Journal Pre-proof

Distributionally Robust Optimal Power Flow with Contextual Information

Adrián Esteban-Pérez, Juan M. Morales

Please cite this article as: Adrián Esteban-Pérez, Juan M. Morales, Distributionally Robust Optimal Power Flow with Contextual Information, *European Journal of Operational Research* (2022), doi: https://doi.org/10.1016/j.ejor.2022.10.024

Highlights

- A novel chance-constrained OPF model that accounts for contextual information

- The dependence between the wind power power forecast and its error is exploited

- The OPF model is robust to ambiguity in the conditional forecast error distribution

- Compared to others, our method produces reliable OPF solutions that are cheaper

1

Distributionally Robust Optimal Power Flow with Contextual Information

Adrián Esteban-Pérez, Juan M. Morales*

Department of Applied Mathematics, University of Málaga, Málaga, 29071, Spain

Abstract

In this paper, we develop a distributionally robust chance-constrained formulation of the Optimal Power Flow problem (OPF) whereby the system operator can leverage contextual information. For this purpose, we exploit an ambiguity set based on probability trimmings and optimal transport through which the dispatch solution is protected against the incomplete knowledge of the relationship between the OPF uncertainties and the context that is conveyed by a sample of their joint probability distribution. We provide a tractable reformulation of the proposed distributionally robust chance-constrained OPF problem under the popular conditional-value-at-risk approximation. By way of numerical experiments run on a modified IEEE-118 bus network with wind uncertainty, we show how the power system can substantially benefit from taking into account the well-known statistical dependence between the point forecast of wind power outputs and its associated prediction error. Furthermore, the experiments conducted also reveal that the distributional robustness conferred on the OPF solution by our probability-trimmings-based approach is superior to that bestowed by alternative approaches in terms of expected cost and system reliability.

---

*Corresponding author

  Email address: `adrianesteban@uma.es, juan.morales@uma.es` (Adrián Esteban-Pérez, Juan M. Morales)

1. Introduction

The Optimal Power Flow (OPF) is a fundamental problem in power system operations. Traditionally, the goal of the OPF problem is to minimize the cost of the power generation dispatch that supplies the electricity demand while complying with some physical and engineering constraints. The growing penetration of electricity generation sources like wind and solar power, which are intrinsically uncertain, has led power system engineers to account for randomness in OPF analyses. Hence, the OPF is to be formulated today as an optimization problem *under uncertainty*.

A common way to cope with uncertainty in the constraints of an optimization problem and, in particular, of an OPF model is by way of the so-called *chance constraints*, which allow the modeler to impose the constraint satisfaction with a certain probability only. Accordingly, chance-constrained optimal power flow models (CC-OPF) have been developed to control the violation probability of, for instance, line and generation capacity limits. In particular, references [13, 23, 28] consider *joint* chance constraints, by which the system operator enforces that all constraints must simultaneously hold with a probability greater than or equal to $1 - \epsilon$, where $\epsilon \in (0, 1)$ is a pre-fixed acceptable tolerance of dispatch infeasibility. This is in contrast to *single* chance constraints, whereby the probability of constraint satisfaction is imposed on each constraint of the OPF separately. Although single chance constraints have been considered in the literature of CC-OPF models due to their attractive tractability properties (see, e.g., [4, 18, 31] and references therein), they do not provide guarantees on the *joint* satisfaction of the OPF constraints in general or, when they do, they usually lead to costly and over-conservative dispatch solutions [23].

2

One of the main challenges in solving CC-OPF problems is that the underlying probability distribution of the random variables affecting the OPF constraints is generally unknown. In fact, in practice, only past historical observations of those variables are available to the system operator. Within this context, Distributionally Robust Optimization (DRO) has emerged as a versatile and powerful paradigm to tackle the ambiguity of the distribution of an optimization problem's random parameters. DRO, therefore considers a set of potential distributions for those parameters, namely, the so-called *ambiguity set* [24]. In the literature on DRO, there exist two distinct ways to specify an ambiguity set, either using statistical moments or probability metrics. The main drawback of using a *moment-based* DRO approach is that the available historical information is only used to estimate the moments of the true data-generating distribution and thus, if new samples of said distribution become available, but the moment estimates are not modified, the ambiguity set does not change. Alternatively, in *metric-based* DRO, the ambiguity set collects all potential distributions whose distance to a particular nominal distribution is lower than or equal to some pre-fixed value. Although there are different choices of probability metrics to construct the ambiguity set, the Wasserstein distance has received much attention within the power systems community. This is partly due to the performance guarantees provided by this distance (see, e.g., [19]) in contrast to the moment-based DRO approach.

Thus, distributionally robust chance-constrained optimal power flow (DRCC-OPF) models seek the optimal dispatch such that all the model constraints are satisfied with a pre-fixed confidence level for all the probability distributions within the ambiguity set built either by means of the Wasserstein metric ([1, 2, 22, 32]), using moments ([15, 16, 17, 29]), or by way of a discrete probability distribution with probability masses and locations varying within a box [12].

Unfortunately, the consideration of *joint* chance constraints in a DRO framework

([10, 30]) renders an intractable optimization problem in general. For this reason, researchers have considered DRO-OPF models based on the well-known conservative, but far more tractable approximation, given by the concept of *conditional-value-at-risk* (**CVaR**), see, for example, [1, 12, 22] and references therein. The authors in [6] show that the **CVaR** offers a tight convex approximation of the chance constraints under the DRO framework, which justifies its popularity. In addition, by way of **CVaR**-based chance-constraints, the power system operator can control not only the violation probability, but also the violation magnitude, which can be important from the standpoint of power system operations.

Yet another key issue is that the system operator is able to not only tune the robustness of the resulting chance-constrained DRO model by adjusting the probability of constraint violation, but also, and very importantly, through the specificity degree of the ambiguity set. In this regard, some previous approaches have focused on producing more meaningful ambiguity sets by incorporating structural information on the underlying true probability distribution (see, e.g., [1, 2, 15, 16, 17]). More specifically, the authors in [1] introduce a DRCC-OPF model with single chance constraints, which considers all distributions within a Wasserstein ball that conform to a given copula-based dependence structure among the random variables. The authors in [2] propose an iterative algorithm for solving a bilinear exact reformulation of a DRCC-OPF model with single chance constraints considering some support information. The authors of [15] and [16] provide and study DRCC-OPF models where some moment and unimodality information is included in the ambiguity set. Both aforementioned approaches are extended in [17] to allow for misspecified modes.

Ideally, one would like to have the smallest ambiguity set that contains the true data-generating distribution. In this vein, if we have some information on the true distribution, we should use it to discard all those other distributions that do not

conform with that information from the ambiguity set. As mentioned above, that information can be, for example, some dependence structure via copulas [1], support information [2] or shape information (such as unimodality) [16, 17]. Our aim is also to leverage a more informed ambiguity set, but, and for the first time to our knowledge, that information refers to a given *context*. This side/contextual information (also known as *features* or *attributes*) is related to outcomes of random variables that may have predictive power on the OPF's uncertainties. Accordingly, we make use of an ambiguity set that accounts for the possible dependence between the uncertainties and these explanatory variables. Thus, the contextual information allows us to discard implausible distributions.

More specifically, in the work we present here, we exploit the contextual information provided by the *point forecasts* of those uncertainties. In the energy forecasting community,it is a well known fact that the power forecast error of a wind farm highly depends on the wind power forecast itself [5, 9]. Within the context of DRCC-OPF, this means that the wind power point forecast constitutes valuable information to build a proper ambiguity set for the wind power forecast error.

The main contributions of this work are thus:

1. We provide a formulation of the DC-OPF problem with *joint* chance constraints as a *conditional stochastic program*, in which both the expected cost of the power dispatch and the chance constraints are conditioned on some contextual information, in particular, the point forecasts of the OPF's uncertainties. To the best of our knowledge, this work is the first to tackle a chance constraint system with a distributionally robust approach *that accounts for contextual information.*

2. To tackle the resulting conditional stochastic program, we propose an entirely data-driven and non-parametric DRO approach, in which no assumption on the

5

relationship between the context and the OPF's uncertainties (i.e., the point forecasts and their error in our case) is made. Our approach makes use of the ambiguity set based on probability trimmings that is introduced in [8] and which is able to integrate contextual information. We prove that, under the widely used **CVaR**-based approximation of the joint chance constraint system [21], it results in a tractable distributionally robust joint chance-constrained OPF model.

3. Finally, by way of numerical experiments in which we compare our DRCC-OPF formulation with alternative approaches in the literature, we show that exploiting the contextual information provided by the wind power point forecasts allows identifying dispatch solutions with a better trade-off between expected cost and system reliability.

The remainder of this paper is organized as follows. Section 2 introduces the mathematical formulation of the distributionally robust joint chance-constrained DC-OPF problem with contextual information that we propose. Then, Section 3 provides a tractable reformulation of the distributionally robust joint chance constraints under the well-known **CVaR** approximation, while Section 4 focuses on reformulating the worst-case expected cost in a manageable way. Results from numerical experiments are presented and discussed in Section 5. Finally, Section 6 concludes the paper with some final remarks. The manuscript also contains appendices with the main notation used throughout the main text and supportive optimization models. The proofs of theoretical results and additional numerical experiments are available in the supplementary material associated with this article.

Notation. Unless otherwise stated, in this paper, we use boldface lowercase letters to represent arrays and boldface capital letters for matrices. Vector $\mathbf{1}$ ($\mathbf{0}$) is a vector

of ones (zeros) of appropriate dimension, and the inner product of two vectors $\mathbf{u}, \mathbf{v}$ will be denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$. The cardinality of a set A will be indicated by $|A|$. The support function of a set $B \subseteq \mathbb{R}^d$, S_B, is defined as $S_B(a) := \sup_{b \in B} \langle a, b \rangle$. Moreover, we reserve the symbol "$\hat{}$" for objects which are dependent on the sample data and denote "expectation" with the symbol $\mathbb{E}$. In addition, throughout the paper we assume that we always have measurability for those objects whose expected values we consider. The reader is referred to Appendix A for a complete list of the notation used throughout this paper.

2. DC-OPF under uncertainty: Mathematical Formulation

Next we introduce the DC-OPF problem under uncertainty. The problem is formulated as a distributionally robust version of the *joint* chance-constrained DC-OPF model described in [23], where we have also accounted for the procurement of reserve capacity and its associated cost, as in [16, 17]. Nonetheless, unlike in [23], where the generators' cost functions are assumed to be quadratic, here we model those costs as convex piecewise linear functions. Furthermore, there exists a number of different variants of the distributionally robust chance-constrained DC-OPF problem (e.g., [2, 12, 16, 22, 32]), which essentially differ in the treatment of the chance constraints (single vs. joint), the cost structure of generators that is assumed, and the ambiguity set used. What makes our formulation unique among those variants is its ability to exploit contextual information.

2.1. Variables and constraints

Consider a power system with a set $\mathcal{L}$ of transmission lines, a set $\mathcal{B}$ of buses, a set $\mathcal{W}$ of wind power plants (or, more generally, weather-dependent renewable generators),

7

and a set $\mathcal{G}$ of conventional generators (i.e., dispatchable units that are not weather-dependent). For ease of formulation, power loads are assumed to be deterministic. Next we introduce each of the main components of the DC-OPF problem.

