}, [\cdot]} \mathbf{K}^\top \mathbf{z} + \sum_{n,d,t} \iota_n \omega_d (\mathbf{S}^\top \mathbf{v}_{ndt} + \mathbf{C}^\top \mathbf{p}_{ndt}), \quad (16)$$

where the annualized total expansion and operational cost is minimized with respect to variables $\mathbf{z}, \mathbf{p}, \mathbf{x}, \mathbf{v}$ and $\boldsymbol{\beta}$ along with auxiliary variables denoted by $[\cdot]$ resulting from the reformulation of probabilistic constraints.

A fully reformulated version of the original model (7) as a mixed-integer SOC program is available in Appendix C.

V. CASE STUDY

We implement the proposed distributionally robust chance- and CVaR-constrained models on an adapted version of the modified IEEE 118-node test system [31]. Data for technical parameters of 19 existing and 22 candidate conventional generating units, two existing wind farms, 99 demands and 186 transmission lines are provided in the online companion of

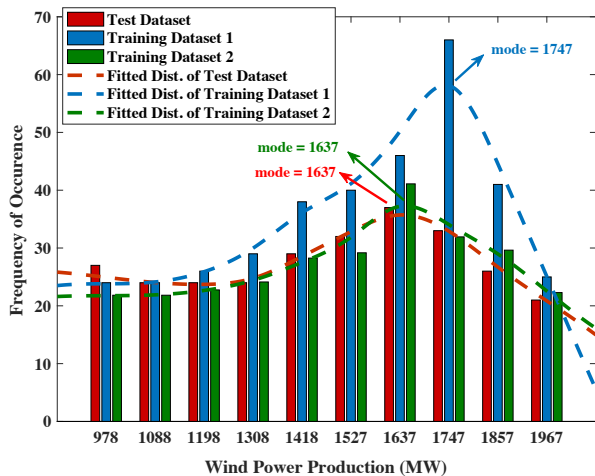


Fig. 2. Frequency of wind power realizations of farm w_1 in an arbitrarily selected operating hour. Test dataset (red) and two training datasets 1 and 2 (blue and green) follow unimodal distributions, but with relatively different places for the mode. The mode and mean of wind power production in the test dataset (red) are 1,637 MW and 1,520 MW; in training dataset 1 (blue) are 1,747 MW and 1,616 MW; and eventually in training dataset 2 (green) are 1,637 MW and 1,591 MW, respectively. Training dataset 2 (green) is more similar to the test dataset (red).

the paper [32]. We consider four types of conventional units, namely, nuclear, coal, combined cycle gas turbine (CCGT), and gas turbines (GT1 and GT2), such that GT2 is more flexible than GT1. Similar to [10], we model the long-term uncertainty of demand growth with 2 long-term scenarios in which the yearly penetration of wind power is 35%. The capacity of wind farms and load level under each scenario are given in [10]. In order to avoid congestion in the system, we increase the capacity of transmission lines by a factor of nine with respect to those in [31].

All simulations are implemented in Matlab using the YALMIP toolbox with Gurobi solver 8.1.1, on a 16-GB RAM personal computer clocking at 3.1 GHz. Relying on the value set for the confidence level ϵ , the distributionally robust CVaR-constrained problem is solved with the computational time of around 4 to 5 hours, whereas the iterative procedure for solving distributionally robust chance-constrained problem terminates in 4 to 6 iterations with the computational time varying from 9 to 12 hours. The computational time grows with increasing the value of confidence level in both distributionally robust chance- and CVaR-constrained models. All source codes are publicly available in the online companion of the paper [32].

A. Data for Wind Power

As in-sample data, two training datasets for wind power production of each farm are taken into consideration such that the distribution of one of them is more similar to that of the test dataset used for out-of-sample analysis. For each training dataset, 5,000 power trajectories are used to construct ambiguity sets, whereas another 5,000-trajectory data is considered as a test dataset to evaluate the capacity expansion solution. Note that each trajectory contains the wind power production over 24 hours.

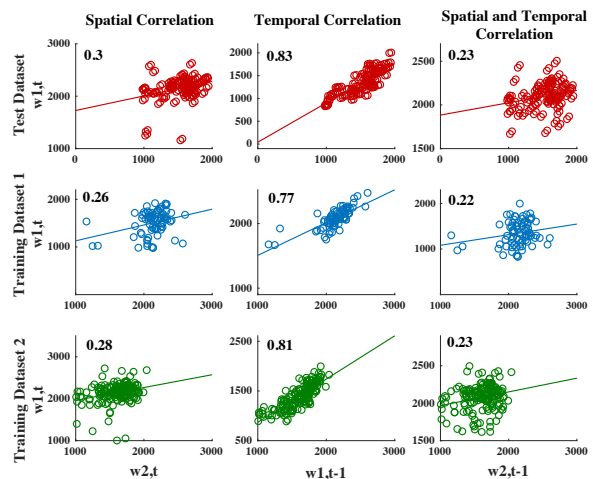


Fig. 3. Illustration of the spatial and temporal correlation in the test dataset (red) and two training datasets 1 and 2 (blue and green) for two wind farms w_1 and w_2 and two arbitrarily selected subsequent hours $t - 1$ and t . The numbers in black within the plots give the correlation coefficients.

For example, Fig. 2 presents the histogram of the production of a wind farm, namely w_1 , at a specific operating hour in a sample representative day pertaining to test and two training datasets. In particular, this figure illustrates the fitted probability distribution of each dataset, the place of mode, and the difference between the distributions of the test and training datasets. The Kullback–Leibler divergence of the test dataset (red) from training datasets 1 (blue) and 2 (green) is 0.3663 and 0.3326, respectively. This indicates that training dataset 2 (green) exhibits more similarity to the test dataset (red). This similarity can also be observed from the distributional characteristics such as mode, mean, and covariance matrix. Fig. 2 illustrates that the training dataset 2 (green) is distributionally closer to the test dataset (red) not only in terms of the mode location, but also in terms of mean wind power production.

Another feature for comparison is the correlation coefficients used to calculate the covariance matrix. The scatter plots in Fig. 3 show the spatial and temporal correlations between two wind farms w_1 and w_2 , and between two arbitrarily selected subsequent hours $t - 1$ and t in the test dataset and two training datasets 1 and 2. Note that the covariance matrix that we use to build the ambiguity set includes both spatial and temporal correlations. From this figure, it is evident that the correlation coefficients of the test dataset (red) are more similar to those of the training dataset 2 (green).

For each of these datasets, we use a K-means clustering approach to draw 10 representative days, representing the target year. With respect to the samples of each cluster, we compute the distributional information of the mode ν_{ndt} , mean μ_{ndt} and covariance matrix Σ_{ndt} to construct the corresponding ambiguity set for each long-term scenario, representative day and hour. Note that we consider different ambiguity sets for different long-term scenarios, representative days and hours, which do not necessarily include the same mean, mode and covariance matrix.

