Distributionally Robust Chance Constrained Optimal Power Flow Assuming Unimodal Distributions with Misspecified Modes

Bowen Li, *Student Member, IEEE,* Ruiwei Jiang, *Member, IEEE,* and Johanna L. Mathieu, *Senior Member, IEEE*

***Abstract*—Chance constrained optimal power flow (CC-OPF) formulations have been proposed to minimize operational costs while controlling the risk arising from uncertainties like renewable generation and load consumption. To solve CC-OPF, we often need access to the (true) joint probability distribution of all uncertainties, which is rarely known in practice. A solution based on a biased estimate of the distribution can result in poor reliability. To overcome this challenge, recent work has explored distributionally robust chance constraints, in which the chance constraints are satisfied over a family of distributions called the ambiguity set. Commonly, ambiguity sets are only based on moment information (e.g., mean and covariance) of the random variables; however, specifying additional characteristics of the random variables reduces conservatism and cost. Here, we consider ambiguity sets that additionally incorporate unimodality information. In practice, it is difficult to estimate the mode location from the data and so we allow it to be potentially misspecified. We formulate the problem and derive a separation-based algorithm to efficiently solve it. Finally, we evaluate the performance of the proposed approach on a modified IEEE-30 bus network with wind uncertainty and compare with other distributionally robust approaches. We find that a misspecified mode significantly affects the reliability of the solution and the proposed model demonstrates a good trade-off between cost and reliability.**

***Index Terms*—Optimal power flow, chance constraint, distributionally robust optimization, misspecified mode, α-unimodality**

I. Introduction

With higher penetrations of renewable generation, uncertainties have increasing influence on power system operation and hence need to be carefully considered in scheduling problems, such as optimal power flow (OPF). To manage the risk arising from uncertainties, different stochastic OPF approaches have been studied. Among these formulations, CC-OPF has been proposed to directly control the constraint violation probability below a pre-defined threshold [1]–[7]. Traditional methods to solve chance constrained programs require knowledge of the joint probability distribution of all uncertainties, which may be unavailable or inaccurate. However, biased estimate may yield poor out-of-sample performance. Randomized techniques such as scenario approximation [8], [9], which provides a priori guarantees on reliability, require the constraints to be satisfied over a large number of uncertainty samples. The solutions from these approaches are usually overly conservative with high costs [7], [10]. Another popular approach is to assume that the uncertainties follow a parametric distribution such as Gaussian [4], [5], [7]. The resulting CC-OPF is often easier to solve but the solution may have low reliability unless the assumed probability distribution happens to be close to the true one.

To achieve better CC-OPF solutions with low operational costs and high reliabilty against uncertainties, researchers have developed distributionally robust chance constrained (DRCC) OPF models [10]–[17]. These models consider a family of distributions, called the ambiguity set, that share certain statistical characteristics. DRCC OPF models require that the chance constraints hold with respect to all distributions within the ambiguity set [18]–[21]. Unlike traditional methods, DRCC OPF solutions are less dependent on the sample outliers and more reliable than assuming a given distribution. Most existing work characterizes the ambiguity set based on moment information obtained from historical data of the uncertainty (see, e.g., [10], [11], [13], [14]). For example, a commonly adopted ambiguity set consists of all distributions whose mean and covariance agree with their corresponding sample estimates [10], [11], [13]. Many uncertainty distributions (e.g., those associated with wind forecast error) are unimodal and so, recently, unimodality has been incorporated to strengthen the ambiguity set and reduce the conservatism of DRCC models [13], [16], [17]. However, as compared to the moments, the mode location is more likely to be misspecified in sample-based estimation.

In this paper, we study a DRCC model with an ambiguity set based on moment and unimodality information with a potentially misspecified mode location. To the best of our knowledge, this paper is the first work discussing misspecification of a value related to a structural property, though others have considered misspecification of moments [10], [14], [18], [19], [21] and misspecification of distributions [12], [22]. Our main theoretical result shows that the distributionally robust chance constraints can be recast as a set of second-order conic (SOC) constraints. Furthermore, we derive an iterative algorithm to accelerate solving the reformulation. In each iteration, we begin by solving an updated relaxed formulation. Then, we efficiently find the most violated SOC constraint, if any, or terminate with a globally optimal solution. To identify the most violated SOC constraint, we need to find the global

This work is supported by the U.S. National Science Foundation Awards CCF-1442495 and CMMI-1662774. B. Li and J. L. Mathieu are with the Department of Electrical Engineering and Computer Science, University of Michigan at Ann Arbor, Ann Arbor, MI 48109 USA (e-mail: libowen@umich.edu; jlmath@umich.edu). R. Jiang is with the Department of Industrial and Operations Engineering, University of Michigan at Ann Arbor, Ann Arbor, MI 48109 USA (e-mail: ruiwei@umich.edu).

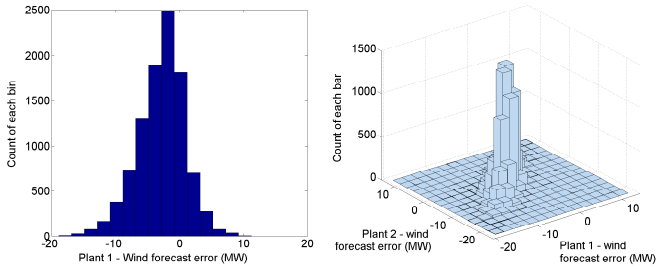

Fig. 1. Histograms of univariate and bivariate wind forecast errors (15 bins).

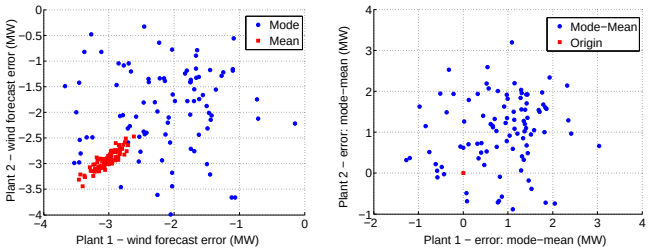

Fig. 2. Scatter plots of mode and mean estimates from data samples (left) and mode vs. mean differences (right).

maximum of a nonlinear optimization problem that is neither convex nor concave. To tackle this, we divide the nonlinear problem into subproblems and then develop algorithms to find the optimal solution of each subproblem. We apply the theoretical results to a direct current (DC) OPF problem and conduct a case study using a modified IEEE 30-bus system with wind power. We compare our results (operational cost, reliability, computational time, and optimal solutions) to those obtained using four alternative ambiguity sets [10], [16], [17], [20].

The remainder of this paper is organized as follows. Section II empirically verifies the (multivariate) unimodality of wind forecast errors and explores misspecification of the mode location. The proposed DRCC model and ambiguity set are introduced in Section III and the main theoretical results are presented in Section IV. Section V includes the case studies and Section VI concludes the paper.

II. Unimodality of Wind Forecast Errors & Error in Mean and Mode Estimates

In this section, we first empirically verify the unimodality of wind forecast error distributions using 10,000 data samples from [6], [7] with statistical outliers omitted (total probability $< 0.1\%$). The samples were generated using a Markov Chain Monte Carlo mechanism [23] based on real data that includes both hourly forecast and actual wind generation in Germany. In Fig. 1, we depict the histograms of univariate and bivariate wind forecast errors with 15 bins. Both histograms empirically justify our assumption that the probability distribution of wind forecast errors is unimodal.

Next, we empirically evaluate the errors of mean and mode estimates (i.e., the peak location in the histogram). We randomly extract 100 groups of samples, each group containing 500 data points, from the wind forecast error data pool. For each group of samples, we estimate the mean by taking sample averages and estimate the mode by identifying the center of the highest bin in the 15-bin histogram. In Fig. 2, we plot all the mean and mode estimates and the differences between them. From the left subfigure, we observe that sampling errors have larger impacts on mode estimates than on mean estimates. From the right subfigure, we observe that the mode estimate can deviate from the corresponding mean estimate in all directions. This indicates the importance of considering the misspecification of mode location in DRCC models, because the mode-mean deviation shows the skewness of the uncertainty. As a result, if we misspecify the mode location (e.g., by modeling a right-skewed distribution as a left-skewed one, see Section III-D for an example), then we may mistakenly relax the chance constraint and get poor out-of-sample performance.

III. DRCC Formulation

A. General Formulation

In this paper, we consider the following physical constraint under uncertainty:

$$a(x)^\top \xi \leq b(x), \tag{1}$$

where $x \in \mathbb{R}^l$ represents an l-dimensional decision variable, and $a(x) : \mathbb{R}^l \to \mathbb{R}^n$ and $b(x) : \mathbb{R}^l \to \mathbb{R}$ represent two affine functions of x. Uncertainty $\xi \in \mathbb{R}^n$ represents an n-dimensional random vector defined on probability space $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_\xi)$ with Borel σ-algebra $\mathcal{B}^n$ and probability distribution $\mathbb{P}_\xi$. The assumption that $a(x)$ and $b(x)$ are affine in x is a standard assumption in existing DRCC models and consistent with the DRCC DC OPF.

To manage constraint violations due to uncertainty, one natural way is to ensure that (1) is satisfied with at least a pre-defined probability threshold $1-\epsilon$, which leads to the following chance constraint [24], [25]:

$$\mathbb{P}_\xi \left(a(x)^\top \xi \leq b(x) \right) \geq 1-\epsilon, \tag{2}$$

where $1-\epsilon$ normally takes a large value (e.g., 0.99).

