Energy and Reserve Dispatch with Distributionally Robust Joint Chance Constraints

Christos Ordoudis, Viet Anh Nguyen,
Daniel Kuhn, Pierre Pinson,

***Abstract*—We develop a two-stage stochastic program for energy and reserve dispatch, which ensures the safe operation of a power system with a high penetration of renewables and a strong interdependence with the natural gas system. Distributionally robust joint chance constraints with Wasserstein ambiguity sets ensure that there is no need for load shedding and renewable spillage with high probability under any distribution compatible with the given statistical data. To make this problem tractable, we solve it in linear decision rules, and we develop a family of conditional value-at-risk (CVaR) approximations for the chance constraints. We show through extensive simulations that the proposed model dominates the corresponding two-stage stochastic program *without* chance constraints that models the consequences of load shedding and renewable spillage explicitly, both in terms of the mean and variability of the out-of-sample cost.**

***Index Terms*—Distributionally robust optimization, energy and reserve dispatch, joint chance constraints, Wasserstein metric.**

I. Introduction

The increased deployment of renewable energy sources such as wind or solar power and tidal energy has dramatically changed the electricity generation mix. While conducive to sustainability, renewable energy sources impair the stability of the transmission system because of their intermittency and limited predictability [1]. Therefore, flexible gas-fired power plants (GFPPs) are widely built to replace retiring thermal or nuclear power plants with the aim to balance the unsteady renewable generation. Thus, a tighter coupling of electricity and natural gas systems is foreseeable [2]. This perspective prompts us to study both systems simultaneously.

In the presence of uncertain renewable energy sources the day-ahead dispatch may not be implementable if the actual renewable production deviates from its forecast. In this case, the conventional power plants need to adjust their production levels in real time. If—in some extreme scenarios—these plants are not flexible enough to restore the integrity of the transmission system, then renewable spillage or load shedding may become necessary. However, such drastic measures incur high economic costs.

The energy and reserve dispatch problem can be addressed with methods from robust optimization [3], stochastic programming [4] and chance-constrained programming [5]. Robust optimization models minimize the cost of the day-ahead dispatch and the corrective recourse actions under the worst-case realization of the renewable forecast error within a prescribed uncertainty set [6], [7]. However, the focus on worst-case scenarios and the necessity to solve larger problem instances in linear decision rules may result in over-conservative solutions. Stochastic programming models, on the other hand, minimize the expected cost under a prescribed discrete distribution [8], [9], but their solutions may display poor out-of-sample performance unless the number of discretization points grows exponentially with the problem dimensions. Finally, chance-constrained programming models do not account for renewable spillage and load shedding but ensure that the system remains stable with high probability in the absence of these drastic recourse measures [10]–[12].

In this paper we formulate the energy and reserve dispatch problem as a data-driven distributionally robust chance-constrained program over a Wasserstein ambiguity set, that is, a ball in the space of probability distributions with respect to the type-1 Wasserstein metric centered at the empirical distribution on a given training dataset [13]. While minimizing the worst-case expected cost, our model prevents violations of the reserve margins and enforces the transmission capacity limits of the electricity and gas networks through distributionally robust joint chance constraints. This means that the underlying classical chance constraints are enforced for all distributions in the ambiguity set. Distributionally robust optimization combines the specificity of stochastic programming (by modeling uncertainty via probability distributions) with the conservatism of robust optimization (by hedging against distributional uncertainty). The proposed approach acknowledges that the available statistical data can often be explained by many different distributions and mitigates the overfitting effects characteristic for classical stochastic programs based on a single distribution, which is invariably corrupted by estimation errors.

Most distributionally robust energy models in the extant literature are based on moment ambiguity sets [14], which contain all distributions that share the same mean vector and covariance matrix or satisfy a set of generalized moment constraints [15]–[20]. Distributionally robust *individual* chance constraints are studied in [19], [20], while *two-sided* chance constraints are addressed in [18]. Genuine *joint* chance constraints enforce simultaneous satisfaction of multiple safety conditions with high probability. They are more expressive and less conservative than multiple individual chance constraints. For instance, an operating regime under which the risk of overloading *any* of 100 transmission lines is at most 1% (joint chance constraint) is preferable to a regime under which each transmission line has an overloading risk of at most 1% (100 individual chance constraints). Indeed, in the latter case the risk of overloading any line can be as high as 100%. Unfortunately, joint chance constraints are less computationally

C. Ordoudis and P. Pinson are with the Technical University of Denmark, Kgs. Lyngby, Denmark (e-mail: chror@dtu.dk,ppin@dtu.dk). V.A. Nguyen and D. Kuhn are with École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (e-mail: viet-anh.nguyen@epfl.ch; daniel.kuhn@epfl.ch).

We gratefully acknowledge financial support from the CRYOS Laboratory at EPFL directed by Prof. Michael Lehning, from the Swiss National Science Foundation under grant BSCGI0_157733 and from the Danish Strategic Research Council under grant 1305-00027B/DSF.

tractable than individual chance constraints. Tractable conservative approximations are obtained by using Bonferroni's inequality to decompose the joint chance constraint into several individual chance constraints or by replacing the non-convex joint chance constraint with a convex conditional value-at-risk (CVaR) constraint [21]. For distributionally robust joint chance constraints with mean covariance ambiguity sets the best CVaR approximation is exact, but identifying this best approximation remains hard [22]. Exact tractable reformulations are only available for a restrictive class of moment ambiguity sets involving first-order dispersion measures [23].

Distributionally robust joint chance-constrained programs over Wasserstein ambiguity sets are NP-hard, but they admit an exact mixed-integer conic reformulation if the decision variables and uncertain parameters are separable [24]–[26]. A tractable robust approximation is proposed in [27].

The contributions of this paper are summarized as follows:

1) We formulate a two-stage stochastic program for energy and reserve dispatch with fuel constraints for the GFPPs. Distributionally robust joint chance constraints with Wasserstein ambiguity sets ensure that there is no need for load shedding and renewable spillage with high probability. To make this problem tractable, we solve it in linear decision rules, and we develop a family of CVaR approximations for the chance constraints. We also devise a sequential convex optimization algorithm for tuning the CVaR approximation.
2) We develop a simulation environment based on a real-time optimal power flow model with load shedding and renewable spillage that enables us to assess the out-of-sample performance of different day-ahead dispatch policies.
3) We show through numerical simulations that the proposed CVaR approximation Pareto dominates the Bonferroni approximation in terms of the out-of-sample cost and the chance constraints' empirical violation probability. We also show that the proposed model is preferable to a two-stage stochastic program *without* chance constraints that models load shedding and renewable spillage explicitly, both in view of the mean and variability of the out-of-sample cost.

The paper is structured as follows. Section II introduces a chance-constrained energy and reserve dispatch model, which is robustified against distributional uncertainty in Section III. Section IV develops efficient approximation schemes for the distributionally robust model under the Wasserstein ambiguity set. Numerical results are reported in Section V.

Notation. We use e to denote the vector of ones. For a logical expression $\mathcal{E}$, the indicator function $\mathbb{I}_{\mathcal{E}}$ evaluates to 1 if $\mathcal{E}$ is true and to 0 otherwise. The Conditional Value-at-Risk (CVaR) at level $\varepsilon \in (0,1)$ of a measurable loss function $\ell(\xi)$ depending on a random vector ξ governed by a distribution $\mathbb{P}$ is defined as $\mathbb{P}\text{-CVaR}_\varepsilon(\ell(\xi)) \triangleq \inf_\tau \left\{\tau + \frac{1}{\varepsilon}\mathbb{E}_{\mathbb{P}}[\max\{0, \ell(\xi)-\tau\}]\right\}$.

