

Stochastic Optimal Power Flow Based on Data-Driven Distributionally Robust Optimization

Yi Guo[†], Kyri Baker[‡], Emiliano Dall’Anese^{*}, Zechun Hu^{**}, Tyler Summers[†]

Abstract—We propose a data-driven method to solve a stochastic optimal power flow (OPF) problem based on limited information about forecast error distributions. The objective is to determine power schedules for controllable devices in a power network to balance operational cost and conditional value-at-risk (CVaR) of device and network constraint violations. These decisions include scheduled power output adjustments and reserve policies, which specify planned reactions to forecast errors in order to accommodate fluctuating renewable energy sources. Instead of assuming the uncertainties across the networks follow prescribed probability distributions, we assume the distributions are only observable through a finite training dataset. By utilizing the Wasserstein metric to quantify differences between the empirical data-based distribution and the real data-generating distribution, we formulate a distributionally robust optimization OPF problem to search for power schedules and reserve policies that are robust to sampling errors inherent in the dataset. A multi-stage closed-loop control strategy based on model predictive control (MPC) is also discussed. A simple numerical example illustrates inherent tradeoffs between operational cost and risk of constraint violation, and we show how our proposed method offers a data-driven framework to balance these objectives.

I. INTRODUCTION

The continued integration of renewable energy sources (RESs) in power systems is making it more complicated for system operators to balance economic efficiency and system reliability and security. As penetration levels of RESs reach substantial fractions of total supplied power, networks will face high operational risks under current operational paradigms. As it becomes more difficult to predict the net load, large forecast errors can lead to power quality and reliability issues causing significant damage and costly outages. Future power networks will require more sophisticated methods for managing these risks, at both transmission and distribution levels.

The flexibility of controllable devices, including power-electronics-interfaced RESs, can be utilized to balance efficiency and risk with optimal power flow methods [1]–[6], which aim to determine power schedules for controllable devices in a power network to optimize an objective function. However, most OPF methods in the research literature and

those widely used in practice are deterministic, assuming point forecasts of exogenous power injections and ignoring forecast errors. Increasing forecast errors push the underlying distributed feedback controllers that must handle the transients caused by these errors closer to stability limits [7].

More recently, research focus has turned to stochastic and robust optimal power flow methods that explicitly incorporate forecast errors, in order to more systematically trade off efficiency and risk and to ease the burden on feedback controllers [8]–[26]. Many formulations assume that uncertain forecast errors follow a prescribed probability distribution (commonly, Gaussian [11], [17], [24]) and utilize analytically tractable reformulations of probabilistic constraints. In practice, forecast error probability distributions are never known; they are only observed indirectly through finite datasets. Sampling-based methods have been applied with a focus on quantifying the probability of constraint violation [12], [13] and for constraining or optimizing conditional value at risk (CVaR) [15], [18], [26]. Distributionally robust approaches use data to estimate distribution parameters (e.g., mean and variance) and aim to be robust to any data-generating distribution consistent with these parameters [18], [21], [22], [25], [26]. Others take a robust approach, assuming only knowledge of bounds on forecast errors and enforcing constraints for any possible realization, e.g., [14], [19]. Overall, this line of recent research has explored tractable approximations and reformulations of difficult stochastic optimal power flow problems. However, none of the existing work explicitly accounts for sampling errors arising from limited data, which in operation can cause poor out-of-sample performance¹. Even with sophisticated recent stochastic programming techniques, decisions can be overly dependent on small amounts of relevant data, a phenomenon akin to overfitting in statistical models.

We propose a data-driven method to solve a stochastic optimal power flow problem based on limited information about forecast error distributions available through a finite training dataset. Our approach is inspired by recent results in the optimization literature on data-driven distributionally robust optimization [27]. We formulate a tractable data-driven distributionally robust optimal power flow problem where the Wasserstein metric [28] is used to quantify differences between an empirical data-based distribution and the real data-generating distribution. In contrast to previous work, we obtain power schedules and reserve policies that are explicitly robust to sampling errors inherent in the dataset.

¹Out-of-sample performance is an evaluation of the optimal decisions using a dataset that is different from the one used to obtain the decision, which can be tested with Monte Carlo simulation

[†]Y. Guo and T. Summers are with the Department of Mechanical Engineering, The University of Texas at Dallas, Richardson, TX, USA, email: {yi.guo2,tyler.summers}@utdallas.edu. [‡]K. Baker is with the Department of Civil, Environmental, and Architectural Engineering, The University of Colorado, Boulder, CO, USA, email:kyri.baker@colorado.edu. ^{*}E. Dall’Anese is with the National Renewable Energy Laboratory, Golden, CO, USA, email: emiliano.dallanese@nrel.gov. ^{**}Z. Hu is with the Department of Electrical Engineering, Tsinghua University, Beijing, P.R. China, email: zechunhu@tsinghua.edu.cn. This material is based on work supported by the National Science Foundation under grant CNS-1566127.

This approach achieves superior out-of-sample performance guarantees in comparison to other stochastic optimization approaches, effectively regularizing against overfitting the decisions to limited available data. A numerical example based on a modified IEEE 118 bus test network illustrates inherent tradeoffs between operational cost and risk of constraint violation, and we show how our proposed method offers a data-driven framework to systematically balance these objectives. The results also point out that making inaccurate assumptions about probability distributions can cause underestimation of risk.

The rest of the paper is organized as follows: Section II describes the network model and the stochastic OPF problem. Section III describes the data-driven distributionally robustness optimization framework and formulates the data-driven stochastic OPF problem. Section IV provides simulation results and discussion. Section V concludes the paper.