B. Distributionally Robust Formulation

In reality, it may be challenging to access the (true) joint probability distribution $\mathbb{P}_\xi$. Oftentimes we may only have a set of historical data and certain domain knowledge of ξ. In this case, we can consider the following distributionally robust chance constraint:

$$\inf_{\mathbb{P}_\xi \in \mathcal{D}_\xi} \mathbb{P}_\xi \left(a(x)^\top \xi \leq b(x) \right) \geq 1-\epsilon. \tag{3}$$

Instead of assuming that $\mathbb{P}_\xi$ takes a specific form, we consider an ambiguity set $\mathcal{D}_\xi$ consisting of plausible candidates of $\mathbb{P}_\xi$. Then, we require that chance constraint (2) holds with respect to all distributions in $\mathcal{D}_\xi$.

C. Ambiguity Sets

In this paper, we consider three ambiguity sets, denoted as $\mathcal{D}_\xi^i$ for $i = 1, 2, 3$, that are defined by a combination of moment and unimodality information. Precisely, we consider a generalized notion of unimodality defined as follows.

Definition 3.1: (α-Unimodality [26]) For any fixed $\alpha > 0$, a probability distribution $\mathbb{P}$ on $\mathbb{R}^n$ is called α-unimodal with mode 0 if $t^\alpha \mathbb{P}(B/t)$ is non-decreasing in $t > 0$ for every Borel set $B \in \mathcal{B}^n$.

From the definition, we notice that α parameterizes the "degree of unimodality." When $\alpha = n = 1$, the definition coincides with the classical univariate unimodality with mode 0. When $\alpha = n > 1$, the density function of ξ (if exists) peaks at the mode and is non-increasing in any directions moving away from the mode. As $\alpha \to \infty$, the requirement of unimodality gradually relaxes and eventually vanishes. Under Definition 3.1, we define the following three ambiguity sets:
Ambiguity set 1: (moment information only)

$$\mathcal{D}_\xi^1 := \left\{ \mathbb{P}_\xi \in \mathcal{P}^n : \mathbb{E}_{\mathbb{P}_\xi}[\xi] = \mu,\ \mathbb{E}_{\mathbb{P}_\xi}[\xi\xi^\top] = \Sigma \right\}, \quad (4)$$

Ambiguity set 2: (moment and α-unimodality, fixed mode)

$$\mathcal{D}_\xi^2 := \left\{ \mathbb{P}_\xi \in \mathcal{P}_\alpha^n \cap \mathcal{D}_\xi^1 :\ \mathcal{M}(\xi) = m_t \right\}, \quad (5)$$

Ambiguity set 3: (moment and α-unimodality, misspecified mode)

$$\mathcal{D}_\xi^3 := \left\{ \mathbb{P}_\xi \in \mathcal{P}_\alpha^n \cap \mathcal{D}_\xi^1 :\ \mathcal{M}(\xi) \in \Xi \right\}, \quad (6)$$

where $\mathcal{P}_\alpha^n$ and $\mathcal{P}^n$ denote all probability distributions on $\mathbb{R}^n$ with and without the requirement of α-unimodality respectively; μ and Σ denote the first and second moments of ξ; and $\mathcal{M}(\xi)$ denotes a function returning the true mode location of ξ with m_t and Ξ representing a single mode value and a connected and compact set. The compact set can be constructed using possible mode estimates calculated from samples of historical data.

Among these three ambiguity sets, we use $\mathcal{D}_\xi^1$ as a benchmark. Set $\mathcal{D}_\xi^2$ is a special case of $\mathcal{D}_\xi^3$, i.e., Ξ only contains a single value m_t. In practice, since the mode estimate is influenced by sampling errors, the mode estimates from data samples are not the same single values but distribute around a certain area. The shape of this area decides the underlying structural skewness in the uncertainty distribution. Hence, we compare $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^3$ to see how misspecified mode estimates affect the DRCC problem. In this paper, we do not additionally consider misspecified moments since this topic has been well-studied [14], [18], [19], [21] and our main results can be easily extended based on these existing works.

D. Numerical Example

We use a simple example to illustrate the impact of an inaccurate mode estimate. We assume random variable ζ follows distribution $\mathbb{P}_{\zeta_1}$. $\mathbb{P}_{\zeta_2}$ is a biased estimate of $\mathbb{P}_{\zeta_1}$ due to sampling errors. Both distributions are illustrated in Fig. 3, where each has zero mean and unit variance. However, $\mathbb{P}_{\zeta_1}$ is right-skewed with mode at -1 and $\mathbb{P}_{\zeta_2}$ is left-skewed with mode at 1. Suppose that we try to reformulate $\mathbb{P}_\zeta(\zeta \le z) \ge 90\%$. Based on the given distributions, we find $z \ge 1.8$ from the correct distribution $\mathbb{P}_{\zeta_1}$ and $z \ge 0.925$ from the biased distribution $\mathbb{P}_{\zeta_2}$. In this example, we observe that a misspecified mode estimate could shrink the 90% confidence bound by almost a half and significantly decrease the reliability of the solution to the chance constraint.

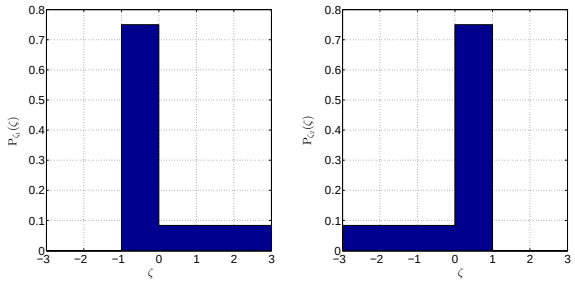

Fig. 3. True estimate $\mathbb{P}_{\zeta_1}$ and biased estimate $\mathbb{P}_{\zeta_2}$

IV. Main Results

A. Assumptions and prior results

To compute the exact reformulation of distributionally robust chance constraints with various ambiguity sets, we make the following assumptions.

Assumption 4.1: For $\mathcal{D}_\xi^2$, we assume that

$$\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) \succ \frac{1}{\alpha^2}(\mu - m_t)(\mu - m_t)^\top.$$

Similarly, for $\mathcal{D}_\xi^3$, we assume that, $\forall m \in \Xi$,

$$\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) \succ \frac{1}{\alpha^2}(\mu - m)(\mu - m)^\top.$$

Assumption 4.2: For $\mathcal{D}_\xi^2$, we assume that $a(x)^\top m_t \le b(x)$. Similarly, for $\mathcal{D}_\xi^3$, we assume that $a(x)^\top m \le b(x)$, $\forall m \in \Xi$.

Both assumptions are standard in the related literature [16], [27]–[29]. Assumption 4.1 ensures that the corresponding $\mathcal{D}_\xi^i \neq \emptyset$. Assumption 4.2 ensures that the constraint is satisfied at the mode. Furthermore, we assume $\epsilon \in (0, 0.5)$ and $\alpha \ge 1$, since in practice the uncertainties will at least be univariate-unimodal.

Reformulations of (3) under $\mathcal{D}_\xi^1$ and $\mathcal{D}_\xi^2$ are derived in previous work.

Theorem 4.1: (Theorem 2.2 in [30]) With $\mathcal{D}_\xi^1$, (3) can be exactly reformulated as

$$\sqrt{\left(\frac{1-\epsilon}{\epsilon}\right) a(x)^\top (\Sigma - \mu\mu^\top) a(x)} \le b(x) - a(x)^\top \mu. \quad (7)$$

Theorem 4.2: (Theorem 1 in [16]) With $\mathcal{D}_\xi^2$, (3) can be exactly reformulated as

$$\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}} \|\Lambda_t a(x)\| \le \tau \left(b(x) - \mu^\top a(x)\right) + \left(\tau - \frac{\alpha+1}{\alpha}\right)(\mu - m_t)^\top a(x), \quad \forall \tau \ge \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}, \quad (8)$$

where $\Lambda_t := \left(\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) - \frac{1}{\alpha^2}(\mu - m_t)(\mu - m_t)^\top\right)^{1/2}$. Since parameter τ has an infinite number of choices, the reformulation in Theorem 4.2 also involves an infinite number of SOC constraints. Here we obtain a similar result for the generalized ambiguity set $\mathcal{D}_\xi^3$.

B. Reformulation for $\mathcal{D}_\xi^3$

We now present the reformation with $\mathcal{D}_\xi^3$, which is based on Theorem 4.2:

$$\sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}}\|\Lambda a(x)\| \le \left(\tau - \frac{\alpha+1}{\alpha}\right)(\mu - m)^\top a(x) + \tau\left(b(x) - \mu^\top a(x)\right),\ \forall \tau \ge \left(\frac{1}{1-\epsilon}\right)^{1/\alpha},\ \forall m \in \Xi, \quad (9)$$

$$a(x)^\top m \le b(x),\ \forall m \in \Xi, \quad (10)$$

where $\Lambda := \left(\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top) - \frac{1}{\alpha^2}(\mu - m)(\mu - m)^\top\right)^{1/2}$ and (10) comes from Assumption 4.2.

Compared to (8), (9) is more complicated with two parameters m and τ each with an infinite number of choices. To solve an optimization problem with (9), we propose an iterative solving algorithm given in Algorithm 1.