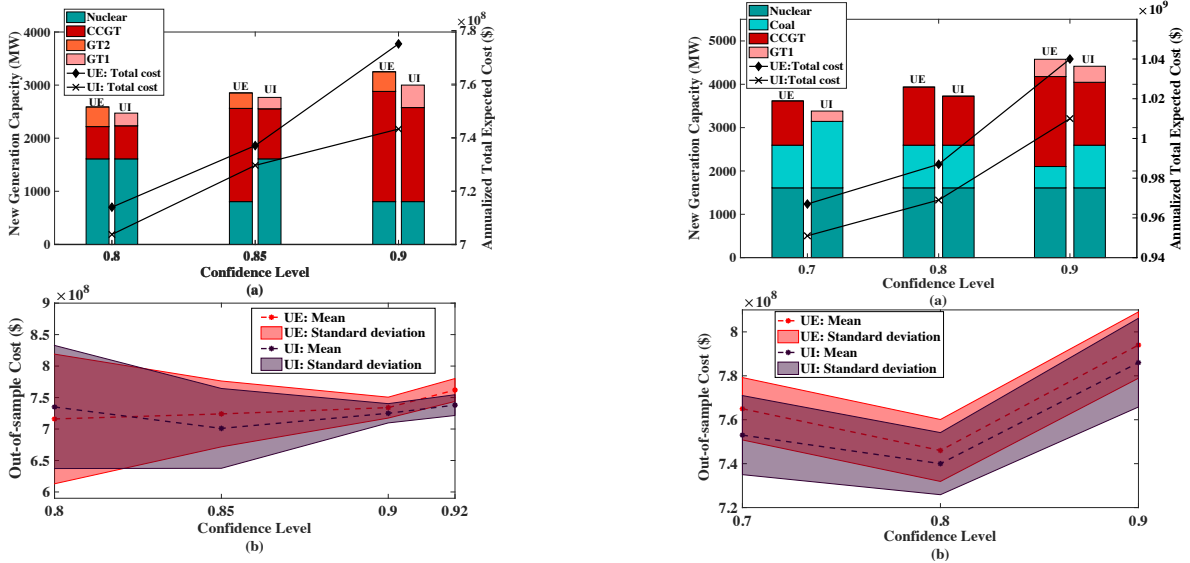


Fig. 4. Distributionally robust chance-constrained results obtained using training dataset 1 (left plots; less similar to the test dataset) and training dataset 2 (right plots; more similar to the test dataset): The upper plots show the new conventional generating units to be built and in-sample system cost. The lower plots present the out-of-sample system cost in terms of mean and standard deviation. UI: unimodality information included. UE: unimodality information excluded. Confidence level: $1 - \epsilon$.

B. Out-of-Sample Simulation

For the out-of-sample simulation, we first solve the tractable reformulated version of the proposed generation expansion planning model (provided in Appendix C), the so-called in-sample model, given the underlying training dataset. We then fix the expansion decisions to those obtained from the in-sample model, and finally re-optimize the operational stage according to the test dataset. For such a re-optimization, given the optimal expansion decisions obtained from the in-sample model, we solve the relaxed unit commitment problem for each representative day separately and deterministically. The out-of-sample problem is always feasible with no operational constraint violation, since we consider wind spillage and involuntarily load shedding as two extreme recourse actions. The former, i.e., wind spillage, occurs when there is wind power excess but conventional generating units cannot decrease their production. The latter, i.e., involuntarily load shedding, happens in the case there is wind power shortage, however conventional generating units cannot increase their production to compensate the deficit. We assume a zero cost for wind spillage since there is no production cost for wind farms, but a comparatively high cost, i.e., \$1,000/MWh, for load shedding. The out-of-sample cost contains the total expansion cost achieved from the in-sample model plus the total mean operational cost obtained from the out-of-sample simulation.

C. Expansion Results With Chance Constraints: Effects of Adding Unimodality Information

We focus on distributionally robust generation expansion planning model with chance constraints. Recall that we use Algorithm 1 in Section IV.A to solve this model. Two cases are considered: Unimodality excluded (UE) and included (UI). The mode of the distribution of wind production forecast error is determined according to the underlying training dataset.

Fig. 4 comprises of four plots illustrating various results, all as a function of confidence level $1 - \epsilon$. The left plots

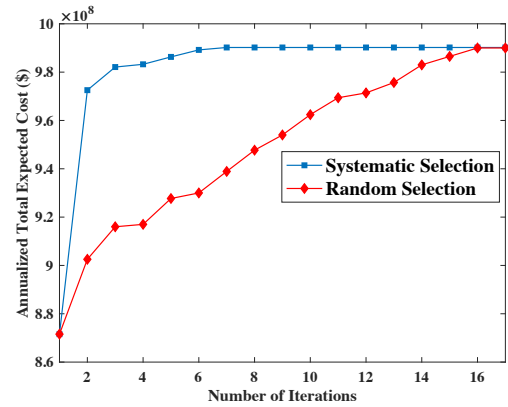


Fig. 5. Distributionally robust chance-constrained model: Evolution of the in-sample system cost with solution strategies based on adding constraints in a systematic way via Algorithm 1 presented in Section IV.A (blue curve) and in a randomized way (red curve).

pertain to training dataset 1, whereas the right plots correspond to training dataset 2, which is comparatively more similar to the test dataset. The upper plots give the optimal units to be built as well as the in-sample cost, i.e., the value obtained for objective function (7a). The lower plots illustrate the out-of-sample cost. The dash lines present the mean cost, while the shaded area around it shows the mean cost plus/minus its standard deviation.

According to the upper plots of Fig. 4 for both training datasets, one can observe that adding unimodality information (case UI) leads to a comparatively lower capacity to be built, as well as a lower in-sample cost. In addition, the type of units to be built in case UI is less flexible, and therefore, less expensive than that in case UE. This implies that adding unimodality information leads to less conservative expansion solutions.

