

# Distribution-Agnostic Stochastic Optimal Power Flow for Distribution Grids

Kyri Baker, *Member, IEEE*, Emiliano Dall'Anese, *Member, IEEE* and Tyler Summers, *Member, IEEE*

**Abstract**—This paper outlines a data-driven, distributionally robust approach to solve chance-constrained AC optimal power flow problems in distribution networks. Uncertain forecasts for loads and power generated by photovoltaic (PV) systems are considered, with the goal of minimizing PV curtailment while meeting power flow and voltage regulation constraints. A data-driven approach is utilized to develop a distributionally robust conservative convex approximation of the chance-constraints; particularly, the mean and covariance matrix of the forecast errors are updated online, and leveraged to enforce voltage regulation with predetermined probability via Chebyshev-based bounds. By combining an accurate linear approximation of the AC power flow equations with the distributionally robust chance constraint reformulation, the resulting optimization problem becomes convex and computationally tractable.

**Index Terms**—Distribution systems; optimal power flow; chance constraints; renewable integration; voltage regulation.

## I. INTRODUCTION

In 2014, installed solar power in the United States grew by 30%, amounting to 6.2 gigawatts of capacity. In the residential sector alone, over a gigawatt of capacity was added, making residential PV the fastest growing portion of the US solar sector [1]. With this massive increase in renewables, many issues in the electric power grid can be observed, such as reverse power flows, voltage fluctuations, and overloaded power lines [2], [3]. With the utilization of advanced optimization and control schemes, however, these issues can be mitigated [4]–[6]. Specifically, with measurements drawn from PV systems and load over time, advanced inverter control, and probabilistic constraints, this paper will develop a control scheme that attempts to minimize the curtailment of PV while adhering to power flow and voltage constraints in a given distribution network.

In this paper, we solve a chance-constrained voltage regulation problem on a five-minute dispatch scale in a distribution grid with a high penetration of PV, where the chance constraints are updated as measurements are taken from the PV and load present in the grid. Here, measurements are made of the actual observed irradiance and load, and the first and second moments of the error distributions are updated over time by using a sample average approximation [7]. Due to the lack of knowledge of the underlying probability distributions

of every renewable source and load in the system, we consider a distribution agnostic, or distribution free, approach to solving the chance constrained OPF problem in distribution grids.

Convex approximations of chance constraints have been developed in the literature, some of which require a priori exact knowledge of the probability distribution or its parameters [8]–[10]. Because of this limitation, data driven chance constraints have been receiving recent attention [11]–[14] for their ability to cope with the lack of exact distributional information present in many real-world instances. In [11] and [15], a distributionally robust chance constrained DC OPF problem is solved at the transmission level for uncertain wind and load. Exact chance constraint reformulations have also been applied to power systems problems by assuming that the underlying probability distributions are Normally distributed [16], [17]; however, in practice this is not the case, or the underlying distribution of the random variable is unknown. Monte-Carlo sampling approaches to incorporate arbitrary probability distributions, on the other hand, can be computationally expensive; it is desirable to have tractable approximations of these chance constraints [18], especially when real-time system operation is considered.

We will utilize a convex approximation of the chance constraint by using the traditional Chebyshev generating function [9], [15], which reformulates the chance constraint into a second-order cone constraint that depends only on the mean and variance of the random variable. We assume the chance constraints are independent; however, distributionally robust joint chance constraints that are close to exact can be considered as in [19]. In addition, linear approximations of the AC power flow equations [20], [21] which greatly improve upon the DC power flow approximation, will be utilized to transform the problem into one that can efficiently be solved in real time. By combining these techniques into a cohesive approach, we see that for enough collected data samples, the voltage is between the prescribed limits with a high probability, and the resulting optimization problem is convex.

The paper is thus organized as follows: In Section II, variable definitions, models for the distribution grid and random variables, and the AC power flow linearization will be discussed. Section III will define the data-driven chance constrained problem, and demonstrate how the resulting convexification of the chance constraint is distribution-agnostic. Simulation results are shown and interpreted in Section IV, and finally, a conclusion and directions for future work are given in Section V.

The work of K. Baker and E. Dall'Anese was supported in part by the Laboratory Directed Research and Development Program at the National Renewable Energy Laboratory.

K. Baker and E. Dall'Anese are with the National Renewable Energy Laboratory, Golden, CO. T. Summers is with the Department of Mechanical Engineering at University of Texas at Dallas. E-mails: kyri.baker@nrel.gov, emiliano.dallanese@nrel.gov, tyler.summers@utdallas.edu.

## II. SYSTEM MODEL AND AC APPROXIMATION

### A. System model

Consider a distribution feeder<sup>1</sup> comprising  $N + 1$  nodes within in the set  $\mathcal{N} \cup \{0\}$ ,  $\mathcal{N} := \{1, \dots, N\}$  (Node 0 is defined as the secondary side of the distribution transformer), and distribution lines represented by the set of edges  $\mathcal{E} := \{(m, n)\} \subset \mathcal{N} \times \mathcal{N}$ . Let  $V_n \in \mathbb{C}$  and  $I_n \in \mathbb{C}$  denote the complex phasors for the line-to-ground voltage and the current injected at node  $n \in \mathcal{N}$ , respectively, and define the vectors  $\mathbf{i} := [I_0, I_1, \dots, I_N]^T \in \mathbb{C}^{N+1}$  and  $\mathbf{v} := [V_0, V_1, \dots, V_N]^T \in \mathbb{C}^{N+1}$ . The system admittance matrix,  $\mathbf{Y} \in \mathbb{C}^{(N+1) \times (N+1)}$ , is formed based on the  $\pi$ -equivalent circuit of the distribution lines and describes the relationship between current and voltage via Ohm's and Kirchoff's laws as  $\mathbf{i} = \mathbf{Y}\mathbf{v}$ . Series and shunt admittances  $y_{mn}$  and  $y_{mn}^{\text{sh}}$ , respectively, of line  $(m, n)$ , the entries of  $\mathbf{Y}$  are formed as:

$$[\mathbf{Y}]_{m,n} := \begin{cases} \sum_{j \in \mathcal{N}_m} y_{mj}^{\text{sh}} + y_{mj}, & \text{if } m = n \\ -y_{mn}, & \text{if } (m, n) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\mathcal{N}_m := \{j \in \mathcal{N} : (m, j) \in \mathcal{E}\}$  is the set of nodes connected to the  $m$ -th node through a single distribution line. For simplicity, assume that shunt admittances  $y_{mn}^{\text{sh}}$  are negligible [22]. A constant- $PQ$  model is adopted for the loads, with  $P_{\ell,n}$  and  $Q_{\ell,n}$  denoting the real and reactive loads at node  $n \in \mathcal{N}$ , respectively. Let  $\mathbf{p}_\ell := [P_{\ell,1}, \dots, P_{\ell,N}]^T$  and  $\mathbf{q}_\ell := [Q_{\ell,1}, \dots, Q_{\ell,N}]^T$  be vectors describing the real and reactive power at all nodes; if no load is present at node  $n \in \mathcal{N}$ , then  $P_{\ell,n} = Q_{\ell,n} = 0$ .

### B. PV and uncertainty model

Define the variable  $P_{\text{av},n}$  as the maximum solar real power generation, or available real power, at node  $n \in \mathcal{N}$ . In order to avoid overvoltages in the grid, real power curtailment can be performed [5], [6], and will be modeled in our framework. To account for the ability of the RES inverters to adjust the output real power, let  $\alpha_n \in [0, 1]$  denote the percentage of available real power curtailed by the PV inverters; i.e., the real power curtailed from PV system  $n$  amounts to  $\alpha_n P_{\text{av},n}$ . If there is no PV system at node  $n$ ,  $\alpha_n = 0$ .

To account for the uncertainty in solar and load in the current five-minute time instance,  $\mathbf{p}_{\text{av}}$ ,  $\mathbf{p}_\ell$ , and  $\mathbf{q}_\ell$  are modeled as random variables [15]. Specifically, the available solar power is modeled as  $\mathbf{p}_{\text{av}} = \bar{\mathbf{p}}_{\text{av}} + \boldsymbol{\delta}_{\text{av}}$ , where  $\bar{\mathbf{p}}_{\text{av}} \in \mathbb{R}^N$

<sup>1</sup>Upper-case (lower-case) boldface letters will be used for matrices (column vectors);  $(\cdot)^T$  for transposition;  $(\cdot)^*$  complex-conjugate; and,  $(\cdot)^H$  complex-conjugate transposition;  $\Re\{\cdot\}$  and  $\Im\{\cdot\}$  denote the real and imaginary parts of a complex number, respectively;  $j := \sqrt{-1}$  the imaginary unit;  $|\cdot|$  denotes the absolute value of a number or the cardinality of a set; and,  $\circ$  denotes Hadamard product. For  $x \in \mathbb{R}$ , function  $[x]_+$  is defined as  $[x]_+ := \max\{0, x\}$ . Further,  $\mathbb{I}_A(x)$  denotes the indicator function over the set  $A \subset \mathbb{R}$ ; that is  $\mathbb{I}_A(x) = 1$  if  $x \in A$  and  $\mathbb{I}_A(x) = 0$  otherwise. For a given  $N \times 1$  vector  $\mathbf{x} \in \mathbb{R}^N$ ,  $\|\mathbf{x}\|_2 := \sqrt{\mathbf{x}^H \mathbf{x}}$ ;  $\text{diag}(\mathbf{x})$  returns a  $N \times N$  matrix with the elements of  $\mathbf{x}$  in its diagonal; and,  $\mathbf{x} \succeq \mathbf{y}$  implies that the inequality  $x_i \geq y_i$  is enforced for all the vector entries  $i = 1, \dots, N$ . Finally,  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix; and,  $\mathbf{0}_N$ ,  $\mathbf{1}_N$  the  $N$ -dimensional vectors with all zeroes and ones, respectively, and  $\mathbf{0}_{N \times M}$ ,  $\mathbf{1}_{N \times M}$  are  $N \times M$  matrices with all zeroes and ones.

is a vector comprising the forecasted values and  $\boldsymbol{\delta}_{\text{av}} \in \mathbb{R}^N$  is a random vector which represents the forecasting error. Similarly, the real and reactive loads can be expressed as  $\mathbf{p}_\ell = \bar{\mathbf{p}}_\ell + \boldsymbol{\delta}_{\mathbf{p},\ell}$  and  $\mathbf{q}_\ell = \bar{\mathbf{q}}_\ell + \boldsymbol{\delta}_{\mathbf{q},\ell}$ , respectively, where  $\bar{\mathbf{p}}_\ell^T$  and  $\bar{\mathbf{q}}_\ell^T$  are the forecasted loads and  $\boldsymbol{\delta}_{\mathbf{p},\ell}, \boldsymbol{\delta}_{\mathbf{q},\ell} \in \mathbb{R}^{J_\ell \cdot N}$  are random vectors describing the forecasting errors for real and reactive loads. In Section III, it will be described in further detail how no information about the underlying probability distribution has to be known in order to include these variables in the optimization formulation, and how measured samples representing the forecasting errors will be used to form the data-driven chance constraints.