Algorithm 1: Iterative solving algorithm

Initialization: $i = 1$, $\tau_0 = \left(\frac{1}{1-\epsilon}\right)^{1/\alpha}$, and $m_0 = \{\text{any singular point in } \Xi\}$;

Iteration i:

Step 1: Solve the reformulated optimization problem with (9) using τ_j and m_j for all $j = 0, \ldots, i-1$ and obtain optimal solution x_i^*. All τ_j and m_j values are collected from previous iterations;

Step 2 (**Separation**): Find worst-case τ_{sep}^* and m^* that result in the largest violation of (9) under x_i^*: **IF** m^* and τ^* does not exist, **STOP** and **RETURN** x_i^* as optimal solution; **ELSE GOTO** Step 3;

Step 3: Set $\tau_i = \tau_{\text{sep}}^*$, $m_i = m^*$, and $i = i + 1$;

Note that the reformulated optimization problem in Step 1 contains only SOC constraints.

C. Step 2 of Algorithm 1

The challenge is how to efficiently perform Step 2 (i.e., the separation problem) of Algorithm 1. In the following, we assume $a(x_i^*) \neq 0$, otherwise (9) is satisfied with x_i^* regardless of the values of τ and m. Next, we define the following terms

$$h = a(x_i^*)^\top(\mu - m)/\alpha,\ \tilde{c} = b(x_i^*) - \mu^\top a(x_i^*),$$
$$\tilde{R} = \sqrt{a(x_i^*)^\top\left(\frac{\alpha+2}{\alpha}\right)(\Sigma - \mu\mu^\top)a(x_i^*)},$$
$$g(\tau) = \sqrt{\frac{1-\epsilon-\tau^{-\alpha}}{\epsilon}},\ f(\tau) = -(\alpha\tau - \alpha - 1).$$

Since $m \in \Xi$, we have $h \in [\underline{h}, \overline{h}]$ where

$$\overline{h} = \max_{m\in\Xi} a(x_i^*)^\top(\mu - m)/\alpha,\ \underline{h} = \min_{m\in\Xi} a(x_i^*)^\top(\mu - m)/\alpha. \quad (11)$$

From Assumption 4.1, we have $\tilde{R} > 0$ and $[\underline{h}, \overline{h}] \subseteq (-\tilde{R}, \tilde{R})$. Hence, (9) can be transformed into

$$\left[g(\tau)\sqrt{\tilde{R}^2 - h^2} + f(\tau)h\right] - \tilde{c}\tau \le 0,\ \forall h \in [\underline{h}, \overline{h}],\ \forall \tau \ge \tau_0, \quad (12)$$

where $\tau_0 = (1/(1-\epsilon))^{1/\alpha}$ (i.e., same as in the Initialization of Algorithm 1).

Under $\mathcal{D}_\xi^2$, the separation problem is an easy-to-solve convex program [16]. However, under $\mathcal{D}_\xi^3$, the new separation problem (i.e., maximizing the left-hand side of (12)) is neither jointly convex nor jointly concave in h and τ (see proof in Appendix A) and hence challenging to solve. The standard solution method, which requires enumerating all boundary points of the domain and all stationary points with respect to (h, τ), is computationally prohibitive because this separation problem has multiple variables with an unbounded domain and a highly-nonlinear objective function. In contrast, we develop a specialized algorithm to efficiently solve this separation problem in two steps: 1) dividing the original problem into solvable subproblems and 2) efficiently solving each subproblem.

1) Dividing the Problem: For given τ, we first note that if $h \in [-\tilde{R}, \tilde{R}]$, then the maximum of $g(\tau)\sqrt{\tilde{R}^2 - h^2} + f(\tau)h$ (i.e., part of the left-hand side of (12) dependent on h) equals $\tilde{R}\sqrt{g(\tau)^2 + f(\tau)^2}$ with maximizer

$$\hat{h}(\tau) = \frac{f(\tau)}{\sqrt{g(\tau)^2 + f(\tau)^2}}\tilde{R}, \quad (13)$$

which always exists on $[\tau_0, \infty)$ because (i) $g(\tau) = 0$ at τ_0 and $f(\tau_0) > 0$ because $\tau_0 < (\alpha+1)/\alpha$ (see proof in Appendix B), and (ii) $f(\tau) = 0$ at $(\alpha+1)/\alpha$ and $g((\alpha+1)/\alpha) > 0$ because $\tau_0 < (\alpha+1)/\alpha$.

Second, by taking the derivative of (13), we observe that $\hat{h}(\tau)$ is strictly decreasing on $[\tau_0, \infty)$ (see proof in Appendix C). In addition, we observe that $\lim_{\tau\to\infty}\hat{h}(\tau) = -\tilde{R}$ and, as $f(\tau_0) > 0$ and $g(\tau_0) = 0$, we have $\hat{h}(\tau_0) = \tilde{R}$. Furthermore, as $[\underline{h}, \overline{h}] \subseteq (-\tilde{R}, \tilde{R})$, there exist $\underline{\tau}, \overline{\tau} \in [\tau_0, \infty)$ such that

$$\hat{h}(\underline{\tau}) = \overline{h} \text{ and } \hat{h}(\overline{\tau}) = \underline{h}. \quad (14)$$

To efficiently solve (14) and search for $\underline{\tau}$ and $\overline{\tau}$, we use a golden section search by first searching for $\overline{\tau}$ on $[\tau_0, \infty)$ and then for $\underline{\tau}$ on $[\tau_0, \overline{\tau}]$. Although the search for $\overline{\tau}$ seems computationally prohibitive because of the unbounded domain $[\tau_0, \infty)$, fortunately, the following lemma derives a finite domain for the search without loss of generality.

Lemma 4.1: If $\underline{h} \ge 0$, then the golden section search for $\overline{\tau}$ can be conducted on $[\tau_0, \frac{\alpha+1}{\alpha}]$. If $\underline{h} < 0$, then the search can be conducted on $[\frac{\alpha+1}{\alpha}, \tau_a]$, where

$$\tau_a = \frac{\alpha+1}{\alpha} - \frac{\underline{h}}{\alpha}\sqrt{\frac{1-\epsilon}{\epsilon(\tilde{R}^2 - \underline{h}^2)}}.$$

The proof is given in Appendix D.

We divide the original separation problem into three subproblems, in which the domains of τ are $[\tau_0, \underline{\tau}]$, $[\underline{\tau}, \overline{\tau}]$, and $[\overline{\tau}, \infty)$, respectively.

2) Solving the Subproblems: We let (h^*, τ^*) represent the maximizer in each subproblem, which is solved to optimality as follows.

Subproblem 1: If $\tau \in [\tau_0, \underline{\tau}]$, then $h^* = \overline{h}$. It follows from (12) that

$$g(\tau)\sqrt{\tilde{R}^2 - \overline{h}^2} + f(\tau)\overline{h} - \tilde{c}\tau \leq 0,$$

which is also equivalent to the following form:

$$F_1(\tau) = C_1 g(\tau) - (\tilde{c} + \alpha\overline{h})\tau + (\alpha+1)\overline{h} \leq 0, \qquad (15)$$

where $C_1 = \sqrt{\tilde{R}^2 - \overline{h}^2} > 0$. Note that $\tilde{c} \geq -\alpha\overline{h}$ from Assumption 4.2. The left-hand side of (15) is concave on τ. Define the derivative of the left-hand side as $F_1'(\tau) = C_1 g'(\tau) - (\tilde{c} + \alpha\overline{h})$, where $g'(\tau) = (\alpha/\epsilon)\tau^{-\alpha-1}/(2g(\tau))$. We observe that $F_1'(\tau_0) > 0$ as $\lim_{\tau \to \tau_0^+} g'(\tau) = +\infty$ and

1) If $F_1'(\underline{\tau}) \leq 0$, then τ^* is the unique solution of $F_1'(\tau) = 0$ within the domain $[\tau_0, \underline{\tau}]$;
2) If $F_1'(\underline{\tau}) > 0$, then $\tau^* = \underline{\tau}$.

Subproblem 2: If $\tau \in [\underline{\tau}, \overline{\tau}]$, then $h^* = \hat{h}(\tau)$. It follows from (12) and (13) that

$$\tilde{R}\sqrt{g(\tau)^2 + f(\tau)^2} - \tilde{c}\tau \leq 0,$$

which is a one-dimensional function of τ and has the following equivalent form:

$$F_2(\tau) = \tilde{R}^2\left(g(\tau)^2 + f(\tau)^2\right) - \tilde{c}^2\tau^2 \leq 0. \qquad (16)$$

First, we observe that $F_2(\tau)$ is differentiable on $[\underline{\tau}, \overline{\tau}]$. Then, we know that the extreme value of $F_2(\tau)$ happens at the critical points (i.e., boundary points $\underline{\tau}$ and $\overline{\tau}$, or τ such that $F_2'(\tau) = 0$). In the following numerical analysis, we present efficient ways to find τ^* that maximizes the left-hand side of (16).

Second, we take the first and second derivatives of the left-hand side of (16) as follows:

$$\begin{aligned} F_2'(\tau) &= \tilde{R}^2\left[\frac{\alpha}{\epsilon}\tau^{-\alpha-1} + 2\alpha(\alpha\tau - \alpha - 1)\right] - 2\tilde{c}^2\tau \\ &= \frac{\alpha\tilde{R}^2}{\epsilon}\tau^{-\alpha-1} + (2\alpha^2\tilde{R}^2 - 2\tilde{c^2})\tau - 2\tilde{R}^2\alpha(\alpha+1), \\ F_2''(\tau) &= -\frac{\alpha\tilde{R}^2(\alpha+1)}{\epsilon}\tau^{-\alpha-2} + (2\alpha^2\tilde{R}^2 - 2\tilde{c^2}). \end{aligned}$$

Third, we analyze the following two cases for find τ^*.