II. Model Formulation

We consider a power system consisting of B buses, L transmission lines, W wind farms, G generators and D demand centers, augmented with P gas pipelines. Each transmission line connects a pair of buses, while each wind farm, generator and demand center is attached to exactly one bus. Each GFPP is served by exactly one gas pipeline. Adopting a DC power flow approximation [28], we denote by $Q^{\mathrm{w}} \in \mathbb{R}^{L\times W}$, $Q^{\mathrm{g}} \in \mathbb{R}^{L\times G}$ and $Q^{\mathrm{d}} \in \mathbb{R}^{L\times D}$ the matrices of power transfer distribution factors. Thus, Q^{w}_{lw} and Q^{g}_{lg} denote the increase in the power flow on line l per unit of power generated by wind farm w and generator g, respectively. Similarly, Q^{d}_{ld} quantifies the *de*crease in the power flow on line l per unit of power extracted at demand center d. We denote by $\Phi \in \mathbb{R}^{P\times G}_+$ the matrix of gas transfer distribution factors. Thus, Φ_{pg} stands for the increase in the gas flow through pipeline p per unit of power produced by generator g. Specifically, we have $\Phi_{pg} = 0$ if g is not a GFPP. All information about the topology and physical properties of the gas and power system is hidden in the transfer distribution factors; see [28], [29].

The power output of the wind farms is modeled as $C(\mu+\xi)$, where $C \in \mathbb{R}^{W\times W}_+$ is the diagonal matrix of the wind farm capacities, $\mu \in \mathbb{R}^W$, $0 \leqslant \mu \leqslant e$, is the relative power output predicted at stage 1, and $\xi \in \mathbb{R}^W$, $-\mu \leqslant \xi \leqslant e-\mu$, is the uncertain deviation from μ, which is revealed at stage 2. We assume that ξ follows a distribution $\mathbb{P}$ centered around 0. Without much loss of generality, we assume that the power consumption $d \in \mathbb{R}^D_+$ at the demand centers is known at stage 1. The operating decisions for the generators are taken in two stages. At stage 1, the day-ahead dispatch $y_1 \in \mathbb{R}^G_+$ as well as the upward and downward reserve margins $r_+, r_- \in \mathbb{R}^G_+$ are chosen for each generator. Upon observation of ξ at stage 2, the real-time adjustments $y_2(\xi) \in \mathbb{R}^G$ to the power production are chosen with the aim to ensure—with high probability—the integrity of the transmission system while respecting the generator capacities. This energy and reserve dispatch problem gives rise to the following two-stage stochastic program

$$
\begin{array}{lll}
\min\limits_{y_1, y_2(\cdot), r_+, r_-} & c^\top y_1 + c_+^\top r_+ + c_-^\top r_- + \mathbb{E}^{\mathbb{P}}[c^\top y_2(\xi)] & \text{(1a)}\\
\text{s.t.} & 0 \leqslant r_+ \leqslant \overline{r}, \quad 0 \leqslant r_- \leqslant \overline{r} & \text{(1b)}\\
& \underline{y} \leqslant y_1 - r_-, \quad y_1 + r_+ \leqslant \overline{y} & \text{(1c)}\\
& e^\top C(\mu+\xi) + e^\top(y_1 + y_2(\xi)) = e^\top d \ \ \mathbb{P}\text{-a.s.} & \text{(1d)}\\
& \mathbb{P}[-r_- \leqslant y_2(\xi) \leqslant r_+] \geqslant 1 - \epsilon^{\mathrm{gen}} & \text{(1e)}\\
& \mathbb{P}[-\overline{f} \leqslant Q^{\mathrm{g}}(y_1 + y_2(\xi)) + Q^{\mathrm{w}} C(\mu+\xi) - Q^{\mathrm{d}} d \leqslant \overline{f}] \geqslant 1 - \epsilon^{\mathrm{grid}} & \text{(1f)}\\
& \mathbb{P}[0 \leqslant \Phi(y_1 + y_2(\xi)) \leqslant \overline{q}] \geqslant 1 - \epsilon^{\mathrm{gas}}. & \text{(1g)}
\end{array}
$$

The objective function (1a) reflects the expected operating costs, where $c \in \mathbb{R}^G_+$ captures the variable costs of the generators, and $c_+, c_- \in \mathbb{R}^G_+$ represent the costs of reserving capacity to balance the system in real time [9]. The constraints (1b) limit the (positive and negative) reserve capacity procurements up to the prescribed maximum $\overline{r} \in \mathbb{R}^G_+$, while (1c) ensures that the day-ahead energy and reserve dispatch obey the production limits $\underline{y}, \overline{y} \in \mathbb{R}^G_+$ of the generators. The almost sure constraint (1d) requires that total production matches total demand with probability 1, while the chance constraints (1e)-(1g) ensure that the real-time adjustments to the power output obey the chosen reserve capacities, the power flows respect the capacity limits $\overline{f} \in \mathbb{R}^L_+$ of the transmission lines, and the power outputs of the GFPPs are limited by the maximum delivery rates $\overline{q} \in \mathbb{R}^P_+$ of the natural gas pipelines, respectively. The

prescribed violation probabilities $\epsilon^{\text{gen}}, \epsilon^{\text{grid}}, \epsilon^{\text{gas}} \in (0,1)$ reflect the risk attitude of the decision maker.

The stochastic program (1) is intractable because it constitutes an infinite-dimensional optimization problem. To mitigate its complexity, it has been proposed to approximate the functional recourse decisions $y_2(\xi)$ by linear decision rules of the form $y_2(\xi) = Y\xi$ for some finite-dimensional coefficient matrix $Y \in \mathbb{R}^{G\times W}$ [3, § 14]. In the context of chance constrained optimal power flow problems, linear decision rules have been used in [18] and [19]. The stochastic program (1) can thus be approximated by the linear decision rule problem

$$\begin{aligned}
\min_{y_1,Y,r_+,r_-} \quad & c^\top y_1 + c_+^\top r_+ + c_-^\top r_- + \mathbb{E}^{\mathbb{P}}\big[c^\top Y\xi\big] && \text{(2a)}\\
\text{s.t.} \quad & \text{Constraints (1b)–(1c)} \\
& e^\top y_1 + e^\top C\mu = e^\top d, \quad e^\top Y + e^\top C = 0 && \text{(2d)}\\
& \mathbb{P}\big[-r_- \leqslant Y\xi \leqslant r_+\big] \geqslant 1-\epsilon^{\text{gen}} && \text{(2e)}\\
& \mathbb{P}\big[-\overline{f} \leqslant (Q^{\text{g}}y_1 + Q^{\text{w}}C\mu - Q^{\text{d}}d) + \\
& \qquad + (Q^{\text{g}}Y + Q^{\text{w}}C)\xi \leqslant \overline{f}\big] \geqslant 1-\epsilon^{\text{grid}} && \text{(2f)}\\
& \mathbb{P}\big[0 \leqslant \Phi(y_1 + Y\xi) \leqslant \overline{q}\big] \geqslant 1-\epsilon^{\text{gas}}, && \text{(2g)}
\end{aligned}$$

where (2d) is obtained by matching the zero- and first-order coefficients of ξ on both sides of (1d), which is allowed because the support of ξ spans $\mathbb{R}^W$ [30, § 2.2]. For ease of exposition, we denote by $x \triangleq (y_1, r_+, r_-) \in \mathbb{R}^{3G}_+$ the collection of all first-stage decisions and by $c_x \triangleq (c, c_+, c_-) \in \mathbb{R}^{3G}_+$ the corresponding aggregate cost vector. Using this notation, problem (2) can be represented more compactly as

