Fast Data-Driven Chance Constrained AC-OPF using Hybrid Sparse Gaussian Processes

Mile Mitrovic, Aleksandr Lukashevich, Petr Vorobev, Vladimir Terzija, Yury Maximov, Deepjyoti Deka

***Abstract*—The alternating current (AC) chance-constrained optimal power flow (CC-OPF) problem addresses the economic efficiency of electricity generation and delivery under generation uncertainty. The latter is intrinsic to modern power grids because of the high amount of renewables. Despite its academic success, the AC CC-OPF problem is highly nonlinear and computationally demanding, which limits its practical impact. For improving the AC-OPF problem complexity/accuracy trade-off, the paper proposes a fast data-driven setup that uses the sparse and hybrid Gaussian processes (GP) framework to model the power flow equations with input uncertainty. We advocate the efficiency of the proposed approach by a numerical study over multiple IEEE test cases showing up to two times faster and more accurate solutions compared to the state-of-the-art methods.**

***Index Terms*—Power Flow, Chance-Constrained Optimization**

I. Introduction

The optimal power flow (OPF) problem is a fundamental tool for the secure and economic operation of power system, used in system security assessments and in electricity markets. Nowadays, the increase of renewable energy generation is seen as a primary approach to reduce greenhouse emissions [1]. Despite the benefits of renewable energy sources (RES), the uncertainty associated with RES generation such as wind and solar leads to significant control challenges for the Transmission System Operators (TSOs) [2], [3]. Optimal optimization and feedback control algorithms should act on second and sub-second levels to prevent the grid from impending energy blackout due to RES uncertainty/fluctuations. Various methods address the accuracy and scalability of the AC OPF problem under uncertainty, including seminal probabilistic optimal power flow [4], [5], and chance-constrained approach [6]–[15]. The chance-constrained setup considered in this paper ensures that the security constraints are satisfied with a high probability and allows TSOs to balance operating costs and system security. Unfortunately, Chance-Constrained (CC) AC-OPF with non-linear Alternating Current-Power flow (AC-PF) equations is known to be computationally demanding and often does not have an analytical solution [16], [17]. A simpler linear version, chance-constraint direct current (DC) approximation (DC CC-OPF) [6], [7], [13] has been widely studied in literature and has a few scalable solutions [8], [18], but suffers from conservative solutions. Furthermore, it requires knowledge of the system parameters such as line impedances that may have errors and uncertainities.

The paper proposes to improve the AC CC-OPF scalability by using a data-driven method that builds upon the framework of Gaussian Process (GP) Regression to replace the power flow equations using prior historical data. It is worth mentioning that GP based power-flow models have been implemented in deterministic OPF formulations [19]–[22], and for CC-OPF in [23]. Compared to vanilla Gaussian process regression in [23], however, our approach uses (1) a **hybrid GP** that combines linear DC-PF with data-driven estimation of the residuals between DC-PF and AC-PF, and (2) uses **sparse** Gaussian regression using few selective data for the estimation of the hybrid GP. Both these steps significantly reduce our approach computational complexity without compromising the solution's accuracy, making it amenable for real-time/online deployment. We demonstrate a practical advantage of our approach against state-of-the-art algorithms on multiple IEEE test-cases.

This paper is organized as follows. Section II describes the problem addressed and the uncertainty modeling. Section III outlines the Gaussian Process Regression and their sparse version. The hybrid approach to solve the AC CC-OPF problem is given in Section IV. Section V presents results of numerical evaluation over various IEEE test cases. Section VI presents a critical discussion, possible extensions and a summary of the paper.

II. Background and Uncertainty Modeling

A. Grid Uncertainty

Let $\mathcal{Q}(\mathcal{B}, \mathcal{E})$ be a transmission grid with a set of buses $\mathcal{B}$, $b = |\mathcal{B}|$, and (directed) lines $\mathcal{E}$, $e = |\mathcal{E}|$. We decompose each power injection at a bus into a controllable generation (referred as g), uncertain part r, and load l:

$$p = p_g - p_l + p_r, \qquad q = q_g - q_l + q_r, \tag{1}$$

where p and q are active and reactive power injections resp. The classical AC-OPF problem assumes that deterministic power injections are known; however, load and renewable generation fluctuate in practice. Let ω be an active power fluctuation so that $p(\omega) = p + \omega$, $q(\omega) = q + \gamma\omega$ where γ is a constant power factor, $\gamma > 0$.

B. Generation and Voltage Control

Following uncertainty realizations ω, controllable generators adapt their generation to maintain the total power balance through the automatic generation control (AGC) [13]. The

AGC assumes that the total power mismatch $\Omega = \sum_{i\in\mathcal{G}} \omega_i$ is divided among the generators according to the participation factors α, $\alpha > 0$, as:

$$p_{g,i}(\omega) = p_{g,i} + \alpha_i \Omega, \quad \sum_{i\in\mathcal{G}} \alpha_i = 1,\ \alpha_i \geq 0,\ \forall i \in \mathcal{G}. \quad (2)$$