Case 1: If $2\alpha^2\tilde{R}^2 - 2\tilde{c^2} \leq 0$, then $F_2'(\tau)$ is monotonically decreasing on τ and $F_2(\tau)$ is concave on τ. Further,

1) if $F_2'(\underline{\tau}) \leq 0$, then $\tau^* = \underline{\tau}$;
2) if $F_2'(\underline{\tau}) > 0$ and $F_2'(\overline{\tau}) \leq 0$, then τ^* is the unique solution of $F_2'(\tau) = 0$ within the domain $[\underline{\tau}, \overline{\tau}]$;
3) if $F_2'(\overline{\tau}) > 0$, then $\tau^* = \overline{\tau}$.

Case 2: If $2\alpha^2\tilde{R}^2 - 2\tilde{c^2} > 0$, $F_2''(\tau)$ is monotonically increasing on τ and $F_2'(\tau)$ is convex on τ. Further,

1) if $F_2''(\overline{\tau}) \leq 0$, then $F_2'(\tau)$ is decreasing within the domain. To find τ^*, we follow the same discussions as in *Case 1*;
2) if $F_2''(\overline{\tau}) > 0$ and $F_2''(\underline{\tau}) \leq 0$, then $F_2'(\tau)$ is first decreasing and then increasing. Define $F_s = F_2'(\tau_s)$ where $F_2''(\tau_s) = 0$ within the domain $[\underline{\tau}, \overline{\tau}]$, $F_l = F_2'(\underline{\tau})$, and $F_u = F_2'(\overline{\tau})$. Further,
   a) if $0 \leq F_s$, then $\tau^* = \overline{\tau}$;
   b) if $F_s < 0 \leq F_l \leq F_u$ or $F_s < 0 \leq F_u < F_l$, then $\tau* = \overline{\tau}$ or the unique solution of $F_2'(\tau) = 0$ within the domain $[\underline{\tau}, \tau_s]$ that maximizes $F_2(\tau)$;
   c) if $F_s \leq F_l < 0 \leq F_u$, then $\tau^* = \underline{\tau}$ or $\overline{\tau}$ that maximizes $F_2(\tau)$;
   d) if $F_s \leq F_u < 0 \leq F_l$, then τ^* equals the unique solution of $F'(\tau) = 0$ within the domain $[\underline{\tau}, \tau_s]$;
   e) if $F_s \leq F_l \leq F_u < 0$ or $F_s \leq F_u < F_l < 0$, then $\tau^* = \underline{\tau}$;
3) if $F_2''(\underline{\tau}) > 0$, then $F_2(\tau)$ is convex on τ. $\tau^* = \underline{\tau}$ or $\overline{\tau}$ that maximizes $F_2(\tau)$.

Subproblem 3: If $\tau \in [\overline{\tau}, \infty]$, then $h^* = \underline{h}$. It follows from (12) that

$$g(\tau)\sqrt{\tilde{R}^2 - \underline{h}^2} + f(\tau)\underline{h} - \tilde{c}\tau \leq 0,$$

which we can transform into the following equivalent form

$$F_3(\tau) = C_3 g(\tau) - (\tilde{c} + \alpha\underline{h})\tau + (\alpha+1)\underline{h} \leq 0, \qquad (17)$$

where $C_3 = \sqrt{\tilde{R}^2 - \underline{h}^2} > 0$. Note that $\tilde{c} \geq -\alpha\underline{h}$ from Assumption 4.2. Define the derivative of the left-hand side of (17) as $F_3'(\tau) = C_3 g'(\tau) - (\tilde{c} + \alpha\underline{h})$, we have $F_3(\tau)$ is concave on τ and as $\tau \to \infty$, $F_3'(\tau) \leq 0$. Further,

1) if $\tilde{c} + \alpha\underline{h} = 0$, $F_3(\tau)$ is an increasing function and $\tau^* = \infty$;
2) if $\tilde{c} + \alpha\underline{h} > 0$, as $\tau \to \infty$, $F_3'(\tau) < 0$. Based on the concavity of $F_3(\tau)$, we have that
   a) if $F_3'(\overline{\tau}) \leq 0$, then $\tau^* = \overline{\tau}$;
   b) if $F_3'(\overline{\tau}) > 0$, then τ^* equals the unique solution of $F_3'(\tau) = 0$ within the domain $[\overline{\tau}, \infty)$.

In the final case 2)-b) of solving **Subproblem 3**, the golden section search for τ with $F_3'(\tau) = 0$ seems computationally prohibitive because of the unbounded domain $[\overline{\tau}, \infty)$. Fortunately, the following lemma provides a finite upper bound τ_b such that

$$F_3'(\tau_b) = C_3 g'(\tau_b) - (\tilde{c} + \alpha\underline{h}) \leq 0.$$

It follows that we only need to perform this search within the bounded domain $[\overline{\tau}, \tau_b]$.

Lemma 4.2: Define

$$\tau_b := \left[\frac{-1 + \sqrt{1 + 4(1-\epsilon)C_2}}{2C_2}\right]^{-\frac{1}{\alpha}},$$

where $C_2 = \frac{\alpha^2 C_3^2}{4\epsilon(\tilde{c} + \alpha\underline{h})^2}$. Then, we have $F_3'(\tau_b) \leq 0$.

The proof is given in Appendix E.

In summary, all the subproblems can be efficiently solved by approaches no more computationally difficult than golden section search. Finally, by combining all three subproblems, we can find the overall worst case τ^*_{sep} and h^* for given x_i^*. If

(12) is satisfied with these parameters, then there is no violated constraint in Step 2 of Algorithm 1. If (12) is not satisfied, we need to use the worst-case τ^*_{sep} and m^* in Step 3 and the iteration continues. Depending on how we define Ξ, m^* are different functions of h^*.

D. Candidates of Ξ

In this section, we demonstrate how the selection of Ξ affects the determination of $\underline{h}$, $\overline{h}$, and m^*. Specifically, we give two examples of Ξ and show how to exactly reformulate (10) (i.e., Assumption 4.2) and how to calculate $\underline{h}$ and $\overline{h}$, given x_i^*. Furthermore, we show how to find the worst case m^* from h^*.

Rectangular Support: We assume that $\Xi = [\underline{k}, \overline{k}]$ and hence we can reformulate (10) as

$$a(x)^\top \left(\frac{\underline{k}+\overline{k}}{2}\right) + |a(x)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right) \le b(x). \tag{18}$$

Furthermore, given x_i^*, we have the following relationships due to (11).

$$\underline{h} = \left[a(x_i^*)^\top \left(\mu - \frac{\underline{k}+\overline{k}}{2}\right) - |a(x_i^*)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right)\right]/\alpha, \tag{19}$$

$$\overline{h} = \left[a(x_i^*)^\top \left(\mu - \frac{\underline{k}+\overline{k}}{2}\right) + |a(x_i^*)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right)\right]/\alpha. \tag{20}$$

Based on (19) and (20), if we have the worst case h^*, we find the worst case m^* by solving (21) for λ_r and substituting in (22):

$$h^* = \left[a(x_i^*)^\top \left(\mu - \frac{\underline{k}+\overline{k}}{2}\right) + \lambda_r |a(x_i^*)|^\top \left(\frac{\overline{k}-\underline{k}}{2}\right)\right]/\alpha, \tag{21}$$

$$m^* = \left(\frac{\underline{k}+\overline{k}}{2}\right) - \lambda_r \textbf{sign}\left(a(x_i^*)\right)\left(\frac{\overline{k}-\underline{k}}{2}\right), \tag{22}$$

where $\textbf{sign}\left(a(x)\right)$ returns a diagonal matrix whose diagonal elements equal the sign of each elements in $a(x)$.