Concerning the out-of-sample results, adding unimodality information is always appealing, provided that an accurate

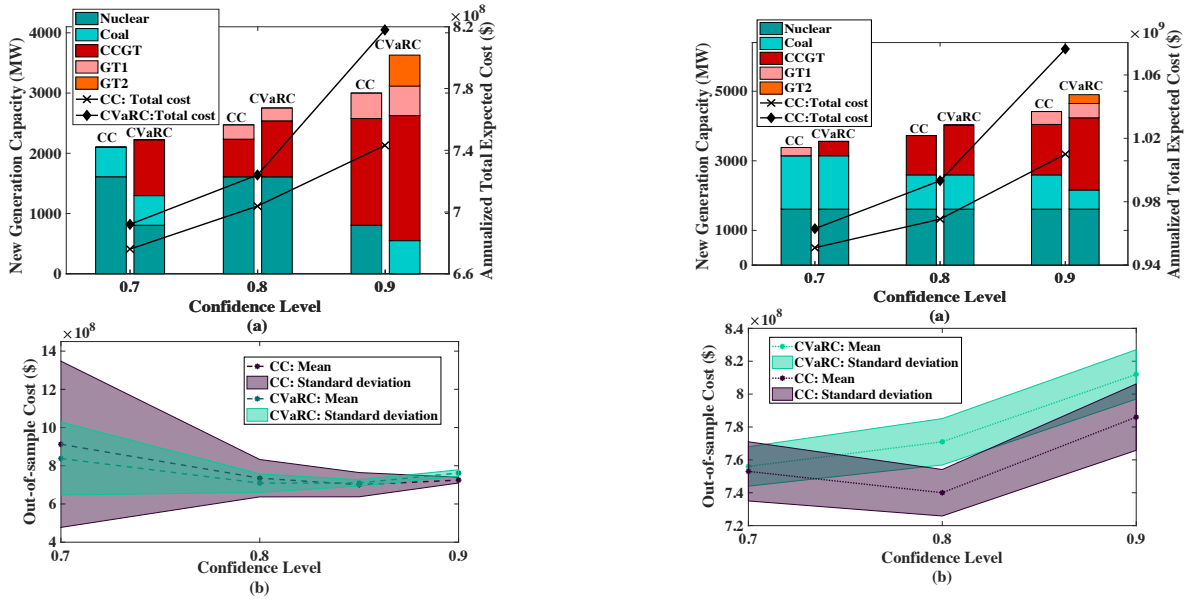


Fig. 6. Distributionally robust chance-constrained (CC) and CVaR-constrained (CVaRC) results obtained using training dataset 1 (left plots; less similar to the test dataset) and training dataset 2 (right plots; more similar to the test dataset): The upper plots show the new conventional generating units to be built and in-sample system cost. The lower plots present the out-of-sample system cost in terms of mean and standard deviation. Unimodality information are included.

training dataset is available – the cost in case UI with training dataset 2 is always lower than that in case UE (see the right lower plot of Fig. 4). However, this observation is less obvious when the training dataset is not accurate enough; see the lower left plot of Fig. 4 related to training dataset 1, where the out-of-sample performance for case UI becomes better than that of UE after the confidence level of 0.85.

Next, we numerically explore the performance of Algorithm 1 in selecting parameters η in a systematic way. To do so, we consider two strategies to select a finite number of values for parameters η . In the first strategy, parameters η take those values that are determined in Step 4 of Algorithm 1. However, in the second strategy, those values are randomly selected. We set the confidence level to be equal to 0.85 and use training dataset 2. The tolerance level in Algorithm 1 is equal to 10^{-7} . Fig. 5 illustrates the evolution of the in-sample system cost, i.e., the value of objective function (7a), for two strategies. For the random strategy, we report the best outcome that we have obtained over 300 efforts, each with different randomly selected values for η . Fig. 5 shows that Algorithm 1 converges to a fixed in-sample cost after 6 iterations, whereas the best outcome of the random selection strategy converges after 16 selections. This numerical observation concludes that the convergence speed of Algorithm 1 is much higher than a random selection strategy.

D. Expansion Results With CVaR Constraints and Unimodality Information: CVaR Constraints Versus Chance Constraints

We keep the unimodality information in the ambiguity sets, and compare the performance of chance and CVaR constraints, as illustrated in Fig. 6. This figure contains four plots with the same structure as in Fig. 4.

As observed in the upper plots of Fig. 6 for both training datasets, the CVaR-constrained model in comparison to the chance-constrained one suggests investing in more new gen-

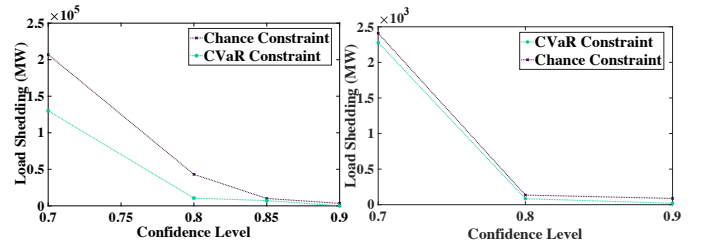


Fig. 7. The annual amount of load shedding obtained from distributionally robust chance-constrained and CVaR-constrained models using training dataset 1 (left plot) and training dataset 2 (right plot).

eration capacity and more flexible units, resulting in a higher in-sample system cost. As an interesting finding, we observe that although the CVaR-constrained model leads to more conservative expansion decisions with respect to the chance-constrained one, it shows a comparatively better out-of-sample performance in the case there is a higher discrepancy between the training and test datasets; see the lower left plot of Fig. 6. The reason for this is that the CVaR-constrained model limits not only the frequency but also the severity of load shedding. Conversely, when the level of similarity between the training and test datasets is high, the chance-constrained model exhibits a comparatively better out-of-sample performance; see the lower right plot of Fig. 6.

Next, we investigate the effect of information accuracy on the out-of-sample performance of the proposed models. The annual amount of load shedding (MW) in the out-of-sample simulation is considered as a measure of reliability. Fig. 7 shows this amount as a function of confidence level for training dataset 1 (left plot) and for training dataset 2 (right plot).

The first observation from Fig. 7 is that using training dataset 2 (more similar to the test dataset) to build the ambiguity sets leads to a much lower load shedding compared to another case wherein training dataset 1 is used. With

training dataset 2 (right plot), the curve for both chance- and CVaR-constrained models is similar, and the amount of annual load shedding varies from around 2,500 MW to near zero by changing the value of confidence level from 0.7 to 0.9. Note that the amount of annual load shedding with training dataset 1 (left plot) for the confidence level of 0.7 is around 2×10^5 MW or 1.3×10^5 MW, depending on which model is used.

The second observation from Fig. 7 is that the CVaR-constrained model does not necessarily outperform the chance-constrained model when the training dataset is accurate enough (right plot). On the contrary, the CVaR-constrained model outperforms the chance-constrained model when the central planner is not confident about the accuracy of the training dataset (left plot).