1) Power Flows: Assume p_{ij} and q_{ij} denote the active and reactive power flows from bus i to bus j along line $(i,j) \in \mathcal{E}$. In the AC power flow formulation, each transmission line's active and reactive power flows depend non-linearly on the voltage magnitudes v and voltage angles θ. The AC power flow equations are given by:

$$p_{ij}(\omega) = v_i(\omega)v_j(\omega)\left[G_{ij}cos\left[\theta_{ij}(\omega)\right] + B_{ij}sin\left[\theta_{ij}(\omega)\right]\right]$$
$$q_{ij}(\omega) = v_i(\omega)v_j(\omega)\left[G_{ij}sin\left[\theta_{ij}(\omega)\right] - B_{ij}cos\left[\theta_{ij}(\omega)\right]\right]$$

where $\theta_{ij} = \theta_i - \theta_j$, G_{ij} and B_{ij} are the entries of the real and imaginary parts of the network admittance. The system is balanced when the power flows leaving each bus are equal to the sum of power injection at that bus. The apparent power flow on the line ij is given by $s_{ij} = \sqrt{(p_{ij})^2 + (q_{ij})^2}$.

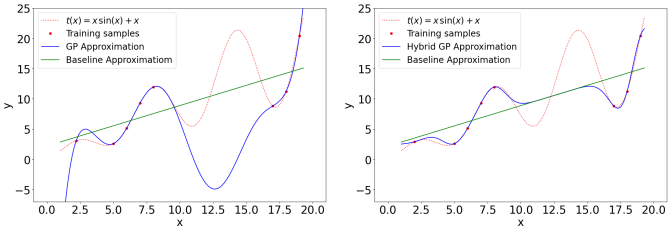
C. Chance Constraints AC OPF

The full setup of the AC chance-constrained OPF problem is as follows [13]:

$$\min_{p_g,q_q,v,\theta} \sum_{i\in\mathcal{G}} \mathbb{E}[c_i(p_{g,i}(w))] \quad (3a)$$
$$\text{s.t. } F(\theta(w), v(w), p(w), q(w)) = 0, \quad \forall w \in W \quad (3b)$$
$$\mathbb{P}(p_{g_i}(\omega) \leq p_{g_i}^{max}) \geq 1 - \epsilon_p, \quad \forall i \in \mathcal{G} \quad (3c)$$
$$\mathbb{P}(p_{g_i}(\omega) \geq p_{g_i}^{min}) \geq 1 - \epsilon_p, \quad \forall i \in \mathcal{G} \quad (3d)$$
$$\mathbb{P}(q_{g_i}(\omega) \leq q_{g_i}^{max}) \geq 1 - \epsilon_q, \quad \forall i \in \mathcal{G} \quad (3e)$$
$$\mathbb{P}(q_{g_i}(\omega) \geq q_{g_i}^{min}) \geq 1 - \epsilon_q, \quad \forall i \in \mathcal{G} \quad (3f)$$
$$\mathbb{P}(v_j(\omega) \leq v_j^{max}) \geq 1 - \epsilon_v, \quad \forall i \in \mathcal{B} \quad (3g)$$
$$\mathbb{P}(v_j(\omega) \geq v_j^{min}) \geq 1 - \epsilon_v, \quad \forall i \in \mathcal{B} \quad (3h)$$
$$\mathbb{P}(s_{ij}(\omega) \leq s_{ij}^{max}) \geq 1 - \epsilon_s, \quad \forall ij \in \mathcal{E} \quad (3i)$$

where expectations $\mathbb{E}$ and probabilities $\mathbb{P}$ are taken over an unknown distribution of uncertainty ω, $c_i(\cdot)$ is a cost function of the i-th generator. The objective function minimizes the expected total cost of active power generations, with cost coefficients c. Eq. (2b) represents AC power balance equations that govern relation between power flows, injections and voltage phasors for all possible fluctuations ω and buses $i \in \mathcal{B}$. The power outputs of conventional generators, voltage magnitudes at buses, and apparent power flow in the lines are constrained using the different chance constraints (2c)–(2i) that must be satisfied with the prescribed probabilities ϵ_p, ϵ_q, ϵ_v, and ϵ_s. In this way, the chance constraints define an ϵ-reliability set within which an operator is safe to execute additional control actions.

In this paper, we consider separate chance constraints, which limit the violation probability for individual constraints and are easy to enforce [14], [18]. The chance constraints (3) are intractable for non-linear AC-PF; however, Gaussian Processes allow to design a more tractable approximation to them. The latter can easily be handled with modern optimization solvers.



(a) Classical Approach: GP learned on historical data;

(b) Hybrid approach: GP learned on residuals between linear model and data.

Fig. 1: Non-linear function $x\sin(x) + x$ (in red) is approximated with GP (blue; left) based on several observations (red dots). The hybrid GP approximation (blue; right) uses a linear proxy (green) to improve the approximation.

III. Model Description

A. Notation

For a matrix X, X_{ij} denotes element ij of a matrix X, and $[x]_i$, $[x]_{:,j}$ its i-th row and j-th column, respectively. With $[x]_{*,j}$, we denote $*$ element of j-th column vector. Further, we will use $\cdot_*$ subscript to denote unseen test data. For the given vector x, $diag(x)$ is used to refer to a diagonal matrix with entries of vector x, while $diag(X)$ represent the vector of diagonal elements of matrix X. The probability density of x is denotes as $p(x)$, $p(x|y)$ representing the probability of x for given y. $\bigtriangledown(\cdot)_{|x_*}$ and $\bigtriangledown^2(\cdot)_{|x_*}$ are the first and second derivative of the function $(\cdot)$ evaluated at x_*.