1. *Wind power plants.* For each wind power plant $m \in \mathcal{W}$, the random power output is modeled as $f_m + \omega_m$, where f_m is the predicted power output and ω_m is the (random) wind forecast error at wind power plant m. We denote the system-wise aggregate wind power forecast error as Ω, i.e., $\Omega := \sum_{m \in \mathcal{W}} \omega_m = \langle \mathbf{1}, \boldsymbol{\omega} \rangle$. Let $\mathbf{f} := (f_m)_{m \in \mathcal{W}}, \boldsymbol{\omega} := (\omega_m)_{m \in \mathcal{W}}$ be the array of predicted power outputs and wind power prediction errors, respectively.

2. *Generators:* For each $j \in \mathcal{G}$, the actual power output of generator j, $\tilde{g}_j(\boldsymbol{\omega})$, is expressed as the sum of the scheduled generation, g_j, and the (random) adjusted power $\tilde{r}_j(\boldsymbol{\omega})$ (also known as *deployed reserve*). As customary, we assume an affine control policy to counterbalance the wind forecast errors by deploying generators' reserves [4], that is,

$$\tilde{g}_j(\boldsymbol{\omega}) := g_j + \tilde{r}_j(\boldsymbol{\omega}) = g_j - \beta_j \Omega = g_j - \beta_j \langle \mathbf{1}, \boldsymbol{\omega} \rangle, \ \ \forall j \in \mathcal{G} \qquad (1)$$

where β_j is the participation factor of generator j. Denote by $\boldsymbol{\beta} := (\beta_j)_{j \in \mathcal{G}}, \mathbf{g} := (g_j)_{j \in \mathcal{G}}$ the array of non-negative participation factors and scheduled generation, respectively. Let $\tilde{\mathbf{g}}(\boldsymbol{\omega}) := (\tilde{g}_j(\boldsymbol{\omega}))_{j \in \mathcal{G}}, \tilde{\mathbf{r}}(\boldsymbol{\omega}) := (\tilde{r}_j(\boldsymbol{\omega}))_{j \in \mathcal{G}} = (-\beta_j \langle \mathbf{1}, \boldsymbol{\omega} \rangle)_{j \in \mathcal{G}}$ be the array of actual power outputs and deployed reserves, in that order.

The following constraints determine the provision of reserve capacities:

$$-\mathbf{r}^D \leqslant \tilde{\mathbf{r}}(\boldsymbol{\omega}) \leqslant \mathbf{r}^U \qquad (2)$$

with $\mathbf{r}^D, \mathbf{r}^U$ being the arrays of downward and upward reserve capacity provided by the generators, respectively.

Naturally, the following technical constraints, which link the generation dispatches and the provision of reserve capacities, must hold:

$$\mathbf{g} + \mathbf{r}^U \leqslant \mathbf{g}^{\max}, \tag{3}$$

$$\mathbf{g} - \mathbf{r}^D \geqslant \mathbf{g}^{\min} \tag{4}$$

where $\mathbf{g}^{\min}, \mathbf{g}^{\max}$ are the arrays of minimum and maximum power output of the generators, respectively.

3. *Network constraints.* The total power generation must equal the total system demand (*power balance constraint*), that is,

$$\langle \mathbf{1}, \tilde{\mathbf{g}}(\boldsymbol{\omega}) \rangle + \langle \mathbf{1}, \mathbf{f} + \boldsymbol{\omega} \rangle = \langle \mathbf{1}, \mathbf{L} \rangle \tag{5}$$

where $\mathbf{L} := (L_b)_{b \in \mathcal{B}}$ denotes the array of nodal loads. Using (1), Eq. (5) is equivalent to:

$$\langle \mathbf{1}, \mathbf{g} \rangle + \langle \mathbf{1}, \mathbf{f} \rangle = \langle \mathbf{1}, \mathbf{L} \rangle \tag{6}$$

$$\langle \mathbf{1}, \boldsymbol{\beta} \rangle = 1, \ \ \boldsymbol{\beta} \geqslant \mathbf{0} \tag{7}$$

which guarantee the power balance both in the dispatch and the real-time stages, respectively.

Finally, we assume that the power flow through the lines is given by a linear function of the nodal power injections, that is, $\mathbf{M}^{\mathcal{G}}(\tilde{\mathbf{g}}(\boldsymbol{\omega})) + \mathbf{M}^{\mathcal{W}}(\mathbf{f} + \boldsymbol{\omega}) - \mathbf{M}^{\mathcal{B}}\mathbf{L}$, based on the DC power flow approximation, where $\mathbf{M}^{\mathcal{G}}, \mathbf{M}^{\mathcal{W}}$ and $\mathbf{M}^{\mathcal{B}}$ are the matrix for generators, wind plants and loads given by the DC power transfer distribution factors [27], in that order. Hence, the constraints

$$-\mathbf{Cap} \leqslant \mathbf{M}^{\mathcal{G}}(\tilde{\mathbf{g}}(\boldsymbol{\omega})) + \mathbf{M}^{\mathcal{W}}(\mathbf{f} + \boldsymbol{\omega}) - \mathbf{M}^{\mathcal{B}}\mathbf{L} \leqslant \mathbf{Cap} \tag{8}$$

enforce the transmission capacity limits where $\mathbf{Cap} := (\mathrm{Cap}_\ell)_{\ell \in \mathcal{L}}$ denotes the array of transmission line capacities.

2.2. Dealing with uncertainty in the DC-OPF problem

In practice, it is often the case that the random vector of forecast errors $\boldsymbol{\omega}$ shows some statistical dependence on some features/covariates, which we can model, in general, by some random vector $\mathbf{z}$. In fact, the forecast wind power output $\mathbf{f}$ serves in itself as an obvious explanatory random vector for the subsequent forecast error $\boldsymbol{\omega}$. In this approach, we want to exploit this statistical dependence to identify a better power generation dispatch and provision of reserve capacity.

Let $\mathbf{z} := (z_m)_{m \in \mathcal{W}}$ be the random vector modeling the features and let $\mathbb{Q}$ be the probability measure of the joint distribution of $\mathbf{z}$ and $\boldsymbol{\omega}$, which is supported on Ξ. For convenience, we define $\boldsymbol{\xi} := (\mathbf{z}, \boldsymbol{\omega})$. Given the array of forecast wind power outputs, $\mathbf{f} := (f_m)_{m \in \mathcal{W}}$, set the contextual information $\boldsymbol{\xi} := (\mathbf{z}, \boldsymbol{\omega}) \in \widetilde{\Xi}$ defined by the event $(\mathbf{z} = \mathbf{f}; \ \boldsymbol{\omega} \in \widetilde{\Xi}_{\boldsymbol{\omega}})$, with $\widetilde{\Xi}_{\boldsymbol{\omega}}$ being the support of $\boldsymbol{\omega}$ conditional on $\mathbf{z} = \mathbf{f}$. The errors of forecasting the power output of a wind farm are naturally bounded. Their lower bound is the forecast value itself, while their upper bound is given by the difference of the capacity of the wind farm and the predicted value. Therefore, $\widetilde{\Xi}_{\boldsymbol{\omega}}$ is the hypercube $\prod_{m \in \mathcal{W}} [-f_m, \overline{C}_m - f_m]$, where $\overline{C}_m$ represents the capacity of wind farm m. Note that the optimal dispatch is, therefore, parametrized on the predicted wind power outputs $\{f_m\}_{m \in \mathcal{W}}$.

In real life, however, neither the joint distribution $\mathbb{Q}$, nor the conditional one $\mathbb{Q}_{\boldsymbol{\omega}/\mathbf{z}=\mathbf{f}}$, are known. The system operator only has access to a finite set of samples of size N (i.e. the training set) of the true joint distribution $\mathbb{Q}$, which we denote as $\widehat{\Xi}_{\omega}^N := \{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N = \{(\widehat{\mathbf{z}}_i, \widehat{\boldsymbol{\omega}}_i)\}_{i=1}^N$. In our context, $\widehat{\Xi}_{\omega}^N$ is made up of N past observations of the predicted wind power outputs and their associated errors. Hence, the system operator needs to infer or construct a proxy of $\mathbb{Q}_{\boldsymbol{\omega}/\mathbf{z}=\mathbf{f}}$ from the sample $\widehat{\Xi}_{\omega}^N$, so that this proxy can be used to compute a reliable and cost-efficient OPF solution. However, the limited information that $\widehat{\Xi}_{\omega}^N$ conveys on $\mathbb{Q}_{\boldsymbol{\omega}/\mathbf{z}=\mathbf{f}}$ makes this inference process

10

ambiguous, and as such, we propose employing the following *distributionally robust* chance-constrained OPF model to protect the system operator's decision against this ambiguity:

$$\min_{\mathbf{x} \in X} \sup_{Q_{\widetilde{\Xi}} \in \widehat{\mathcal{U}}} \mathbb{E}_{Q_{\widetilde{\Xi}}} \left[C(\tilde{\mathbf{g}}(\boldsymbol{\omega})) + \langle \mathbf{c}^D, \mathbf{r}^D \rangle + \langle \mathbf{c}^U, \mathbf{r}^U \rangle \right] \qquad (9)$$

$$\text{s.t. } \inf_{Q_{\widetilde{\Xi}} \in \widehat{\mathcal{U}}} Q_{\widetilde{\Xi}} \left(\begin{array}{c} -\mathbf{r}^D \leqslant \tilde{\mathbf{r}}(\boldsymbol{\omega}) \leqslant \mathbf{r}^U \\ -\mathbf{Cap} \leqslant \mathbf{M}^{\mathcal{G}}(\tilde{\mathbf{g}}(\boldsymbol{\omega})) + \mathbf{M}^{\mathcal{W}}(\mathbf{f} + \boldsymbol{\omega}) - \mathbf{M}^{\mathcal{B}}\mathbf{L} \leqslant \mathbf{Cap} \end{array} \right) \geqslant 1 - \epsilon \quad (10)$$

where X is the deterministic feasible set for the array of decision variables $\mathbf{x} = (\mathbf{g}, \boldsymbol{\beta}, \mathbf{r}^D, \mathbf{r}^U)$ defined by the constraints (3), (4), (6) and (7).

The set $\widehat{\mathcal{U}}$ in (9)-(10) stands for an ambiguity set for the true conditional distribution $\mathbb{Q}_{\boldsymbol{\omega}/\mathbf{z}=\mathbf{f}}$. Here we will use the ambiguity set based on probability trimmings and optimal transport introduced in [8], which allows us to exploit the side information within a DRO framework in a fully data-driven sense. This ambiguity set is formally introduced in Section 2.3.