E. Relaxing Unimodality Requirement

Our next out-of-sample analysis explores the effects of relaxing unimodality requirement on the performance of chance- and CVaR-constrained models. Recall that two wind farms are considered, therefore, $\alpha = 2$ for cases with the unimodality information. In the following, we consider two cases wherein the unimodality degree α is equal to 2 and ∞ , respectively. Note that when parameter α goes towards infinity, i.e., $\alpha \rightarrow \infty$, the unimodality requirement is relaxed. In such a case, constraint (10b) boils down to

$$\sqrt{\frac{1-\epsilon}{\epsilon}} \|\beta_i \mathbf{1}^\top (\Sigma - \mu \mu^\top)^{\frac{1}{2}}\| \leq P_i^{\max} x_i - p_i - \beta_i \mathbf{1}^\top \mu, \forall i. \quad (17)$$

In addition, (13) transforms to

$$\left\| \begin{array}{l} \theta_i - \beta_i \mathbf{1}^\top \mu \\ (\Sigma - \mu \mu^\top)^{\frac{1}{2}} \beta_i \mathbf{1} \end{array} \right\| \leq 2\epsilon (P_i^{\max} x_i - p_i) - (2\epsilon - 1)\theta_i - \beta_i \mathbf{1}^\top \mu, \quad \forall i. \quad (18)$$

The out-of-sample system costs are given in Fig. 8 for the case of $\alpha = 2$ (left plot) and the case of $\alpha \rightarrow \infty$ (right plot). According to this figure, as α goes to infinity, the out-of-sample performance of both models becomes similar. In other words, the proposed distributionally robust chance- and CVaR-constrained models are asymptotically equivalent if the unimodality information is excluded from the ambiguity sets.

VI. CONCLUSION

This paper proposes a distributionally robust risk-aware generation expansion planning model to manage the violation risk of operational limits arising from the uncertainties pertaining to the wind power production. Since the empirical observations illustrate the unimodality properties of potential distributions of wind forecast error uncertainty, the inclusion of distributions with non-unimodal structures in the ambiguity set leads to unnecessarily conservative expansion decisions. Aiming to build more realistic moment-based ambiguity sets, this paper incorporates structural feature of unimodality, eliminating non-structured distributions from the ambiguity sets. Two risk measures, namely chance and CVaR constraints, are used to adjust the violation of operational limits with a specific confidence level. Out-of-sample analyses are used to highlight the importance of incorporating structural information into the model. We numerically find out when training

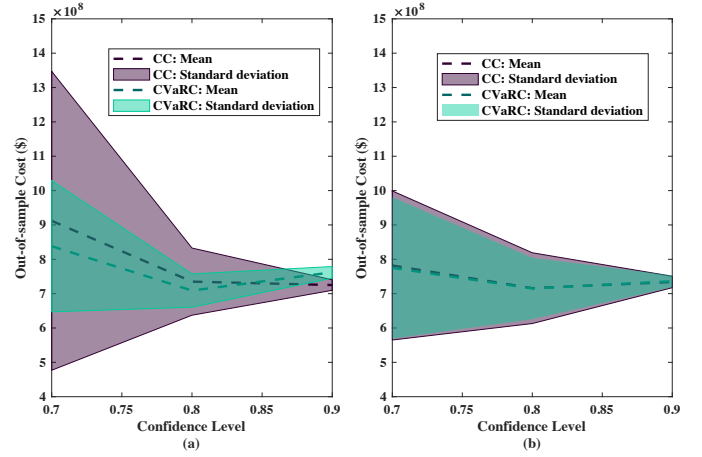


Fig. 8. The out-of-sample performance of distributionally robust chance-constrained (CC) and CVaR-constrained (CVaRC) models for training dataset 1 when (a): $\alpha = 2$ and (b): $\alpha \rightarrow \infty$.

datasets achieved from historical observations cannot give a clear insight into the distributional information of uncertainty, the CVaR-constrained model provides a comparatively better performance in terms of out-of-sample cost. Conversely, when an accurate estimation of distributions is available, the chance-constrained model outperforms the CVaR-constrained one. These two models, however, give a similar performance when unimodality features are excluded.

As potential future works, it is of interest to explore the effects of having inexact information about moments and unimodality features. From the application perspective, it is of interest to extend the current model and conduct a capacity expansion planning study with coordinated investment in generation, storage and transmission assets.

APPENDIX A: NOMENCLATURE

A. Indices and Sets

- $d \in \{1, \dots, \mathcal{D}\}$ Index for representative days.
- $i \in \{1, \dots, \mathcal{G}\}$ Index for conventional generating units.
- $j \in \{1, \dots, \mathcal{J}\}$ Index for transmission lines.
- $n \in \{1, \dots, \mathcal{N}\}$ Index for long-term scenarios.
- $t \in \{1, \dots, 24\}$ Index for operating hours.
- I^G Set of conventional generating units.
- $I^{G^C} \subset I^G$ Set of candidate conventional generating units.

B. Uncertainty modeling

- $\mu \in \mathbb{R}^Z$ Mean vector of wind forecast error ξ .
- $\nu \in \mathbb{R}^Z$ Mode vector of wind forecast error ξ .
- $\Sigma \in \mathbb{R}^{Z \times Z}$ Covariance matrix of wind forecast error ξ .
- Ω^α Set of potential α -unimodal distributions.
- $\mathbb{P} \in \Pi$ Worst-case distribution of the short-term uncertainty selected from the ambiguity set Π .
- L Number of loads.
- Z Number of random variables (wind farms).

C. Parameters

- $\iota_n \in \mathbb{R}_+$ Probability of long-term scenario n .
- $\mathbf{C} \in \mathbb{R}_+^G$ Vector of production cost of conventional generating units [\$/MWh].
- $\mathbf{K} \in \mathbb{R}_+^{G^C}$ Vector of annualized capital cost of candidate conventional generating units [\$/].

$\mathbf{l}_{ndt} \in \mathbb{R}_+^L$	Vector of load level under long-term scenario n in representative day d at hour t [MW].
$\mathbf{M}^D \in \mathbb{R}^{\mathcal{J} \times L}$	Matrix of power transfer distribution factor for loads.
$\mathbf{M}^G \in \mathbb{R}^{\mathcal{J} \times G}$	Matrix of power transfer distribution factor for conventional generating units.
$\mathbf{M}^W \in \mathbb{R}^{\mathcal{J} \times Z}$	Matrix of power transfer distribution factor for wind farms.
$\mathbf{S} \in \mathbb{R}_+^G$	Vector of start-up cost of conventional generating units [\$].
$\mathbf{w}_{ndt} \in \mathbb{R}_+^Z$	Mean vector of power production of wind farms under long-term scenario n in representative day d at hour t [MW].
$\omega_d \in \mathbf{N}$	Number of days in cluster of representative day d [day].
$F_j^{\max} \in \mathbb{R}_+$	Capacity of transmission line j [MW].
$P_i^{\max} \in \mathbb{R}_+$	Capacity of conventional generating unit i [MW].
$P_i^{\min} \in \mathbb{R}_+$	Lower bound for production level of conventional generating unit i [MW].
$R_i^{\text{st}} \in \mathbb{R}_+$	Start-up and shut-down ramp rate capability of conventional generating unit i [MW].
$U_i^{\text{dn}} \in \mathbb{R}_+$	Minimum down time of conventional generating unit i [hour].
$U_i^{\text{up}} \in \mathbb{R}_+$	Minimum up time of conventional generating unit i [hour].
E. Variables	
$\beta_{ndt} \in \mathbb{R}^G$	Vector of participation factor of conventional generating units under long-term scenario n in representative day d at hour t [per-unit].
η, θ	Auxiliary variables.
$\mathbf{p}_{ndt} \in \mathbb{R}_+^G$	Vector of nominal (tentative) dispatch of conventional generating units under long-term scenario n in representative day d at hour t [MW].
$\mathbf{v}_{ndt} \in \{0, 1\}^G$	Vector of start-up status of conventional generating units under long-term scenario n in representative day d at hour t .
$\mathbf{x}_{ndt} \in \{0, 1\}^G$	Vector of on/off commitment status of conventional generating units under long-term scenario n in representative day d at hour t .
$\mathbf{z} \in \{0, 1\}^G$	Vector of binary variables indicating whether the candidate conventional generating units are selected to be built.