B. Problem Formulation

In the classical AC-OPF modeling with Gaussian processes [19], [23], the Gaussian process regression model infers the full AC-OPF function in Eq. (3) from previously collected input and output measurement samples as

$$y_i = g(x_i) + \omega_i, \qquad i = 1, ..., N$$

where N is the total number of samples, $\omega \sim \mathcal{N}(0, \Sigma_\omega)$ is the Gaussian noise with covariance Σ_ω and g is the GP function. The input x is the vector of controllable active power generators p_g, and uncertain active power injections p_l and p_r. Note that the reactive power injections q_l and q_r are not considered as inputs to the GPR model as they are given by a fixed power factor ratio from the active power injections, increasing redundancy. The output vector y corresponds to all dependent output variables, such as voltages at non-generator buses, reactive power at generator buses, and apparent power flow in the lines. Thus, $x = [p_g^T, p_l^T, p_r^T]$ and $y = [v^T, q_g^T, s^T]$.

The disadvantage of this approach is its high complexity that requires a lot of samples for an accurate approximation over a linearized fit of the non-linear function. Fig. 1(a) illustrates this phenomenon. Our proposed *hybrid* GPR for AC-PF avoids this by combining a linear DC approximated physical model, and

an additive GP-part that is learned on the residuals between the non-linear AC and the linear DC model:

$$y_i - z(x_i) = g(x_i) + \omega_i, \qquad i = 1, ..., N, \tag{4}$$

where z is a known or estimated linear part, and g is an additive GP-term that describes unknown non-linearities of the system. As shown for a candidate non-linear function in Fig. 1(b), this approach has better fit. IN the next section, we describe how the hybrid GP is estimated using available historical data.

C. Gaussian Process Regression

Gaussian process regression (GPR) [24] is a non-parametric Bayesian approach used for non-linear regression problems. We employ GPR, as a supervised machine learning model, to infer a vector-valued residual function $g(x_i)$ in (4) using a dataset of N samples generated from previously collected measurements of residual outputs r_i and inputs x_i as:

$$r_i = y_i - z(x_i) = g(x_i) + \omega_i, \qquad i = 1, ..., N \tag{5}$$

where the measurement noise on data points ω_i corresponds to the process noise in (4). Historical dataset of inputs x and residual outputs r for learning GPR model is given by:

$$S = \{(x_i, r_i)_{i=1}^N, x_i \in \mathbb{R}^{n_x}, r_i \in \mathbb{R}^{n_y}\}$$

where n_x and n_y are total number of input and output variables.

Each output dimension r_a, $a = 1, ..., n_y$ is learned independently, given the input data $X = [x_i^T]_{1 \le i \le N} \in \mathbb{R}^{N \times n_x}$. Thus, specifying a GP-prior on g in each output dimension $a \in \{1, ..., n_y\}$ with defined a prior mean function $m_a(\cdot)$, chosen kernel $k_a(\cdot, \cdot)$ and conditioning on the data S results in a normally (Gaussian) posterior distribution along dimension a at a test point x_*, defined with mean and variance:

$$\mu_a(x_*) = k_*^T (K_a + \sigma_{n,a}^2 I)^{-1} [r]_{:,a} \tag{6a}$$

$$\sigma_a^2(x_*) = K_{**} - k_*^T (K_a + \sigma_{n,a}^2 I)^{-1} k_* \tag{6b}$$

where $\sigma_{n,a}^2$ is a variance of *process* noise corresponding to *measurement* noise in (5); I is an identity matrix; $[K_a]_{ij} = k_a(x_i, x_j)$ is the Gram matrix of the sampled training data points; $k_* = k_a(x_i, x_*)$ is the vector of covariance between unseen point and training samples; $K_{**} = k_a(x_*, x_*)$ is scalar; and $m_a(X) = [m_a(x_1), m_a(x_2), ..., m_a(x_N)]$. According to [24], a good prior leads to good data interpolation in GPR. We consider a commonly used zero prior mean $m(X) = 0$. Below we use the squared exponential (SE) kernel for the covariance:

$$k_a(x_i, x_j) = \sigma_{f,a}^2 \exp\left(-\frac{1}{2}(x_i - x_j)^T \Lambda_a^{-1} (x_i - x_j)\right)$$

in which $\sigma_{f,a}^2$ is the signal variance and $\Lambda_a = diag[l_{a,1}^2, l_{a,2}^2, ..., l_{a,n_x}^2]$ is a positive diagonal length scale matrix. The predictive distribution of the multivariate output dimension is given by:

$$\mu(x_*) = [\mu_1(x_*), \mu_2(x_*), ..., \mu_{n_y}(x_*)]^T$$
$$\Sigma(x_*) = diag([\sigma_1^2(x_*), \sigma_2^2(x_*), ..., \sigma_{n_y}^2(x_*)])$$

Cholesky decomposition is applied to minimize computational efforts in inverting $(K_a + \sigma_{n,a}^2 I)$ during prediction.