### C. AC Power Flow Approximation

The power balance at node  $n$  can be written as the following:

$$V_n \sum_{m \in \mathcal{N}_n} y_{nm}^* (V_n - V_m)^* = S_n \quad (2)$$

where  $S_n = (1 - \alpha_n)P_{\text{av},n} - P_{\ell,n} - jQ_{\ell,n}$ . This well-known power balance equation typically appears in the constraint sets of standard formulations of optimal power flow problems, and renders the underlying optimization problem nonconvex [23]. Another source of nonconvexity can be derived from the voltage constraint  $V_{\min} \leq |V_n| \leq V_{\max}$ , where  $V_{\min}$  and  $V_{\max}$  represent a pre-determined lower and upper limits on the voltage magnitude (e.g., ANSI C.84.1 limits). Non-convexity of the problem can result in no guarantee of global optimality, and from a computational standpoint, the problem complexity may become prohibitive when the problem size grows [24]. While semidefinite relaxation techniques can sometimes be successfully employed to bypass the nonconvexity of these constraints while achieving globally optimal solutions of the nonconvex OPF [5], [23], in order to develop a computationally efficient OPF formulation, linear approximations of (2) and voltage-regulation constraints will be utilized in this work. These approximate power-flow relationships will also facilitate the application of convex approximation of chance constraints to the problem that will be formulated in Section III.

Hence, let  $\mathbf{v}_{\mathcal{N}} := [V_1, \dots, V_N]^T$  and  $\mathbf{i}_{\mathcal{N}} := [I_1, \dots, I_N]^T$  be the vectors of voltages and currents, respectively, at all nodes  $\mathcal{N}$ , and rewrite the current-voltage relationship  $\mathbf{i} = \mathbf{Y}\mathbf{v}$  as follows:

$$\begin{bmatrix} I_0 \\ \mathbf{i}_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} y_{0j} & \mathbf{y}_{0\mathcal{N}}^T \\ \mathbf{y}_{0\mathcal{N}} & \mathbf{Y}_{\mathcal{N}\mathcal{N}} \end{bmatrix} \begin{bmatrix} V_0 \\ \mathbf{v}_{\mathcal{N}} \end{bmatrix} \quad (3)$$

where  $\mathbf{y}_{0\mathcal{N}} \in \mathbb{C}^{1 \times N}$  is comprising elements  $\{[\mathbf{Y}]_{0,n}\}_{n \in \mathcal{N}}$  of the admittance matrix  $\mathbf{Y}$  and, similarly,  $\mathbf{Y}_{\mathcal{N}\mathcal{N}} \in \mathbb{C}^{N \times N}$  is comprising elements  $\{[\mathbf{Y}]_{m,n}\}_{m,n \in \mathcal{N}}$ . Without loss of generality, node 0 is taken to be the slack node, and the voltage at the secondary  $V_0 = \rho_0 e^{j\theta_0^0}$  is assumed known. Let  $\mathbf{s} := (\mathbf{I}_N - \text{diag}\{\boldsymbol{\alpha}\})\mathbf{p}_{\text{av}} - \mathbf{p}_\ell - j\mathbf{q}_\ell$  denote the vector of complex power injections. With  $V_0$  known, Theorem 1 and Corollary 2 in [25] assert that, if  $(\rho_0)^2 > 4\|\mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}\|^\dagger \|\mathbf{s}\|_2$ ,

then the voltages across nodes  $\mathcal{N}$  can be approximated as the following:

$$\mathbf{v}_{\mathcal{N}} \approx V_0 \left( \mathbf{1}_N + \frac{1}{(\rho_0)^2} \mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}(\mathbf{s})^* \right) \quad (4)$$

with an approximation error per node  $n \in \mathcal{N}$  that is bounded by  $(4/(\rho_0^t)^3) \|\mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}\|_2 \|\mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}\|^\dagger \|\mathbf{s}\|_2^2$ . This approximation error is inversely proportional to  $\rho_0$ , and decreases with respect to the decreasing of the network net load [25].

Equation (4) provides a convenient way to bypass the non-convexity of the balance constraint (2) in the OPF problem. However, a second layer of approximation is sought next, in order to establish a linear (and approximate) relationship between injected complex powers and voltage magnitudes. This relationship will be utilized subsequently to derive convex constraints for voltage magnitudes. To this end, express the voltage phasors  $\mathbf{v}_{\mathcal{N}}$  as  $\mathbf{v}_{\mathcal{N}} = \rho_{\mathcal{N}} \circ e^{j\theta_{\mathcal{N}}}$ , where  $\rho_{\mathcal{N}} \in \mathbb{R}^N$  collects the voltage magnitudes  $\{\rho_n\}^{n \in \mathcal{N}}$  and  $\theta_{\mathcal{N}}$  the voltage phases  $\{\theta_n\}_{n \in \mathcal{N}}$ . Assuming that the nominal voltage  $\rho_0$  dominates the voltage drops on the lines, i.e.  $\|\mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}(\mathbf{s})^*\|_2 / (\rho_0^t)^2 \ll 1$ , a first-order approximation to voltage magnitudes and angles across the distribution network can be described as:

$$\begin{aligned} \rho_{\mathcal{N}} \approx \mathbf{g}_\rho(\boldsymbol{\alpha}, \mathbf{p}_{\text{av}}, \mathbf{p}_\ell, \mathbf{q}_\ell) &:= \rho_0 \mathbf{1}_N - \frac{1}{\rho_0} \mathbf{C}_{\mathcal{N}\mathcal{N}} \mathbf{q}_\ell \\ &+ \frac{1}{\rho_0} \mathbf{R}_{\mathcal{N}\mathcal{N}} ((\mathbf{I}_N - \text{diag}\{\boldsymbol{\alpha}\}) \mathbf{p}_{\text{av}} - \mathbf{p}_\ell) \end{aligned} \quad (5)$$

$$\begin{aligned} \theta_{\mathcal{N}} \approx \mathbf{g}_\theta(\boldsymbol{\alpha}, \mathbf{p}_{\text{av}}, \mathbf{p}_\ell, \mathbf{q}_\ell) &:= \theta_0 \mathbf{1}_N + \frac{1}{(\rho_0)^2} \mathbf{R}_{\mathcal{N}\mathcal{N}} \mathbf{q}_\ell \\ &+ \frac{1}{(\rho_0)^2} \mathbf{C}_{\mathcal{N}\mathcal{N}} ((\mathbf{I}_N - \text{diag}\{\boldsymbol{\alpha}\}) \mathbf{p}_{\text{av}} - \mathbf{p}_\ell) \end{aligned} \quad (6)$$

where the network-related matrices  $\mathbf{R}_{\mathcal{N}\mathcal{N}}$  and  $\mathbf{C}_{\mathcal{N}\mathcal{N}}$  are defined as  $\mathbf{R}_{\mathcal{N}\mathcal{N}} := \Re\{\mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}\}$  and  $\mathbf{C}_{\mathcal{N}\mathcal{N}} := \Im\{\mathbf{Y}_{\mathcal{N}\mathcal{N}}^{-1}\}$ , respectively. Equation (5) will be utilized next to develop a computationally efficient solution to the chance-constrained OPF problem.

### III. PROBLEM FORMULATION

In this section, we will describe the problem formulation first with generic chance constraints and then with their data-driven convex approximation. Consider the following optimization problem to compute the optimal curtailment levels for the PV inverters:

$$(P0) \min_{\boldsymbol{\rho}, \boldsymbol{\alpha}} \mathbb{E}(f(\boldsymbol{\rho}, \boldsymbol{\alpha}, \mathbf{p}_\ell, \mathbf{q}_\ell)) \quad (7a)$$

subject to

$$\boldsymbol{\rho} = \mathbf{g}_\rho(\boldsymbol{\alpha}, \mathbf{p}_{\text{av}}, \mathbf{p}_\ell, \mathbf{q}_\ell) \quad (7b)$$

$$\Pr\{V_{\min} \leq \rho_n\} \geq 1 - \epsilon \quad (7c)$$

$$\Pr\{\rho_n \leq V_{\max}\} \geq 1 - \epsilon \quad (7d)$$

$$0 \leq \alpha_n \leq 1 \quad (7e)$$

for all  $n \in \mathcal{N}$ , where  $g_{\rho,n}(\cdot)$  denotes the  $n$ -th element of the vector-valued function  $\mathbf{g}_\rho(\cdot)$  in (5). Constraint (7b) represents

a surrogate for the power balance equation; constraints (7d) and (7c) are the chance constraints that require the voltage to be within its upper and lower limits with at least  $1 - \epsilon$  probability; and constraint (7e) limits the curtailment percentage from 0–100%. The cost function  $f(\boldsymbol{\rho}, \boldsymbol{\alpha}, \mathbf{p}_\ell, \mathbf{q}_\ell)$  is convex and can consider a sum of penalties on curtailment, penalties on power drawn from the substation, penalties on voltage violations, among other objectives. In the data-driven framework, the expected value of  $f(\cdot)$  is formed from the sample average over all data measurements.

In the current formulation, however, constraints (7d)–(7c) can prove to be problematic. For non-Gaussian distributions, Monte Carlo-based methods may be used; however, these can lead to high computational burdens. In addition, even if the function  $g_{\rho,n}(\cdot)$  is affine, the feasible set of (7d)–(7c) may be nonconvex. To account for a variety of possible distributions of the forecasting errors  $\delta$  and yet derive a computationally efficient solution method for the optimization, a data-driven, convex approximation of the chance constraints is described next.