APPENDIX B: RELAXATION OF UNIT COMMITMENT INTEGRALITY CONSTRAINTS

In order to tightly relax the operational-stage binary variables x_{indt} and v_{indt} to lie between zero and one, we use the convex relaxation approach proposed in [29]. By relaxing the binary variables and adding extra constraints, the feasible set of the unit commitment problem is substituted by an approximation of its convex hull. Additional constraints for tightening the relaxation are considered as probabilistic constraints. As the first alternative, they can be written in the form of distributionally robust chance constraints as

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{ind(t-1)}(\boldsymbol{\xi}) \leq R_i^{\text{st}} x_{ind(t-1)}\}$$

$$+ (P_i^{\max} - R_i^{\text{st}})(x_{indt} - v_{indt}) \geq 1 - \epsilon, \forall i, n, d, t \quad (19a)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{indt}(\boldsymbol{\xi}) \leq P_i^{\max} x_{isrt} - (P_i^{\max} - R_i^{\text{st}})v_{indt}\} \geq 1 - \epsilon, \quad \forall i, n, d, t \quad (19b)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{indt}(\boldsymbol{\xi}) - p_{ind(t-1)}(\boldsymbol{\xi}) \leq (P_i^{\min} + R_i^{\text{up}})x_{indt} - P_i^{\min} x_{ind(t-1)} - (P_i^{\min} + R_i^{\text{up}} - R_i^{\text{st}})v_{indt}\} \geq 1 - \epsilon, \quad \forall i, n, d, t \quad (19c)$$

$$\min_{\mathbb{P} \in \Pi} \mathbb{P}\{p_{isr(t-1)}(\boldsymbol{\xi}) - p_{indt}(\boldsymbol{\xi}) \leq R_i^{\text{st}} x_{ind(t-1)} - (P_i^{\min} + R_i^{\text{dn}} - R_i^{\text{st}})v_{indt} - (R_i^{\text{st}} - R_i^{\text{dn}})x_{indt}\} \geq 1 - \epsilon, \quad \forall i, n, d, t. \quad (19d)$$

As the second alternative, they can be added in the form of distributionally robust CVaR constraints as

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{ind(t-1)}(\boldsymbol{\xi})] \leq R_i^{\text{st}} x_{ind(t-1)} + (P_i^{\max} - R_i^{\text{st}})(x_{indt} - v_{indt}), \quad \forall i, n, d, t \quad (20a)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{indt}(\boldsymbol{\xi})] \leq P_i^{\max} x_{isrt} - (P_i^{\max} - R_i^{\text{st}})v_{indt}, \quad \forall i, n, d, t \quad (20b)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{indt}(\boldsymbol{\xi}) - p_{ind(t-1)}(\boldsymbol{\xi})] \leq (P_i^{\min} + R_i^{\text{up}})x_{indt} - P_i^{\min} x_{ind(t-1)} - (P_i^{\min} + R_i^{\text{up}} - R_i^{\text{st}})v_{indt} \geq 1 - \epsilon_i, \quad \forall i, n, d, t \quad (20c)$$

$$\max_{\mathbb{P} \in \Pi} \text{CVaR}_{\mathbb{P}}^{\epsilon}[p_{isr(t-1)}(\boldsymbol{\xi}) - p_{indt}(\boldsymbol{\xi})] \leq R_i^{\text{st}} x_{ind(t-1)} - (P_i^{\min} + R_i^{\text{dn}} - R_i^{\text{st}})v_{indt} - (R_i^{\text{st}} - R_i^{\text{dn}})x_{indt}, \quad \forall i, n, d, t. \quad (20d)$$

APPENDIX C: FINAL MODEL

The final tractable model for distributionally robust generation expansion planning problem (7) is a mixed-integer second order cone program as given below:

Objective function: (16)

Subject to:

$$(7b) - (7f), (15)$$

$$0 \leq x_{indt} \leq 1; \quad 0 \leq v_{indt} \leq 1, \quad \forall i, n, d, t$$

Reformulated probabilistic constraints.

As mentioned earlier, probabilistic constraints can be enforced in the form of either distributionally robust chance constraints, i.e., (8) and (19), or in the form of distributionally robust CVaR constraints, i.e., (9) and (20). In the following, we present their resulting reformulations. Owing to the ramping constraints of conventional generating units that enforce intertemporal coupling between hours t and $t-1$, we define a new uncertainty parameter vector $\hat{\boldsymbol{\xi}}_{ndt} \in \mathbb{R}^{2Z}$ for these constraints as

$$\hat{\boldsymbol{\xi}}_{ndt} = \begin{bmatrix} \boldsymbol{\xi}_{ndt} \\ \boldsymbol{\xi}_{nd(t-1)} \end{bmatrix}. \quad (21)$$

Parameters $\hat{\boldsymbol{\mu}}_{ndt} \in \mathbb{R}^{2Z}$ and $\hat{\boldsymbol{\Sigma}}_{ndt} \in \mathbb{R}^{2Z \times 2Z}$ represent the mean and covariance of $\hat{\boldsymbol{\xi}}_{ndt}$, respectively. In addition, $\hat{\mathbf{v}}_{ndt} \in \mathbb{R}^{2Z}$ refers to the mode location. Similar to the definition of Φ_{ndt} , matrix $\hat{\Phi}_{ndt}$ is defined as

$$\hat{\Phi}_{ndt} = \frac{\alpha + 2}{\alpha} \left(\hat{\boldsymbol{\Sigma}}_{ndt} - \hat{\boldsymbol{\mu}}_{ndt} \hat{\boldsymbol{\mu}}_{ndt}^{\top} \right)$$

$$- \frac{1}{\alpha^2} (\hat{\boldsymbol{\mu}}_{ndt} - \hat{\boldsymbol{\nu}}_{ndt}) (\hat{\boldsymbol{\mu}}_{ndt} - \hat{\boldsymbol{\nu}}_{ndt})^\top. \quad (22)$$

For notational clarity, we drop indices n and d from $\{p_{indt}, x_{indt}, \beta_{indt}, \mathbf{p}_{ndt}, \mathbf{x}_{ndt}, \boldsymbol{\beta}_{ndt}, \mathbf{l}_{ndt}, \mathbf{w}_{ndt}\}$ and from all auxiliary variables. We also drop indices n , d and t from $\{\boldsymbol{\mu}_{ndt}, \boldsymbol{\nu}_{ndt}, \boldsymbol{\Sigma}_{ndt}, \boldsymbol{\Phi}_{ndt}, \hat{\boldsymbol{\mu}}_{ndt}, \hat{\boldsymbol{\nu}}_{ndt}, \hat{\boldsymbol{\Sigma}}_{ndt}, \hat{\boldsymbol{\Phi}}_{ndt}\}$.