GPR training comes down to optimization of the hyperparameters $\theta = [\{\Lambda_a\}, \sigma_{f,a}^2, \sigma_{n,a}^2]$, where $\sigma_{n,a}^2$ is a process noise and $\{\Lambda_a\}$ is the set of length scale hyperparameters in the symmetric matrix Λ_a. The Maximum Likelihood Estimate with the log marginal likelihood is utilized to optimize hyperparameters as follows [25]:

$$\begin{aligned} \log\ p([r]_{:,a}|X, \Theta) = &\frac{1}{2}[r]_{:,a}^T (K_a + \sigma_{n,a}^2 I)[r]_{:,a} \\ &-\frac{1}{2}\log|K_a + \sigma_{n,a}^2 I| - \frac{N}{2}\log 2\pi \end{aligned} \tag{7}$$

The negative log likelihood is minimized to find optimal hyperparameters values using SLSQP solver [26].

D. Uncertainty Propagation

In the GPR model, for the given deterministic inputs, the predicted outputs have a normal (Gaussian) distribution described with mean and variance (6). Since we consider a stochastic problem, the inputs x are also random. Evaluating the outputs of a GPR from an input distribution is generally intractable, and the resulting outputs' distribution is not normal. Although some approximations of the resulting distribution exist [27], most of them are computationally demanding and very conservative. Thus, we consider computationally cheap and a practical approximation where the uncertain inputs are modelled via Gaussian distribution:

$$x_* \sim \mathcal{N}(\mu_{x_*}, \Sigma_{x_*}) = \mathcal{N}\left(\begin{bmatrix} p_g \\ p_d \end{bmatrix}, \begin{bmatrix} \Sigma_g & \Sigma_{gd} \\ \Sigma_{gd}^T & \Sigma_d \end{bmatrix}\right) \tag{8}$$

where p_g, $p_d = [p_l^T, p_r^T]^T$ are vectors of optimized control variables and forecasted operating points; $\Sigma_g \in \mathbb{R}^{n_g x n_g}$, $\Sigma_d \in \mathbb{R}^{n_d x n_d}$ and $\Sigma_{gd} \in \mathbb{R}^{n_g x n_d}$ are input covariance sub-matrices.

The covariance Σ_d models variances of fluctuations of loads and renewable energy based generators and correlations between them. We assume Σ_d to be diagonal, so $(\Sigma_d)_{ii} = \sigma_i^2, (\Sigma_d)_{ij} = 0\ \forall i \neq j$.

The sub-matrix Σ_g refers to uncertainties in generators which are caused due to the fluctuations in loads and RES via Automated Generation Control (AGC) (see Eq. 2). Following the AGC, the covariance in the generator's response is given as

$$(\Sigma_g)_{ij} = \alpha_i \alpha_j tr(\Sigma_d), \qquad i, j \in \mathcal{G}$$

Similarly, the correlation between the generation and the uncertain inputs d are given by the sub-matrix Σ_{gd}:

$$(\Sigma_{gd})_{ij} = \alpha_i \sigma_j^2, \quad i \in \mathcal{G}, j \in \mathcal{B} \setminus \mathcal{G},$$

where r_t is given by Eq. (5).

To determine $\mu_a(x_*)$ and $\sigma_a^2(x_*)$ from a Gaussian input $x_* \sim \mathcal{N}(\mu_{x_*}, \Sigma_{x_*})$, various approximation methods of the posterior of the GPR model have been proposed in the literature [28], [29]. This work considers a first-order Taylor approximation (TA1) for fitting nonlinearities. Applying the

first order Taylor approximation (TA1) in Eq. (6), the mean value and variance of output distribution are:

$$\mu_{TA1}(x_*) = \mu_a(\mu_{x_*}) \tag{9a}$$
$$\sigma^2_{TA1}(x_*) = \sigma^2_a(\mu_{x_*}) + \bigtriangledown\mu(\mu_{x_*})\Sigma_{x_*}(\bigtriangledown\mu(\mu_{x_*}))^T \tag{9b}$$

The mean prediction on a random input x_* does not provide any correction over the zero-order. At the same time, the variance takes into account the gradient of the posterior mean (6a) and the covariance Σ_{x_*} of the input of the GPR.

E. Sparse Approximation

Large data sets make the GPR model intractable since the inversion of the $N \times N$ covariance matrix K_a requires $O(N^3)$ time. Sparse Gaussian Processes reduce complexity by selecting the inducing inputs and the kernel hyperparameters by maximizing a lower bound of the marginal log-likelihood [30]. The idea behind sparse Gaussian Processes is to use a small set of points, often referred to as inducing points, $X_m = [x_i^T]_{1\le i\le M} \in \mathbb{R}^{M\times n_x},\ M \le N$, to better approximate marginal likelihood. For such a subsample, the approximate posterior GPR mean and covariance function depends on a mean vector μ_m and a covariance matrix A_m as:

$$\begin{aligned}\widehat{\mu}_a(x_*) &= k_{m*}^T(K_{mm,a} + \sigma^2_{m,a}I)^{-1}\mu_{m,a}\\ \widehat{\sigma}^2_a(x_*) &= K_{**} - k_{m*}^T(K_{mm,a} + \sigma^2_{m,a}I)^{-1}k_{m*}\\ &\quad + k_{m*}^T K_{mm,a} A_m K_{mm,a} k_{m*}\end{aligned}$$

where

$$\mu_{m,a} = \sigma^{-2}_{m,a}K_{mm,a}\Gamma K_{mn,a}[r]_{:,a}, \quad A_m = K_{mm,a}\Gamma K_{mm,a}$$
$$\Gamma = (K_{mm,a} + \sigma^{-2}_{m,a}K_{mn,a}K_{nm,a})^{-1}$$

where K_{mm} is $M \times M$ covariance matrix of $\{X_i\}_{i=1}^m$; K_{mn} is $M \times N$ covariance matrix between $\{X_i\}_{i=1}^M$ and the training set X; k_{m*} is the cross-covariance vector.