$$\begin{aligned}
\min_{(x,Y)\in\Theta} \quad & c_x^\top x + \mathbb{E}^{\mathbb{P}}\big[c^\top Y\xi\big] && \text{(3a)}\\
\text{s.t.} \quad & \mathbb{P}\left[A^j(Y)\xi \leqslant b^j(x)\right] \geqslant 1-\epsilon^j \quad \forall j \in \mathcal{J}, && \text{(3b)}
\end{aligned}$$

where Θ stands for the set of all $(x,Y) \in \mathbb{R}^{3G}_+ \times \mathbb{R}^{G\times W}$ satisfying (1b), (1c) and (2d). The joint chance constraint (3b) is indexed by $j \in \mathcal{J} \triangleq \{\text{gen}, \text{grid}, \text{gas}\}$ and thus encodes the capacity constraints (2e) through

$$A^{\text{gen}}(Y) \triangleq \begin{bmatrix} Y \\ -Y \end{bmatrix}, \quad b^{\text{gen}}(x) \triangleq \begin{bmatrix} r_+ \\ -r_- \end{bmatrix},$$

the line capacity constraints (2f) through

$$\begin{aligned}
A^{\text{grid}}(Y) &\triangleq \begin{bmatrix} Q^{\text{g}}Y + Q^{\text{w}}C \\ -Q^{\text{g}}Y - Q^{\text{w}}C \end{bmatrix},\\
b^{\text{grid}}(x) &\triangleq \begin{bmatrix} \overline{f} - Q^{\text{w}}C\mu + Q^{\text{d}}d - Q^{\text{g}}y_1 \\ \overline{f} + Q^{\text{w}}C\mu - Q^{\text{d}}d + Q^{\text{g}}y_1 \end{bmatrix},
\end{aligned}$$

and the pipeline capacity constraints (2g) through

$$A^{\text{gas}}(Y) \triangleq \begin{bmatrix} \Phi Y \\ -\Phi Y \end{bmatrix}, \quad b^{\text{gas}}(x) \triangleq \begin{bmatrix} \overline{q} - \Phi y_1 \\ \Phi y_1 \end{bmatrix}.$$

In spite of the decision rule approximation, problem (3) remains intractable. In fact, only checking feasibility of the chance constraint (3b) is already #P-hard even if ξ follows a uniform distribution on a box [31]. Moreover, the distribution $\mathbb{P}$, which is needed to evaluate both the expectation in (3a) and the probabilities in (3b), is not even observable in practice but must be inferred from data. Unfortunately, the available data is often scarce, and the procurement of additional samples is either infeasible or expensive. Indeed, any wind power time series is invariably restricted to the service life of the corresponding wind farm. However, if problem (3) is fitted to a small training dataset, and the resulting optimal decisions are evaluated on a (different) test dataset, then the test performance is often disappointing, even if the training and test datasets are governed by the same (unknown) distribution $\mathbb{P}$ [13].

III. Data-driven Distributionally Robust Optimization

Assume now that the decision maker is ignorant of $\mathbb{P}$ but has access to finitely many training samples $\hat{\xi}_i$, $i \leqslant N$, drawn independently from $\mathbb{P}$ (a wind power time series). As $\mathbb{P}$ is unknown, a fundamental input of problem (3) is thus lacking. A naïve remedy would be to replace the unknown $\mathbb{P}$ with the discrete empirical distribution $\widehat{\mathbb{P}}_N$, that is, the uniform distribution on the (known) training samples. This amounts to solving the sample average approximation of problem (3), which is prone to yield biased decisions that perform poorly in out-of-sample tests for small sample sizes N. Hence, it makes sense to reformulate (3) as a distributionally robust optimization problem that hedges against all distributions in a neighborhood of $\widehat{\mathbb{P}}_N$ with respect to the Wasserstein metric.

Definition 1 (Wasserstein metric)**.** *The type-1 Wasserstein distance between two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ on $\mathbb{R}^W$ is defined as*

$$\mathbb{W}(\mathbb{P}_1,\mathbb{P}_2) \triangleq \left\{ \begin{array}{ll} \min\limits_{\Pi} & \displaystyle\int_{\mathbb{R}^W\times\mathbb{R}^W} \|\xi_1 - \xi_2\|\, \Pi(\mathrm{d}\,\xi_1, \mathrm{d}\,\xi_2) \\ \text{s.t.} & \Pi \textit{ is a distribution on } \mathbb{R}^W \times \mathbb{R}^W \\ & \textit{with marginals } \mathbb{P}_1 \textit{ and } \mathbb{P}_2\textit{, respectively.} \end{array} \right.$$

The Wasserstein distance between $\mathbb{P}_1$ and $\mathbb{P}_2$ can be viewed as the cost of an optimal mass transportation plan Π that minimizes the cost of moving $\mathbb{P}_1$ to $\mathbb{P}_2$, where $\|\xi_1 - \xi_2\|$ is the cost of moving a unit mass from ξ_1 to ξ_2. In the following we denote by $\mathcal{M}(\Xi)$ the set of all distributions on the polyhedron $\Xi = \{\xi \in \mathbb{R}^W : H\xi \leqslant h\}$, where $H = \begin{bmatrix} I & -I \end{bmatrix}^\top$ and $h = \begin{bmatrix} (e-\mu)^\top & \mu^\top \end{bmatrix}^\top$, and we define the ambiguity set

$$\mathcal{P} \triangleq \Big\{\mathbb{P} \in \mathcal{M}(\Xi) : \mathbb{W}(\mathbb{P}, \widehat{\mathbb{P}}_N) \leqslant \rho\Big\}$$

as the family of all distributions on Ξ that have a Wasserstein distance of at most $\rho \geqslant 0$ from the empirical distribution $\widehat{\mathbb{P}}_N$. The hope is that, for a judiciously chosen radius ρ, the ambiguity set $\mathcal{P}$ contains the unknown true distribution with high confidence. Following [13], we can then recast (3) as a distributionally robust optimization problem of the form

$$\begin{aligned}
\min_{(x,Y)\in\Theta} \quad & c_x^\top x + \max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}\big[c^\top Y\xi\big] && \text{(4a)}\\
\text{s.t.} \quad & \min_{\mathbb{P}\in\mathcal{P}} \ \mathbb{P}\left[A^j(Y)\xi \leqslant b^j(x)\right] \geqslant 1-\epsilon^j \quad \forall j \in \mathcal{J}, && \text{(4b)}
\end{aligned}$$

which minimizes the worst-case expected operating costs and requires that the joint chance constraints are satisfied for all distributions in the ambiguity set $\mathcal{P}$. If the true distribution belongs to $\mathcal{P}$, then the optimal value of (4) overestimates the true expected cost of the optimal decisions. Moreover, the optimal decisions satisfy the true chance constraints. Modelling the ambiguity set as a Wasserstein ball in the space of distributions has several benefits that may appeal to decision makers, i.e., it provides rigorous finite-sample and asymptotic consistency guarantees and offers computational tractability [13, § 2].

It is known that the empirical distribution $\widehat{\mathbb{P}}_N$ converges in Wasserstein metric (and thus also weakly) to the unknown true distribution as N tends to infinity. One can thus show that for any given significance level $\beta \in (0,1)$ there is a sequence $\rho_N(\beta) \geqslant 0$, $N \in \mathbb{N}$, that converges to 0 such that the Wasserstein ball of radius $\rho_N(\beta)$ around $\widehat{\mathbb{P}}_N$ contains the unknown true distribution with confidence $1-\beta$ for every N [13, Theorem 3.4]. In practice, the best Wasserstein radius for a given sample size is determined in a data-driven manner, e.g., via cross validation; see Section V. Moreover, the distributionally robust chance constrained program (4) admits several tractable conservative approximations.