#### A. Distribution Agnostic Chance Constraints

First, consider a generic chance constraint  $\Pr\{g(\mathbf{x}, \boldsymbol{\delta}) > 0\} \leq \epsilon$ , where function  $g(\mathbf{x}, \boldsymbol{\delta})$  is convex in the optimization variables  $\mathbf{x}$  for given values of the random vector  $\boldsymbol{\delta}$ . In order to develop a convex approximation of this constraint, next consider a function  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  that is nonnegative valued, nondecreasing, and convex. Assume that  $\psi(\cdot)$  – referred to as the (one-dimensional) generating function – satisfies the conditions  $\psi(x) > \psi(0) \forall x > 0$  and  $\psi(0) = 1$ . Given a positive scalar  $z > 0$  and a random variable  $\delta$ , it follows that  $\psi(\cdot)$  is such that the following holds:  $\mathbb{E}_\delta\{\psi(z\delta)\} \geq \mathbb{E}_\delta\{\mathbb{I}_{[0,+\infty)}(z\delta)\} = \Pr\{z\delta \geq 0\} = \Pr\{\delta \geq 0\}$ . Thus, by taking  $\delta = g(\mathbf{x}, \boldsymbol{\delta})$  and replacing  $z$  with  $z^{-1}$ , one has that the following bound holds for all  $z > 0$  and  $\mathbf{x}$  [8]:

$$\Pr\{g(\mathbf{x}, \boldsymbol{\delta}) > 0\} \leq \mathbb{E}_\delta \{ \psi(z^{-1}g(\mathbf{x}, \boldsymbol{\delta})) \}. \quad (8)$$

Thus, the constraint

$$\inf_{z>0} \{ z \mathbb{E}_\delta \{ \psi(z^{-1}g(\mathbf{x}, \boldsymbol{\delta})) \} - z\epsilon \} \leq 0 \quad (9)$$

represents a sufficient condition for  $\Pr\{g(\mathbf{x}, \boldsymbol{\delta}) > 0\} \leq \epsilon$ . This implies that (9) is also a conservative convex approximation of the chance constraint  $\Pr\{g(\mathbf{x}, \boldsymbol{\delta}) \geq 0\} \leq 1 - \epsilon$ . With regards to the convexity of (9), notice that since  $\psi(\cdot)$  is nondecreasing and convex and  $g(\cdot, \boldsymbol{\delta})$  is convex, it follows that the mapping  $(\mathbf{x}, z) \rightarrow z\psi(z^{-1}g(\mathbf{x}, \boldsymbol{\delta}))$  is convex. In addition, if  $g$  is biaffine in  $\mathbf{x}$  and  $\boldsymbol{\delta}$ , and  $\psi$  is quadratic, then constraint (9) is also convex.

Specifically, for the data-driven model that will be formulated without knowledge of underlying probability distributions, consider the traditional Chebyshev generating function  $\psi(x) = (1+x)^2$  [15], and let  $\bar{\boldsymbol{\delta}}$  and  $\boldsymbol{\Sigma}$  denote the mean and covariance matrix, respectively, of  $\boldsymbol{\delta}$  formed from all samples  $S$ . That is,  $\bar{\boldsymbol{\delta}} := \frac{1}{S} \sum_{s=1}^S \boldsymbol{\delta}[s]$  and  $\boldsymbol{\Sigma} := \frac{1}{S} \sum_{s=1}^S ((\boldsymbol{\delta}[s] - \bar{\boldsymbol{\delta}})(\boldsymbol{\delta}[s] - \bar{\boldsymbol{\delta}})^\top)$ . Note that these moments can be computed

offline using the available collected data and updated whenever additional samples are measured; thus, the number of samples does not affect the computation time.

Notice further that function  $g_{\rho,n}(\boldsymbol{\alpha}, \boldsymbol{\delta})$  expressing the voltage magnitude at node  $n$  can be re-written as  $g_{\rho,n}(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \mathbf{h}_n^T(\boldsymbol{\alpha})\boldsymbol{\delta} + \rho_0$ , where  $\mathbf{a}_n := \frac{1}{\rho_0}[\mathbf{R}_{\mathcal{N}\mathcal{N}}^T]_n$  and  $\mathbf{h}_n(\boldsymbol{\alpha})$  is an affine function of  $\boldsymbol{\alpha}$  defined as:

$$\mathbf{h}_n(\boldsymbol{\alpha}) := \begin{bmatrix} \frac{1}{\rho_0}(\mathbf{I}_N - \text{diag}\{\boldsymbol{\alpha}\})[\mathbf{R}_{\mathcal{N}\mathcal{N}}^T]_n \\ -\frac{1}{\rho_0}[\mathbf{R}_{\mathcal{N}\mathcal{N}}^T]_n \\ -\frac{1}{\rho_0}[\mathbf{C}_{\mathcal{N}\mathcal{N}}^T]_n \end{bmatrix}. \quad (10)$$

Then, by using the the generating function  $\psi(x) = (1+x)^2$ , we can write (9) as the following [9]:

$$g_{\rho,n}(\boldsymbol{\alpha}^\tau, \bar{\boldsymbol{\delta}}^\tau) - V_{\max} + \bar{\epsilon} \left\| (\boldsymbol{\Sigma}^\tau)^{\frac{1}{2}} \mathbf{h}_n(\boldsymbol{\alpha}^\tau) \right\|_2 \leq 0 \quad (11a)$$

$$V_{\min} - g_{\rho,n}(\boldsymbol{\alpha}^\tau, \bar{\boldsymbol{\delta}}^\tau) + \bar{\epsilon} \left\| (\boldsymbol{\Sigma}^\tau)^{\frac{1}{2}} \mathbf{h}_n(\boldsymbol{\alpha}^\tau) \right\|_2 \leq 0 \quad (11b)$$

where  $\bar{\epsilon} := \sqrt{\frac{1-\epsilon}{\epsilon}}$ . Notice that (11) are second-order cone constraints, and so forming the final optimization problem by replacing constraints (7d) and (7c) with (11), the result is a second-order cone problem. It can be seen that from this approach, equation (11) is a distributionally-robust constraint; that is, (11) ensures that the original chance constraint is satisfied for any distribution of the forecast errors as long as data samples are measured to form the first and second moments. If the first and second moments are known exactly, the original chance constraints (7c) and (7d) are guaranteed to be satisfied. However, in the following section it is shown how accurately the chance constraints can be satisfied by only using estimates of the mean and covariance that are drawn from samples.