Objective function (9) minimizes the expected total operational cost over the worst-case distribution from the ambiguity set $\widehat{\mathcal{U}}$, including the sum of the (random) total generation cost, $C(\tilde{\mathbf{g}}(\boldsymbol{\omega}))$, and the (deterministic) cost of providing up- and down-reserve capacities, $\langle \mathbf{c}^D, \mathbf{r}^D \rangle + \langle \mathbf{c}^U, \mathbf{r}^U \rangle$. The total generation cost function $C(\cdot)$ is assumed to be given by the sum of $|\mathcal{G}|$ convex piecewise linear cost functions with S_j pieces/blocks, i.e., $C(\tilde{\mathbf{g}}(\boldsymbol{\omega})) := \sum_{j \in \mathcal{G}} \max_{s=1,\ldots,S_j} \{m_{js}\tilde{g}_j(\boldsymbol{\omega}) + n_{js}\}$, where m_{js}, n_{js} stand for the slope and the intercept of the s-th piece for generator j, respectively. Parameters $\mathbf{c}^D, \mathbf{c}^U$ are the arrays of downward and upward reserve procurement cost of the generators, respectively. Note that the wind power production cost is assumed to be zero. Finally, constraint (10) establishes a tolerance ϵ in terms of the joint violation probability of the OPF constraints under any conditional probability distribution $Q_{\widetilde{\Xi}}$ for $\boldsymbol{\omega}$ (given some contextual information) within the ambiguity set $\widehat{\mathcal{U}}$, which will be

11

defined right after the following remark.

Remark 1. *Estimating the probability distribution of the wind power forecast error conditional on the point prediction, i.e.,* $\mathbb{Q}_{\boldsymbol{\omega}/\mathbf{z=f}}$ *(a task that is typically referred to as* probabilistic forecasting*) requires* postulating a statistical model for that distribution, *following the scheme* "from data to decision through prediction" *. The "prediction" step in that scheme may negatively affect the decision value in the end due to model misspecification, the estimation error and/or simply because of the so-called "optimizer's curse" or "optimization bias" [19].*

Our approach, in contrast, is fully data driven in the sense that it follows an alternative scheme by which the decision is directly inferred from the data without being forced to use a specific predictive model (which could be the wrong choice). Furthermore, as we explain below, our approach protects the decision against the uncertainty intrinsic to the process of inferring the conditional *forecast error distribution from samples of the* joint *distribution of the pair (point prediction, prediction error). This can be particularly important when the sample size is small and said uncertainty is high as a result.*

2.3. The ambiguity set based on probability trimmings

To formalize the ambiguity set based on probability trimmings that will play the role of $\widehat{\mathcal{U}}$ in (9)-(10), we first need to introduce a series of probability-related concepts. We start with the Wasserstein metric of order 1, a well-known metric for probability distributions with finite first moment closely linked to the optimal transport problem [26]:

Definition 1 (1-Wasserstein metric)**.** *Given two probability measures P, Q with finite first moment supported on Ξ, that is, $P, Q \in \mathcal{P}_1(\Xi)$, the Wasserstein metric of order 1*

12

between P and Q, $W_1\big(P,Q\big)$, is given by the value

$$\inf_{\Pi} \left\{ \int_{\Xi^2} \|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|_1 \, \Pi(\mathrm{d}\boldsymbol{\xi}_1, \mathrm{d}\boldsymbol{\xi}_2) \;\; : \;\; \begin{array}{l} \Pi \text{ is a joint distribution of } \boldsymbol{\xi}_1 \text{ and } \boldsymbol{\xi}_2 \\ \text{with marginals } P \text{ and } Q, \text{ respectively} \end{array} \right\}$$

Next, we introduce the definition of trimmings and trimming sets of an empirical probability measure. Essentially, trimming a probability measure allows to play down the weight of some regions of the measurable space without completely removing them from the support set.

Definition 2 (($1-\alpha$)-empirical trimmings sets)**.** *Consider the sample data* $\{\widehat{\boldsymbol{\xi}}_i\}_{i=1}^N$ *and their associated empirical measure* $\widehat{\mathbb{Q}}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{\boldsymbol{\xi}}_i}$*. If* $\alpha > 0$*, the set of all* ($1-\alpha$)-*trimmings of* $\widehat{\mathbb{Q}}_N$ *is given by all probability distributions in the form* $\sum_{i=1}^N b_i \delta_{\widehat{\boldsymbol{\xi}}_i}$ *such that* $0 \leq b_i \leq \frac{1}{N\alpha}$*,* $\forall i = 1, \ldots, N$*, and* $\sum_{i=1}^N b_i = 1$*.*

For ease of understanding, we provide below an example of a ($1-\alpha$)-empirical trimmings set.

Example 1. *Suppose the empirical joint measure* $\widehat{\mathbb{Q}}_N := \sum_{i=1}^3 \delta_{(\widehat{z}_i, \widehat{\omega}_i)} = \frac{1}{3}(\delta_{(1,0)} + \delta_{(0,5)} + \delta_{(2,3)})$ *($N = 3$). If* $\alpha = 0.5$*, then* $\frac{1}{N\alpha} = \frac{1}{3 \cdot 0.5} = \frac{2}{3}$*. Therefore, the* 0.5-*trimmings set of* $\widehat{\mathbb{Q}}_N$ *is given by*

$$\mathcal{R}_{0.5}(\widehat{\mathbb{Q}}_N) := \left\{ \sum_{i=1}^3 b_i \delta_{\widehat{\boldsymbol{\xi}}_i} : 0 \leq b_i \leq \frac{2}{3}, \forall i = 1, \ldots, 3; \sum_{i=1}^3 b_i = 1 \right\}$$

The following statements hold thus true:

$$\widehat{\mathbb{Q}}_N = \frac{1}{3}(\delta_{(1,0)} + \delta_{(0,5)} + \delta_{(2,3)}) \in \mathcal{R}_{0.5}(\widehat{\mathbb{Q}}_N)$$

$$\mathcal{Q} = \frac{2}{3}\delta_{(1,0)} + \frac{1}{3}\delta_{(0,5)} \in \mathcal{R}_{0.5}(\widehat{\mathbb{Q}}_N)$$

$$\mathcal{P} = \delta_{(1,0)} \notin \mathcal{R}_{0.5}(\widehat{\mathbb{Q}}_N) \text{ (the trimming must retain one point and a half at least)}$$

$$\mathcal{S} = \frac{2}{3}\delta_{(1,0)} + \frac{1}{6}\delta_{(0,5)} + \frac{1}{6}\delta_{(2,3)} \in \mathcal{R}_{0.5}(\widehat{\mathbb{Q}}_N)$$

13

$$\mathcal{V} = \frac{3}{4}\delta_{(1,0)} + \frac{1}{12}\delta_{(0,5)} + \frac{2}{12}\delta_{(2,3)} \notin \mathcal{R}_{0.5}(\widehat{\mathbb{Q}}_N) \ (because \ b_1 > 2/3)$$

Our ambiguity set also relies on the concept of *minimum transportation budget*, which is provided below.

Definition 3 (Minimum transportation budget)**.** *Given $\alpha > 0$, the* minimum transportation budget*, which we denote as $\underline{\rho}_{N\alpha}$, is the 1-Wasserstein distance between the set of probability distributions $\mathcal{P}_1(\widetilde{\Xi})$ and the $(1 - \alpha)$-trimming of the empirical distribution $\widehat{\mathbb{Q}}_N$ that is the* closest *to that set, that is,*

$$\underline{\rho}_{N\alpha} = \frac{1}{N\alpha} \sum_{k=1}^{\lfloor N\alpha \rfloor} dist(\boldsymbol{\xi}_{k:N}, \widetilde{\Xi}) + \left(1 - \frac{\lfloor N\alpha \rfloor}{N\alpha}\right) dist(\boldsymbol{\xi}_{\lceil N\alpha \rceil:N}, \widetilde{\Xi}) \tag{11}$$

where $\boldsymbol{\xi}_{k:N}$ is the k-th nearest data point from the sample to set $\widetilde{\Xi}$ and $dist(\boldsymbol{\xi}_j, \widetilde{\Xi}) := \inf_{\boldsymbol{\xi} \in \widetilde{\Xi}} dist(\boldsymbol{\xi}_j, \boldsymbol{\xi}) = \inf_{\boldsymbol{\xi} \in \widetilde{\Xi}} ||\boldsymbol{\xi}_j - \boldsymbol{\xi}||$.

The *minimum transportation budget* is a minimum threshold to ensure the non-emptiness of the trimmings-based ambiguity set, which is finally defined as follows.

Definition 4 (Ambiguity set based on probability trimmings)**.** *Assume that $\mathbb{Q} \in \mathcal{P}_1(\Xi)$, $\alpha > 0$, and consider the set $\mathcal{R}_{1-\alpha}(\widehat{\mathbb{Q}}_N)$ of $(1 - \alpha)-$trimmings of the empirical joint distribution $\widehat{\mathbb{Q}}_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{(\widehat{\mathbf{z}}_i, \widehat{\boldsymbol{\omega}}_i)}$. If $\rho \geqslant \underline{\rho}_{N\alpha}$, where $\underline{\rho}_{N\alpha}$ is the minimum transportation budget introduced in Definition 3, then the ambiguity set based on trimmings, $\widehat{\mathcal{U}}_N(\alpha, \rho)$, is defined as the set of all distributions $Q_{\widetilde{\Xi}}$ supported on $\widetilde{\Xi}$ such that $W_1(\mathcal{R}_{1-\alpha}(\widehat{\mathbb{Q}}_N), Q_{\widetilde{\Xi}}) \leq \rho$.*

The ambiguity set based on probability trimmings collects all distributions with support set $\widetilde{\Xi}$ that result from a *partial optimal transport* of mass α from $\widehat{\mathbb{Q}}_N$ to $\widetilde{\Xi}$ within a budget ρ. The parameters α and ρ tune the *shape* and the *size* of the ambiguity set, respectively. Indeed, α can be seen as the minimum amount of mass

14

of $\widehat{\mathbb{Q}}_N$ which plays a role in the conditional inference and ρ controls the degree of robustness/conservativeness. Hereinafter, we refer to α as the *trimming level* and to ρ as the *transportation budget* or *robustness parameter*. Further technical information on the ambiguity set $\widehat{\mathcal{U}}_N(\alpha, \rho)$ can be obtained from [8].

Remark 2. *In order to account for contextual information in a DRO framework, the authors in [3] propose an ambiguity set $\widehat{\mathcal{U}}$ different to the one based on probability trimmings (i.e., $\widehat{\mathcal{U}}_N(\alpha, \rho)$) that we advocate here. More specifically, the ambiguity set they suggest is a Wasserstein ball centered at the discrete distribution supported on the $\widehat{\omega}$-coordinates of the K data points from the sample $\widehat{\Xi}_{\omega}^N$ that are the closest to $\widetilde{\Xi}$. In the supplementary material, we use an example based on a small three-node system which has been taken from [20] to illustrate that our trimmings-based ambiguity set generally delivers better dispatch solutions in terms of expected cost and system reliability than the one introduced in [3] for the DC-OPF problem under uncertainty.*

In the next section, we introduce a tractable and conservative approximation of the distributionally robust joint chance constraints (10) using the notion of Conditional Value at Risk (**CVaR**). As previously mentioned, the use of the **CVaR** in the context of chance-constrained distributionally robust OPF is very popular in the literature (see, for example, the recent publications [12] and [22]).