IV. Tractable Approximations

The distributionally robust chance constrained program (4) is still hard. Indeed, for $\rho = 0$ it reduces to a classical chance constrained program under the discrete empirical distribution. Such problems are known to be NP-hard even in the simplest settings [32, Theorem 1]. Leveraging results from [13], we now derive tractable conservative approximations for (4). In Section IV-A we first discuss an exact reformulation for the objective function (4a). In Sections IV-B and IV-C we then provide two conservative approximations for the feasible set

$$\Omega_{\mathrm{CC}} \triangleq \left\{ (x,Y) : \min_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left[A(Y)\xi \leqslant b(x)\right] \geqslant 1-\epsilon \right\}$$

of a generic joint chance constraint of the form (4b), where the superscript j is omitted to avoid clutter. In Section IV-D we finally assess the computational tractability of the two approaches. All proofs are relegated to the e-companion [33].

A. Reformulation of the Objective Function

Evaluating the objective function (4a) for a fixed $Y \in \mathbb{R}^{G\times W}$ necessitates the solution of a worst-case expectation problem of a linear function in ξ over the Wasserstein ball $\mathcal{P}$. By [13, Corollary 5.1] this problem is equivalent to the conic program

$$\begin{aligned}
&\max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}\left[c^\top Y \xi\right] \\
&= \left\{ \begin{array}{lll} \min\limits_{\lambda^o, s^o, \gamma^o} & \lambda^o \rho + \frac{1}{N}\sum_{i=1}^N s_i^o & \\ \text{s.t.} & c^\top Y \widehat{\xi}_i + \gamma_i^{o\top}(h - H\widehat{\xi}_i) \leqslant s_i^o & \forall i \leqslant N \\ & \|H^\top \gamma_i^o - Y^\top c\|_* \leqslant \lambda^o & \forall i \leqslant N \\ & \gamma_i^o \in \mathbb{R}^{2W}_+ & \forall i \leqslant N \\ & \lambda^o \in \mathbb{R}_+,\ s^o \in \mathbb{R}^N, & \end{array} \right.
\end{aligned}$$

where $\|\cdot\|_*$ stands for the dual norm of $\|\cdot\|$.

B. Combined Bonferroni and CVaR Approximation

If the joint chance constraint involves K linear inequalities, we can decompose the matrix $A(Y)$ and the vector $b(x)$ as

$$A(Y) = \left[a_1(Y) \cdots a_K(Y)\right]^\top, \quad b(x) = \left[b_1(x) \cdots b_K(x)\right]^\top.$$

The joint chance constraint is thus equivalent to

$$\min_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left[a_k(Y)^\top \xi \leqslant b_k(x)\, \forall k \leqslant K\right] \geqslant 1-\epsilon. \tag{5}$$

Given a set of individual violation tolerances $\epsilon_k \geqslant 0$, $k \leqslant K$, with $\sum_{k=1}^K \epsilon_k = \epsilon$, one can exploit Bonferroni's inequality to split the original joint chance constraint up into a family of K simpler but more conservative individual chance constraints. This amounts to approximating the feasible set Ω_{CC} by

$$\Omega_{\mathrm{B}} \triangleq \left\{ (x,Y) : \min_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left[a_k(Y)^\top \xi \leqslant b_k(x)\right] \geqslant 1-\epsilon_k \forall k \leqslant K \right\}.$$

Bonferroni's inequality implies that $\Omega_{\mathrm{B}} \subseteq \Omega_{\mathrm{CC}}$, see [5, § 6.1]. Optimizing over Ω_{B} remains hard even for $\rho = 0$, which prompts us to approximate the individual worst-case chance constraints by worst-case CVaR constraints. Thus, Ω_{B} is conservatively approximated by

$$\begin{aligned}
\Omega_{\mathrm{BC}} \triangleq \Big\{ (x,Y) :\ & \\
& \max_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-}\,\mathrm{CVaR}_{\epsilon_k}\big[a_k(Y)^\top \xi - b_k(x)\big] \leqslant 0\, \forall k \leqslant K \Big\}.
\end{aligned}$$

One can show that Ω_{BC} constitutes the best convex inner approximation of Ω_{B} in a sense made precise in [21], and thus $\Omega_{\mathrm{BC}} \subseteq \Omega_{\mathrm{B}}$. Moreover, we have $\Omega_{\mathrm{BC}} = \Omega_{\mathrm{B}}$ if $\epsilon_k \leqslant N^{-1}$ for all $k \leqslant K$ [24, Corollary 2]. The following proposition further guarantees that optimizing over Ω_{BC} is easy.

Proposition 1. *The set Ω_{BC} admits the conic reformulation*

$$\Omega_{\mathrm{BC}} \triangleq \left\{ \begin{array}{ll} (x,Y) \in \mathbb{R}^{3G} \times \mathbb{R}^{G\times W} : & \\ \lambda_k \rho + N^{-1}\sum_{i=1}^N s_{ik} \leqslant 0 & \forall k \leqslant K \\ \tau_k \leqslant s_{ik} & \forall i \leqslant N, k \leqslant K \\ a_k(Y)^\top \widehat{\xi}_i - b_k(x) + (\epsilon_k - 1)\tau_k & \\ \quad + \epsilon_k \gamma_{ik}^\top (h - H\widehat{\xi}_i) \leqslant \epsilon_k s_{ik} & \forall i \leqslant N, k \leqslant K \\ \|\epsilon_k H^\top \gamma_{ik} - a_k(Y)\|_* \leqslant \epsilon_k \lambda_k & \forall i \leqslant N, k \leqslant K \\ \gamma_{ik} \in \mathbb{R}^{2W}_+ & \forall i \leqslant N, k \leqslant K \\ \tau \in \mathbb{R}^K,\ \lambda \in \mathbb{R}^K,\ s \in \mathbb{R}^{N\times K} & \end{array} \right\}.$$

C. Optimized CVaR Approximation

The Bonferroni approximation is inadequate when the sets of violating wind power scenarios for different individual chance constraints in Ω_{B} have significant overlap [34]. In this case, one may convert the original (linear) joint chance constraint to an equivalent (nonlinear) individual chance constraint before deploying the CVaR approximation [22], [34]. To do so, denote by $\Delta_{++} \triangleq \{\delta \in \mathbb{R}^K_{++} : e^\top \delta = 1\}$ the relative interior of the probability simplex, and note that (5) is equivalent to

$$\min_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\left[\max_{k\leqslant K}\left\{\delta_k\left[a_k(Y)^\top \xi - b_k(x)\right]\right\} \leqslant 0\right] \geqslant 1-\epsilon \tag{6}$$

for any fixed $\delta \in \Delta_{++}$. Note that the overall scale of δ is immaterial, and thus the normalization $e^\top\delta = 1$ does not restrict generality. Note also that (6) constitutes a distributionally robust individual chance constraint, which is immediately susceptible to the CVaR approximation. To see this, denote by

$$\mathcal{R}_\delta(x,Y) \triangleq \max_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-}\,\mathrm{CVaR}_\epsilon\left[\max_{k\leqslant K}\left\{\delta_k\left[a_k(Y)^\top \xi - b_k(x)\right]\right\}\right]$$

the worst-case CVaR function, and define

$$\Omega_{\mathrm{C}}(\delta) \triangleq \left\{(x,Y) : \mathcal{R}_\delta(x,Y) \leqslant 0\right\}.$$

As in Section IV-B, one can show that $\Omega_{\mathrm{C}}(\delta) \subseteq \Omega_{\mathrm{CC}}$ for every $\delta \in \Delta_{++}$ [21]. We emphasize that $\Omega_{\mathrm{C}}(\delta)$ depends nontrivially on δ even though the worst-case probability on the left hand side of (6) is manifestly constant in δ. Hence, δ constitutes a vector of scaling parameters that can be tuned to optimize the

quality of the CVaR approximation. The following proposition further guarantees that $\mathcal{R}_\delta(x,Y)$ can be evaluated efficiently, which implies that optimizing over $\Omega_{\mathrm{C}}(\delta)$ is easy.