Notation: The inner product of two vectors $a, b \in \mathbb{R}^m$ is denoted by $\langle a, b \rangle := a^T b$. The N -fold product of distribution \mathbb{P} on a set Ξ is denoted by \mathbb{P}^N , which represents a distribution on the Cartesian product space $\Xi_N = \Xi \times \dots \times \Xi$. Superscript “ $\hat{\cdot}$ ” is reserved for the objects that depend on a training dataset $\hat{\Xi}_N$. We use $(\cdot)'$ to denote vector or matrix transpose and A^T to denote matrix A to the power T .

II. NETWORK MODEL AND OPTIMAL POWER FLOW

We consider N_d devices connected via a transmission network within a planning time horizon T discrete time steps. The elements connected to the transmission system may include 1) traditional generators; 2) fixed, deferrable, and curtailable loads; 3) storage devices like batteries and plug-in electric vehicles, which are able to act as both generators and loads. In this case, we model two types of devices: devices (e.g., conventional thermal and hydro generators, deferrable/curtailable loads and storage devices) with controllable power flow affected by decision variables; and devices (e.g., flexible and intermittent renewable energy or fixed loads) with fixed or uncertain power flow which will not be affected by decision variables.

A. Devices with Controllable Power Flow

The power flow of each controllable device is modeled with a discrete-time linear dynamical system

$$x_{t+1}^j = \bar{A}_j x_t^j + \bar{B}_j u_t^j, \quad (1)$$

where device j at time t has internal state $x_t^j \in \mathbb{R}^{n_j}$, dynamics matrix $\bar{A}_j \in \mathbb{R}^{n_j \times n_j}$, input matrix $\bar{B}_j \in \mathbb{R}^{n_j \times m_j}$, and control input $u_t^j \in \mathbb{R}^{m_j}$. The first element of x_t^j corresponds to the power injection of device j at time t into the network at its bus, and other elements describe internal dynamics such as state-of-charge (SOC) of energy storage devices. States and inputs over the planning horizon could be expressed in the following notation: $\mathbf{x}^j = [(x_1^j)', \dots, (x_T^j)']' \in \mathbb{R}^{n_j T}$ and $\mathbf{u}^j = [(u_0^j)', \dots, (u_{T-1}^j)']' \in \mathbb{R}^{m_j T}$. We can then write compactly:

$$\mathbf{x}^j = A_j x_0^j + B_j \mathbf{u}^j, \quad (2)$$

where

$$A_j := \begin{bmatrix} \bar{A}_j \\ \bar{A}_j^2 \\ \vdots \\ \bar{A}_j^T \end{bmatrix}, B_j := \begin{bmatrix} \bar{B}_j & 0 & \dots & 0 \\ \bar{A}_j \bar{B}_j & \bar{B}_j & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \bar{A}_j^{T-1} \bar{B}_j & \dots & \bar{A}_j \bar{B}_j & \bar{B}_j \end{bmatrix}.$$

B. Devices with Uncontrollable Power Flow

The fixed power flow for device j is given by $r_j + G_j \xi$ with positive values denoting net power injection into the network. The vector $r_j \in \mathbb{R}^T$ denotes the predicted power injection over the planning horizon, and the matrix $G_j \in \mathbb{R}^{T \times N_\xi T}$ maps the random vector $\xi \in \Xi \subseteq \mathbb{R}^{N_\xi T}$ to a prediction error of the power injection or extraction for device j . If uncertainty of device j is not explicitly modeled, then $G_j = 0$.

We will assume that the probability distribution \mathbb{P} of ξ is unknown and observed only through a finite dataset $\hat{\Xi}_N$.

C. Cost Functions and Constraints

An operating objective of our problem is to minimize a sum of cost functions associated with controllable devices $J_j : \mathbb{R}^{n_j T} \times \mathbb{R}^{m_j T} \rightarrow \mathbb{R}$. The cost functions are modeled as convex quadratic:

$$J_j(\mathbf{x}^j, \mathbf{u}^j) := f_{jx}' \mathbf{x}^j + \frac{1}{2} \mathbf{x}^{j'} H_{jx} \mathbf{x}^j + f_{ju}' \mathbf{u}^j + \frac{1}{2} \mathbf{u}^{j'} H_{ju} \mathbf{u}^j + c_j, \quad (3)$$

with H_{jx} and H_{ju} being positive semidefinite matrices.

We consider: 1) power balance constraints, 2) power line flow constraints and 3) local device constraints. The power balance constraints are

$$\sum_{j=1}^{N_d} (r_j + G_j \xi + C_j \mathbf{x}^j) = 0, \quad (4)$$

which ensures that the generation and consumption of power inside the transmission system are balanced over the planning horizon. The matrix C_j selects the first element of \mathbf{x}^j . In this paper, we employ a widely used linearization of the nonlinear AC power flow equations, in which it is assumed that lines are lossless, normalized voltage magnitudes are close to unity, and voltage angle differences are small [14], [18]. In this case, the line flow are linear functions of nodal injections.