3. A tractable approximation of the distributionally robust joint chance constraints

The distributionally robust joint chance constraints (10) can be written equivalently as the following distributionally robust single chance constraint, where we have already replaced the generic $\widehat{\mathcal{U}}$ with the ambiguity set based on probability trimmings

15

$\widehat{\mathcal{U}}_N(\alpha, \rho)$:

$$\inf_{Q_{\widetilde{\Xi}} \in \widehat{\mathcal{U}}_N(\alpha, \rho)} Q_{\widetilde{\Xi}} \left(\max_{k \leq K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k} \leqslant 0 \right) \geqslant 1 - \epsilon \qquad (12)$$

Functions $\langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k}$, $k \leq K$, represent the OPF constraints involved in the joint chance constraint (10) expressed as inequalities lower than or equal to zero. These constraints are all linear with respect to the random vector $\boldsymbol{\omega}$.

In practice, the system operator is not only concerned about the joint violation of the OPF constraints, but also about the magnitude of this violation. Indeed, the joint chance constraint (12) does not offer guarantees on how positive $\max_{k \leq K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k}$ is. This is one of the main reasons to adopt a risk-averse approach to handle the joint chance constraint via the well-known concept of Conditional-Value-at-Risk (**CVaR**), which quantifies the conditional expectation of $\max_{k < K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k}$ on its right ϵ-tail. More specifically, the **CVaR** at level $\epsilon \in (0, 1)$ of an univariate random variable $\phi(\boldsymbol{\omega})$ under the probability measure Q, $Q - \mathbf{CVaR}_\epsilon(\phi(\boldsymbol{\omega}))$, is defined as the value $\inf_{\tau \in \mathbb{R}} \{ \tau + \frac{1}{\epsilon} \mathbb{E}_Q[(\phi(\boldsymbol{\omega}) - \tau)^+] \}$ and when the infimum is attained, τ represents the Value-at-Risk with confidence level $1 - \epsilon$ [25].

In this paper, in lieu of (12), we consider the following tractable (convex) approximation:

$$\sup_{Q_{\widetilde{\Xi}} \in \widehat{\mathcal{U}}_N(\alpha, \rho)} Q_{\widetilde{\Xi}} - \mathbf{CVaR}_\epsilon \left(\max_{k \leq K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k} \right) \leqslant 0 \qquad (13)$$

which is, in addition, conservative, because (13) implies (12).

Constraint (13) can be equivalently reformulated as follows [22], [25]:

$$\inf_{\tau \in \mathbb{R}} \left\{ \tau + \frac{1}{\epsilon} \sup_{Q_{\widetilde{\Xi}} \in \widehat{\mathcal{U}}_N(\alpha, \rho)} \mathbb{E}_{Q_{\widetilde{\Xi}}} \left[\left(\max_{k \leq K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k} - \tau \right)^+ \right] \right\} \leqslant 0 \qquad (14)$$

The next proposition states a tractable reformulation of (14). For ease of exposition,

16

we first need to recast function $(\max_{k \leq K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k} - \tau)^+$ as

$$\left(\max_{k \leq K} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a_{2k} - \tau \right)^+ := \max_{k \leqslant K+1} \langle \mathbf{a}_{1k}, \boldsymbol{\omega} \rangle + a'_{2k} \tag{15}$$

where $a'_{2k} = a_{2k} - \tau$ for $k \leq K$, $\mathbf{a}_{1K+1} = \mathbf{0}$ and $a'_{2K+1} = 0$.

Proposition 1 (Reformulation of the **CVaR**-based distributionally robust joint chance constraints)**.** *Set $\alpha > 0$. Then, for any value of $\rho \geqslant \underline{\rho}_{N\alpha}$, the **CVaR**-based distributionally robust joint chance constraints defined by (13) can be equivalently reformulated as follows:*

$$\inf_{\tau \in \mathbb{R}, \lambda_2 \geqslant 0, \mu_i \geqslant 0, \theta_2 \in \mathbb{R}, \boldsymbol{\gamma}_{ik}, \mathbf{v}_{ik}} \left\{ \tau + \frac{1}{\epsilon} \left[\lambda_2 \rho + \theta_2 + \frac{1}{N\alpha} \sum_{i=1}^{N} \mu_i \right] \right\} \leqslant 0 \tag{16a}$$

$$s.t. \;\; \mu_i + \theta_2 + \lambda_2 \|\mathbf{z}^* - \widehat{\mathbf{z}}_i\|_1 \geqslant a'_{2k} + S_{\widetilde{\Xi}_\omega}(\mathbf{v}_{ik})$$

$$- \langle \boldsymbol{\gamma}_{ik}, \widehat{\boldsymbol{\omega}}_i \rangle, \forall i \leqslant N, \forall k \leqslant K+1 \tag{16b}$$

$$\boldsymbol{\gamma}_{ik} - \mathbf{v}_{ik} = -\mathbf{a}_{1k}, \forall i \leqslant N, \forall k \leqslant K+1 \tag{16c}$$

$$\|\boldsymbol{\gamma}_{ik}\|_\infty \leqslant \lambda_2, \forall i \leqslant N, \forall k \leqslant K+1 \tag{16d}$$

where $S_{\widetilde{\Xi}_\omega}(\cdot)$ stands for the support function of $\widetilde{\Xi}_\omega$ and $\langle \cdot, \cdot \rangle$ represents the dot product (see Appendix A.5).

Once we have reformulated the **CVaR**-based distributionally robust joint chance constraint (13), we only need to reformulate the DRO problem defined by the inner supremum in (9). Since this requires a careful and independent analysis, we consider it in the following section.

4. An exact tractable reformulation of the worst-case expected cost

In what follows, we provide an exact and tractable reformulation of the objective function (9) as a continuous linear program.

17

The term

$$C(\tilde{\mathbf{g}}(\boldsymbol{\omega})) = \sum_{j \in \mathcal{G}} \max_{s=1,\ldots,S_j} \left\{ m_{js} \left[g_j - \beta_j \Omega \right] + n_{js} \right\} \tag{17}$$

is the sum of the maximum of *univariate* linear functions in terms of Ω, which is, moreover, convex in Ω. This observation is key to reformulating (9) in a tractable way. In fact, the ambiguity set $\widehat{\mathcal{U}}_N(\alpha, \rho)$ for the worst-case probability distribution in the inner supremum of (9) can be equivalently replaced with the following one, which is also expressed in terms of Ω only:

$$\widehat{\mathcal{U}}_N^{\Omega}(\alpha, \rho) := \{ P_{\widetilde{\Xi}_\Omega} : W_1(\mathcal{R}_{1-\alpha}(\widehat{\mathbb{P}}_N), P_{\widetilde{\Xi}_\Omega}) \leq \rho, \; P_{\widetilde{\Xi}_\Omega}(\widetilde{\Xi}_\Omega) = 1 \} \tag{18}$$

where $\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{(\widehat{\mathbf{z}}_i, \widehat{\Omega}_i)}$ is the empirical distribution supported on the samples $(\widehat{\mathbf{z}}, \widehat{\Omega}_i), i = 1, \ldots, N$; and $\widetilde{\Xi}_\Omega$ stands for the event

$$(\mathbf{z} = \mathbf{f}; \Omega \in [\underline{\Omega}, \overline{\Omega}]), \text{ with } [\underline{\Omega}, \overline{\Omega}] = \left[-\sum_{m \in \mathcal{W}} f_m, \sum_{m \in \mathcal{W}} (\overline{C}_m - f_m) \right]$$

The interval $[\underline{\Omega}, \overline{\Omega}]$ is the conditional support for the random variable Ω (that is, the support set for the system-wise aggregate wind power forecast error, given the predicted power outputs of the wind farms). Essentially, what we have done above is to map the original probability space for the random vector $(\mathbf{z}, \boldsymbol{\omega})$ onto a new probability space for the random vector $(\mathbf{z}, \Omega)$ by the linear map $\boldsymbol{\omega} \mapsto \sum_{m \in \mathcal{W}} \omega_m$, $\Omega = \sum_{m \in \mathcal{W}} \omega_m$, which leaves the objective cost function unaltered. In doing so, the inner supremum in (9) can be fully recast in terms of Ω only as follows:

$$\sup_{P_{\widetilde{\Xi}_\Omega} \in \widehat{\mathcal{U}}_N^{\Omega}(\alpha, \rho)} \mathbb{E}_{P_{\widetilde{\Xi}_\Omega}} \left[\sum_{j \in \mathcal{G}} \max_{s=1,\ldots,S_j} \left\{ m_{js} \left[g_j - \beta_j \Omega \right] + n_{js} \right\} + \langle \mathbf{c}^D, \mathbf{r}^D \rangle + \langle \mathbf{c}^U, \mathbf{r}^U \rangle \right] \tag{19}$$

The proposition below presents a tractable reformulation of (19) as a continuous linear program.

18

Proposition 2 (LP reduction of the worst-case expected cost). *Set $\alpha > 0$ and assume that $\|(\mathbf{z}, \Omega)\| := \|\mathbf{z}\| + |\Omega|$ for some norm $\|\cdot\|$ in $\mathbb{R}^{d_{\mathbf{z}}}$. Then, for any value of $\rho \geqslant \underline{\rho}_{N\alpha}$, the DRO problem defined by* (19) *can be reformulated as the following continuous linear program:*

$$\inf_{\lambda \geqslant 0; \theta \in \mathbb{R}, \overline{\mu}_i, t_i, \underline{t}_{ij}, \overline{t}_{ij}, \widehat{t}_{ij} \forall i \leqslant N, \ \forall j \in \mathcal{G}} \quad \lambda\rho + \theta + \frac{1}{N\alpha} \sum_{i=1}^{N} \overline{\mu}_i + \langle \mathbf{c}^D, \mathbf{r}^D \rangle + \langle \mathbf{c}^U, \mathbf{r}^U \rangle \tag{20a}$$

$$s.t. \ \overline{\mu}_i + \theta + \lambda\|\mathbf{z}^* - \widehat{\mathbf{z}}_i\| \geqslant t_i, \forall i \leqslant N \tag{20b}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \underline{t}_{ij} - \lambda(\underline{\Omega} - \widehat{\Omega}_i), \ \forall i \in \underline{I} \tag{20c}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \overline{t}_{ij} - \lambda(\overline{\Omega} - \widehat{\Omega}_i), \ \forall i \in \underline{I} \tag{20d}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \underline{t}_{ij} + \lambda(\underline{\Omega} - \widehat{\Omega}_i), \ \forall i \in \overline{I} \tag{20e}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \overline{t}_{ij} + \lambda(\overline{\Omega} - \widehat{\Omega}_i), \ \forall i \in \overline{I} \tag{20f}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \overline{t}_{ij} - \lambda(\overline{\Omega} - \widehat{\Omega}_i), \ \forall i \in I \tag{20g}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \underline{t}_{ij} + \lambda(\underline{\Omega} - \widehat{\Omega}_i), \ \forall i \in I \tag{20h}$$

$$t_i \geqslant \sum_{j \in \mathcal{G}} \widehat{t}_{ij}, \ \forall i \in I \tag{20i}$$

$$\overline{\mu}_i \geqslant 0, \ \forall i \leqslant N \tag{20j}$$

$$\underline{t}_{ij} \geqslant m_{js} \left[g_j - \beta_j \underline{\Omega} \right] + n_{js}, \ \forall i \leqslant N, \ \forall j \in \mathcal{G}, \ \forall s \leqslant S_j \tag{20k}$$

$$\overline{t}_{ij} \geqslant m_{js} \left[g_j - \beta_j \overline{\Omega} \right] + n_{js}, \ \forall i \leqslant N, \ \forall j \in \mathcal{G}, \ \forall s \leqslant S_j \tag{20l}$$

$$\widehat{t}_{ij} \geqslant m_{js} \left[g_j - \beta_j \widehat{\Omega}_i \right] + n_{js}, \ \forall i \leqslant N, \ \forall j \in \mathcal{G}, \ \forall s \leqslant S_j \tag{20m}$$

where $\underline{I} := \{i \in \{1, \dots N\} : \widehat{\Omega}_i < \underline{\Omega}\}, I := \{i \in \{1, \dots N\} : \widehat{\Omega}_i \in [\underline{\Omega}, \overline{\Omega}]\}$, and $\overline{I} := \{i \in \{1, \dots N\} : \widehat{\Omega}_i > \overline{\Omega}\}$.