Proposition 2. *For any fixed $(x,Y) \in \mathbb{R}^{3G} \times \mathbb{R}^{G\times W}$ and $\delta \in \Delta_{++}$, the worst-case CVaR $\mathcal{R}_\delta(x,Y)$ coincides with the optimal value of the conic program*

$$\begin{array}{lll} \inf & \lambda\rho + \frac{1}{N}\sum_{i=1}^N s_i & \\ \text{s.t.} & \tau \in \mathbb{R},\ \lambda \in \mathbb{R},\ s \in \mathbb{R}^N,\ \gamma_{ik} \in \mathbb{R}^{2W}_+ & \forall i \leqslant N, k \leqslant K \\ & \tau \leqslant s_i & \forall i \leqslant N \\ & \delta_k \left[a_k(Y)^\top \widehat{\xi}_i - b_k(x) \right] + (\epsilon - 1)\tau & \\ & \qquad + \epsilon \gamma_{ik}^\top (h - H\widehat{\xi}_i) \leqslant \epsilon s_i & \forall i \leqslant N, k \leqslant K \\ & \| \epsilon H^\top \gamma_{ik} - \delta_k a_k(Y) \|_* \leqslant \epsilon\lambda & \forall i \leqslant N, k \leqslant K. \end{array}$$

Returning to problem (4), we denote the number of inequalities in the j-th chance constraint by K^j and define Δ^j_{++} as the relative interior of the K^j-dimensional probability simplex. Moreover, for any $\delta^j \in \Delta^j_{++}$, we denote by $\mathcal{R}^j_{\delta^j}(x,Y)$ the worst-case CVaR function corresponding to the j-th chance constraint, $j \in \mathcal{J}$. The previous discussion implies that

$$\min_{(x,Y)\in\Theta} c_x^\top x + \max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[c^\top Y \xi] \tag{7a}$$

$$\text{s.t. } \mathcal{R}^j_{\delta^j}(x,Y) \leqslant 0 \quad \forall j \in \mathcal{J} \tag{7b}$$

constitutes a tractable conic program and provides an upper bound on (4) for every *fixed* set of scaling parameters. In principle, the *best* upper bound can be found by minimizing (7) over all $\delta^j \in \Delta^j_{++}$, $j \in \mathcal{J}$. We emphasize that this best upper bound generally exceeds the optimal value of (4); see [24, § 3]. Moreover, unfortunately, the variant of problem (7) that treats the δ^j as additional decision variables is nonconvex, thus resisting efficient solution. This motivates us to devise an iterative algorithm that optimizes sequentially over (x,Y) and δ^j, $j \in \mathcal{J}$, which is inspired by [22]. In the following we denote by $\eta > 0$ the minimum relative improvement per iteration and by $\bar{t} \in \mathbb{N}$ the maximum iteration count.

0) **Initialization.** Set $g^0 \leftarrow +\infty$, $t \leftarrow 1$, $\delta^j_t \leftarrow e/K^j$ $\forall j \in \mathcal{J}$.
1) **Step 1.** Find a solution (x_t, Y_t, v_t) of

$$g_t = \left\{ \begin{array}{ll} \min\limits_{(x,Y)\in\Theta, v\geqslant 0} & c_x^\top x + \max\limits_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[c^\top Y\xi] + \mathrm{M} e^\top v \\ \text{s.t.} & \mathcal{R}^j_{\delta^j_t}(x,Y) \leqslant v^j \quad \forall j \in \mathcal{J}, \end{array} \right.$$

which is a tractable conic program thanks to the results of Sections IV-A and IV-C. If $|(g_t - g_{t-1})/g_t| < \eta$ or $t \geqslant \bar{t}$, then stop and report (x_t, Y_t, v_t), else go to Step 2.
2) **Step 2.** Find a solution δ_{t+1} of

$$\min_\delta \Big\{ \sum_j \mathcal{R}^j_{\delta^j}(x_t, Y_t) : \delta^j \in \Delta^j_{++} \ \forall j \in \mathcal{J} \Big\},$$

which is a tractable conic program by virtue of Proposition 2. Set $t \leftarrow t+1$, and return to Step 1.

The sequence $\{g_t\}_{t\in\mathbb{N}}$ of objective values generated by the algorithm is non-increasing and thus guaranteed to converge.

The auxiliary slack variables $v \geqslant 0$ in the optimization problem of Step 1 are penalized with a big-M constant in the objective. They ensure feasibility in case of poor initialization of the scaling parameters. If M is chosen sufficiently large, then the algorithm is guaranteed to terminate with $v = 0$ and thus outputs a decision (x_t, Y_t) that is feasible in (7).

D. Discussion

The results of this section give rise to two tractable conservative approximations for the chance constrained program (4). Under the joint Bonferroni and CVaR approximation, the worst-case expectation in the objective function (4a) is replaced by the conic program derived in Section IV-A, while each joint chance constraint is conservatively approximated by its corresponding Bonferroni feasible set Ω_{BC}, which admits a conic representation by virtue of Proposition 1. This results in a single tractable conic program that can be solved efficiently with off-the-shelf software. Under the optimized CVaR approximation, on the other hand, a feasible (and hopefully near-optimal) solution to (4) is found by the efficient sequential convex optimization algorithm from Section IV-C. We emphasize that all conic programs underlying the two approaches reduce to simple linear programs if the Wasserstein metric is defined in terms of the 1-norm or the ∞-norm.

The sequential convex optimization algorithm underlying the optimized CVaR approximation enjoys several benefits. First, it bypasses the necessity to solve a nonconvex optimization problem with bilinear terms, which emerge in the exact reformulations of joint chance constraints derived in [13, § 5.1]. Moreover, in contrast to the approaches proposed in [18], [24], it remains applicable even when there are $K > 2$ inequalities in the chance constraint and when these inequalities involve products of decision variables and uncertain parameters.

In retrospect, we conclude that the optimized CVaR approximation from Section IV-C is superior to the joint Bonferroni and CVaR approximation from Section IV-B because the feasible set $\Omega_{\mathrm{C}}(\delta)$ involves fewer constraints and auxiliary variables than Ω_{BC}. Indeed, Ω_{BC} introduces $2K + (N \times K) + (N \times K \times 2W)$ auxiliary variables and $K + 4(N \times K)$ constraints, while $\Omega_{\mathrm{C}}(\delta)$ creates only $2 + N + (N \times K \times 2W)$ auxiliary variables and $1 + N + 3(N \times K)$ constraints. The relative advantage of $\Omega_{\mathrm{C}}(\delta)$ over Ω_{BC} in terms of complexity of representation becomes increasingly significant for higher dimensions N, K and W. We emphasize that the parsimonious representation of $\Omega_{\mathrm{C}}(\delta)$ comes at the expense of solving a sequence of conic programs. However, as we will demonstrate through numerical experiments in Section V, the sequential convex optimization algorithm usually terminates after only a few iterations and outputs superior decisions.