#### IV. SIMULATION RESULTS

##### A. Test feeder and input data

The proposed optimization and control scheme is tested using a modified version of the IEEE 37-node test feeder shown in Figure 1. Twenty-one photovoltaic (PV) systems are located at nodes 4, 7, 9, 10, 11, 13, 16, 17, 20, 22, 23, 26, 28, 29, 30, 31, 32, 33, 34, 35, and 36, indicated in the figure with boxes around the Node number. The modified network is obtained by considering a single-phase equivalent, and by replacing the loads specified in the original dataset with real load data measured from feeders in Anatolia, CA during the first week of August 2012 [26], which coincide with the solar irradiance data from the same location.

The total loading of the feeder for a typical day during this week can be seen in Figure 2, where data granularity is of 5 minutes, and a 24-hour simulation period is considered. Line impedances, shunt admittances, as well as active and reactive loads are adopted from the IEEE 37-node dataset.<sup>2</sup> The aggregate available power  $\sum_n P_{av,n}$  during the course of the day is plotted in Figure 2. The voltage limits  $V_{\max}$  and  $V_{\min}$  are set to 1.05 pu and 0.95 pu, respectively. It can be seen

<sup>2</sup>Available at: [ewh.ieee.org/soc/pes/dsacom/testfeeders](http://ewh.ieee.org/soc/pes/dsacom/testfeeders).

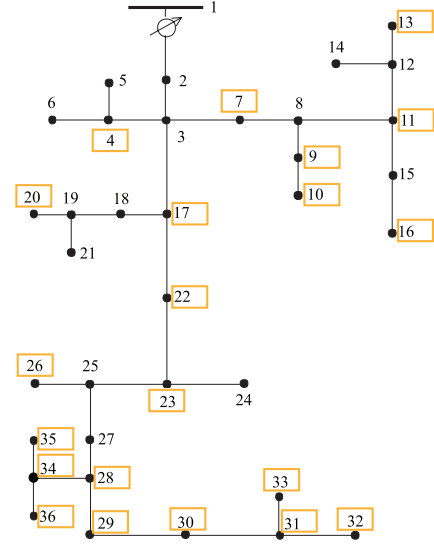


Figure 1: IEEE 37-node test feeder considered in the test cases.

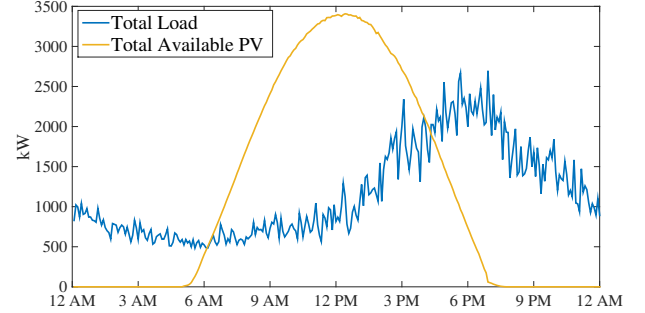


Figure 2: Total available solar generation and feeder loading.

from the system figure that many PV systems are placed at the end of the feeder, which can result in overvoltage conditions [27]. In this particular case, with PV inverters injecting all available power into the feeder, overvoltage conditions are observed at multiple nodes during solar peak irradiation hours, as seen in Figure 3. For these simulations, the cost function aimed to minimize the cost of PV curtailment; i.e.:

$$f(\boldsymbol{\rho}, \boldsymbol{\alpha}, \mathbf{p}_\ell, \mathbf{q}_\ell) = \sum_{i \in \mathcal{N}} b_i \alpha_i P_{av,n}, \quad (12)$$

where the cost of curtailing power at each node is set to be  $b_i = \$6/p.u.$

##### B. Varying the amount of data

Varying the number of samples  $S$  used to form the mean and covariance matrix and its impact on the solution is now analyzed. In Figure 4, the voltage envelopes are shown for the hours of 6 AM through 6 PM and compared for  $S = 5, 100, \text{ and } 500$  samples, and a chance constraint violation parameter of  $\epsilon = 0.05$  (a 5% violation of the constraints is allowed). Both the load and solar error samples are drawn from a truncated Gaussian distribution with the tails ending

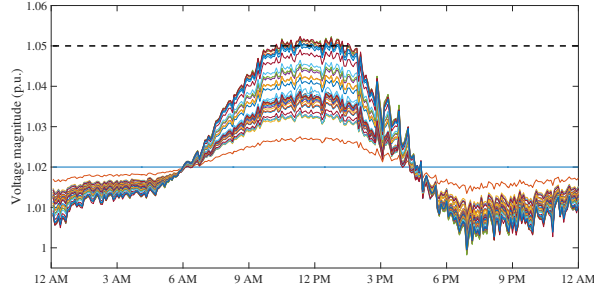


Figure 3: Voltages without advanced inverter control.