The L bi-directional transmission lines power flow constraints can be expressed as

$$\sum_{j=1}^{N_d} \Gamma_j (r_j + G_j \xi + C_j \mathbf{x}^j) \leq \bar{p}, \quad (5)$$

where $\Gamma_j \in \mathbb{R}^{2LT \times T}$ maps the power injection or extraction of each device to its contribution to each connected line and can be constructed from network line impedances, and \bar{p} denotes nominal line flow limits. The local constraints are also modeled as linear inequalities of the form:

$$T_j \mathbf{x}^j + U_j \mathbf{u}^j + Z_j \xi \leq w_j. \quad (6)$$

where $T_j \in \mathbb{R}^{l_j \times n_j T}$, $U_j \in \mathbb{R}^{l_j \times m_j T}$, and $Z_j \in \mathbb{R}^{l_j \times N_\xi T}$, and $w_j \in \mathbb{R}^{l_j}$ is a local constraint parameter vector. These constraints can be used to model allowable power injection ranges and other device limits.

D. Reserve Policies

Deterministic OPF formulations ignore the prediction error ξ and compute an open-loop input sequence for each device. In a stochastic setting, one must optimize over causal *policies* $\mathbf{u}^j = \pi_j(\xi)$, where $\pi_j : \mathbb{R}^{N_s T} \rightarrow \mathbb{R}^{m_j T}$ is a measurable function that specifies how each device should respond to forecast errors as they are discovered. We can now formulate a finite horizon stochastic optimization problem

$$J_{Cost} = \inf_{\pi_j \in \Pi_c} \mathbb{E} \sum_{j=1}^{N_d} J_j(A_j x_0^j + B_j \pi_j(\xi), \pi_j(\xi)),$$

subject to

$$\sum_{j=1}^{N_d} (r_j + G_j \xi + C_j(A_j x_0^j + B_j \pi_j(\xi))) = 0, \forall \xi,$$

$$\mathbb{E} f_1 \left(\sum_{j=1}^{N_d} \Gamma_j (r_j + G_j \xi + C_j(A_j x_0^j + B_j \pi_j(\xi))) - \bar{p} \right) \leq 0,$$

$$\mathbb{E} f_2 \left(T_j(A_j x_0^j + B_j \pi_j(\xi)) + U_j \pi_j(\xi) - w_j \right) \leq 0, j = 1, \dots, N_d. \quad (7)$$

where f_1 and f_2 denote general constraint risk functions, which in our case will depend on forecast error data and auxiliary optimization variables. Also, the cost function is proportional to the first and second moments of the uncertainties ξ , because each device cost function are convex quadratic. The problem (7) is infinite dimensional, so we restrict attention to affine policies:

$$\mathbf{u}^j = D_j \xi + e_j. \quad (8)$$

so that each participant device j (e.g., traditional generators, flexible loads or energy storage) power schedule \mathbf{u}^j is parameterized by a nominal schedule $e_j = [e_0^j, \dots, e_{T-1}^j]'$ plus a linear function D_j of prediction error realizations. To obtain causal policies, D_j must be lower-triangular. The D_j matrices can be interpreted in terms of planned Automatic Generation Control (AGC) parameters [14]. Under affine policies, the power balance constraints are linear functions of the distribution of ξ , which are equivalent to

$$\sum_{j=1}^{N_d} (r_j + C_j(A_j x_0^j + B_j e_j)) = 0, \sum_{j=1}^{N_d} (G_j + C_j B_j D_j) = 0. \quad (9)$$

III. DATA-DRIVEN DISTRIBUTIONALLY ROBUST OPTIMIZATION VIA THE WASSERSTEIN METRIC

There is a variety of ways to reformulate the general stochastic OPF problem (7) to obtain tractable problems that can be solved by standard convex optimization solvers. These include assuming specific functional forms for the forecast error distribution (e.g., Gaussian) and using specific constraint risk functions, such as those encoding value at risk (i.e., chance constraints), conditional value at risk, distributional robustness, and support robustness. In all cases, the out-of-sample performance of the resulting decisions in operational practice ultimately relies on the quality of data describing

the forecast errors and the validity of assumptions made about probability distributions. Many existing approaches make either too strong or too weak assumptions that lead to underestimation or overestimation of risk. In this section, we review a recently proposed tractable method for data-driven distributionally robust optimization [27]. We then use it to formulate a data-driven distributionally robust OPF problem that is based exclusively on a finite training dataset $\hat{\Xi}_N$, is robust to sampling errors, gives explicit control of conservativeness, and offers superior out-of-sample performance guarantees. This is accomplished by optimizing over the worst-case distribution within a Wasserstein ball in the space of probability distributions centered at the uniform distribution on the set of training data. This turns out to admit a tractable reformulation for certain risk functions, including CVaR.

A. Data-Driven Stochastic Programming

Consider the stochastic program

$$J^* := \inf_{y \in \mathbb{Y}} \left\{ \mathbb{E}^{\mathbb{P}} [h(y, \xi)] := \int_{\Xi} h(y, \xi) \mathbb{P}(d\xi) \right\}, \quad (10)$$

with decision variable $y \in \mathbb{Y} \subseteq \mathbb{R}^n$, random vector ξ with probability distribution \mathbb{P} supported on $\Xi \subseteq \mathbb{R}^m$, and cost function $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$. In virtually all practical settings, the distribution \mathbb{P} is unknown, so the problem is in some sense ill-posed, since it is not even possible to evaluate the cost function for a given decision. We typically have only a finite training dataset of N independent samples observed from \mathbb{P} , which we denote by $\hat{\Xi}_N := \{\hat{\xi}_i\}_{i \leq N} \subseteq \Xi$.