We highlight that the choice of the individual violation tolerances ϵ_k critically affects the performance of the joint Bonferroni and CVaR approximation. Unfortunately, however, finding the optimal values of ϵ_k is hard [21, Remark 2.1], and optimizing separately over the ϵ_k and the decision variables as in the algorithm of Section IV-C is also impractical because of bilinear terms. In the numerical experiments we thus set $\epsilon_k = \epsilon/K$ for all $k \leqslant K$ as recommended in [21] even though this choice is known to be conservative when ϵ is small or when the inequalities in the joint chance constraint are positively correlated [22, Example 3.1].

V. Numerical Results and Concluding Remarks

We assess the quality of the two approximations described in Section IV on an extended variant of the IEEE 24-bus Reliability Test System [35]. The original system accommodates 24 buses, 34 transmission lines, 12 generators (6 of which are

GFPPs) and 17 demand centers. For further details see [35]. We augment this system with 6 wind farms connected to buses 1, 2, 11, 12, 12 and 16, respectively, all of which have a capacity of 250 MW. We also add 3 gas pipelines that serve a pair of GFPPs each. The pipelines serving the pairs $\{1,2\}$, $\{3,4\}$ and $\{5,6\}$ have capacity 10,000 kcf, 5,500 kcf and 7,000 kcf, respectively. Under this parameterization, the total installed wind capacity adds up to 55% of the system demand. We assume that each generator can use at most 40% of its capacity for reserve provision (i.e., $\overline{r} = 0.4\overline{y}$) at a cost equal to 20% of the variable production cost (i.e., $c_+ = c_- = 0.2c$).

Throughout the experiments we work with the Wasserstein metric induced by the 1-norm on $\mathbb{R}^W$, and thus all arising optimization problems are equivalent to tractable linear programs. Moreover, we set $H = 0$ and $h = 0$, which amounts to approximating Ξ by $\mathbb{R}^W$. We have observed that this approximation greatly accelerates the computations but has no significant impact on the results for the Wasserstein radii of interest. Finally, we set $\epsilon^j = 5\%$ for all $j \in \mathcal{J}$.

Recall that the distribution $\mathbb{P}$ of the uncertain deviation ξ from the relative wind power output forecast μ is only indirectly observable through independent and identically distributed (i.i.d.) samples from $\mathbb{P}$. In our experiments we construct μ and synthetic samples from $\mathbb{P}$ as in [36] using relative wind power output data for 6 wind farms in southeastern Australia from 2012 to 2013.

We first pass the raw (percentage) data through the inverse logistic function to obtain data on the real line and compute the corresponding sample mean $\hat{\mu}$ and sample covariance matrix $\hat{\Sigma}$. We then construct provisional datapoints in $[0,1]$ by applying the logistic function componentwise to $N+M$ i.i.d. samples from the normal distribution $\mathcal{N}(\hat{\mu}, \hat{\Sigma})$. Finally, we set the predicted wind power output μ to the sample average of the first N provisional datapoints, and we construct N training samples and M test samples by subtracting μ from the provisional datapoints. Note that the resulting training samples are asymptotically independent for large N, while the test samples are independent conditional on the training samples.

The full input data of all numerical experiments is disclosed in the e-companion [33]. All simulations are run on a 4 core 3.4 GHz desktop computer running Windows 8. All optimization problems are implemented in MATLAB using the YALMIP interface [37] and solved via Gurobi 8.0.

A. Operation without Reoptimization

In the first experiment we assess the candidate solutions of (4) obtained with the methods from Section IV, assuming that the system operator implements the linear decision rules without reoptimizing. To this end, we generate $N \in \{50, 200\}$ training samples $\hat{\xi}_i$, $i \leqslant N$, and $M = 10^3$ test samples $\hat{\xi}_{N+i}$, $i \leqslant M$, with the procedure described above. Using the training data, we then solve (4) both with the combined Bonferroni and CVaR approximation as well as the optimized CVaR approximation for $\epsilon = 5\%$ and for different Wasserstein radii $\rho \in \mathbb{R}_+$. The quality of an optimal solution $\hat{x}(\rho)$ and $\hat{Y}(\rho)$ (which constitutes an implicit function of the training samples) is assessed by its empirical out-of-sample cost

$$\widehat{\mathcal{C}}(\rho) \triangleq c_x^\top \hat{x}(\rho) + \tfrac{1}{M}\sum_{i=1}^M c^\top \hat{Y}(\rho)\hat{\xi}_{N+i}$$

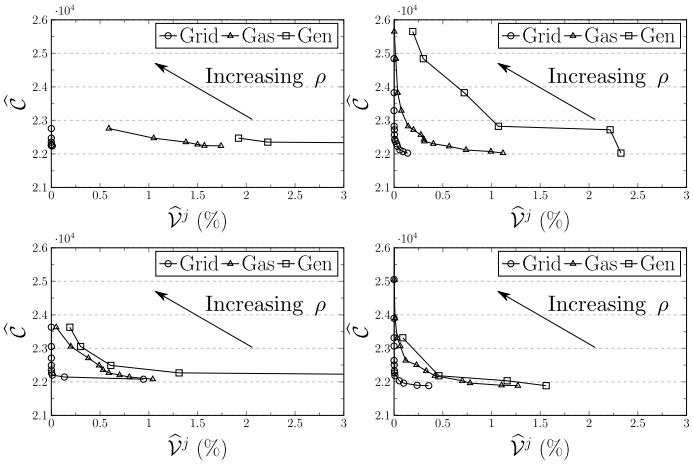

Fig. 1. Pareto frontiers of the out-of-sample costs versus the out-of-sample violation probabilities for the joint Bonferroni and CVaR (left) and optimized CVaR (right) approximations with $N = 50$ (top) and $N = 200$ (bottom).

and its empirical out-of-sample violation probabilities

$$\widehat{\mathcal{V}}^j(\rho) \triangleq \tfrac{1}{M}\sum_{i=1}^M \mathbb{I}_{A(\hat{Y}(\rho))\hat{\xi}_{N+i} > b(\hat{x}(\rho))}, \quad j \in \mathcal{J}.$$

Finally, all results are averaged across 100 independent simulation runs in order to increase their statistical robustness.

Figure 1 visualizes the trade-off between the out-of-sample costs and violation probabilities under the two approximations. Note that with increasing Wasserstein radius ρ the costs increase, while all three violation probabilities decrease. This is expected as larger Wasserstein radii result in more conservative solutions that dispatch more expensive generators. The curves in Figure 1 can thus be interpreted as Pareto frontiers. Note that for sufficiently large values of ρ, both approximations can guarantee that all chance constraints are satisfied out of sample (i.e., the empirical violation probabilities are smaller than $\epsilon = 5\%$). We also observe that, as N increases, the violation probabilities for fixed ρ tend to decrease. Finally, for the majority of all Wasserstein radii that result in sufficiently small out-of-sample violation probabilities $\leqslant 5\%$, the optimized CVaR approximation generates lower out-of-sample costs than the combined Bonferroni and CVaR approximation.