19

5. Numerical results

In this section, we present and discuss results from a series of numerical experiments that have been run on a modified version of the IEEE 118-bus system considered in [12]. All the data and codes needed to reproduce those experiments are available for download in the GITHUB repository [7]. The experiments have been carried out on a Linux-based server using up to 13200 CPUs running in paralell, each clocking at 2.6 GHz with 200 GB of RAM. We have employed CPLEX 20.1.0 under DOcplex Python Modeling API to solve the associated continuous linear programs with the barrier algorithm using up to 22 threads. In addition, we have set the CPLEX parameter `preprocessing_dual` to 1.

We solve the CC-DRO OPF problem (9)–(10) using the **CVaR**-based approximation stated in Section 3, but with different ambiguity sets, namely: (i) The ambiguity set based on probability trimmings, introduced in this paper, which we refer to as DROTRIMM; and (ii) a Wasserstein ball centered at the empirical distribution supported on the $\widehat{\boldsymbol{\omega}}$-coordinates of the N samples in $\widehat{\Xi}_\omega^N$, i.e., $\{(\widehat{\mathbf{z}}_i, \widehat{\boldsymbol{\omega}}_i)\}_{i=1}^N$. This leads to the distributionally robust chance-constrained OPF model proposed in [22], which we call DROW. Importantly, this is a DRO model that fully ignores the contextual information, since the center of the Wasserstein ball it uses is made up of *all* past samples of wind power forecast errors (regardless of the current wind power point predictions). Roughly speaking, DROTRIMM also works with all the past N samples of wind power forecast errors, but only those that lead to the worst-case conditional distribution of the prediction errors are moved onto the conditional support. However, this movement must entail a transportation cost smaller than or equal to a given budget ρ and the computation of that cost is directly contingent on the current context (that is, the current wind power point forecasts).

20

In addition, we benchmark the previous two distributionally robust methods with an alternative approach that is commonly used in the literature for solving optimization problems with probabilistic constraints, known as the *scenario approach*, but adapted to account for contextual information. We have taken the required adaptation from [14, Chapter 4], which, in our setting, involves solving a DC-OPF problem in which the uncertain constraints are enforced for the wind power forecast errors associated with the K samples nearest to the context. We refer to this adaptation of the popular scenario approach as SCENA.

The training data consist of a set of pairs $\{(\widehat{\mathbf{z}}_i, \widehat{\boldsymbol{\omega}}_i)\}_{i=1}^N$, from which we can directly obtain the collection of pairs $\{(\widehat{\mathbf{z}}_i, \widehat{\Omega}_i)\}_{i=1}^N$, where $\widehat{\Omega}_i := \sum_{m \in \mathcal{W}} \widehat{\omega}_{i,m}$. For ease of computation and to simplify the analysis below, we have considered the same radius or transportation budget for the two ambiguity sets in both the objective and the chance constraints of the DRO OPF problem (9)–(10).

5.1. Evaluation of the out-of-sample performance via re-optimization

Given a context (in the form of point forecasts of the power outputs of the wind farms), a training dataset, and a robustness parameter ρ, each method *met* (either DROW or DROTRIMM in our case) provides a forward generation dispatch and reserve capacity provision $\mathbf{y}^{met} := (\mathbf{g}, \mathbf{r}^D, \mathbf{r}^U)$. To evaluate the actual or out-of-sample performance of that $\mathbf{y}^{met}$, we draw a sample of wind power forecast errors $\widehat{\boldsymbol{\omega}}_j$ from a test dataset, and the vector of recourse variables $\mathbf{r}$ (that is, the real-time power adjustments) is computed by solving the deterministic optimal power flow available in Appendix B. In this deterministic OPF problem, wind spillage (with a cost equal to 0) and involuntary load curtailment (with a cost equal to \$500/MWh) are considered as feasible recourse actions, aside from the deployment of reserves by generators. In this way, the *out-of-sample performance* of a method *met*, which produces the forward

21

dispatch $\mathbf{y}^{met}$, $J(\mathbf{y}^{met})$, is computed by the empirical out-of-sample cost averaged over the test set formed by a certain number of samples of $\mathbb{Q}_{\boldsymbol{\omega}/\mathbf{z}=\mathbf{f}}$. In addition, in order to measure the *reliability* of a solution (that is, if $\mathbf{y}^{met}$ is feasible or not in real time), the *violation probability* is estimated over the test set. In this estimation, we count as a violation every time a recourse action involving load curtailment or wind spillage is to be taken in real time to restore the power balance. This is equivalent to counting (over the test set) the number of times a constraint is violated in the original affine-policy-based OPF model.

5.2. A 118-bus case study

As previously mentioned, we consider a modified version of the IEEE 118-bus system used in [12]. The system includes 54 conventional generators and eight wind power plants that we have added and placed at buses 2, 16, 33, 37, 55, 67, 83, and 116. In addition, the piecewise linear cost functions of all generators are comprised of three pieces or blocks. All the data pertaining to the network, generators, and transmission lines are available at the GITHUB repository [7].

We analyze two scenarios, which differ in the level of wind power penetration in the system. Below we explain how we have generated samples for the joint distribution of the wind power forecast and its error at each wind power plant. The so generated samples are also available online at the GITHUB repository [7]:

1. Let $\widetilde{f}_m$ be the per-unit point forecast of the power output at wind plant $m \in \mathcal{W}$. A sample of $\widetilde{f}_m$, for all $m \in \mathcal{W}$, is randomly drawn with replacement from a collection of $16\,694$ p.u. wind power data recorded in several zones and made available by the Global Energy Forecasting Competition 2014 [11]. We have selected zones 1, 2, 3, 4, 5, 6, 9, and 10 of the aforementioned data set and

22

assigned them to the eight wind power plants located at buses 16, 116, 83, 2, 55, 67, 33, and 37, respectively.

2. For each wind farm $m \in \mathcal{W}$, we have assumed that the (nominal, normalized) random variable W_m, which represents the nominal actual power generated at wind plant m, follows a Beta distribution with mean $\widetilde{f}_m$ and standard deviation σ. This standard deviation depends on both physical parameters and the quality of the forecasting model, following the model proposed in [9]. For simplicity, in all numerical experiments, given $\widetilde{f}_m$, we determine σ as the value of the following function $\sigma(\widetilde{f}_m) := 0.2\widetilde{f}_m + 0.02$, empirically obtained in [9] for the case of a lead time of six hours. Therefore, the actual wind power production, and hence, the forecast error are conditional on the forecast power output issued. More specifically, the forecast error is given as the difference of a realization $\widehat{W}_m$ of the r.v. $W_m \sim \text{Beta}(A, B)$ and the point forecast $\widetilde{f}_m$, where $A, B > 0$ are the solution (if it exists) of the following system of non-linear equations:

$$\widetilde{f}_m = \frac{A}{A + B} \tag{21a}$$

$$\sigma^2(\widetilde{f}_m) = \frac{AB}{(A + B)^2(A + B + 1)} \tag{21b}$$

To ensure that this non-linear system of equations has a solution, the samples $\widetilde{f}_m$ from the dataset that are less than or equal to 0.05 p.u. are set to 0.05, and the ones greater than or equal to 0.95 p.u. are set to 0.95. Thus, for each wind power plant, the per-unit point forecast lies in the interval $[0.05, 0.95]$.

3. To work with MW, we multiply the per-unit realized power output and the point prediction $\widetilde{f}_m$ by the wind plant capacity $\overline{C}_m$, thus getting a pair of predicted power output and its error $(\overline{C}_m \widetilde{f}_m, \overline{C}_m(\widehat{W}_m - \widetilde{f}_m))$ for wind farm m.

4. Steps 1, 2 and 3 are repeated N times to get the desired sample size.

23

Each independent run in our simulations involves repeating the above process.

Finally, the test set used to compute the out-of-sample performance of a data-driven solution via re-optimization (i.e., the actual probability of violating the uncertain OPF constraints and the actual expected operational cost) is constructed by drawing 1000 samples from the wind-power-data generating model based on the beta distribution presented above, with mean equal to the point prediction acting as the selected context. Therefore, this test set constitutes a discrete approximation of the forecast error distribution conditional on a given context, which will be specified later. Importantly, the shape and size of the ambiguity set that DROTRIMM uses is to be changed with the sample size N (which is indicative of the amount of information on the joint distribution of $(\mathbf{z}, \boldsymbol{\omega})$ we have). Consequently, the trimming level α defining this set is to be dependent on N. Accordingly, we have set $\alpha_N := K_N/N$, where K_N is the number of nearest neighbors used by SCENA. We have specifically taken $K_N := \lfloor N^{0.9} \rfloor$ so that the resulting α_N is consistent with the convergence results included in [8]. Observe that both α and K have been augmented with the subscript N to make their dependence on the sample size explicit.

5.2.1. Medium wind penetration case

In this case, all eight wind farms in the system have a capacity of 200 MW and the context is given by $\mathbf{z}^* = 180 \cdot \mathbf{1}$ MW, that is, the point forecast is 180 MW for all the wind power plants. Hence, the level of wind power penetration in the system (i.e., the ratio of the predicted system-wise wind power production to the total system demand) is approximately 63%.