The out-of-sample performance of a decision $\hat{y} \in \mathbb{Y}$ is $\mathbb{E}^{\mathbb{P}} [h(\hat{y}, \xi)]$, which can be interpreted as the expected cost of \hat{y} under a new sample of ξ that is independent of the training dataset. In any stochastic optimization problem, we seek a decision with good out-of-sample performance. A common approach is to approximate \mathbb{P} based on the training data $\hat{\Xi}_N$, effectively replacing \mathbb{P} with a discrete empirical distribution defined by the uniform distribution on $\hat{\Xi}_N$, which is denoted $\hat{\mathbb{P}}_N$. However, this can cause decisions to be overly attuned to the data and lead to poor out-of-sample performance, especially when data is high-dimensional, limited in quantity, and expensive to obtain. The same is true for other approaches that utilize compressed summaries of $\hat{\Xi}_N$ (e.g., mean and variance) or assume a specific distribution that generated $\hat{\Xi}_N$.

Instead, we adopt an approach that explicitly accounts for our ignorance of true data-generating distribution \mathbb{P} , is based explicitly on the available data $\hat{\Xi}_N$, and provides out-of-sample performance guarantees [27]. A set $\hat{\mathcal{P}}_N$ of distributions is constructed from $\hat{\Xi}_N$ that contains those that could most plausibly have generated $\hat{\Xi}_N$. One can then formulate a distributionally robust optimization (DRO) problem:

$$\hat{J}_{DRO} := \inf_{y \in \mathbb{Y}} \sup_{\mathbb{Q} \in \hat{\mathcal{P}}_N} \mathbb{E}^{\mathbb{Q}} [h(y, \xi)], \quad (11)$$

where the objective is to minimize the expected cost function with respect to the *worst-case* data-generating distribution contained in $\hat{\mathcal{P}}_N$.

B. The Wasserstein Metric

The ambiguity set $\hat{\mathcal{P}}_N$ is defined using the Wasserstein metric, which defines a distance in the space $\mathcal{M}(\Xi)$ of all probability distributions \mathbb{Q} supported on Ξ with $\mathbb{E}^{\mathbb{Q}}[\|\xi\|] = \int_{\Xi} \|\xi\| \mathbb{Q}(d\xi) < \infty$.

Definition (Wasserstein Metric). Denote by \mathcal{L} the space of all Lipschitz continuous functions $f: \Xi \rightarrow \mathbb{R}$ with Lipschitz constant less than or equal to 1. The Wasserstein metric $d_W: \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \rightarrow \mathbb{R}$ is defined as

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) = \sup_{f \in \mathcal{L}} \left(\int_{\Xi} f(\xi) \mathbb{Q}_1(d\xi) - \int_{\Xi} f(\xi) \mathbb{Q}_2(d\xi) \right), \quad \forall \mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{M}(\Xi). \quad (12)$$

Intuitively, the Wasserstein metric quantifies the minimum “transportation” cost to move mass from one distribution to another. We can now use the Wasserstein metric to define the ambiguity set

$$\hat{\mathcal{P}}_N := \left\{ \mathbb{Q} \in \mathcal{M}(\Xi) : d_w(\hat{\mathbb{P}}_N, \mathbb{Q}) \leq \varepsilon \right\}, \quad (13)$$

which contains all distributions within a Wasserstein ball of radius ε centered at the empirical distribution $\hat{\mathbb{P}}_N$. The radius ε can be chosen so that the ball contains the true distribution \mathbb{P} with a prescribed confidence level and leads to performance guarantees [27]. The radius ε also explicitly controls the conservativeness of the resulting decision; large ε will produce decisions that rely less on the specific features of the dataset $\hat{\Xi}_N$ and give better robustness to sampling errors.

C. Distributionally Robust Optimization of CVaR

The distributionally robust optimization problem (11) does not seem any easier than (10), and seems quite possibly even worse. However, it was shown in [27] that the (11) admits a tractable reformulation and in fact for piecewise affine cost functions of the form $h(y, \xi) = \max_{k \leq K} \langle a_k(y), \xi \rangle + b_k$ reduces to a linear program. One important special case is when the cost function corresponds to the conditional value at risk of a linear function $\langle \bar{a}(y), \xi \rangle + \bar{b}$ with confidence level $\alpha \in (0, 1]$ [29], where $K = 2$ written as the piecewise affine function formation $h(y, \xi)$. Then the cost function takes the form

$$\begin{aligned} J^* &= \inf_{y \in \mathbb{Y}} \left\{ \text{CVaR}_{\alpha}(\langle \bar{a}(y), \xi \rangle + \bar{b}) \right\}, \\ &= \inf_{y \in \mathbb{Y}} \left\{ \inf_{\tau \in \mathbb{R}} \mathbb{E}^{\mathbb{P}} \left[-\tau + \frac{1}{\alpha} \max\{\langle \bar{a}(y), \xi \rangle + \bar{b} + \tau, 0\} \right] \right\}, \\ &= \inf_{y \in \mathbb{Y}, \tau \in \mathbb{R}} \mathbb{E}^{\mathbb{P}} \left[\max_{k=1,2} \langle a_k(y), \xi \rangle + b_k(\tau) \right], \end{aligned} \quad (14)$$