Ellipsoidal Support: We assume that $\Xi = \{m : m = m_c + P^{1/2}u, ||u||_2 \le 1\}$, where $P \succ 0$. Then we can reformulate (10) as

$$a(x)^\top m_c + \left\|P^{1/2}a(x)\right\|_2 \le b(x). \tag{23}$$

Furthermore, due to (11), we have the following relationships:

$$\underline{h} = \left[a(x_i^*)^\top(\mu - m_c) - \left\|P^{1/2}a(x_i^*)\right\|_2\right]/\alpha, \tag{24}$$

$$\overline{h} = \left[a(x_i^*)^\top(\mu - m_c) + \left\|P^{1/2}a(x_i^*)\right\|_2\right]/\alpha. \tag{25}$$

Next, if we have the worst case h^*, we find the worst case m^* directly by solving (26) for λ_e and substituting in (27):

$$h^* = \left[a(x_i^*)^\top(\mu - m_c) - \lambda_e \left\|P^{1/2}a(x_i^*)\right\|_2\right]/\alpha, \tag{26}$$

$$m^* = m_c + \lambda_e \frac{Pa(x_i^*)}{\left\|P^{1/2}a(x_i^*)\right\|_2}. \tag{27}$$

V. Case Study

A. Simulation Setup

We consider the DC OPF problem from [16] (specifically, (36a)–(36g) in [16] where the problem is referred to as a risk-constrained economic dispatch problem, but it is actually a DC OPF problem because it includes the DC power flow equations as constraints). The formulation is similar to those in [3], [6], [7], [10]. We assume that the system has two wind power plants with wind forecast error $\tilde{w} = [\tilde{w}_1, \tilde{w}_2]^\top$. With N_G generators and N_B buses, the design variables are generation $P_G \in \mathbb{R}^{N_G}$, up and down reserve capacities $R_G^{up} \in \mathbb{R}^{N_G}, R_G^{dn} \in \mathbb{R}^{N_G}$, and a distribution vector $d_G \in \mathbb{R}^{N_G}$, which determines the real-time reserve provision from each generator used to balance the wind forecast error. The full problem formulation is as follows.

$$\min\ P_G^T[C_1]P_G + C_2^T P_G + C_R^T(R_G^{up} + R_G^{dn}) \tag{28a}$$

$$\text{s.t.} -P_l \le AP_{\text{inj}} \le P_l \tag{28b}$$

$$R_G = -d_G(\tilde{w}_1 + \tilde{w}_2) \tag{28c}$$

$$P_{\text{inj}} = C_G(P_G + R_G) + C_W(P_W^f + \tilde{w}) - C_L P_L \tag{28d}$$

$$\underline{P}_G \le P_G + R_G \le \overline{P}_G \tag{28e}$$

$$-R_G^{dn} \le R_G \le R_G^{up} \tag{28f}$$

$$\mathbf{1}_{1\times N_G} d_G = 1 \tag{28g}$$

$$\mathbf{1}_{1\times N_B}(C_G P_G + C_W P_W^f - C_L P_L) = 0 \tag{28h}$$

$$P_G \ge \mathbf{0}_{N_G\times 1},\ d_G \ge \mathbf{0}_{N_G\times 1} \tag{28i}$$

$$R_G^{up} \ge \mathbf{0}_{N_G\times 1},\ R_G^{dn} \ge \mathbf{0}_{N_G\times 1} \tag{28j}$$

where $[C_1] \in \mathbb{R}^{N_G\times N_G}$, $C_2 \in \mathbb{R}^{N_G}$, and $C_R \in \mathbb{R}^{N_G}$ are cost parameters. Constraint (28b) bounds the power flow, which is calculated from the power injections P_{inj} defined in (28d) and the parameter matrix A, by the line limits P_l. Constraint (28c) computes the real-time reserve usage R_G for each generator. In (28d) P_W^f is the wind forecast, P_L is the load, and C_G, C_W, and C_L are matrices that map generators, wind power plants, and loads to buses; (28e) restricts generation to within its limits $[\underline{P}_G, \overline{P}_G]$; (28f) restricts R_G by the reserve capacity; (28g), (28h) enforce power balance with and without wind forecast error; and (28i), (28j) ensure all decision variables are non-negative. Note that this formulation is similar to that in [4] in that it assumes wind forecast errors are affinely compensated by generator reserves. The difference between the formulations is the treatment of the reserve costs. Reference [4] uses the expected cost of the total generation (scheduled generation plus uncertain reserve actions) and our formulation uses the cost of the reserve capacity.

We test our approach on a modified IEEE 30-bus system with network and cost parameters from [31]. We set $C_R = 10C_2$. We add the wind power plants to buses 22 and 5 and set $P_W^f = [66.8, 68.1]$ MW. We use the same wind power forecast uncertainty data (10000 scenarios) as in Section II. We congest the system by increasing each load by 50% and reducing the limit of the line connecting buses 1 and 2 to 30 MW. All optimization problems are solved using CVX with the Mosek solver [32], [33].

To construct the ambiguity sets, unlike in Section II, the outliers are used when estimating the statistical parameters

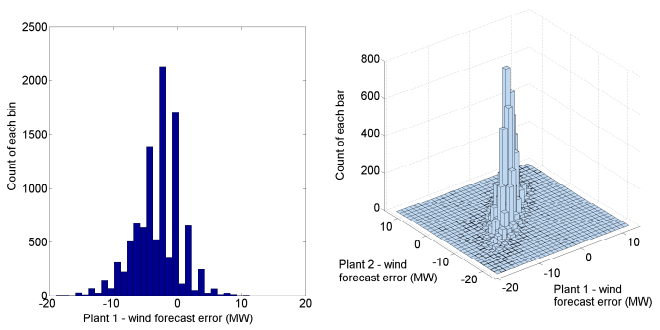

Fig. 4. Histogram of univariate and bivariate wind forecast errors (30 bins).

(first moment μ, second moment Σ, and the set of the mode Ξ) and evaluating the reliability of the solution. We set $\epsilon = 5\%$, $\alpha = 1$, and assume Ξ is a rectangular set.

B. Additional Ambiguity Sets

We benchmark against two additional ambiguity sets from related work.

Ambiguity set 4: (moment and unimodality with fixed mode at the mean [20])

$$\mathcal{D}_\xi^4 := \left\{\mathbb{P}_\xi \in \mathcal{P}_\alpha^n \cap \mathcal{D}_\xi^1 : \ \mathcal{M}(\xi) = \mu\right\}. \tag{29}$$

Ambiguity set 5: (moment and unimodality with $\alpha = 1$ and arbitrary mode [17])

$$\mathcal{D}_\xi^5 := \left\{\mathbb{P}_\xi \in \mathcal{P}_1^n \cap \mathcal{D}_\xi^1\right\}. \tag{30}$$

Set $\mathcal{D}_\xi^4$ is a special case of $\mathcal{D}_\xi^2$ with the mode at the mean, while $\mathcal{D}_\xi^5$ is a special case of $\mathcal{D}_\xi^3$ with $\alpha = 1$ and $\Xi := \{\text{all possible values of } \mathcal{M}(\xi)\}$ that is an ellipsoidal set based on μ and Σ as shown in Assumption 4.1. In other words, our $\mathcal{D}_\xi^3$ is more general than $\mathcal{D}_\xi^2$, $\mathcal{D}_\xi^4$, and $\mathcal{D}_\xi^5$. The reformulations of $\mathcal{D}_\xi^4$ and $\mathcal{D}_\xi^5$ are simpler than $\mathcal{D}_\xi^3$ with a single SOC constraint

$$\mathcal{K}\sqrt{a(x)^\top(\Sigma - \mu\mu^\top)a(x)} \le b(x) - a(x)^\top\mu, \tag{31}$$

where $\mathcal{K}$ can be found in [20] for $\mathcal{D}_\xi^4$ and in [17] for $\mathcal{D}_\xi^5$.

C. Simulation Results

1) Estimation of Ξ: We next analyze how the data size of each sample N_{data} and the number of bins within the histogram N_{bin} affect the estimate of the mode support. Figure 4 shows that if we change N_{bin} from 15 to 30 the histograms no longer show a unimodal distribution, as compared to Fig. 1. The problem is exacerbated as N_{bin} grows.

We next explore the impact of the size of the data pool. We first use the entire data pool to select 100 samples with different data sizes N_{data} (100 and 1000) and number of bins N_{bin} (15 and 30) and show scatter plots of the mode values in Fig. 5. As N_{data} gets larger, the mode values are more condensed and hence more accurate. When $N_{bin} = 30$ and $N_{data} = 100$ mode values appear in several disjoint regions, but this disjointness is mitigated as N_{data} increases to 1000. Based on the scatter plots, we determined the parameters $\underline{k}, \overline{k}$ of the four rectangular sets Ξ used in $\mathcal{D}_\xi^3$. The results are given in Table I.

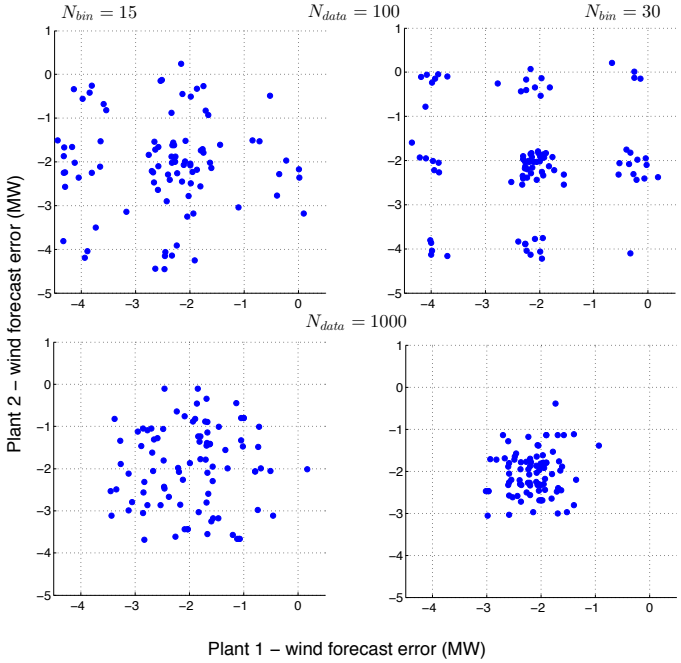

Fig. 5. Mode values from samples with different N_{data} and N_{bin}. Data is sampled from the full data pool.

TABLE I
FULL POOL: $\underline{k}$ AND $\overline{k}$ (MW) OF FOUR RECTANGULAR SETS Ξ.