The sparse approximation reduce the time complexity of the GPR prediction to $O(NM^2)$ from N^3 preserving information from original sample span.

IV. Hybrid CC-OPF based on Gaussian Process Regression

A. Setup

The chance-constrained optimization problem given by Eq. (3) is computationally hard, and solving a deterministic problem with over uncertainty realizations instead of the stochastic one is a common way to decrease its complexity. Although the letter helps to get rid of uncertainty, the non-linearity of the power flow equations remains an issue.

In contrast to the classical scenario optimization approach mentioned above, modeling power flow equations with uncertainty through Gaussian processes lead to a more tractable solution. In our hybrid approach, we substitute the power flow balance equation with a combination of a linear approximation and an additive GPR model learned on the residuals:

$$\mu_y = f(\mu_{x_*}) + \mu(x_*) \tag{10a}$$
$$\sigma^2_y = diag(\Sigma_f(\Sigma_{x_*})) + diag(\Sigma(x_*)) \tag{10b}$$

where $f(\mu_{x_*}) = [v_{dc}, A\mu_{x_*}^T + b]$; $v_{dc} = 1$ p.u. is constant DC output voltage, A and b are the parameters; $diag(\Sigma_f(\Sigma_{x_*}))$ and $diag(\Sigma(x_*))$ are column vectors of diagonal matrix elements where $\Sigma_f = (0, 0; 0A\Sigma_{x_*}A^\top)$. To improve tractability, one can reformulate chance constraints for both output and input decision variables. According to [31], one can express individual output probabilities in terms of the mean μ_y and variance σ^2_y as:

$$\begin{cases}\mathbb{P}(\mu_{y_z} + \sqrt{\sigma^2_{y_z}} \le y_z^{max}) \le 1 - \epsilon_{y_z}\\ \mathbb{P}(\mu_{y_z} - \sqrt{\sigma^2_{y_z}} \ge y_z^{min}) \le 1 - \epsilon_{y_z}.\end{cases} \tag{11a}$$

equivalently $y_z^{\min} + \tau_{y_z}\sqrt{\sigma^2_{y_z}} \le \mu_{y_z} \le y_z^{\max} - \tau_{y_z}\sqrt{\sigma^2_{y_z}}$, where τ_{y_z} is the quantile function $\Phi^{-1}(1-\epsilon_{y_z})$ of the standard normal distribution and $z = 1, ..., n_y$.

Similarly, one can reformulate constraints on the input variables as in [6]:

$$\begin{cases}\mathbb{P}(p_{g_k} + \alpha_k\, tr(\Sigma_d) \le p_{g_k}^{max}) \le 1 - \epsilon_{p_{g_k}}\\ \mathbb{P}(p_{g_k} - \alpha_k\, tr(\Sigma_d) \ge p_{g_k}^{min}) \le 1 - \epsilon_{p_{g_k}}\end{cases} \tag{12a}$$

equivalently, as in (11):

$$p_{g_k}^{min} + \tau_{p_{g_k}}\alpha_k\, tr(\Sigma_d) \le p_{g_k} \le p_{g_k}^{max} - \tau_{p_{g_k}}\alpha_k\, tr(\Sigma_d) \tag{13}$$

where $\tau_{p_{g_k}} = \Phi^{-1}(1 - \epsilon_{p_{g_k}})$ and $k = 1, ..., n_g$.

The cost function depends quadratically on the participation factors α and generator active powers p_g:

$$\begin{aligned}&\sum_{k\in\mathcal{G}} \mathbb{E}_\omega[c_k(p_{g,k}(w))]\\ &\qquad = \sum_{k\in\mathcal{G}}\{c_{2k}(p_{g_k}^2 + tr(\Sigma_d)\alpha_k^2) + c_{1k}p_{g_k} + c_{0k}\}\end{aligned} \tag{14}$$

where $\{c_{2k}, c_{1k}, c_{0k}\}_{k=1}^{n_u} \ge 0$ are scalar cost coefficients.