Suppose the support of \mathbb{P} is defined by $\Xi := \{\xi \in \mathbb{R}^m : H\xi \leq d\}$. Then the general form of the distributionally robust optimization problem for CVaR

$$\hat{J}_{DRO} = \inf_{y \in \mathbb{Y}, \tau \in \mathbb{R}} \sup_{\mathbb{Q} \in \hat{\mathcal{P}}_N} \mathbb{E}^{\mathbb{Q}} \left[\max_{k=1,2} \langle a_k(y), \xi \rangle + b_k \right], \quad (15)$$

as shown in [27], can be equivalently reformulated as the linear program

$$\inf_{\lambda, s_i, \gamma_{ik}, y, \tau} \lambda \varepsilon + \frac{1}{N_s} \sum_{i=1}^{N_s} s_i, \quad (16)$$

subject to

$$\begin{aligned} b_k + \langle a_k(y), \hat{\xi}_i \rangle + \langle \gamma_{ik}, d - H\hat{\xi} \rangle &\leq s_i, \quad \forall i \leq N_s, \forall k = 1, 2, \\ \|H' \gamma_{ik} - a_k(y)\|_{\infty} &\leq \lambda, \quad \forall i \leq N_s, \forall k = 1, 2, \\ \gamma_{ik} &\geq 0, \quad \forall i \leq N_s, \forall k = 1, 2, \end{aligned}$$

where $\varepsilon \geq 0$. For $\varepsilon \rightarrow 0$, the optimization results of (16) correspond to the expectation of the loss function $h(y, \hat{\xi})$ under the empirical distribution, where $\frac{1}{N_s} \sum_{i=1}^{N_s} s_i$ represents the sample average of the loss function $h(y, \hat{\xi})$.

D. Distributionally Robust Stochastic OPF with CVaR

We now use the above developments to formulate a data-driven distributionally robust OPF problem to balance an operating efficiency metric with CVaR values of line flow and local device constraint violations. Considering the stochastic OPF shown in (7), the m -th power line flow constraints (5) can be written in the form

$$g_m(D, e, \xi) = [\Phi(D)]_m \xi + [b(e)]_m, \quad (17)$$

where $[\cdot]_m$ denotes the m -th elements of a vector or m -th row of a matrix. A similar form for the n -th local devices constraint can be obtained, which we denote as $h_{jq}(D, e, \xi)$. The decisions variables D and e both appear linearly in g_m and h_{jq} . The CVaR of the line flow and local devices constraints are then given by [18]

$$\begin{aligned} \mathbb{E}[g_m(D, e, \xi) + \tau_m]_+ - \tau_m \alpha &\leq 0, \quad m = 1, \dots, 2LT \\ \mathbb{E}[h_{jq}(D, e, \xi) + \tau_{jq}]_+ - \tau_{jq} \alpha &\leq 0, \quad j = 1, \dots, N_d, q = 1, \dots, l_j. \end{aligned} \quad (18)$$

We collect the line flow and local device constraints into a list of $V = 2LT + \sum_{j=1}^{N_d} l_j$ elements. The first $2LT$ elements are line flow constraints, and the remaining $\sum_{j=1}^{N_d} l_j$ elements are local device constraints. The CVaR value of the v -th constraint in the above list is denoted by $\Theta_v(D, e, \xi, \tau) = \max_{k=1,2} \langle a_{vk}(y), \xi \rangle + b_{vk}(\tau)$, where decision vector y consists of optimization variables D, e . For $v = 1, \dots, 2LT$, the parameters of the line constraints CVaR Θ_v are as follows, with ξ redefined here as $[\xi, 1]'$

$$\begin{aligned} a_{v1}(y) &= \left[\sum_{j=1}^{N_d} \Gamma_j C_j B_j D_j, \sum_{j=1}^{N_d} \Gamma_j C_j B_j e_j \right] \\ b_{v1}(\tau) &= -\bar{p} + \sum_{j=1}^{N_d} \Gamma_j (r_j + G_j \xi + C_j A_j x_0^j) + \tau_v - \tau_v \alpha, \\ a_{v2}(y) &= [0, 0], \quad b_{v2}(\tau) = -\tau_v \alpha \end{aligned}$$

For $v = 2LT + 1, \dots, V$, the CVaR of the local device constraints Θ_v can be defined analogously to obtain similar forms for $a_{v1}(y), b_{v1}(\tau)$ and $a_{v2}(y), b_{v2}(\tau)$.

Based on the above, we now formulate a distributionally robust OPF problem (19), which minimizes a weighted sum

of operational cost and the CVaR of the line and local device constraint violations:

$$\begin{aligned}
\hat{J}_{OPF} &= \hat{J}_{Cost} + \rho \hat{J}_{DRO} \\
&= \inf_{\substack{y \in \mathbb{Y} \\ \tau \in \mathbb{R}}} \left\{ \mathbb{E} \sum_{j=1}^{N_d} J_j(\hat{\xi}) + \rho \sup_{Q \in \hat{\mathcal{P}}_N} \sum_{v=1}^V \mathbb{E}^Q[\Theta_v(D, e, \xi, \tau)] \right\}, \\
&= \inf_{\substack{y \in \mathbb{Y}, \tau \in \mathbb{R}, \\ \lambda_v, s_{iv}, \gamma_{ikv}}} \left\{ \mathbb{E} \sum_{j=1}^{N_d} J_j(\hat{\xi}) + \sum_{v=1}^V \left(\lambda_v \varepsilon + \frac{1}{N_s} \sum_{i=1}^{N_s} s_{iv} \right) \right\},
\end{aligned} \tag{19}$$

subject to:

$$\begin{aligned}
\sum_{j=1}^{N_d} (r_j + C_j(A_j x_0^j + B_j e_j)) &= 0, \\
\sum_{j=1}^{N_d} (G_j + C_j B_j D_j) &= 0, \\
\rho(b_{kv}(\tau) + \langle a_{kv}(y), \hat{\xi}_i \rangle + \langle \gamma_{ikv}, d - H \hat{\xi}_i \rangle) &\leq s_{iv}, \\
\|H' \gamma_{ikv} - \rho a_{kv}(y)\|_{\infty} &\leq \lambda_v, \\
\gamma_{ikv} &\geq 0, \\
\forall i \leq N_s, \forall v \leq V, k = 1, 2
\end{aligned}$$

where $\rho \in \mathbb{R}_+$ quantifies the power system operators' risk aversion. This is a quadratic program that explicitly uses the training dataset $\hat{\Xi}_N := \{\hat{\xi}_i\}_{i \leq N}$. The risk aversion parameter ρ and the Wasserstein radius ε allow us to explicitly balance tradeoffs between efficiency, risk, and sampling errors inherent in $\hat{\Xi}_N$.