N_{bin}		$N_{data} = 100$ $\underline{k}$	$\overline{k}$	$N_{data} = 1000$ $\underline{k}$	$\overline{k}$
15	Plant 1	-4.44	0.10	-3.45	0.17
	Plant 2	-4.45	0.24	-3.69	-0.11
30	Plant 1	-4.36	0.19	-3.02	-0.93
	Plant 2	-4.22	0.22	-3.06	-0.39

TABLE II
PARTIAL POOL: $\underline{k}$ AND $\overline{k}$ (MW) OF FOUR RECTANGULAR SETS Ξ

N_{bin}		$N_{data} = 50$ $\underline{k}$	$\overline{k}$	$N_{data} = 200$ $\underline{k}$	$\overline{k}$
10	Plant 1	-4.77	0.58	-3.52	0.09
	Plant 2	-5.05	0.44	-4.43	0.06
20	Plant 1	-5.82	0.06	-4.36	-0.09
	Plant 2	-5.76	0.04	-3.86	0.19

We repeated the analysis using only a partial data pool, specifically, we randomly selected 1000 data from the full pool to comprise the partial pool. We also use different choices of N_{data} and N_{bin}. The scatter plots are shown in Fig. 6 and parameter values for Ξ are given in Table II.

2) Construction of the Ambiguity Sets: In all case studies, since we focus on mode misspecification not moment misspecification, moments are calculated using the full or partial data pool and all ambiguity sets use the same moments.

For ambiguity set $\mathcal{D}_\xi^2$, we perform tests with the following six fixed mode estimates:

- M1: mode determined using the full (partial) data pool with histogram of 15 (10) bins. This case demonstrates the performance of $\mathcal{D}_\xi^2$ with an accurate mode estimate.

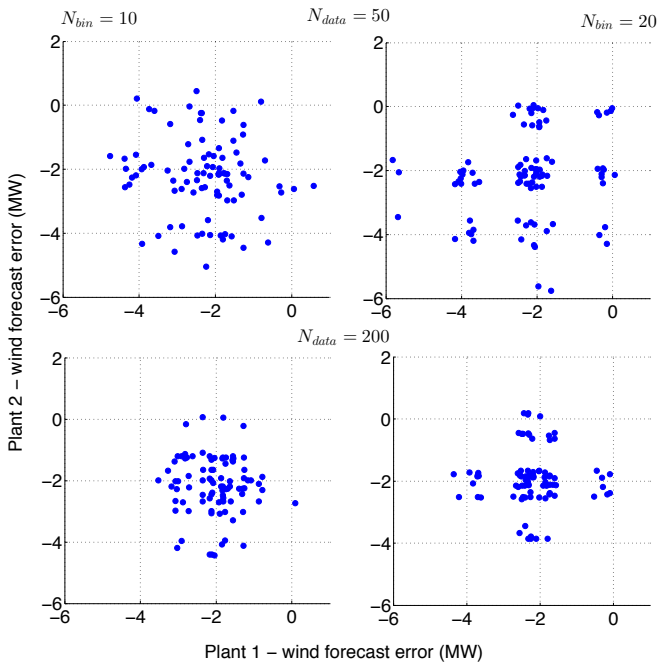

Fig. 6. Mode values from samples with different N_{data} and N_{bin}. Data is sampled from the partial data pool.

- M2: mode determined using the full (partial) data pool with histogram of 30 (20) bins. This case shows how N_{bin} affects the result.
- M3-6: combinations of the largest $\overline{k}$ and the smallest $\underline{k}$ of both plants from Table I (full pool) and Table II (partial pool). These cases demonstrate the affect of outlying data samples.

For ambiguity set $\mathcal{D}_\xi^3$, we perform tests with different Ξ, specifically, $\Xi_1 : 100 \times 15, \Xi_2 : 1000 \times 15, \Xi_3 : 100 \times 30, \Xi_4 : 1000 \times 30, \Xi_5 : 50 \times 10, \Xi_6 : 200 \times 10, \Xi_7 : 50 \times 20$, and $\Xi_8 : 200 \times 20$, where the first number refers to N_{data} and the second number refers to N_{bin}. In each case we use the parameters $\underline{k}, \overline{k}$ from Tables I and II.

3) Objective Costs: We next analyze the objective costs and the optimal reserve capacities using different ambiguity sets. The results are summarized in Table III.

From Table III, we see that $\mathcal{D}_\xi^1$ has the highest objective cost since it does not include the assumption of unimodality. The cost of $\mathcal{D}_\xi^2$ varies with the mode estimate. We observe opposite variations on the total up and down reserve capacities since different mode estimates lead to different estimates of the skewness of the uncertainty distribution. Comparing M1 and M2 to M3-6 we see that inaccurate estimation of the mode could lead to either higher or lower costs. Furthermore, results for M1 and M2 are significantly different demonstrating the effect of different choices of N_{bin}.

The costs of $\mathcal{D}_\xi^3$ are higher than those of $\mathcal{D}_\xi^2$ since the solution is designed to cope with mode misspecification. The costs do not vary significantly as a function of N_{bin} and N_{data}. For a given N_{bin}, as N_{data} increases, the costs decrease since the mode estimates are more closely clustered.

The cost of $\mathcal{D}_\xi^4$ is higher than the costs of $\mathcal{D}_\xi^2$ with M1, demonstrating the benefit in allowing the mode to be different than the mean. The cost of $\mathcal{D}_\xi^5$ is close to that of $\mathcal{D}_\xi^3$ with Ξ_3 since the mode estimates are widely distributed in this case; however, the cost of all other $\mathcal{D}_\xi^3$ is below that of $\mathcal{D}_\xi^5$. As expected, $\mathcal{D}_\xi^3$ is lower bounded by the fixed mode ambiguity sets $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^4$, and upper bounded by $\mathcal{D}_\xi^5$.

4) Reliability: Using the solutions we generated, we run out-of-sample test with 20 samples of 5000 wind forecast errors to evaluate the joint reliability of each optimal solution. We define the joint reliability as the percentage of wind forecast errors for which all chance constraints are satisfied. Then, we compare the reliability results with our pre-defined probability level ($1 - \epsilon = 95\%$). The results are summarized in Table IV.

We observe that reliability ranking almost always matches the cost ranking. Ambiguity sets $\mathcal{D}_\xi^1$ and $\mathcal{D}_\xi^5$ have the most conservative solutions and hence higher reliability and costs. The reliability of $\mathcal{D}_\xi^3$ is lower bounded by the reliability of $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^4$, and upper bounded by the reliability of $\mathcal{D}_\xi^5$. It also shows robustness against the selection of N_{data} and N_{bin}. For the full pool, all ambiguity sets achieve constraint satisfaction above 95%. For the partial pool, $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^4$ fail to meet the threshold, while ambiguity sets with misspecified modes $\mathcal{D}_\xi^3$, arbitrary modes $\mathcal{D}_\xi^5$, or no unimodality assumptions $\mathcal{D}_\xi^1$ achieve constraint satisfaction above 95%.

In this example, $\mathcal{D}_\xi^5$ can be use to approximate $\mathcal{D}_\xi^3$ since they have similar reliability. However, $\mathcal{D}_\xi^3$ is less conservative than $\mathcal{D}_\xi^5$ if Ξ does not include the global worst case mode. Set $\mathcal{D}_\xi^3$ is also more applicable to multivariate unimodality as $\mathcal{D}_\xi^5$ is only defined for $\alpha = 1$.

5) Computational Effort: Table V shows the iteration count, percent of total solve time used for solving the separation problem, and computational time for $\mathcal{D}_\xi^2$ and $\mathcal{D}_\xi^3$. The problems can be solved within 10 iterations and the computational time grows linearly with the number of iterations. Set $\mathcal{D}_\xi^3$ requires more iterations than $\mathcal{D}_\xi^2$. Problems using ambiguity sets $\mathcal{D}_\xi^1$, $\mathcal{D}_\xi^4$, and $\mathcal{D}_\xi^5$ can each be solved in a single run, and each takes less than one second. The time percent of the separation problem using $\mathcal{D}_\xi^3$ is much smaller than that using $\mathcal{D}_\xi^2$. We also observe that the time percent drops as the iteration number increases since more violated constraints are included in Step 1 of Algorithm 1 over the iterations.

To qualitatively understand the scalability of the Algorithm 1, we note that the number of separation problems solved in each iteration is always the same as the number of distributionally robust chance constraints in the original problem formulation. The time percent of the separation problem will slightly increase over iterations as more violated constraints are considered. Specifically, in the first iteration, the number of the constraints is the same to the reformulation using $\mathcal{D}_\xi^1$.