Applying the GP-balance equation (10), an analytical reformulation of chance-constraints with uncertainty margins (11) (12), the hybrid GP CC-OPF problem is:

$$\min_{p_g,\alpha,\mu_y,\sigma_y} \sum_{k\in\mathcal{G}}\{c_{2k}(p_{g_k}^2 + tr(\Sigma_d)\alpha_k^2) + c_{1k}p_{g_k} + c_{0k}\} \tag{15a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{G}} \alpha_k = 1,\ \alpha_k \ge 0 \tag{15b}$$

$$\sum_{k\in\mathcal{G}} p_{g_k} = \sum_{i\in n_l} p_{l_i} - \sum_{j\in n_r} p_{r_j} \tag{15c}$$

$$\mu_y = f(\mu_{x_*}) + \mu(x_*) \tag{15d}$$

$$\sigma^2_y = diag(\Sigma_f(\Sigma_{x_*})) + diag(\Sigma(x_*)) \tag{15e}$$

$$y_z^{min} + \lambda_{y_z} \le \mu_{y_z} \le y_z^{max} - \lambda_{y_z} \tag{15f}$$

$$u_k^{min} + \lambda_{p_{g_k}} \le p_{g_k} \le u_k^{max} - \lambda_{p_{g_k}} \tag{15g}$$

where $\lambda_{y_z} = \tau_{y_z}\sqrt{\sigma^2_{y_z}}$ for $z \in n_y$ and $\lambda_{p_{g_k}} = \tau_{p_{g_k}}\alpha_k\sqrt{tr(\Sigma_d)}$ are the uncertainty margins; $\sqrt{tr(\Sigma_d)}$ represents the standard deviation of the total active power imbalance Ω. The equations (15c)–(15e) describe the hybrid GP-balance equations by replacing a AC power flow equations with a data-driven regression vector-function. These equations represent

the balance between generation and consumption in the power systems by optimizing the decision variables of the controllable generators. However, the proposed OPF optimization method (15) is a non-convex optimization problem. Since the SE kernel is twice differentiable, the problem can be solved using a nonlinear optimization method. Accordingly, we used a primal-dual interior-point algorithm with filter line-search to solve (15). This optimization problem is solved within the CasADi [32] nonlinear optimization framework using the available IPOPT [33] solver.

B. Evaluation Metrics

To evaluate the performance of the GPR regression model, the root mean square error (RMSE) metric is used for each output variable a. If the predicted output vector $\hat{y}$ (GP-mean values) and ground truth y (AC-PF output values), the RMSE is calculated as follows:

$$RMSE_a = \sqrt{\frac{1}{N_*}\sum_{i=1}^{N_*}([y]_{i,a} - [\hat{y}]_{i,a})^2} \tag{16}$$

where N_* is number of test samples. The average RMSE score of all outputs $RMSE = n_y^{-1}\sum_{a=1}^{n_y} RMSE_a$ is the accuracy measure we use in our experiments.

V. Results

We used Intel Core i7-5500U CPU @ 2.40GHz with 8GB RAM for numerical computations and `pandapower` [34] package to validate the results. The source code for experiments is accessible online[1].

A. Case Study

Performance and scalability of the proposed hybrid GP CC-OPF method are evaluated over 9-bus and 39-bus IEEE test cases. We model the forecast errors as multivariate Gaussian random variables with zero mean and standard deviation corresponding to 15% of (forecasted) loads ($\sigma_l = 0.15\, p_l$) and 30% for the renewable generations ($\sigma_r = 0.3\, p_r$). Following [35], we set the acceptable violation probabilities of generator limits to $\epsilon_{p_g} = 0.1\%$ for controllable generators, and $\epsilon_y = 2.5\%$ for all other constraints.

B. Model Performance

The hybrid and full GPR models are trained on residuals between datasets collected from full AC-OPF and DC-OPF approximation. The synthetic data of the IEEE 9-bus system consists of 8 input features and 15 outputs, while for IEEE 39, the model has 37 inputs and 74 outputs. We generated 75 and 200 random training samples for IEEE 9 and 39-bus systems and considered a different number of inducing points of 10, 30, and 50 for both full and hybrid GPR models. The iterative interior-point algorithm to solve both hybrid and full GPR CC-OPF and find a feasible solution with convergence tolerance of $\epsilon_{tol} = 10^{-5}$.

[1]https://github.com/mile888/hybrid_gp

Models with the TA1 approximation method taking into account the full training dataset and a different number of inducing points are compared using RMSE as an accuracy score and CPU time to estimate computational complexity.

The results in Table I indicate that the hybrid GPR model with TA1 outperforms both RMSE and CPU time with respect to full TA1 [23]. Similarly, sparse approximation allows to significantly reduce the time complexity for both hybrid and full approaches. Notice, that RMSE computed between the DC approximation and the set of samples is RMSE = $6.65e^{-2}$ p.u. (IEEE 9) and $3.04e^{-2}$ p.u. (IEEE 39). In other words, applying the GPR model with a few samples leads to at least 6 times more accurate approximation than the DC baseline.

Test Case	**IEEE 39**		**IEEE 9**	
Parameters	RMSE 10^{-2} p.u.	Time sec.	RMSE 10^{-2} p.u.	Time sec.
hybrid TA1	**0.53**	10.82	**0.30**	0.24
hybrid-sparse 50 TA1	1.12	2.93	0.56	0.10
hybrid-sparse 30 TA1	1.22	1.23	0.79	0.06
hybrid-sparse 10 TA1	1.41	**0.51**	1.22	**0.02**
full TA1	1.37	19.38	0.65	0.35
full-sparse 50 TA1	2.69	4.12	0.97	0.17
full-sparse 30 TA1	2.80	1.42	1.67	0.13
full-sparse 10 TA1	3.25	**0.51**	2.43	0.03

TABLE I: GP CC-OPF approximation method results for IEEE 9 and IEEE 39 bus systems

C. Comparison of GP CC-OPF to Sample-Based Methods

In this section, we compare solution cost, computational time, and reliability of the proposed hybrid and full GPR CC-OPF approaches with Scenario Approximation (SA) CC-OPF [36]. In all cases, we use the first-order Taylor approximation (TA1) as a function linearization to simplify the computations. The empirical constraint violation is computed using 1000 Monte-Carlo samples that follow (Gaussian) uncertainty distribution.