E. Data-Driven Distributionally Robust Model Predictive Control

The proposed distributionally robust optimization framework can be utilized as a component of a multi-stage closed-loop model predictive control (MPC) strategy. This strategy allows explicit incorporation of historical datasets and forecasting techniques that can update predictions during operation. MPC is a feedback control technique that implements optimal open-loop control decisions at the current time step t -th while considering future system states over a fixed time horizon \mathcal{H}_t [30]–[33]. In the above present problem, the distributionally robust stochastic OPF (19) could be regarded as a building block for a MPC-based closed-loop control strategy, which includes the current and future system states based on the forecasts of the renewable energy availability and the local devices dynamics. The MPC-based control strategy involves the following steps:

- At time step t , forecast the uncertainties across the power network over horizon \mathcal{H}_t .
- Solve (19) over horizon \mathcal{H}_t .
- Implement the reserve control policies for each device at current time step t .
- Move to time step $t + 1$, and return to the first step.

The optimal reserve control policies are computed for the entire time horizon \mathcal{H}_t , but only the decisions for the current

time step are implemented. Then we shift to next time step with the forecasts updated. To simplify the exposition, our numerical results will focus on solving (19) in one particular time step. We are pursuing and elaborating on a multi-stage data-driven distributionally robust MPC algorithm and properties in ongoing work.

IV. NUMERICAL EXAMPLE

To illustrate the proposed framework, we use a modified IEEE 118 bus test network as defined in [34] shown in Fig.1. A wind farm is connected to bus 9 with nominal infeed of 1000 MW. To understand the trade offs of the proposed distributionally robust optimization methodology, we consider only the line flow constraint between bus 8 and bus 9, marked by a cross in Fig.1. The maximum rating on the power line between bus 8 and 9 is 950 MW. Other line flow constraints, security constraints, and voltage constraints can also be easily incorporated, and we are pursuing this in ongoing work. The wind power forecast errors are derived from real wind data from the hourly wind power measurements provided in the 2012 Global Energy Forecasting competition (GEFCom2012) [35]. The wind power forecast errors are based on the so-called persistence forecast, which predicts the wind power output at the next time step to be equal to that at the previous time step. It can be seen that the forecast errors are highly leptokurtic, i.e., that the errors are have outliers that make the distribution tails much heavier than Gaussian tails. Since the wind power data from GEFCom2012 are normalized in $[0,1]$, we scale the forecast errors to have zero mean and the standard deviation $\sigma = 300$ MW. In this case study, we only consider a single line constraint, hence no other local device constraint is included. Additionally, we do not assume a bound on the wind power forecast errors ξ , therefore, the H and d in (19) are equal to zero. α is set to be 0.05 in these case studies.

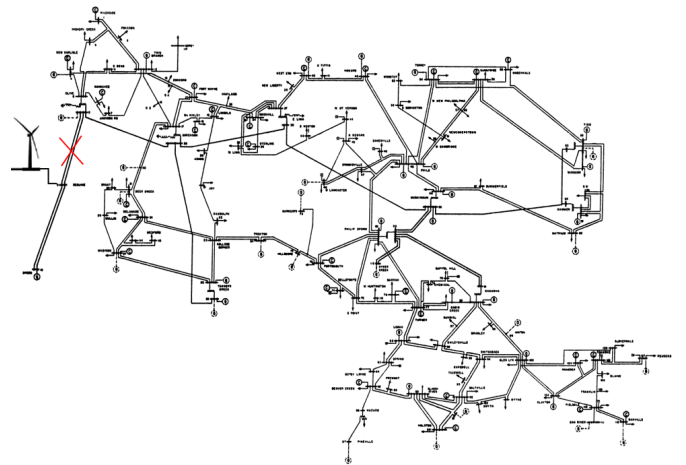


Fig. 1. IEEE 118 bus test network.

For comparison purposes, we also introduce two other methods which we call CVaR OPF and Gaussian OPF to evaluate the constraint violations. In CVaR OPF, we use a basic sample average to approximate CVaR of the

line constraint violations with training dataset $\hat{\Xi}_{N_s}$, which corresponds to setting $\varepsilon = 0$ in (19). In Gaussian OPF, we assume the data are generated by a Gaussian distribution with the sample mean and variance from dataset $\hat{\Xi}_{N_s}$. We subsample $N_s = 100$ elements from the training dataset $\hat{\Xi}_{N_s}$.

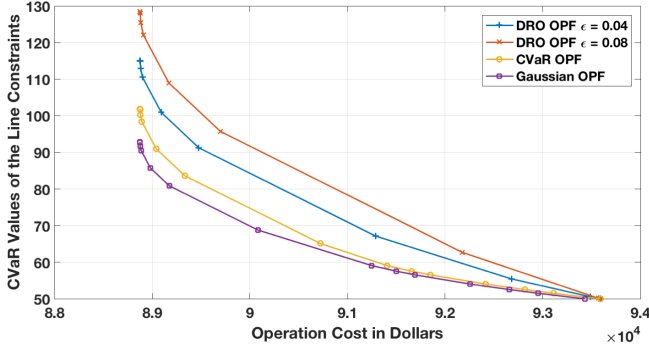


Fig. 2. Predicted tradeoff between operating cost and CVaR of line constraint violations, based on different models of data uncertainty. The Gaussian model underestimates the risk, as demonstrated by the out-of-sample evaluations in Fig.3.