Figures 1 and 2 illustrate the box plots corresponding to the total downward and upward reserve capacity that is scheduled, the violation probability and the expected cost delivered out of sample by SCENA, DROTRIMM and DROW as a function of

24

(a) Total downward reserves

(b) Total upward reserves

(c) Violation probability (out of sample)

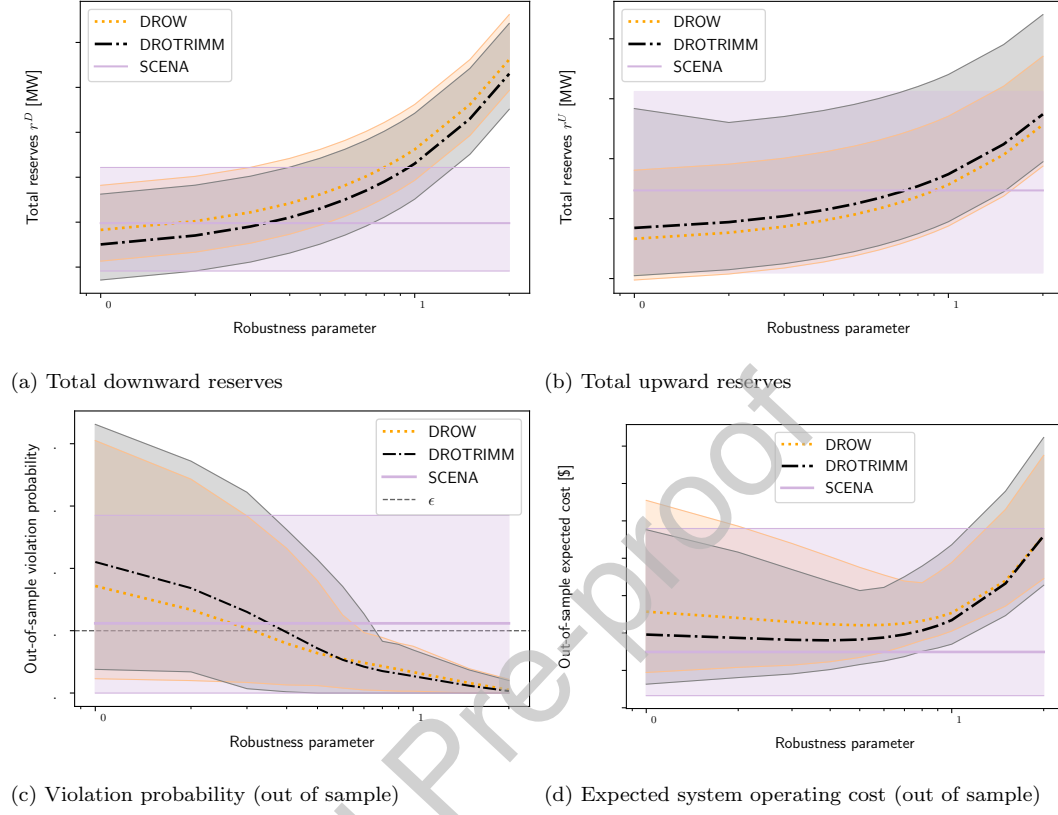(d) Expected system operating cost (out of sample)

Figure 1. Medium level of wind penetration, $N = 100$ and $\epsilon = 0.1$: Total downward and upward reserve capacity and performance metrics

their corresponding robustness parameter for sample sizes $N = 100$ and $N = 300$, respectively. Naturally, the results provided by SCENA do not change along the x-axis in the plots, because this method is not based on *distributional robustness*. The box plots have been obtained from 200 independent runs for each sample size. We have set $\epsilon = 0.1$. The robustness parameter of DROW is the radius of the Wasserstein ball, while the robustness parameter for DROTRIMM is the budget excess over the minimum transportation budget (see Definition 4) [8].
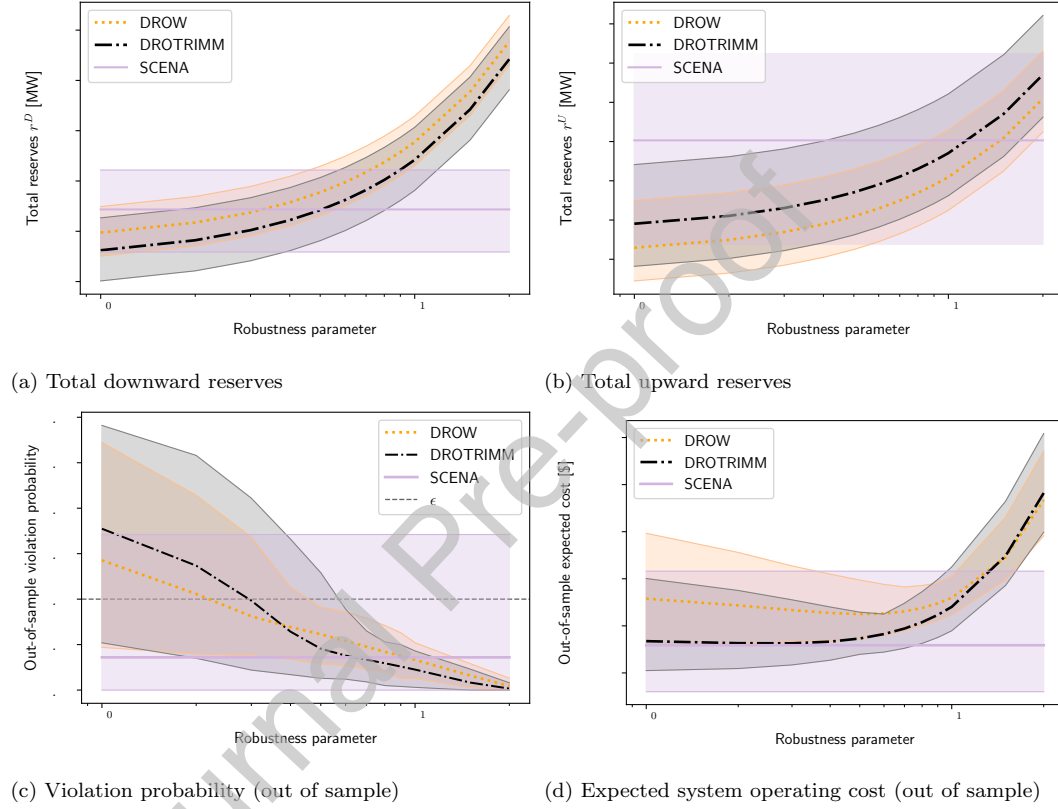
25

(a) Total downward reserves

(b) Total upward reserves

(c) Violation probability (out of sample)

(d) Expected system operating cost (out of sample)

Figure 2. Medium level of wind penetration, $N = 300$ and $\epsilon = 0.1$: Total downward and upward reserve capacity and performance metrics

26

The color-shaded areas have been obtained by joining the minimum and maximum edge cases of the box plots, while the associated bold colored lines link their means. These figures allow us to check which of the methods provides the most cost-efficient dispatch solutions on average without exceeding the threshold ϵ. As expected, the reliability of the OPF solution given by DROW and DROTRIMM increases as the value of their robustness parameter is augmented, because more reserve capacity is procured. In turn, as more reserve capacity is scheduled, the magnitude and frequency of expensive load shedding events tend to diminish, which explains why the expected system operating cost may also decrease with the robustness parameter. This justifies the use of Distributionally Robust Optimization to tackle the chance-constrained OPF problem. However, when said parameter reaches a large enough value, the expected cost starts to grow quickly, because the cost of procuring additional reserve capacity no longer compensates for the cost savings entailed by the reduction in the amount of curtailed load.

While SCENA provides OPF solutions that are competitive in terms of expected cost, these solutions do not comply with the specified reliability threshold in many of the runs. In addition, the performance of the OPF solutions obtained from SCENA exhibit a high variability, which is clearly due to the fact that this method is a non-robust approach and as such, is highly negatively affected by the uncertainty associated with the conditional inference it must perform.

On the other hand, when comparing DROW and DROTRIMM, whereas the former needs a lower value of the robustness parameter to attain the desired level of solution reliability, DROTRIMM gets to identify OPF solutions that are also reliable, while systematically cheaper on average. This phenomenon becomes even more evident when we increase the sample size N from 100 to 300. Indeed, a richer joint data sample contains more information on the statistical dependence of the wind power

27

forecast error on the associated point prediction, which our approach manages to take advantage of. To give some numbers, if we just consider the range of values for the robustness parameters for which the violation probability is kept below the tolerance ϵ, the average expected cost savings of DROTRIMM with respect to DROW go from 0.82%, when $N = 100$ to 1.82%, when $N = 300$. From a technical point of view, DROW tends to produce OPF solutions with a higher cost because it underestimates the amount of upward reserve capacity that should be procured, clearly because this method is oblivious to the context and therefore, plans for the *marginal* distribution of the wind power forecast errors and not for the conditional one.

To elaborate further on the differences among the three methods, Table 1 includes the maximum, average, and minimum out-of-sample expected cost[1] under the value of the robustness parameter that is *optimal* for methods DROW and DROTRIMM, i.e., which leads to reliable OPF solutions with the minimum average expected cost for each of these two approaches. The standard deviation of this cost is also provided in the last row of Table 1. When $N = 100$, the (exacerbated) robustness of DROW produces OPF solutions with low average cost and variance, although DROTRIMM manages to find OPF solutions that are more economical in expectation. When $N = 300$, DROTRIMM clearly beats DROW on all metrics, because the excessive robustness of DROW (which is the result of ignoring the context) no longer pays off. Again, SCENA provides the cheapest OPF solutions on average, but these are useless because they do not satisfy the reliability requirement.

---

[1]These statistics are computed over the 200 hundred independent runs.

Table 1. Medium level of wind penetration, summary data for total expected cost [\$] under the optimal value of the robustness parameter for methods DROW and DROTRIMM.

	$N = 100$			$N = 300$		
	DROTRIMM	DROW	SCENA	DROTRIMM	DROW	SCENA
max	24790	24388	25795	23250	23972	24160
avg	23068	23258	22493	22826	23250	22588
min	22511	22649	21325	22443	22728	21602
std	300	295	742	159	246	496

5.2.2. High wind penetration case

In this alternative setting, all the wind farms have a capacity of 250 MW and the context is given by $\mathbf{z}^* = 225 \cdot \mathbf{1}$ MW. Hence, the level of wind power penetration in the system is approximately 80%.

Figures 3 and 4, and Table 2 are analogous to Figures 1 and 2, and Table 1 of the previous case, respectively. The higher level of wind power penetration in this new instance implies a higher level of uncertainty in the system. This accentuates the difference in performance between DROW and DROTRIMM when $N = 100$, that is, in a small sample regime. More specifically, the relative difference between the out-of-sample average expected cost achieved by DROW and DROTRIMM increases from 0.82% in the previous case to 2.27% in this new one. It is true, though, that DROW offers reliable OPF solutions with the lowest variance in expected cost when $N = 100$, provided that its robustness parameter is optimally tuned, see Table 2. However, its superiority in this respect ends when N grows to 300, at which point DROTRIMM provides the most cost-efficient OPF solutions in every respect[2]. Again

---

[2]Note in Table 2 that, while the standard deviation of the expected cost is a bit higher under

Table 2. High level of wind penetration, summary data for total expected cost [\$] under the optimal value of the robustness parameter for DROW and DROTRIMM.

	$N = 100$			$N = 300$		
	DROTRIMM	DROW	SCENA	DROTRIMM	DROW	SCENA
max	19794	18804	21699	18101	18255	19863
avg	17334	17737	16483	17025	17274	16779
min	16490	16795	15040	16486	16703	15639
std	488	381	1093	296	292	767

the reason for this difference in performance has to do with the different provision of upward and downward reserve capacity that DROW and DROTRIMM prescribe.

For its part, the SCENA method keeps on providing cheap, but unreliable OPF solutions under a higher level of wind power penetration. In fact, the variability in cost, violation probability and reserves of the OPF solutions given by this method is remarkably high in contrast with that of DROTRIMM and DROW, even higher than in the case of a medium wind power penetration level (compare the range of the box plots in Figure 4).