VI. Conclusion

In this paper, we proposed a distributionally robust chance constrained optimal power flow formulation considering uncertainty distributions with known moments and generalized unimodality with misspecified modes. We derived an efficient solving algorithm by iteratively constructing new optimization

TABLE III
OBJECTIVE COSTS AND RESERVE CAPACITIES

Full pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_1	Ξ_2	Ξ_3	Ξ_4		
Total Cost	26160	19440	19546	18993	19547	19526	19542	19949	19818	19982	19896	19818	19982
Generation Cost	13032	11515	11504	11506	11481	11491	11506	11522	11522	11522	11522	11514	11522
Reserve Cost	13129	7925	8042	7488	8065	8035	8036	8427	8296	8460	8373	8304	8460
Up Reserve (MW)	38.8	26.8	26.2	26.3	25.1	26.1	26.1	27.1	26.8	27.1	27.0	26.7	27.1
Down Reserve (MW)	26.9	12.9	14.0	11.1	15.2	14.1	14.1	15.1	14.7	15.2	14.9	14.8	15.2

Partial pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_5	Ξ_6	Ξ_7	Ξ_8		
Total Cost	21845	16759	16740	15250	15915	16735	16718	17883	17909	17876	17909	16791	17949
Generation Cost	11713	11376	11376	11324	11295	11336	11367	11420	11420	11420	11420	11371	11420
Reserve Cost	10132	5383	5364	3926	4620	5399	5351	6463	6489	6456	6489	5420	6529
Up Reserve (MW)	31.1	20.1	20.1	17.5	16.1	19.4	19.1	22.0	22.0	21.9	22.0	19.2	22.1
Down Reserve (MW)	19.6	6.8	6.7	2.1	7.0	7.6	7.7	10.3	10.4	10.4	10.4	7.9	10.6

TABLE IV
JOINT RELIABILITY (%) FOR $1-\epsilon = 95\%$

Full pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_1	Ξ_2	Ξ_3	Ξ_4		
min	99.78	98.22	98.08	97.72	98.14	98.02	98.06	98.54	98.48	98.58	98.52	98.44	98.58
avg	99.87	98.61	98.53	98.21	98.48	98.47	98.54	98.94	98.86	98.96	98.91	98.81	98.96
max	99.94	98.84	98.84	98.42	98.74	98.84	98.86	99.14	99.10	99.14	99.12	99.04	99.14

Partial pool	$\mathcal{D}_\xi^1$	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$				$\mathcal{D}_\xi^4$	$\mathcal{D}_\xi^5$
		M1	M2	M3	M4	M5	M6	Ξ_5	Ξ_6	Ξ_7	Ξ_8		
min	99.46	92.32	92.16	82.60	88.20	91.58	92.42	95.56	95.70	95.64	95.70	92.72	95.78
avg	99.64	93.13	93.01	83.42	88.92	92.20	93.16	96.24	96.29	96.24	96.29	93.48	96.40
max	99.78	93.68	93.62	84.20	89.54	92.80	93.58	96.64	96.64	96.62	96.64	93.80	96.78

TABLE V
ITERATION COUNT, TIME PERCENT OF SEPARATION PROBLEM, AND COMPUTATIONAL TIME FOR $\mathcal{D}_\xi^2$ AND $\mathcal{D}_\xi^3$

Full pool	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$			
	M1	M2	M3	M4	M5	M6	Ξ_1	Ξ_2	Ξ_3	Ξ_4
Iterations	4	4	8	8	4	4	9	8	9	6
Time Percent (%)	67.17	69.23	61.19	61.52	69.05	69.40	31.75	32.03	31.15	38.90
Time (s)	16.73	16.65	40.25	39.34	17.39	16.93	33.53	31.64	33.58	19.53

Partial pool	$\mathcal{D}_\xi^2$						$\mathcal{D}_\xi^3$			
	M1	M2	M3	M4	M5	M6	Ξ_5	Ξ_6	Ξ_7	Ξ_8
Iterations	4	4	6	7	4	4	9	9	9	9
Time Percent (%)	70.92	72.31	64.72	62.70	70.79	69.60	32.46	32.27	30.80	32.18
Time (s)	16.78	17.05	27.30	33.35	16.79	16.91	34.08	36.72	36.21	36.08

problems based on the separation step. In each iteration of the algorithm, the resulting optimization problem contains only SOC constraints and hence can be solved with commercial solvers. Using wind forecast errors, we found that the distribution of mode estimates are highly dependent on the data pool size, the data size of each sample, and the number of bins used in the histogram. We tested our approach on a modified IEEE 30-bus system and compared our results to those generated with other ambiguity sets. Without the assumption of unimodality, we obtain overly conservative results as unrealistic distributions are included in the ambiguity set. Considering unimodality, but with fixed mode, the results are highly dependent on the quality of the mode estimate. Considering unimodality with misspecified mode, the results are relatively consistent across different mode supports and the performance is bounded by that of the fixed-mode model and that of the arbitrary-mode model. With univariate unimodality and large mode deviations, the misspecified-mode model can be well approximated by the arbitrary-mode model.

In future work, we will first extend the current results by considering more accurate descriptions of the mode support. For example, we could represent the mode support as a union of disjoint sets that matches the mode profile. Secondly, we will evaluate the scalability of the approach on a more realistic system. Specifically, we will determine the computational effort of solving second-order cone programs (i.e., Step 1 of Algorithm 1) and separation problems (i.e., Step 2 of Algorithm 1) over the iterations and explore the application of techniques such as cutting plane algorithms [4] and parallelization to reduce the computational time. Thirdly, we will study how the current approach works in the cases with other misspecified information such as moments.

APPENDIX A
CONVEXITY AND CONCAVITY OF (12)

Here we prove the left side of (12) is neither jointly convex nor concave in h and τ through counter examples. We first pick $\alpha = \tilde{R} = 1$, $\epsilon = 0.05$, and $\tilde{c} = 0$ without loss of generality. Then we select two groups of points and calculate the left-side values v. Group 1: $[h, \tau, v] = (0.1, 2, 2.985)$ and $(0.3, 3, 3.05)$, then the midpoint $(0.2, 2.5, 3.15)$ has a value higher than line segment value 3.0175 (concave). Group 2: $[h, \tau, v] = (0.4, 11, 0.1990)$ and $(0.6, 10, -1.5015)$, then the midpoint $(0.5, 10.5, -0.6693)$ has a value lower than line segment value -0.65125 (convex).

APPENDIX B
BOUNDS OF τ_0

First, $\tau_0 = (1/(1-\epsilon))^{1/\alpha} > 1$ because $\epsilon \in (0, 0.5)$ and $\alpha \geq 1$. Second, we show that $\tau_0 = (1/(1-\epsilon))^{1/\alpha} < (\alpha+1)/\alpha$. To this end, we note that $1/(1-\epsilon) < 2$ and prove that

$$2 \leq H(\alpha) := \left(\frac{\alpha+1}{\alpha}\right)^{\alpha}$$

for all $\alpha \geq 1$. As $H(1) = 2$, it suffices to show that $H(\alpha)$ is strictly increasing when $\alpha \geq 1$. Equivalently, we prove that

$$G(\alpha) = \ln(H(\alpha)) = \alpha(\ln(\alpha+1) - \ln(\alpha))$$

is strictly increasing on $[1, \infty)$. Taking the first two derivatives yields

$$G'(\alpha) = \ln\left(\frac{\alpha+1}{\alpha}\right) - \frac{1}{\alpha+1}, \quad G''(\alpha) = -\frac{1}{\alpha(\alpha+1)^2} < 0.$$

As $\lim_{\alpha\to\infty} G'(\alpha) = 0$, we have $G'(\alpha) > 0$ on $[1, \infty)$. It follows that $G(\alpha)$ is strictly increasing on $[1, \infty)$.

APPENDIX C
MONOTONICITY OF $\hat{h}(\tau)$

By taking the derivative of (13), we have

$$\hat{h}'(\tau) = \Big[-g(\tau)\big(\alpha g(\tau) + f(\tau)g'(\tau)\big)\Big] \cdot \frac{\tilde{R}}{\big(f(\tau)^2 + g(\tau)^2\big)^{\frac{3}{2}}},$$

where $\tilde{R}/\big(f(\tau)^2 + g(\tau)^2\big)^{3/2} > 0$. To show that $\hat{h}(\tau)$ is strictly decreasing on $[\tau_0, \infty)$, we let $J(\tau) := -g(\tau)\left(\alpha g(\tau) + f(\tau)g'(\tau)\right)$ and show that $J(\tau_0) = 0$ and $J(\tau) < 0$ on (τ_0, ∞) by discussing the following three cases:

1) If $\tau = \tau_0$, then $J(\tau) = 0$ since $g(\tau_0) = 0$;
2) If $\tau \in (\tau_0, (\alpha+1)/\alpha)$, then $J(\tau) < 0$ since $g(\tau) > 0$, $f(\tau) > 0$, and
$$g'(\tau) = (\alpha/\epsilon)\tau^{-\alpha-1}/(2g(\tau)) > 0;$$
3) If $\tau \geq (\alpha+1)/\alpha$, then we have $g(\tau) > 0$, $f(\tau) \leq 0$ and
$$\big(\alpha g(\tau) + f(\tau)g'(\tau)\big)' = f(\tau)g''(\tau) \geq 0$$
because $g(\tau)$ is concave (i.e., $g''(\tau) \leq 0$). Then, the minimum of $\big(\alpha g(\tau) + f(\tau)g'(\tau)\big)$ over $[(\alpha+1)/\alpha, \infty)$ is attained at $\tau = (\alpha+1)/\alpha$ and equals $\alpha g(\tau)$, which is positive. Hence, we conclude that $J(\tau) < 0$ on $[(\alpha+1)/\alpha, \infty)$.

APPENDIX D
PROOF OF LEMMA 4.1

First, if $\underline{h} \geq 0$, then we have $\overline{\tau} \in \left[\tau_0, \frac{\alpha+1}{\alpha}\right]$ because $\hat{h}\left(\frac{\alpha+1}{\alpha}\right) = 0$ and $\hat{h}(\tau)$ is strictly decreasing on $[\tau_0, \infty)$. It follows that the golden section search for $\overline{\tau}$ can be conducted on $[\tau_0, \frac{\alpha+1}{\alpha}]$.