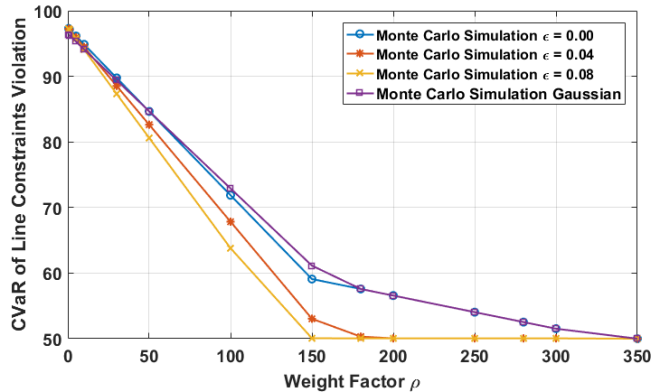


Fig. 3. Out-of-sample performance is superior to other approaches and conservativeness is explicitly controllable, as demonstrated by Monte Carlo simulations. For each value of ρ we subsampled new values from the training dataset and computed the empirical CVaR of the line constraint.

Fig.2 visualizes the solutions of the proposed stochastic OPF by three methods: CVaR OPF, Gaussian OPF and DRO OPF ($\varepsilon = 0.04, 0.08$). The results illustrate the tradeoffs between optimal operational cost and CVaR of line constraint violations by varying the risk aversion parameter ρ . The tradeoffs could be explicitly adjusted by ρ according to the system operators' efficiency and risk tradeoff preferences.

Fig.2 also demonstrates the underestimation of the risk of constraint violations by assuming a Gaussian distribution. Distributionally robust optimization considers the worst-case scenario of the uncertainties to produce more conservative decisions on generator power adjustments and reserve policies. The conservativeness of the results are controllable by adjusting the Wasserstein distance ε . In this paper, we presents two cases $\varepsilon = 0.04, \varepsilon = 0.08$. To emphasize, the

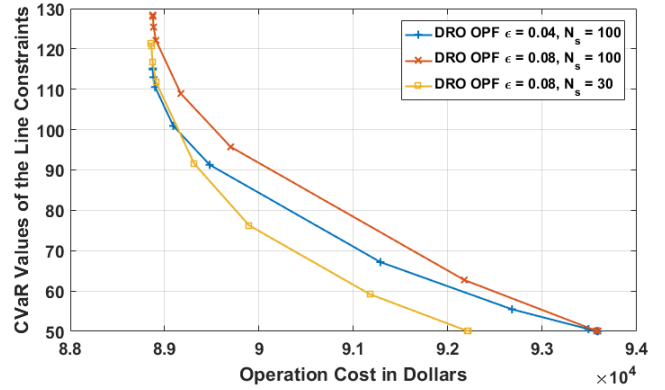


Fig. 4. Sample errors can lead to the underestimation of the risk of constraint violations when the data contain less variance than the real distribution, but their effect is diminished by the data-driven distributionally robust approach.

results of DRO OPF ($\varepsilon = 0.00$) overlap with the curve given by CVaR OPF in Fig.2. Increasing ε provides better robustness to sampling errors and yields superior out-of-sample performance guarantees.

Fig.3 visualizes the out-of-sample performance guarantee of stochastic OPF solution offered by distributionally robust model based on Monte Carlo simulation. The number of scenarios in Monte Carlo Simulation is 1000. For each value of ρ we subsampled new values from the training dataset and computed the empirical CVaR of the line constraint. The decisions from DRO OPF method ensure smaller line constraint violation for all values of the risk aversion ρ . Again, we also see that increasing the Wasserstein radius ε provides lower risk. The benefits saturate for small ρ .

In the DRO OPF case, we also re-solved stochastic OPF problem with reduced size of training dataset to $N_s = 30$. Fig.4 visualizes the effects of sampling errors. We see in this case that the smaller dataset has caused an underestimation of risk, since the reduced data contains less variance than the full data. Sampling errors can cause either underestimation or overestimation of actual risks depending on whether the dataset features more or less variation than the real data generating distribution, but their effect is diminished with the proposed distributionally robust approach.

V. CONCLUSIONS

In this paper, we have proposed a distributionally robust optimization method to solve a stochastic OPF problem with the limited knowledge of the uncertain forecast error across the power system network. By explicitly quantifying the differences between the empirical data-based distributions and real data-generating distributions using the Wasserstein metric, we computed power schedules and reserve policies which were robust to sampling errors. This approach has superior out-of-sample performance guarantees in comparison with other approaches and allows a more systematic trade off of efficiency and risk. A simple numerical example illustrated these basic tradeoffs. The results also demonstrated that making inaccurate assumptions about probability distributions

can underestimate the risk of constraint violations compared to the distributionally robust optimization. Ongoing and future work includes case studies in larger networks and adapting the approach to handle voltage constraints in distribution networks using alternative linearizations of the power flow equations, as in [36].