We conclude this section with a remark on computational time. DROTRIMM and DROW have the same complexity (essentially, the number of constraints grows linearly with the sample size N). The continuous linear program that results from tackling the chance-constrained DRO OPF problem by way of DROTRIMM and the **CVaR** approximation takes around 15 minutes to be solved on average, for a sample

---

DROTRIMM than under DROW when $N = 300$, the maximum and minimum values reached by the expected cost under each method reveals that DROTRIMM produces a distribution of the expected cost displaced towards cheaper OPF solutions.
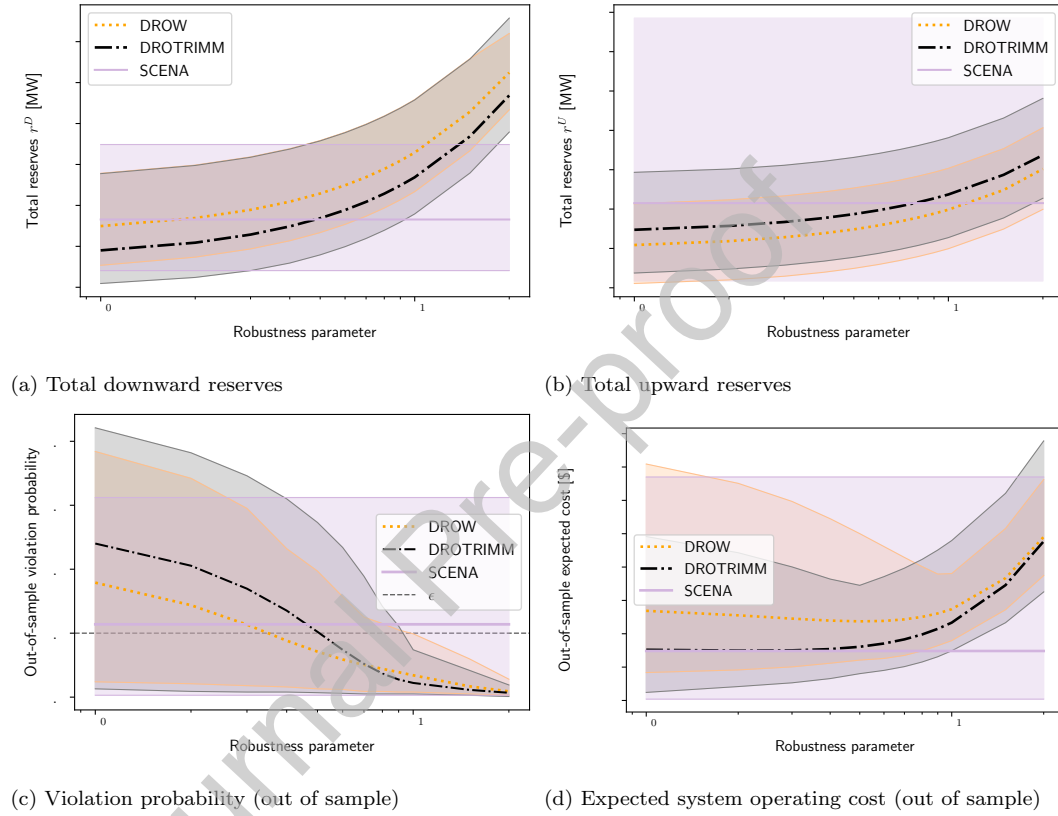
(a) Total downward reserves

(b) Total upward reserves

(c) Violation probability (out of sample)

(d) Expected system operating cost (out of sample)

Figure 3. High level of wind penetration, $N = 100$ and $\epsilon = 0.1$: Total downward and upward reserve capacity and performance metrics
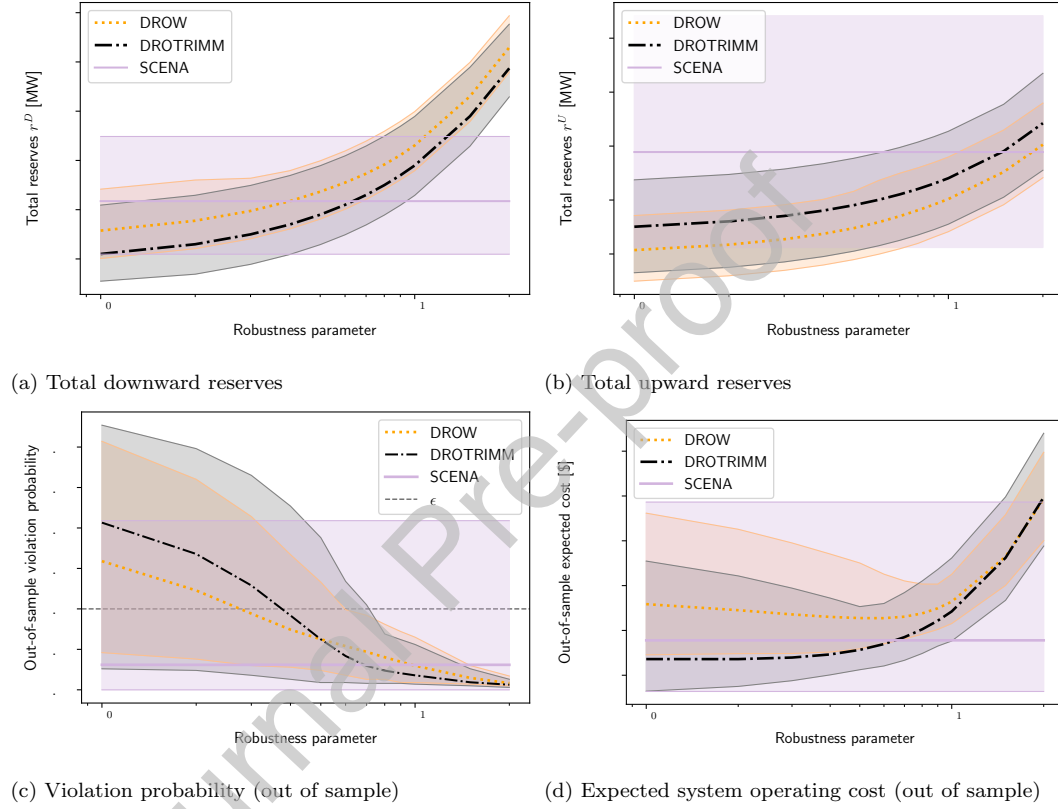
31

(a) Total downward reserves

(b) Total upward reserves

(c) Violation probability (out of sample)

(d) Expected system operating cost (out of sample)

Figure 4. High level of wind penetration, $N = 300$ and $\epsilon = 0.1$: Total downward and upward reserve capacity and performance metrics

32

size $N = 300$, using CPLEX 20.1.0 on a Linux-based server with 22 CPUs clocking at 2.6 GHz and 200 GB of RAM in total.

6. Conclusion

In this paper, we have developed a distributionally robust chance-constrained OPF model that is able to exploit contextual information through an ambiguity set based on probability trimmings. We have provided a reformulation of this model as a continuous linear program using the well-known **CVaR** approximation. By way of a series of numerical experiments conducted on a modified 118-bus power network with wind uncertainty, we have shown that, by exploiting the statistical dependence between the point forecast of the wind power outputs and its associated forecast error, our approach can identify dispatch solutions that, while satisfying the required system reliability, lead to costs savings of up to several percentage points with respect to the OPF solutions provided by an alternative DRO method that ignores said statistical dependence.

In future work, we plan to devise data-driven schemes for appropriately tuning the robustness parameter in our distributionally robust chance-constrained OPF model in accordance with the risk preferences of the system operator (for instance, by resorting to cross-validation or bootstrapping). We also want to extend this model to account for intertemporal constraints, which, among other things, would involve adapting our probability-trimming-based ambiguity set to deal with stochastic processes and time series data.

Acknowledgments

33

References

[1] Arrigo, A., Kazempour, J., De Grève, Z., Toubeau, J.-F., & Vallée, F. (2021). Embedding dependencies between wind farms in uncertainty-aware optimal power flow. *arXiv preprint*, *arXiv:2104.08101*.

[2] Arrigo, A., Ordoudis, C., Kazempour, J., De Grève, Z., Toubeau, J.-F., & Vallée, F. (2022). Wasserstein distributionally robust chance-constrained optimization for energy and reserve dispatch: An exact and physically-bounded formulation. *European Journal of Operational Research*, *296*, 304–322.

[3] Bertsimas, D., McCord, C., & Sturt, B. (2023). Dynamic optimization with side information. *European Journal of Operational Research*, *304*, 634–651.

[4] Bienstock, D., Chertkov, M., & Harnett, S. (2014). Chance-constrained optimal power flow: Risk-aware network control under uncertainty. *SIAM Review*, *56*, 461–495.

[5] Bludszuweit, H., Dominguez-Navarro, J., & Llombart, A. (2008). Statistical analysis of wind power forecast error. *IEEE Transactions on Power Systems*, *23*, 983–991.

[6] Chen, Z., Kuhn, D., & Wiesemann, W. (2022). Data-Driven Chance Constrained Programs over Wasserstein Balls. *Operations Research*, *Ahead of Print*. doi:`10.1287/opre.2022.2330`.

[7] Esteban-Pérez, A., & Morales, J. M. (2021). Distributionally Robust Optimal Power Flow with Contextual Information – Codes and Data. *GitHub repository*, *https://github.com/groupoasys/DRO_DCOPF_CONTEXTUAL*.

[8] Esteban-Pérez, A., & Morales, J. M. (2021). Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming*, (pp. 1–37).

[9] Fabbri, A., Gomez San Roman, T., Rivier Abbad, J., & Mendez Quezada, V. (2005). Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Transactions on Power Systems*, *20*, 1440–1446.

[10] Hanasusanto, G. A., Roitch, V., Kuhn, D., & Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information. *Operations Research*, *65*, 751–767.

[11] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, *32*, 896–913.

[12] Jabr, R. A. (2020). Distributionally robust CVaR constraints for power flow optimization. *IEEE Transactions on Power Systems*, *35*, 3764–3773.

[13] Jia, M., Hug, G., & Shen, C. (2021). Iterative decomposition of joint chance constraints in OPF. *IEEE Transactions on Power Systems*, *36*, 4836–4839.

35

[14] Koduri, N. (2021). *Essays on Decision Making Under Uncertainty*. Ph.D. thesis Massachusetts Institute of Technology.

[15] Li, B., Jiang, R., & Mathieu, J. L. (2016). Distributionally robust risk-constrained optimal power flow using moment and unimodality information. In *2016 IEEE 55th Conf. on Decision and Control (CDC)* (pp. 2425–2430).

[16] Li, B., Jiang, R., & Mathieu, J. L. (2019). Ambiguous risk constraints with moment and unimodality information. *Mathematical Programming.*, *173*, 151–192.

[17] Li, B., Jiang, R., & Mathieu, J. L. (2019). Distributionally robust chance-constrained optimal power flow assuming unimodal distributions with misspecified modes. *IEEE Transactions on Control of Network Systems*, *6*, 1223–1234.

[18] Lubin, M., Dvorkin, Y., & Backhaus, S. (2016). A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Transactions on Power Systems*, *31*, 3840–3849.

[19] Mohajerin Esfahani, P., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming.*, *171*, 115–166.

[20] Morales, J. M., Conejo, A. J., Liu, K., & Zhong, J. (2012). Pricing electricity in pools with wind producers. *IEEE Transactions on Power Systems*, *27*, 1366–1376.