A. Reformulated Distributionally Robust Chance Constrains

According to the reformulation algorithm explained in Section IV-A, the resulting constraints from the reformulation of (8) and (19) are written as

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \eta_{its}^{-\alpha}}{\epsilon}} \|\beta_{it} \mathbf{1}^\top \boldsymbol{\Phi}\| &\leq \eta_{its} [P_i^{\max} x_{it} - p_{it} - \beta_{it} \mathbf{1}^\top \boldsymbol{\nu}] \\ &- \left(\frac{\alpha+1}{\alpha}\right) (\beta_{it} \mathbf{1}^\top) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall i, t, s \end{aligned} \quad (23a)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \gamma_{its}^{-\alpha}}{\epsilon}} \|\beta_{it} \mathbf{1}^\top \boldsymbol{\Phi}\| &\leq \gamma_{its} [p_{it} - P_i^{\min} x_{it} + \beta_{it} \mathbf{1}^\top \boldsymbol{\nu}] \\ &+ \left(\frac{\alpha+1}{\alpha}\right) (\beta_{it} \mathbf{1}^\top) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall i, t, s \end{aligned} \quad (23b)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \delta_{its}^{-\alpha}}{\epsilon}} \|\hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\Phi}}\| &\leq \delta_{its} [p_{i(t-1)} - p_{it} + R_i^{\text{up}} x_{i(t-1)}] \\ &+ R_i^{\text{st}} (1 - x_{i(t-1)}) - \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\nu}} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top (\hat{\boldsymbol{\mu}} - \boldsymbol{\nu}), \quad \forall i, t, s \end{aligned} \quad (23c)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \zeta_{its}^{-\alpha}}{\epsilon}} \|\hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\Phi}}\| &\leq \zeta_{its} [p_{it} - p_{i(t-1)} + R_i^{\text{dn}} x_{it}] \\ &+ R_i^{\text{st}} (1 - x_{it}) + \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\nu}} + \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top (\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\nu}}), \quad \forall i, t, s \end{aligned} \quad (23d)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \varphi_{jts}^{-\alpha}}{\epsilon}} \|\mathbf{M}_j^{\text{G}} \boldsymbol{\beta}_t \mathbf{1}^\top + \mathbf{M}_j^{\text{W}} \boldsymbol{\Phi}\| &\leq \\ \varphi_{jts} [F_j^{\max} + \mathbf{M}_j^{\text{D}} \mathbf{1}_t - \mathbf{M}_j^{\text{G}} \mathbf{p}_t - \mathbf{M}_j^{\text{W}} \mathbf{w}_t - (\mathbf{M}_j^{\text{G}} \boldsymbol{\beta}_t \mathbf{1}^\top + \mathbf{M}_j^{\text{W}}) \mathbf{1}^\top \boldsymbol{\nu}] \\ &- \left(\frac{\alpha+1}{\alpha}\right) (\mathbf{M}_j^{\text{G}} \boldsymbol{\beta}_t \mathbf{1}^\top + \mathbf{M}_j^{\text{W}}) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall j, t, s \end{aligned} \quad (23e)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \sigma_{jts}^{-\alpha}}{\epsilon}} \|\mathbf{M}_j^{\text{G}} \boldsymbol{\beta}_t \mathbf{1}^\top + \mathbf{M}_j^{\text{W}} \boldsymbol{\Phi}\| &\leq \\ \sigma_{jts} [F_j^{\max} - \mathbf{M}_j^{\text{D}} \mathbf{1}_t + \mathbf{M}_j^{\text{G}} \mathbf{p}_t + \mathbf{M}_j^{\text{W}} \mathbf{w}_t + (\mathbf{M}_j^{\text{G}} \boldsymbol{\beta}_t \mathbf{1}^\top + \mathbf{M}_j^{\text{W}}) \boldsymbol{\nu}] \\ &+ \left(\frac{\alpha+1}{\alpha}\right) (\mathbf{M}_j^{\text{G}} \boldsymbol{\beta}_t \mathbf{1}^\top + \mathbf{M}_j^{\text{W}}) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall j, t, s \end{aligned} \quad (23f)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \lambda_{its}^{-\alpha}}{\epsilon}} \|\beta_{i(t-1)} \mathbf{1}^\top \boldsymbol{\Phi}\| &\leq \lambda_{its} [R_i^{\text{st}} x_{i(t-1)}] \\ &+ (P_i^{\max} - R_i^{\text{st}}) (x_{it} - v_{it}) - p_{i(t-1)} - \beta_{i(t-1)} \mathbf{1}^\top \boldsymbol{\nu}] \\ &- \left(\frac{\alpha+1}{\alpha}\right) (\beta_{i(t-1)} \mathbf{1}^\top) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall i, t, s \end{aligned} \quad (23g)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \pi_{its}^{-\alpha}}{\epsilon}} \|\beta_{it} \mathbf{1}^\top \boldsymbol{\Phi}\| &\leq \pi_{its} [(P_i^{\max} - R_i^{\text{st}}) v_{it}] \\ &- p_{it} - \beta_{it} \mathbf{1}^\top \boldsymbol{\nu}] - \left(\frac{\alpha+1}{\alpha}\right) (\beta_{it} \mathbf{1}^\top) (\boldsymbol{\mu} - \boldsymbol{\nu}), \quad \forall i, t, s \end{aligned} \quad (23h)$$

$$\begin{aligned} \sqrt{\frac{1 - \epsilon - \rho_{its}^{-\alpha}}{\epsilon}} \|\hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\Phi}}\| &\leq \rho_{its} [p_{i(t-1)} - p_{it}] \\ &+ (P_i^{\min} + R_i^{\text{up}}) x_{it} - (P_i^{\min} + R_i^{\text{up}} - R_i^{\text{st}}) v_{it} \\ &- P_i^{\min} x_{i(t-1)} - \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\nu}} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top (\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\nu}}), \quad \forall i, t, s \end{aligned} \quad (23i)$$

$$\sqrt{\frac{1 - \epsilon - \psi_{its}^{-\alpha}}{\epsilon}} \|\hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\Phi}}\| \leq \psi_{its} [p_{it} - p_{i(t-1)} + R_i^{\text{st}} x_{i(t-1)}]$$