B. Operation with Reoptimization

In the second experiment we assess the candidate solutions $(\hat{x}, \hat{Y})$ of (4) under the premise that only the first-stage decision $\hat{x} = (\hat{y}_1, \hat{r}_+, \hat{r}_-)$ is implemented and that the real-time adjustments $y_2(\xi)$ to the power production are determined by solving a deterministic optimal power flow problem

$$\begin{array}{ll} \min\limits_{y_2,r,l,w} & c^\top y_2 + c_r e^\top r + c_l e^\top l + c_w e^\top w \\ \text{s.t.} & 0 \leqslant \hat{y}_1 + y_2 \leqslant \overline{y}, \;\; -(\hat{r}_- + r) \leqslant y_2 \leqslant \hat{r}_+ \\ & e^\top y_2 + e^\top(C\xi - w) + e^\top l = 0 \\ & -\overline{f} \leqslant Q^{\mathrm{g}}(\hat{y}_1 + y_2) \\ & \qquad + Q^{\mathrm{w}}(C\mu + C\xi - w) - Q^{\mathrm{d}}(d - l) \leqslant \overline{f} \\ & 0 \leqslant \Phi(\hat{y}_1 + y_2) \leqslant \overline{q} \\ & 0 \leqslant r \leqslant \hat{y}_1 - \hat{r}_-, \; 0 \leqslant l \leqslant d, \; 0 \leqslant w \leqslant C(\mu + \xi), \end{array} \tag{8}$$

which optimizes over the power adjustments $y_2 \in \mathbb{R}^G$ and the reserve increments $r \in \mathbb{R}^G_+$ of the conventional generators, the load shedding quantities $l \in \mathbb{R}^D_+$ at the different demand

centers and the wind spills $w \in \mathbb{R}^W_+$ at the different wind farms. We impose penalties $c_w = 0$, $c_l = 1{,}000$ and $c_r = 2{,}000$ to prioritize wind spilling over load shedding and reserve adjustments. The constraints in the last line of (8) ensure that the reserve increments are compatible with the generator capacities and that the load shedding and wind spilling quantities do not exceed the actual demands and wind power output realizations, respectively. All other constraints have natural counterparts in (1), and thus their meaning is evident. Note that in contrast to the day-ahead scheduling problem (1), the real-time optimal power flow problem (8) enforces the capacity constraints for the reserves, the transmission lines and the gas pipelines deterministically. This is possible because (8) is always feasible irrespective of $\widehat{x}$ and ξ. A trivial feasible solution is obtained, for instance, by setting $y_2 = \widehat{y}_1$, $r = \widehat{y}_1 - \widehat{r}_-$, $l = d$ and $w = C(\mu + \xi)$.

The second experiment relies on $N \in \{25, 50, 100, 200\}$ training samples $\widehat{\xi}_i$, $i \leqslant N$, and $M = 10^3$ test samples $\widehat{\xi}_{N+i}$, $i \leqslant M$, generated with the same procedure as before. Using the training data, we solve the day-ahead scheduling problem (4) both with the combined Bonferroni and CVaR approximation as well as the optimized CVaR approximation for $\epsilon = 5\%$ and for different Wasserstein radii $\rho \in \mathbb{R}_+$. The solution $(\widehat{x}(\rho), \widehat{Y}(\rho))$ of (4) then serves as an input for the real-time dispatch problem (8), whose solution $(\widehat{y}_2(\rho,\xi), \widehat{r}(\rho,\xi), \widehat{l}(\rho,\xi), \widehat{w}(\rho,\xi))$ inherits the dependence on ρ. The quality of these solutions (all of which depend on the training data) is measured by the empirical out-of-sample cost

$$\widehat{\mathcal{C}}(\rho) \triangleq c_x^\top \widehat{x}(\rho) + \tfrac{1}{M}\sum_{i=1}^M c^\top y_2(\rho, \widehat{\xi}_{N+i}) + c_r e^\top r(\rho, \widehat{\xi}_{N+i}) + \tfrac{1}{M}\sum_{i=1}^M c_l e^\top l(\rho, \widehat{\xi}_{N+i}) + c_w e^\top w(\rho, \widehat{\xi}_{N+i}).$$

Figure 2 illustrates $\widehat{\mathcal{C}}(\rho)$ as a function of ρ, computed using the optimized CVaR approximation and averaged across 100 simulation runs. A closer inspection of the results for each simulation run reveals that $\widehat{\mathcal{C}}(\rho)$ attains a distinct minimum at a critical Wasserstein radius $\widehat{\rho}^\star > 0$ for all tested sample sizes N. This shows that a naïve operator who ignores ambiguity (by setting $\rho = 0$) faces higher out-of-sample costs than a more sophisticated operator who acknowledges the presence of ambiguity (by setting $\rho = \widehat{\rho}^\star$). The benefits of an ambiguity-aware model are more substantial for lower sample sizes N. For example, accounting for ambiguity can reduce the out-of-sample costs by 1.4% when there are $N = 200$ training samples, while a reduction of up to 6.3% is possible when there are only $N = 25$ training samples. Figure 3 visualizes the interquantile range between the empirical 10% and 90% quantiles of $\widehat{\mathcal{C}}(\rho)$ with respect to 100 simulation runs. In analogy to Figure 2, we observe that the variability of the out-of-sample cost $\widehat{\mathcal{C}}(\rho)$ displays a sharp minimum at a strictly positive Wasserstein radius. Thus, accounting for ambiguity has the dual benefit of reducing both the expectation and (to a larger extent) the variability of the out-of-sample costs. In both cases, the potential benefits are more significant when there are fewer training samples. Qualitatively similar results are obtained under the joint Bonferroni and CVaR approximation. Further details are provided in the e-companion [33].

By solving the chance constrained program (4) we ignore a fraction ϵ of the scenarios in which the constraints are most

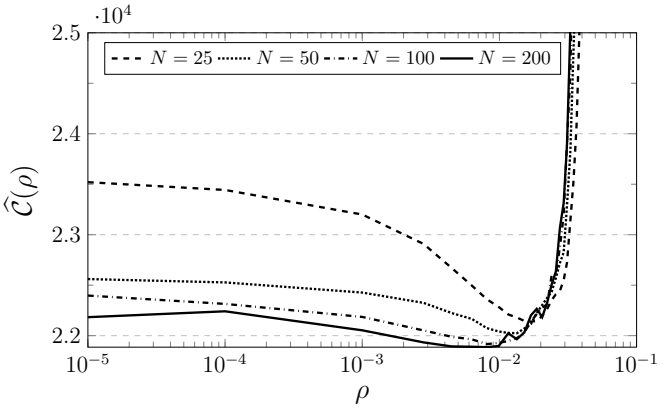

Fig. 2. Average out-of-sample cost $\widehat{\mathcal{C}}(\rho)$ using the optimized CVaR approximation and reoptimization.

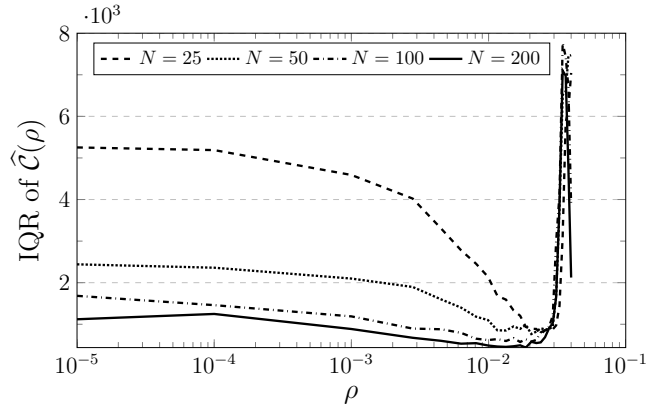

Fig. 3. Interquantile range between the empirical 10% and 90% quantiles of $\widehat{\mathcal{C}}(\rho)$ using the optimized CVaR approximation and reoptimization.

difficult to satisfy. Once the first-stage decisions of (4) have been implemented and the uncertainty ξ has been revealed, the real-time dispatch problem (8) determines a feasible recourse action (which may involve drastic measures such as load shedding etc. that are not modeled in (4)). Instead of this indirect reoptimization approach, one could alternatively require all constraints in (4) to hold with probability 1 while directly allowing for wind spilling, load shedding and reserve adjustments. This amounts to solving a combination of models (4) and (8), which we will henceforth refer to as the *collective optimization model* (see the e-companion [33] for details).