Second, if $\underline{h} < 0$, we have $\overline{\tau} > \frac{\alpha+1}{\alpha}$. We define

$$h_2(\tau) := \frac{f(\tau)}{\sqrt{\frac{1-\epsilon}{\epsilon} + f(\tau)^2}}\tilde{R} \tag{32}$$

and observe that $\hat{h}(\tau) < h_2(\tau)$ for all $\tau \geq \overline{\tau}$, because $f(\tau) < f((\alpha+1)/\alpha) = 0$ and $0 < g(\tau)^2 < \frac{1-\epsilon}{\epsilon}$. It follows that there exists a $\tau_a \in (\frac{\alpha+1}{\alpha}, \infty)$ such that $h_2(\tau_a) = \underline{h}$ because (i) $h_2(\tau)$ is continous on $\left[\frac{\alpha+1}{\alpha}, \infty\right)$, (ii) $h_2\left(\frac{\alpha+1}{\alpha}\right) = 0$, and (iii) $\lim_{\tau\to\infty} h_2(\tau) = -\tilde{R} < \underline{h}$. Hence, $\overline{\tau} \leq \tau_a$ because $\hat{h}(\tau_a) < h_2(\tau_a) = \underline{h} = \hat{h}(\overline{\tau})$ and $\hat{h}(\tau)$ is strictly decreasing on $[\tau_0, \infty)$.

Third, we solve the equation $h_2(\tau_a) = \underline{h}$ as follows to find τ_a.

$$\begin{aligned} & h_2(\tau_a) = \underline{h} \\ \Rightarrow \quad & f(\tau_a) = \underline{h}\sqrt{\frac{1-\epsilon}{\epsilon(\tilde{R}^2 - \underline{h}^2)}} \\ \Rightarrow \quad & \tau_a = \frac{\alpha+1}{\alpha} - \frac{\underline{h}}{\alpha}\sqrt{\frac{1-\epsilon}{\epsilon(\tilde{R}^2 - \underline{h}^2)}}. \end{aligned}$$

APPENDIX E
PROOF OF LEMMA 4.2

First, for all $\tau \geq \overline{\tau} > \tau_0 > 1$, we have $g(\tau) > 0$ and so the following inequality holds:

$$g'(\tau) = \frac{\frac{\alpha}{\epsilon}\tau^{-\alpha-1}}{2g(\tau)} < g_2(\tau) = \frac{\frac{\alpha}{\epsilon}\tau^{-\alpha}}{2g(\tau)}. \tag{33}$$

Then, based on (33) and the fact that $C_3 > 0$, we have

$$C_3 g'(\tau) - (\tilde{c} + \alpha\underline{h}) = F_3'(\tau) \leq F_4(\tau) = C_3 g_2(\tau) - (\tilde{c} + \alpha\underline{h})$$

for all $\tau > \tau_0$.

Second, there exists a $\tau_b \in [\overline{\tau}, \infty)$ such that $F_4(\tau_b) = 0$ because (i) $F_4(\tau)$ is continuous on $[\overline{\tau}, \infty)$, (ii) $F_4(\overline{\tau}) \geq F_3'(\overline{\tau}) > 0$, and (iii) as $\tau \to \infty$, $F_4(\tau) \to -(\tilde{c} + \alpha\underline{h}) < 0$. Let $\tau^* \in [\overline{\tau}, \infty)$ be such that $F_3'(\tau^*) = 0$. Then, it follows that $\tau_b \geq \tau^*$ because $F_3'(\tau_b) \leq F_4(\tau_b) = 0$.

Third, by transforming the equation $F_4(\tau_b) = 0$, we obtain

$$C_2(\tau_b^{-\alpha})^2 + \tau_b^{-\alpha} - (1-\epsilon) = 0,$$

where $C_2 = \frac{\alpha^2 C_3^2}{4\epsilon(\tilde{c}+\alpha\underline{h})^2}$. This is a quadratic equation of $\tau_b^{-\alpha}$ and we find that there exists a unique positive solution τ_b with

$$\tau_b = \left[\frac{-1 + \sqrt{1 + 4(1-\epsilon)C_2}}{2C_2}\right]^{-\frac{1}{\alpha}}.$$

REFERENCES

[1] H. Zhang and P. Li, "Chance constrained programming for optimal power flow under uncertainty," *IEEE Trans Power Systems*, vol. 26, no. 4, pp. 2417–2424, 2011.

[2] R. A. Jabr, "Adjustable robust OPF with renewable energy sources," *IEEE Trans Power Systems*, vol. 28, no. 4, pp. 4742–4751, 2013.

[3] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration," *IEEE Trans Power Systems*, vol. 28, no. 4, 2013.

[4] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: risk-aware network control under uncertainty," *SIAM Review*, vol. 56, no. 3, pp. 461–495, 2014.

[5] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson, "Analytical reformulation of security constrained optimal power flow with probabilistic constraints," in *IEEE PowerTech Conference*, Grenoble, France, 2013.

[6] M. Vrakopoulou, B. Li, and J. Mathieu, "Chance constrained reserve scheduling using uncertain controllable loads Part I: Formulation and scenario-based analysis," *IEEE Trans Smart Grid (in press)*, 2017.

[7] B. Li, M. Vrakopoulou, and J. Mathieu., "Chance constrained reserve scheduling using uncertain controllable loads Part II: Analytical reformulation," *IEEE Trans Smart Grid (in press)*, 2017.

[8] M. Campi, G. Calafiore, and M. Prandini, "The scenario approach for systems and control design," *Annual Reviews in Control*, vol. 33, no. 2, pp. 149–157, 2009.

[9] K. Margellos, P. Goulart, and J. Lygeros, "On the road between robust optimization and the scenario approach for chance constrained optimization problems," *IEEE Trans Automatic Control*, vol. 59, no. 8, pp. 2258–2263, 2014.

[10] Y. Zhang, S. Shen, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Trans Power Systems*, vol. 32, no. 2, pp. 1378–1388, 2017.

[11] W. Xie and S. Ahmed, "Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation," *IEEE Trans Power Systems*, vol. 33, no. 2, pp. 1860–1867, 2018.

[12] Y. Guo, K. Baker, E. Dall'Anese, Z. Hu, and T. Summers, "Stochastic optimal power flow based on data-driven distributionally robust optimization," *arXiv preprint arXiv:1706.04267*, 2017.

[13] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on conditional value at risk and distributional robustness," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 116 – 125, 2015.

[14] M. Lubin, Y. Dvorkin, and S. Backhaus, "A robust approach to chance constrained optimal power flow with renewable generation," *IEEE Trans Power Systems*, vol. 31, no. 5, pp. 3840–3849, 2016.

[15] B. Li, R. Jiang, and J. L. Mathieu, "Distributionally robust chance constrained optimal power flow assuming log-concave distributions (accepted)," in *IEEE Conference on Decision and Control*, Dublin, Ireland, 2018.

[16] B. Li, R. Jiang, and J. Mathieu, "Ambiguous risk constraints with moment and unimodality information," *Mathematical Programming (Accepted)*, 2017.

[17] L. Roald, F. Oldewurtel, B. V. Parys, and G. Andersson, "Security constrained optimal power flow with distributionally robust chance constraints," *arXiv preprint arXiv:1508.06061*, 2015.

[18] L. E. Ghaoui, M. Oks, and F. Oustry, "Worst-case value-at-risk and robust portfolio optimization: A conic programming approach," *Operations Research*, vol. 51, no. 4, pp. 543–556, 2003.

[19] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to data-driven problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.

[20] B. Stellato, *Data-driven chance constrained optimization*. Master thesis, ETH Zurich, 2014.

[21] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Mathematical Programming*, vol. 158, no. 1, pp. 291–327, 2016.

[22] S. H. Tseng, E. Bitar, and A. Tang, "Random convex approximations of ambiguous chance constrained programs," in *IEEE Conference on Decision and Control*, Las Vegas, NV, 2016.

[23] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," *IEEE Trans Energy Conversion*, vol. 23, no. 1, pp. 234–240, 2008.

[24] A. Charnes, W. Cooper, and G. Symonds, "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil," *Management Science*, vol. 4, no. 3, pp. 235–263, 1958.

[25] B. Miller and H. Wagner, "Chance constrained programming with joint constraints," *Operations Research*, vol. 13, no. 6, pp. 930–945, 1965.

[26] S. W. Dharmadhikari and K. Joag-Dev, *Unimodality, convexity, and applications*. Academic Press, 1988.

[27] G. Hanasusanto, *Decision Making under Uncertainty: Robust and Data-Driven Approaches*. PhD thesis, Imperial College London, 2015.

[28] B. V. Parys, P. Goulart, and D. Kuhn, "Generalized Gauss inequalities via semidefinite programming," *Mathematical Programming*, vol. 156, no. 1, pp. 271–302, 2016.

[29] B. V. Parys, P. Goulart, and M. Morari, "Distributionally robust expectation inequalities for structured distributions," *Mathematical Programming (in press)*, 2017.

[30] M. Wagner, "Stochastic 0–1 linear programming under limited distributional information," *Operations Research Letters*, vol. 36, no. 2, pp. 150–156, 2008.

[31] R. D. Zimmerman, C. E. Murillo-Sanchez, R. J. Thomas, and L. Fellow, "MATPOWER: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[32] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, 2014.

[33] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, 2008.