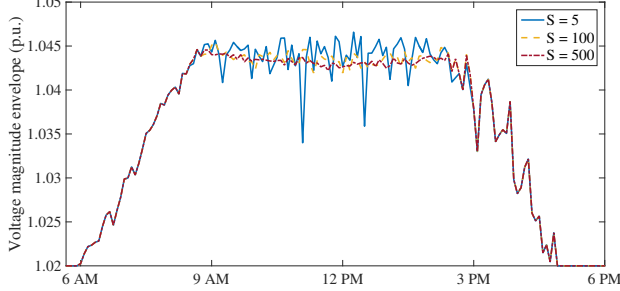


Figure 4: Voltage envelope for different values of  $S$ .

at  $\pm 3\sigma$  of the distribution, where  $\sigma$ , the standard deviation, is assumed to be 3% of the actual value of load and solar irradiance.

For a small amount of measured data ( $S = 5$ ), the voltage limits are still satisfied due to the conservative nature of the distributionally robust constraints, but the voltage magnitude fluctuates dramatically at multiple points (around 11 AM and 1 PM, for instance) due to the inaccuracy of the chance constraints. For example, this could be due to the measurements indicating that a higher level of solar irradiance is more likely, so the controller curtails more power compared to when more information is obtained, as seen in Figure 5. As the number of samples increases, the voltage profile flattens, and due to the conservativeness of the chance constraints, remains below the maximum threshold of 1.05 p.u. A more detailed analysis on the “value of data” with regards to conservatism of data-driven chance constraints and amount of historical data can be found in [14].

### C. Restrictions on chance-constraint fulfillment

Depending on the level of chance constraint fulfillment parameter  $\epsilon$ , the constraint may be more or less conservative with respect to regulating the voltage at each node. Relaxing the constraint somewhat may improve the total system cost, but if the constraint is relaxed too far, the voltage limits risk causing a violation. In Figure 6, the total system cost is plotted for the hours of 6 AM to 6 PM for  $\epsilon = 0.05, 0.10$ , and  $0.40$ . As expected, the looser the chance constraint violation parameter, the lower the cost. However, it can be seen from Figure 7 that as this parameter increases, the voltage profile flattens and the system is operating closer to the maximum voltage limits, indicating that a large disturbance may cause overvoltage

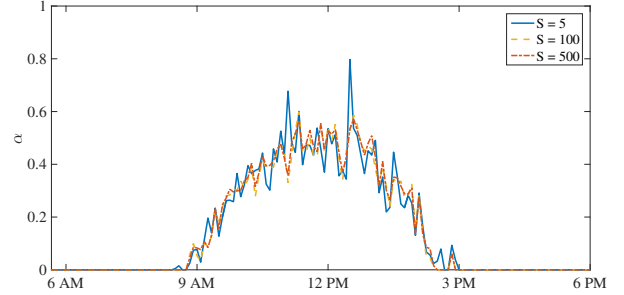


Figure 5: Total PV curtailment for different values of  $S$ .

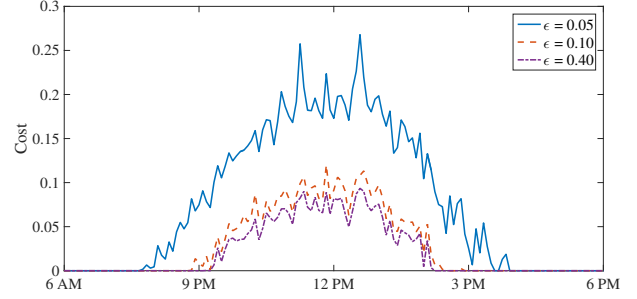


Figure 6: Total cost for different values of  $\epsilon$ .

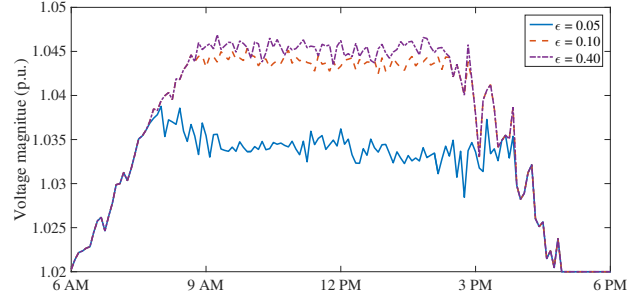


Figure 7: Voltage envelopes for different values of  $\epsilon$ .

conditions to occur. The cost and benefit from increasing this parameter must be determined for each individual application; here, it appears that  $\epsilon = 0.05$  may be too restrictive, but  $\epsilon = 0.40$  may be too loose for the small tradeoff of cost decrease from  $\epsilon = 0.10$ . In addition, in model predictive or multi-timestep approaches,  $\epsilon$  could be varied throughout the prediction horizon.

## V. CONCLUSION

In this paper, we demonstrated a data-driven approach to solving the chance-constrained voltage regulation problem in a distribution network under high levels of PV generation. The resulting chance constraints are distribution-agnostic; i.e., no knowledge about the underlying probability distributions needs to be known a priori. It was shown that the amount of measured data can improve the accuracy of the constraints and does not greatly impact the speed of the method, due to the fact that the first and second moments used in the chance constraints can be computed offline.

Future work will extend this framework into the joint chance constraint case, as well as include other sources of uncertainty

such as price, user behavior, and demand responsive loads. A model predictive approach could be considered in future work in order to incorporate energy storage models and other intertemporal constraints into the framework as well.