[21] Nemirovski, A., & Shapiro, A. (2007). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, *17*, 969–996.

[22] Ordoudis, C., Nguyen, V. A., Kuhn, D., & Pinson, P. (2021). Energy and reserve dispatch with distributionally robust joint chance constraints. *Operations Research Letters*, *49*, 291–299.

[23] Peña-Ordieres, A., Molzahn, D. K., Roald, L. A., & Wächter, A. (2020). DC optimal power flow with joint chance constraints. *IEEE Transactions on Power Systems*, *36*, 147–158.

[24] Rahimian, H., & Mehrotra, S. (2022). Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, *3*, 1–85.

[25] Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, *2*, 21–41.

[26] Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians*. doi:`10.1007/978-3-319-20828-2`.

[27] Stott, B., Jardim, J., & Alsac, O. (2009). DC power flow revisited. *IEEE Transactions on Power Systems*, *24*, 1290–1300.

[28] Vrakopoulou, M., Margellos, K., Lygeros, J., & Andersson, G. (2013). A probabilistic framework for reserve scheduling and $n-1$ security assessment of systems with high wind power penetration. *IEEE Transactions on Power Systems*, *28*, 3885–3896.

[29] Xie, W., & Ahmed, S. (2018). Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation. *IEEE Transactions on Power Systems*, *33*, 1860–1867.

[30] Xie, W., & Ahmed, S. (2020). Bicriteria approximation of chance-constrained covering problems. *Operations Research*, *68*, 516–533.

[31] Zhang, H., & Li, P. (2011). Chance constrained programming for optimal power flow under uncertainty. *IEEE Transactions on Power Systems*, *26*, 2417–2424.

[32] Zhou, A., Yang, M., Wang, M., & Zhang, Y. (2020). A linear programming approximation of distributionally robust chance-constrained dispatch with Wasserstein distance. *IEEE Transactions on Power Systems*, *35*, 3366–3377.

Appendix A. Notation

The main notation used throughout the text is stated below for quick reference. Other symbols are defined as required.

Appendix A.1. Sets, numbers and indices

$\mathcal{B}$ Set of buses, indexed by b.

$\mathcal{L}$ Set of lines, indexed by ℓ.

$\mathcal{G}$ Set of generators (dispatchable units), indexed by j.

$\mathcal{W}$ Set of wind power plants, indexed by m.

Appendix A.2. Parameters and functions

f   Array of forecasted power outputs [MW].

$\widetilde{\mathbf{f}}$ Array of nominal (p.u.) forecasted power outputs.

L   Array of loads [MW].

$\mathbf{g}^{\min}, \mathbf{g}^{\max}$ Array of upper and lower capacity limits of generators [MW].

Cap   Array of line capacities [MW].

$\overline{\mathbf{C}}$ Array of installed capacities of the wind power plants [MW].

$\mathbf{M}^{\mathcal{G}/\mathcal{W}/\mathcal{B}}$ Matrix of DC power transfer distribution factors, which maps nodal power injections to line flows for generators/wind farms/loads.

$\mathbf{c}^{D}, \mathbf{c}^{U}$ Array of downward and upward reserve capacity costs [\$/MW].

$C(\cdot)$ Total production cost function, which is given by the sum of $|\mathcal{G}|$ convex piecewise linear cost functions with S_j pieces for generator j, i.e., $C(\tilde{\mathbf{g}}(\boldsymbol{\omega})) := \sum_{j \in \mathcal{G}} \max_{s=1,\ldots,S_j} \{m_{js}\tilde{g}_j(\boldsymbol{\omega}) + n_{js}\}$, where m_{js}, n_{js} are the slope and the intercept of the s-th piece for generator j, respectively [\$].

Appendix A.3. Random variables and uncertain parameters

$\mathbf{z}$ Random vector of features/covariates.

$\boldsymbol{\omega}$ Random vector representing the wind power forecast errors of the $|\mathcal{W}|$ wind power plants [MW].

$\boldsymbol{\xi}$ Random vector representing the pair of features/covariates and the wind power forecast errors of the $|\mathcal{W}|$ wind farms, that is, $\boldsymbol{\xi} := (\mathbf{z}, \boldsymbol{\omega})$.

W_m Actual power output at wind power plant $m \in \mathcal{W}$ in per unit.

Ξ_ω Support set of the random vector $\boldsymbol{\omega}$.

$\widetilde{\Xi}_{\boldsymbol{\omega}}$ Support set of the random vector $\boldsymbol{\omega}$ conditional on $\mathbf{z} = \mathbf{f}$, which is given by the hypercube $\prod_{m \in \mathcal{W}} [-f_m, \overline{C}_m - f_m]$.

Ξ Support set of the random vector $(\mathbf{z}, \boldsymbol{\omega})$.

$\widetilde{\Xi}$ Contextual information, that is, the event $(\mathbf{z} = \mathbf{f}; \; \boldsymbol{\omega} \in \widetilde{\Xi}_{\boldsymbol{\omega}})$.

Ω Random variable defined as $\sum_{m \in \mathcal{W}} \omega_m$, which describes the system-wise aggregate wind power forecast error [MW].

$\widetilde{\Xi}_\Omega$ Contextual information linked to the random vector $(\mathbf{z}, \Omega)$, that is, the event $(\mathbf{z} = \mathbf{f}; \Omega \in [\underline{\Omega}, \overline{\Omega}])$, with $[\underline{\Omega}, \overline{\Omega}] = \left[-\sum_{m \in \mathcal{W}} f_m, \sum_{m \in \mathcal{W}} (\overline{C}_m - f_m) \right]$.

$\tilde{\mathbf{g}}(\boldsymbol{\omega})$ Array of power generation outputs of generators (random vector) [MW].

40

$\tilde{\mathbf{r}}(\boldsymbol{\omega})$ Array of reserves deployed by generators (random vector) [MW].

$\mathbb{E}_Q$ Expectation operator with respect to the probability measure Q.

δ_ξ Dirac distribution at ξ.

Appendix A.4. Variables

$\mathbf{g}$ Generators' power dispatch [MW].

$\boldsymbol{\beta}$ Array of generators' participation factors.

$\mathbf{r}^D, \mathbf{r}^U$ Array of downward/upward reserve capacities provided by generators [MW].

$\mathbf{x}$ Vector of decision variables, that is, $\mathbf{x} := (\mathbf{g}, \boldsymbol{\beta}, \mathbf{r}^D, \mathbf{r}^U)$.

$\mathbf{y}$ Vector of first-stage decision variables (power dispatch and reserve capacity provision), that is, $\mathbf{y} := (\mathbf{g}, \mathbf{r}^D, \mathbf{r}^U)$.

Appendix A.5. Other symbols

$\mathbf{1}$ Array of ones (of appropriate dimension).

$\mathbf{0}$ Array of zeros (of appropriate dimension).

$|A|$ Cardinality of a set A.

$(x)^+$ Positive part of x, i.e., $\max\{x, 0\}$.

$\lfloor x \rfloor$ Floor function of x, given by $\max\{m \in \mathbb{Z} \,/\, m \leqslant x\}$.

$\langle \cdot, \cdot \rangle$ Dot product.

W_1 1-Wasserstein distance.

$\mathcal{P}_1(\Xi), \mathcal{P}_1(\widetilde{\Xi})$ The set of all probability distributions with finite first moment supported on $\Xi, \widetilde{\Xi}$, respectively.

$\mathcal{R}_{1-\alpha}(P)$ The set of all $(1-\alpha)$-trimmings of the probability distribution P.

ρ Robustness parameter.

$\underline{\rho}_{N\alpha}$ Minimum transportation budget.

S_B Support function of a set $B \subseteq \mathbb{R}^d$, defined as $S_B(a) := \sup_{b \in B} \langle a, b \rangle$.

$Q - \mathbf{CVaR}_\epsilon(\phi(\boldsymbol{\omega}))$ Conditional Value at Risk at level $\epsilon \in (0, 1)$ of $\phi(\boldsymbol{\omega})$ under the probability measure Q; that is, the value $\inf_{\tau \in \mathbb{R}}\{\tau + \frac{1}{\epsilon}\mathbb{E}_Q[(\phi(\boldsymbol{\omega}) - \tau)^+]\}$.

Appendix B. Real-time re-dispatch problem

This appendix contains the optimization program used to evaluate the out-of-sample performance of a given solution of the chance-constrained DC-OPF problem. Given N, a data-driven solution $\mathbf{y}_N := (\mathbf{g}, \mathbf{r}^D, \mathbf{r}^U)_N$ and a realization of the forecast error $\widehat{\boldsymbol{\omega}}_i$, the operator of the system solves the following deterministic linear program:

$$\min_{\mathbf{r}, \boldsymbol{\Delta d}, \boldsymbol{\Delta \omega}} C(\mathbf{g}_N + \mathbf{r}) + \langle \mathbf{c}^{\text{shed}}, \boldsymbol{\Delta d} \rangle + \langle \mathbf{c}^D, \mathbf{r}_N^D \rangle + \langle \mathbf{c}^U, \mathbf{r}_N^U \rangle \tag{B.1}$$

$$\text{s.t.} \ \ \mathbf{0} \leqslant \boldsymbol{\Delta d} \leqslant \mathbf{L} \tag{B.2}$$

$$\mathbf{0} \leqslant \boldsymbol{\Delta \omega} \leqslant \mathbf{f} + \widehat{\boldsymbol{\omega}}_i \tag{B.3}$$

$$-\mathbf{r}_N^D \leqslant \mathbf{r} \leqslant \mathbf{r}_N^U \tag{B.4}$$

$$\langle \mathbf{1}, \mathbf{r} \rangle + \langle \mathbf{1}, \boldsymbol{\Delta d} \rangle + \langle \mathbf{1}, \widehat{\boldsymbol{\omega}}_i - \boldsymbol{\Delta \omega} \rangle = 0 \tag{B.5}$$

$$-\mathbf{Cap} \leqslant \mathbf{M}^{\mathcal{G}}(\mathbf{g}_N + \mathbf{r}) + \mathbf{M}^{\mathcal{W}}(\mathbf{f} + \widehat{\boldsymbol{\omega}}_i - \boldsymbol{\Delta \omega}) - \mathbf{M}^{\mathcal{B}}(\mathbf{L} - \boldsymbol{\Delta d}) \leqslant \mathbf{Cap} \tag{B.6}$$

where $\mathbf{r}$, $\boldsymbol{\Delta d}$ and $\boldsymbol{\Delta \omega}$ are the deployed reserves, load shedding and wind spillage vector of decision variables; and the parameter c_b^{shed} is the load shedding cost at bus b. The objective function in (B.1) minimizes the total operational cost of the system, which comprises the electricity generation cost, the load shedding cost and the total cost of up- and down-reserve capacities. The latter is known and constant and thus, does not intervene in the minimization. Constraints (B.2) and (B.3) limit the amount of load involuntarily curtailed and the amount of wind power unused to the actual realizations of the load and the wind power production, respectively. Constraint (B.4) ensures that the deployed reserves are kept within the reserve capacities scheduled in the forward stage. Constraint (B.5) constitutes the real-time power balance equation and, finally, constraints (B.6) enforce the transmission capacity limits.

43