We compare our approach to two baselines. (A) we solve a deterministic AC-OPF problem for each uncertainty realization and take a corresponding quantile in the output distribution. Although computationally efficient, this approach is not guaranteed to provide a (probabilistically) feasible solution; (B) once solving a deterministic AC-OPF, one gets we get optimal participation factors α^* and power distribution. Later on, we find a solution to the AC power flow balance equation for each uncertainty realization and run the Automatic Generation Control with fixed participation factors.

Notice that (A) considers generalized nonlinear feedback, while (B) does not optimize the participation factor α and uses the same participation factors for each sample. Thus they give the upper and the lower bound on the feasible solution value. Numerical evaluation of the bounds in given in Table II.

D. Comparison of Uncertainty Margins

Figure 2 compares uncertainty margins [12] calculated with the hybrid and full GP CC-OPF (red and blue), the Monte Carlo simulation of AC-OPF for base-case (yellow and green),

System	IEEE 39			IEEE 9		
Parameters	Cost, 10^6	Failure Prob., %	Time sec.	Cost, 10^3	Failure Prob., %	Time sec.
A (full recourse)	7.87	n/a	0.93	4.06	-	0.86
B (base case)	7.53	35.36	0.93	3.47	9.76	0.86
Full GP CC-OPF	7.75	2.36	19.38	4.04	0.44	0.35
Hybr. GP CC-OPF	7.75	2.48	9.58	4.02	0.8	0.12
CC-OPF (20 scenarios)	n/a	n/a	n/a	3.48	7.52	4.4
CC-OPF (50 scenarios)	7.64	14.20	96.4	3.84	4.92	12.1
CC-OPF (100 scenarios)	7.70	8.24	184.70	3.99	1.24	22.80
CC-OPF (200 scenarios)	7.81	0.16	505.00	n/a	n/a	n/a

TABLE II: Cost function values, CPU time and probability of a failure for various setups.

and the Monte Carlo simulation for CC-OPF solution (brown and salmon) with 100 (IEEE 9) and 200 (IEEE39) scenarios. The bar plots represent an example of voltage, reactive power, and apparent power flow uncertainty margins for each system.

The 3 standard deviation distance for uncertainty margins is analytically derived using the estimated GP-variance (15e) around the estimated GP-mean (15d). The uncertainty margins from Monte Carlo are asymmetrical due to MC-based distributions from non-linear AC-OPF. Thus, Monte Carlo uncertainty margins determine upper $(1-\epsilon)$ = 99.73% and lower (ϵ) = 0.27% quantiles (equal to 3 std) of the output distribution around the AC-PF solution $y(x)$, denoted by $y^{1-\epsilon}$ and y^{ϵ}. Thus, the constraint tightenings are:

$$\lambda^{upper} = y^{1-\epsilon} - y(x) \qquad \lambda^{lower} = y(x) - y^{\epsilon} \tag{17a}$$

The power margins are larger in absolute terms than the voltage margins. The GP CC-OPF margins of both approaches are symmetric, leading to larger or smaller margins compared with the upper and lower empirical quantiles obtained from the Monte Carlo AC-OPF and CC-OPF simulations in (17). This indicates that the GP uncertainty margins are well approximated.

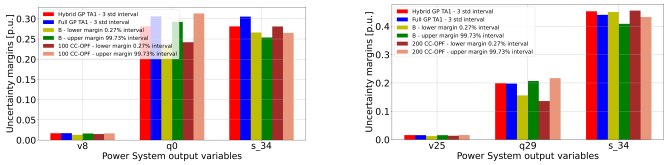


Fig. 2: IEEE 9 case (left) and IEEE 39 case (right). Output uncertainty margins λ, for example, voltage, reactive and apparent power constraints.

E. Distributional Robustness

In operational practice, a joint distribution of renewables and loads is not known exactly, and one needs to account for it when solving a problem. Furthermore, distribution mismatch might result from broken sensors or a cyberattack. This section shows that the proposed approach is robust against a sufficiently large mismatch between the expected distribution and the real one. Figure 1 illustrates the case of a corrupted distribution. Indeed, the interval $[7.5, 17]$ has no input data, leading to poor performance of simple GPR. However, the Hybrid GPR presented in this paper handles such intervals better. We also observe a similar result in the case of power grids.

Our experiment consists of three major steps: (1) change the data distribution adversarially, (2) train over the corrupted data, and (3) test the predictions of models. We compare the classical GPR [23] and the proposed hybrid GRR model. For an adversarial change of the input data distribution, we throw out a part of the input data within a certain interval; an example of this is shown in Fig. III.

We use IEEE 9 case with the dataset generated in Section IV for our experiment. In each of the experiments, only one of the buses (bus 4, bus 6, and bus 8) is subjected to whipping out a 15 MW interval. See Fig. III for details. Later on, we train the classical and the hybrid GPR on this data. Table III illustrates a superior performance of the hybrid GPR on this data.