## REFERENCES

- [1] D. Cusick, "Solar power sees unprecedented boom in U.S." *Scientific American*, March 2015.
- [2] A. Woyte, V. Van Thong, R. Belmans, and J. Nijs, "Voltage fluctuations on distribution level introduced by photovoltaic systems," *IEEE Trans. on Energy Conv.*, vol. 21, no. 1, pp. 202–209, 2006.
- [3] B. Palmintier, R. Broderick, B. Mather, M. Coddington, K. Baker, F. Ding, M. Reno, M. Lave, and A. Bharatkumar, "On the path to sunshot: Emerging issues and challenges in integrating solar with the distribution system." National Renewable Energy Laboratory, Tech. Rept. NREL/TP-5D00-65331, May 2016.
- [4] E. Liu, J. Bebic, B. Kroposki, J. de Bedout, and W. Ren, "Distribution system voltage performance analysis for high-penetration pv," in *IEEE Energy 2030 Conf.*
- [5] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Optimal dispatch of photovoltaic inverters in residential distribution systems," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 487–497, Apr. 2014.
- [6] A. Samadi, R. Eriksson, L. Soder, B. G. Rawn, and J. C. Boemer, "Coordinated active power-dependent voltage regulation in distribution grids with PV systems," *IEEE Trans. on Power Del.*, vol. 29, no. 3, pp. 1454–1464, June 2014.
- [7] W. Wang and S. Ahmed, "Sample average approximation of expected value constrained stochastic programs," *Operations Research Letters*, vol. 36, no. 5, pp. 515–519, 2008.
- [8] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. on Optimization*, vol. 17, no. 4, pp. 969–996, 2007.
- [9] G. Calafiore and L. El Ghaoui, "On distributionally robust chance constrained linear programs," *Journal of Optimization Theory and Applications*, vol. 130, no. 1, pp. 1–22, 2006.
- [10] E. Dall'Anese, S. V. Dhople, B. B. Johnson, and G. B. Giannakis, "Optimal dispatch of residential photovoltaic inverters under forecasting uncertainties," *IEEE J. of Photovoltaics*, vol. 5, no. 1, pp. 350–359, Jan 2015.
- [11] S. S. Y. Zhang and J. Mathieu, "Data-driven optimization approaches for optimal power flow with uncertain reserves from load control," in *Proceedings of the American Control Conference (ACC)*, Chicago, IL, 2015.
- [12] E. Delage and Y. Ye, "Distributionally robust optimization under moment uncertainty with application to the problems," *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.
- [13] B. Stellato, "Data-driven chance constrained optimization," *Master's thesis*, ETH Zürich, Sept. 2014.
- [14] R. Jiang and Y. Guan, "Data-driven chance constrained stochastic program," *Optimization-Online*, 2012, [Online] Available at [http://www.optimization-online.org/DB\\_HTML/2012/07/3525.html](http://www.optimization-online.org/DB_HTML/2012/07/3525.html).
- [15] 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, Nov. 2015.
- [16] 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.
- [17] L. Roald, S. Misra, M. Chertkov, and G. Andersson, "Optimal power flow with weighted chance constraints and general policies for generation control," 2015, [Online] Available at: <http://arxiv.org/pdf/1504.00057>.
- [18] J. Goh and M. Sim, "Distributionally robust optimization and its tractable approximations," *Operations Research*, vol. 58, no. 4-part-1, pp. 902–917, 2010.
- [19] S. Zymler, D. Kuhn, and B. Rustem, "Distributionally robust joint chance constraints with second-order moment information," *Mathematical Programming*, vol. 137, no. 1, pp. 167–198, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s10107-011-0494-7>
- [20] S. Dhople, S. Guggilam, and Y. Chen, "Linear approximations to ac power flow in rectangular coordinates," *Allerton Conference on Communication, Control, and Computing*, in Press, 2015.
- [21] S. Guggilam, E. Dall'Anese, Y. Chen, S. Dhople, and G. B. Giannakis, "Scalable optimization methods for distribution networks with high PV integration," *IEEE Transactions on Smart Grid*, 2015, in press.
- [22] W. H. Kersting, *Distribution System Modeling and Analysis*. 2nd ed., Boca Raton, FL: CRC Press, 2007.
- [23] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Trans. Power Syst.*, vol. 1, no. 1, pp. 92–107, Feb. 2012.
- [24] S. Paudyal, C. A. Canizares, and K. Bhattacharya, "Three-phase distribution OPF in smart grids: Optimality versus computational burden," in *2nd IEEE PES Intl. Conf. and Exhibition on Innovative Smart Grid Technologies*, Manchester, UK, Dec. 2011.
- [25] S. Bolognani and S. Zampieri, "On the existence and linear approximation of the power flow solution in power distribution networks," *IEEE Trans. on Power Systems*, vol. 31, no. 1, pp. 163–172, 2016.
- [26] J. Bank and J. Hambrick, "Development of a high resolution, real time, distribution-level metering system and associated visualization modeling, and data analysis functions," National Renewable Energy Laboratory, Tech. Rept. NREL/TP-5500-56610, May 2013.
- [27] F. Ding, B. Mather, N. Ainsworth, P. Gotseff, and K. Baker, "Locational sensitivity investigation on PV hosting capacity and fast track PV screening," in *IEEE PES T&D*, Dallas, TX, May 2016.