REFERENCES

- [1] J. Carpentier, "Contribution to the economic dispatch problem," *Bulletin de la Societe Francaise des Electriciens*, vol. 3, no. 8, pp. 431–447, 1962.
- [2] H. W. Dommel and W. F. Tinney, "Optimal power flow solutions," *IEEE Transactions on power apparatus and systems*, no. 10, pp. 1866–1876, 1968.
- [3] O. Alsac and B. Stott, "Optimal load flow with steady-state security," *IEEE transactions on power apparatus and systems*, no. 3, pp. 745–751, 1974.
- [4] R. Baldick, B. H. Kim, C. Chase, and Y. Luo, "A fast distributed implementation of optimal power flow," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 858–864, 1999.
- [5] S. H. Low, "Convex relaxation of optimal power flow part i: Formulations and equivalence," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.
- [6] S. H. Low, "Convex relaxation of optimal power flow part ii: Exactness," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 2, pp. 177–189, 2014.
- [7] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*, vol. 7. McGraw-hill New York, 1994.
- [8] T. Yong and R. H. Lasseter, "Stochastic optimal power flow: formulation and solution," in *Power Engineering Society Summer Meeting, 2000. IEEE*, vol. 1, pp. 237–242, IEEE, 2000.
- [9] F. Capitanescu, S. Fliscounakis, P. Panciatici, and L. Wehenkel, "Cautious operation planning under uncertainties," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1859–1869, 2012.
- [10] A. J. Conejo, M. Carrion, and J. M. Morales, *Decision making under uncertainty in electricity markets*, vol. 1. Springer, 2010.
- [11] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM Review*, vol. 56, no. 3, pp. 461–495, 2014.
- [12] M. Vrakopoulou, M. Katsampani, K. Margellos, J. Lygeros, and G. Andersson, "Probabilistic security-constrained ac optimal power flow," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*, pp. 1–6, IEEE, 2013.
- [13] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "A probabilistic framework for reserve scheduling and n-1 security assessment of systems with high wind power penetration," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3885–3896, 2013.
- [14] J. Warrington, P. Goulart, S. Mariéthoz, and M. Morari, "Policy-based reserves for power systems," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4427–4437, 2013.
- [15] Y. Zhang and G. B. Giannakis, "Robust optimal power flow with wind integration using conditional value-at-risk," in *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*, pp. 654–659, IEEE, 2013.
- [16] M. Perninge and C. Hamon, "A stochastic optimal power flow problem with stability constraints-part ii: the optimization problem," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1849–1857, 2013.
- [17] L. Roald, F. Oldewurtel, T. Krause, and G. Andersson, "Analytical reformulation of security constrained optimal power flow with probabilistic constraints," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*, IEEE, 2013.
- [18] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on conditional value at risk and distributional robustness," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 116–125, 2015.
- [19] R. A. Jabr, S. Karaki, and J. A. Korbane, "Robust multi-period opf with storage and renewables," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2790–2799, 2015.
- [20] L. Roald, S. Misra, T. Krause, and G. Andersson, "Corrective control to handle forecast uncertainty: A chance constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1626–1637, 2017.
- [21] L. Roald, F. Oldewurtel, B. Van Parys, and G. Andersson, "Security constrained optimal power flow with distributionally robust chance constraints," *arXiv preprint arXiv:1508.06061*, 2015.
- [22] B. Li, R. Jiang, and J. L. Mathieu, "Distributionally robust risk-constrained optimal power flow using moment and unimodality information," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pp. 2425–2430, IEEE, 2016.
- [23] H. Zhang and P. Li, "Chance constrained programming for optimal power flow under uncertainty," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2417–2424, 2011.
- [24] M. Lubin, Y. Dvorkin, and S. Backhaus, "A robust approach to chance constrained optimal power flow with renewable generation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3840–3849, 2016.
- [25] K. Baker, E. Dall'Anese, and T. Summers, "Distribution-agnostic stochastic optimal power flow for distribution grids," in *North American Power Symposium (NAPS), 2016*, pp. 1–6, IEEE, 2016.
- [26] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed optimal power flow for smart microgrids," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, 2013.
- [27] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations," *arXiv preprint arXiv:1505.05116*, 2015.
- [28] C. R. Givens and R. M. Shortt, "A class of wasserstein metrics for probability distributions," *The Michigan Mathematical Journal*, vol. 31, no. 2, pp. 231–240, 1984.
- [29] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.
- [30] E. F. Camacho and C. B. Alba, *Model predictive control*. Springer Science & Business Media, 2013.
- [31] P. Fortenbacher, A. Ulbig, S. Koch, and G. Andersson, "Grid-constrained optimal predictive power dispatch in large multi-level power systems with renewable energy sources, and storage devices," in *Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES*, pp. 1–6, IEEE, 2014.
- [32] F. Adamek, M. Arnold, and G. Andersson, "On decisive storage parameters for minimizing energy supply costs in multicarrier energy systems," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 1, pp. 102–109, 2014.
- [33] A. Parisio, E. Rikos, and L. Glielmo, "Stochastic model predictive control for economic/environmental operation management of microgrids: An experimental case study," *Journal of Process Control*, vol. 43, pp. 24–37, 2016.
- [34] Electrical and T. R. Computer Engineering Department, Illinois Institute of Technology, "Ieee 118-bus, 54-unit, 24-hour system."
- [35] "Global energy forecasting competition 2012 wind energy competition," 2012.
- [36] E. Dall'Anese, K. Baker, and T. Summers, "Chance-constrained ac optimal power flow for distribution systems with renewables," *IEEE Transactions on Power Systems*, 2017.