$$\begin{aligned} &- (P_i^{\min} + R_i^{\text{dn}} - R_i^{\text{st}}) v_{it} - (R_i^{\text{st}} - R_i^{\text{dn}}) x_{it} \\ &+ R_i^{\text{st}} (1 - x_{it}) + \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\nu}} + \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top (\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\nu}}), \quad \forall i, t, s, \end{aligned} \quad (23j)$$

where index s indicates iteration s and auxiliary variables $\{\eta_{its}, \gamma_{its}, \delta_{its}, \zeta_{its}, \varphi_{jts}, \sigma_{jts}, \lambda_{its}, \pi_{its}, \rho_{its}, \psi_{its}\}$ are used to reformulate constraints (8a)-(8f), (19a)-(20d), respectively. In (23c)-(23d) and (23i)-(23j), variable vector $\hat{\boldsymbol{\beta}}_{it}$ is defined as $\hat{\boldsymbol{\beta}}_{it} = [\beta_{it} \mathbf{1}^\top \quad -\beta_{i(t-1)} \mathbf{1}^\top]^\top$.

B. Reformulated Distributionally Robust CVaR Constraints

In the case that probabilistic constraints are considered in the form of distributionally robust CVaR constraints, i.e. (9) and (20), they are approximately reformulated as

$$\begin{aligned} \left\| \theta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (P_i^{\max} x_{it} - p_{it}) \\ &- \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1 \right] \theta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i, t \end{aligned} \quad (24a)$$

$$\begin{aligned} \left\| \theta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (P_i^{\max} x_{it} - p_{it}) \\ &- \left[\frac{(2\epsilon-1)(\alpha+1)-1}{\alpha} \right] \theta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i, t \end{aligned} \quad (24b)$$

$$\begin{aligned} \left\| \left(\frac{\alpha+1}{\alpha}\right) \theta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (P_i^{\max} x_{it} - p_{it}) \\ &- \left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha} \right] \theta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i, t \end{aligned} \quad (24c)$$

$$\begin{aligned} \left\| \gamma_{it} + \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (p_{it} - P_i^{\min} x_{it}) \\ &- \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1 \right] \gamma_{it} + \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i, t \end{aligned} \quad (25a)$$

$$\begin{aligned} \left\| \gamma_{it} + \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (p_{it} - P_i^{\min} x_{it}) \\ &- \left[\frac{(2\epsilon-1)(\alpha+1)-1}{\alpha} \right] \gamma_{it} + \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i, t \end{aligned} \quad (25b)$$

$$\begin{aligned} \left\| \left(\frac{\alpha+1}{\alpha}\right) \gamma_{it} + \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (p_{it} - P_i^{\min} x_{it}) \\ &- \left[\frac{(2\epsilon-1)(\alpha+1)}{\alpha} \right] \gamma_{it} + \left(\frac{\alpha+1}{\alpha}\right) \beta_{it} \mathbf{1}^\top \boldsymbol{\mu}, \quad \forall i, t \end{aligned} \quad (25c)$$

$$\begin{aligned} \left\| \delta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\mu}} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (p_{i(t-1)} - p_{it}) \\ &+ R_i^{\text{up}} x_{i(t-1)} + R_i^{\text{st}} (1 - x_{i(t-1)}) \\ &- \left[\frac{2\epsilon(\alpha+1)}{\alpha} - 1 \right] \delta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\mu}}, \quad \forall i, t \end{aligned} \quad (26a)$$

$$\begin{aligned} \left\| \delta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\mu}} \right\| &\leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (p_{i(t-1)} - p_{it}) \\ &+ R_i^{\text{up}} x_{i(t-1)} + R_i^{\text{st}} (1 - x_{i(t-1)}) \\ &- \left[\frac{(2\epsilon-1)(\alpha+1)-1}{\alpha} \right] \delta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\mu}}, \quad \forall i, t \end{aligned} \quad (26b)$$

$$\left\| \left(\frac{\alpha+1}{\alpha}\right) \delta_{it} - \left(\frac{\alpha+1}{\alpha}\right) \hat{\boldsymbol{\beta}}_{it}^\top \hat{\boldsymbol{\mu}} \right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] (p_{i(t-1)} - p_{it})$$

$$- \left[\frac{(2\epsilon - 1)(\alpha + 1)}{\alpha} \right] \rho_{it} - \left(\frac{\alpha + 1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu}, \quad \forall i, t \quad (32c)$$

$$\begin{aligned} & \left\| \psi_{it} + \left(\frac{\alpha+1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu} - \hat{\Phi} \hat{\beta}_{it} \right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] [p_{it} + p_{i(t-1)} + R_i^{\text{st}} x_{i(t-1)}] \\ & - (P_i^{\text{min}} + R_i^{\text{dn}} - R_i^{\text{st}}) v_{it} - (R_i^{\text{st}} - R_i^{\text{dn}}) x_{it} + R_i^{\text{st}} (1 - x_{it}) \\ & - \left[\frac{2\epsilon(\alpha + 1)}{\alpha} - 1 \right] \psi_{it} + \left(\frac{\alpha + 1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu}, \quad \forall i, t \quad (33a) \end{aligned}$$

$$\begin{aligned} & \left\| \psi_{it} + \left(\frac{\alpha+1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu} - \hat{\Phi} \hat{\beta}_{it} \right\| \leq \left[\frac{2\epsilon(\alpha + 1)}{\alpha} \right] [p_{it} + p_{i(t-1)} + R_i^{\text{st}} x_{i(t-1)}] \\ & - (P_i^{\text{min}} + R_i^{\text{dn}} - R_i^{\text{st}}) v_{it} - (R_i^{\text{st}} - R_i^{\text{dn}}) x_{it} + R_i^{\text{st}} (1 - x_{it}) \\ & - \left[\frac{(2\epsilon - 1)(\alpha + 1) - 1}{\alpha} \right] \psi_{it} + \left(\frac{\alpha + 1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu}, \quad \forall i, t \quad (33b) \end{aligned}$$

$$\begin{aligned} & \left\| \left(\frac{\alpha+1}{\alpha} \right) \psi_{it} + \left(\frac{\alpha+1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu} - \hat{\Phi} \hat{\beta}_{it} \right\| \leq \left[\frac{2\epsilon(\alpha+1)}{\alpha} \right] [p_{it} + p_{i(t-1)}] \\ & + R_i^{\text{st}} x_{i(t-1)} - (P_i^{\text{min}} + R_i^{\text{dn}} - R_i^{\text{st}}) v_{it} - (R_i^{\text{st}} - R_i^{\text{dn}}) x_{it} \\ & + R_i^{\text{st}} (1 - x_{it}) - \left[\frac{(2\epsilon - 1)(\alpha + 1)}{\alpha} \right] \psi_{it} + \left(\frac{\alpha + 1}{\alpha} \right) \hat{\beta}_{it}^\top \hat{\mu}, \quad \forall i, t, \quad (33c) \end{aligned}$$

where auxiliary variables $\{\theta_{it}, \gamma_{it}, \delta_{it}, \zeta_{it}, \varphi_{jt}, \sigma_{jt}, \lambda_{it}, \pi_{it}, \rho_{it}, \psi_{it}\}$ are used to reformulate constraints (9a)-(9f) and (20a)-(20d), resulting in constraints (24)-(33), respectively.