If $\rho > 0$, then the constraints of the collective model must hold $\mathbb{P}$-almost surely for each $\mathbb{P} \in \mathcal{P}$, that is, they must hold for all $\xi \in \Xi$. In this case we solve the problem in linear decision rules $y_2(\xi) = Y\xi$ and use the results of Section IV-A to reformulate the objective function. If $\rho = 0$, on the other hand, then the constraints of the collective model must only hold for all training samples $\widehat{\xi}_i$, $i \leqslant N$. In this case we solve the problem exactly by assigning a separate real-time adjustment decision to each training sample and replacing the expectation in the objective function with the sample average.

For any fixed $\rho \geqslant 0$, the out-of-sample cost $\widehat{C}(\rho)$ of the collective model is computed as before via reoptimization. Table I reports the average as well as the interquantile range between the 10% and 90% quantiles of the empirical out-of-sample cost $\widehat{\mathcal{C}}(\rho)$ with respect to 100 independent simulation runs for the collective model (for $\rho = 0$ and $\rho = \widehat{\rho}^\star$) as well as the chance constrained model with the combined Bonferroni and CVaR

TABLE I
AVERAGE AND INTERQUANTILE RANGE BETWEEN THE 10% AND 90% QUANTILES (IN BRACKETS) OF THE EMPIRICAL OUT-OF-SAMPLE COST OBTAINED WITH THE COLLECTIVE AND CHANCE CONSTRAINED MODELS. THE LOWEST MEANS AND INTERQUANTILE RANGES FOR EACH TRAINING SAMPLE SIZE ARE HIGHLIGHTED IN BOLD.

	$N = 25$	$N = 50$	$N = 100$	$N = 200$
Collective model with $\rho = 0$	24,081 (4,566)	22,718 (2,756)	22,313 (1,003)	22,176 (743)
Collective model with $\rho = \widehat{\rho}^\star$	24,532 (**297**)	24,540 (**301**)	24,531 (**268**)	24,522 (**267**)
Bonferroni & CVaR with $\rho = \widehat{\rho}^\star$	22,275 (2,083)	22,042 (1,206)	21,927 (589)	21,888 (522)
Optimized CVaR with $\rho = \widehat{\rho}^\star$	**22,139** (1,204)	**22,024** (850)	**21,916** (653)	**21,886** (545)

approximation and with the optimized CVaR approximation (both for $\rho = \widehat{\rho}^\star$). Recall that $\widehat{\rho}^\star \in \arg\min_{\rho \geqslant 0} \widehat{C}(\rho)$. Maybe surprisingly, we observe that the out-of-sample cost is smallest for the chance constrained programming model (4) with optimized CVaR approximation even though wind spillage, load shedding and reserve increments, all of which impact the out-of-sample cost, are disregarded in (4). In fact, the chance constraints in (4) only ensure that the necessity of resorting to such drastic measures arises with a probability of at most ε. Thus, (4) provides better first-stage decisions than the (seemingly more realistic) collective model. While the collective model with $\rho = 0$ disregards distributional uncertainty and thus leads to the largest variability in the out-of-sample cost, the collective model with $\rho = \widehat{\rho}^\star$ achieves a lower cost variability than both chance constrained models.

For $N = 200$, problem (4) is solved in 20 seconds on average under the combined Bonferroni and CVaR approximation. The algorithm for computing the optimized CVaR approximation from Section IV-C converges in 38 seconds on average and typically requires less than 3 iterations.

REFERENCES

[1] L. Söder *et al.*, "Experience and challenges with short-term balancing in European systems with large share of wind power," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 853–861, 2012.

[2] U.S. Energy Information Administration, "International energy outlook 2016," Tech. Rep., 2016.

[3] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.

[4] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, 2009.

[5] A. Prékopa, *Stochastic Programming*. Springer, 1995.

[6] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, 2013.

[7] M. Zugno and A. J. Conejo, "A robust optimization approach to energy and reserve dispatch in electricity markets," *Eur. J. Oper. Res.*, vol. 247, no. 2, pp. 659–671, 2015.

[8] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2197–2206, 2011.

[9] J. M. Morales, A. J. Conejo, K. Liu, and J. Zhong, "Pricing electricity in pools with wind producers," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1366–1376, 2012.

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

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

[12] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson, "Analytical reformulation of security constrained optimal power flow with probabilistic constraints," in *2013 IEEE PowerTech Conference*, June 2013, pp. 1–6.

[13] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 171, no. 1, pp. 115–166, 2018.

[14] 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.

[15] F. Alismail, P. Xiong, and C. Singh, "Optimal wind farm allocation in multi-area power systems using distributionally robust optimization approach," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 536–544, 2018.

[16] W. Wei, J. Wang, and S. Mei, "Dispatchability maximization for co-optimized energy and reserve dispatch with explicit reliability guarantee," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3276–3288, 2016.

[17] P. Xiong, P. Jirutitijaroen, and C. Singh, "A distributionally robust optimization model for unit commitment considering uncertain wind power generation," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 39–49, 2017.

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

[19] 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 Syst.*, vol. 32, no. 2, pp. 1378–1388, 2017.

[20] K. Baker, E. Dall'Anese, and T. Summers, "Distribution-agnostic stochastic optimal power flow for distribution grids," in *2016 North Am. Power Symp. (NAPS)*, 2016, pp. 1–6.

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

[22] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Math. Program.*, vol. 137, no. 1-2, pp. 167–198, 2013.

[23] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, "Ambiguous joint chance constraints under mean and dispersion information," *Oper. Res.*, vol. 65, no. 3, pp. 751–767, 2017.

[24] Z. Chen, D. Kuhn, and W. Wiesemann, "Data-driven chance constrained programs over Wasserstein balls," *Optimization Online*, 2018.

[25] W. Xie, "On distributionally robust chance constrained program with Wasserstein distance," *Optimization Online*, 2018.

[26] R. Ji and M. Lejeune, "Data-driven distributionally robust chance-constrained optimization with Wasserstein metric," *Optimization Online*, 2018.

[27] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4924–4936, 2018.

[28] R. D. Christie, B. F. Wollenberg, and I. Wangensteen, "Transmission management in the deregulated environment," *Proc. of the IEEE*, vol. 88, no. 2, pp. 170–195, 2000.

[29] B. Zhao, A. J. Conejo, and R. Sioshansi, "Unit commitment under gas-supply uncertainty and gas-price variability," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2394–2405, 2016.

[30] D. Kuhn, W. Wiesemann, and A. Georghiou, "Primal and dual linear decision rules in stochastic and robust optimization," *Math. Program.*, vol. 130, no. 1, pp. 177–209, 2011.

[31] M. Dyer and A. Frieze, "On the complexity of computing the volume of a polyhedron," *SIAM J. Comput.*, vol. 17, no. 5, pp. 967–974, 1988.

[32] J. Luedtke, S. Ahmed, and G. L. Nemhauser, "An integer programming approach for linear programs with probabilistic constraints," *Math. Program.*, vol. 122, no. 2, pp. 247–272, 2010.

[33] C. Ordoudis, V. A. Nguyen, D. Kuhn, and P. Pinson, "Electronic companion - Energy and reserve dispatch with distributionally robust joint chance constraints," https://github.com/nvietanh/DR_JCC, 2018.

[34] W. Chen, M. Sim, J. Sun, and C.-P. Teo, "From CVaR to uncertainty set: Implications in joint chance-constrained optimization," *Oper. Res.*, vol. 58, no. 2, pp. 470–485, 2010.

[35] C. Grigg *et al.*, "The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, 1999.

[36] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 763–770, 2016.

[37] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *2004 IEEE Int. Conf. Robot. Autom.*, 2004, pp. 284–289.