REFERENCES

- [1] S. Pineda, J. M. Morales, and T. K. Boomsma, "Impact of forecast errors on expansion planning of power systems with a renewables target," *Eur. J. Oper. Res.*, vol. 248, no. 3, pp. 1113–1122, Feb. 2016.
- [2] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Oper. Res.*, vol. 58, no. 3, pp. 595–612, 2010.
- [3] A. Shapiro, "Distributionally robust stochastic programming," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2258–2275, 2017.
- [4] H. Rahimian and S. Mehrotra, "Distributionally robust optimization: A review," 2019. [Online]. Available: <https://arxiv.org/abs/1908.05659>
- [5] W. Wiesemann, D. Kuhn, and M. Sim, "Distributionally robust convex optimization," *Oper. Res.*, vol. 62, no. 6, pp. 1358–1376, 2014.
- [6] Z. Hu and L. J. Hong, "Kullback-leibler divergence constrained distributionally robust optimization," *Available at Optimization Online*, 2013.
- [7] D. Kuhn, P. Mohajerin Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh, "Wasserstein distributionally robust optimization: Theory and applications in machine learning," *INFORMS Tutorials in Operations Research*, pp. 130–166, 2019.
- [8] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Prog.*, vol. 171, pp. 115–166, 2018.
- [9] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Energy and reserve dispatch with distributionally robust joint chance constraints," *Available at Optimization Online*, 2018.
- [10] F. Pourahmadi, J. Kazempour, C. Ordoudis, P. Pinson, and S. H. Hosseini, "Distributionally robust chance-constrained generation expansion planning," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2888–2903, Jul. 2020.
- [11] A. Zare, C. Y. Chung, J. Zhan, and S. O. Faried, "A distributionally robust chance-constrained MILP model for multistage distribution system planning with uncertain renewables and loads," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5248–5262, Sep. 2018.
- [12] D. Alvarado, A. Moreira, R. Moreno, and G. Strbac, "Transmission network investment with distributed energy resources and distributionally robust security," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 5157–5168, Nov. 2019.
- [13] A. Velloso, D. Pozo, and A. Street, "Distributionally robust transmission expansion planning: A multi-scale uncertainty approach," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3353–3365, Sep. 2020.

- [14] A. Hajebrahimi, I. Kamwa, E. Delage, and M. Abdelaziz, "Adaptive distributionally robust optimization for electricity and electrified transportation planning," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4278–4289, Sep. 2020.
- [15] B. Li, R. Jiang, and J. L. Mathieu, "The value of including unimodality information in distributionally robust optimal power flow," 2018. [Online]. Available: <https://arxiv.org/abs/1811.10217>
- [16] B. Li, R. Jiang, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow assuming unimodal distributions with misspecified modes," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 3, pp. 1223–1234, Sep. 2019.
- [17] A. Schwele, J. Kazempour, and P. Pinson, "Do unit commitment constraints affect generation expansion planning? A scalable stochastic model," *Energy Systems*, vol. 11, no. 2, pp. 247–282, May 2020.
- [18] A. Moreira, G. Strbac, and B. Fanzere, "An ambiguity averse approach for transmission expansion planning," *2019 IEEE Milan PowerTech Conf.*, pp. 1–6, Jun. 2019, Milan, Italy.
- [19] A. J. Conejo, L. Baringo, S. J. Kazempour, and A. S. Siddiqui, *Investment in Electricity Generation and Transmission: Decision Making Under Uncertainty*. Springer, Switzerland, 2016.
- [20] S. Dharmadhikari and K. Joag-Dev, *Unimodality, Convexity, and Applications*. Elsevier, 1988.
- [21] B. P. Van Parys, P. J. Goulart, and D. Kuhn, "Generalized Gauss inequalities via semidefinite programming," *Math. Prog.*, vol. 156, no. 1–2, pp. 271–302, 2016.
- [22] B. P. Van Parys, P. J. Goulart, and M. Morari, "Distributionally robust expectation inequalities for structured distributions," *Math. Prog.*, vol. 173, no. 1–2, pp. 251–280, 2019.
- [23] B. Li, R. Jiang, and J. L. Mathieu, "Ambiguous risk constraints with moment and unimodality information," *Math. Prog.*, vol. 173, no. 1–2, pp. 151–192, 2019.
- [24] R. A. Jabr, "Distributionally robust CVaR constraints for power flow optimization," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3764–3773, Sep. 2020.
- [25] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *J. Bank. Finance*, vol. 26, no. 7, pp. 1443–1471, Jul. 2002.
- [26] B. Li, R. Jiang, and J. L. Mathieu, "Distributionally robust risk-constrained optimal power flow using moment and unimodality information," in *Proc. 55th IEEE Conf. Decision and Control*, Dec. 2016, Las Vegas, NV, USA.
- [27] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Prog.*, vol. 137, no. 1–2, pp. 167–198, 2013.
- [28] K. W. Ding, M. H. Wang, and N. J. Huang, "Distributionally robust chance constrained problem under interval distribution information," *Optimization Letters*, vol. 12, no. 6, pp. 1315–1328, 2018.
- [29] B. Hua and R. Baldick, "A convex primal formulation for convex hull pricing," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3814–3823, Sep. 2017.
- [30] D. Kuhn, W. Wiesemann, and A. Georghiou, "Primal and dual linear decision rules in stochastic and robust optimization," *Math. Prog.*, vol. 130, no. 1, pp. 177–209, Nov. 2011.
- [31] I. Pena, C. B. Martinez-Anido, and B.-M. Hodge, "An extended IEEE 118-bus test system with high renewable penetration," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 281–289, Jan. 2018.
- [32] [Online]. Available: <https://github.com/farzanehpourahmadi/DRCC.git>



Farzaneh Pourahmadi (M'20) received the M.S. and Ph.D. degrees in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2014 and 2019, respectively. She is currently a Postdoctoral Fellow at the Department of Mathematical Sciences, University of Copenhagen, Denmark. From June 2018 to June 2019, she was a Visiting Researcher with the Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark. Her research interests include power system planning and operation, as well as optimization and its application to power systems.



Jalal Kazempour (SM'18) received the Ph.D. degree in electrical engineering from the University of Castilla-La Mancha, Ciudad Real, Spain, in 2013. He is currently an Associate Professor with the Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark. His focus area is the intersection of multiple fields, including power and energy systems, electricity markets, optimization, game theory, and machine learning.