Experiment	RMSE GPR	RMSE Hybrid GPR	% dropped
Load 4	$3.65 \cdot 10^{-1}$	$\mathbf{9.11 \cdot 10^{-3}}$	55.0%
Load 6	$3.89 \cdot 10^{-1}$	$\mathbf{11.60 \cdot 10^{-3}}$	38.0%
Load 8	$9.25 \cdot 10^{-2}$	$\mathbf{8.34 \cdot 10^{-3}}$	33.5%

TABLE III: RMSE comparison for GPR and hybrid GPR over corrupted distributions. The last column specifies the portion of data distribution of the corresponding load that has been dropped.

VI. Discussion

The paper studies various machine learning approaches to chance-constrained AC optimal power flow problems. Because of the high non-linearity of the problem, we focused mostly on Gaussian Process Regression, one of the most popular model-free approaches to approximate non-linear stochastic data. Furthermore, it allows for avoiding spurious local minimums thus improving tractability of the AC-OPF problem. As the Gaussian Process Regression has poor scalability, we utilize a sparsity trick that allowed us to significantly reduce the complexity. Finally, we proposed a hybrid GRP approach that demonstrates higher robustness to missing and corrupted data than the classical GPR.

References

[1] L. Hockstad and L. Hanel, "Inventory of us greenhouse gas emissions and sinks," Environmental System Science Data Infrastructure for a Virtual Ecosystem, Tech. Rep., 2018.

[2] Z. Liu, L. v. d. Sluis, and W. Winter, "Challenges, experiences and possible solutions in transmission system operation with large wind integration," *OSTI*, 2012.

[3] D. Koutsoyiannis, "The unavoidable uncertainty of renewable energy and its management," in *Proceedings of the EGU General Assembly Conference, Vienna, Austria*, 2016, pp. 17–22.

[4] M. J. Morshed, J. B. Hmida, and A. Fekih, "A probabilistic multi-objective approach for power flow optimization in hybrid wind-pv-pev systems," *Applied energy*, vol. 211, pp. 1136–1149, 2018.

[5] S. Shargh, B. Mohammadi-Ivatloo, H. Seyedi, M. Abapour *et al.*, "Probabilistic multi-objective optimal power flow considering correlated wind power and load uncertainties," *Renewable Energy*, vol. 94, pp. 10–21, 2016.

[6] 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.

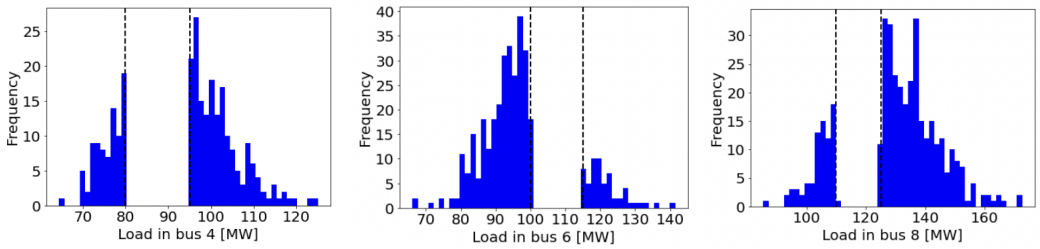


Fig. 3: Histograms of buses' corrupted loads for each of the three experiments.

[7] X. Du, X. Lin, Z. Peng, S. Peng, J. Tang, and W. Li, "Chance-constrained optimal power flow based on a linearized network model," *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106890, 2021.

[8] N. Viafora, S. Delikaraoglou, P. Pinson, and J. Holbøll, "Chance-constrained optimal power flow with non-parametric probability distributions of dynamic line ratings," *International Journal of Electrical Power & Energy Systems*, vol. 114, p. 105389, 2020.

[9] A. Lukashevich, V. Gorchakov, P. Vorobev, D. Deka, and Y. Maximov, "Importance sampling approach to chance-constrained dc optimal power flow," *arXiv preprint arXiv:2111.11729*, 2021.

[10] A. B. Owen, Y. Maximov, and M. Chertkov, "Importance sampling the union of rare events with an application to power systems analysis," *Electronic Journal of Statistics*, vol. 13, no. 1, pp. 231–254, 2019.

[11] 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.

[12] J. Schmidli, L. Roald, S. Chatzivasileiadis, and G. Andersson, "Stochastic ac optimal power flow with approximate chance-constraints," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[13] L. Roald and G. Andersson, "Chance-constrained AC optimal power flow: Reformulations and efficient algorithms," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2906–2918, 2017.

[14] M. Lubin, Y. Dvorkin, and L. Roald, "Chance constraints for improving the security of ac optimal power flow," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1908–1917, 2019.

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

[16] B. Cousins and S. Vempala, "A cubic algorithm for computing gaussian volume," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on discrete algorithms*. SIAM, 2014, pp. 1215–1228.

[17] L. Khachiyan, "Complexity of polytope volume computation," in *New trends in discrete and computational geometry*. Springer, 1993, pp. 91–101.

[18] A. Lukashevich and Y. Maximov, "Power grid reliability estimation via adaptive importance sampling," *IEEE Control Systems Letters*, vol. 6, pp. 1010–1015, 2021.

[19] P. Pareek and H. D. Nguyen, "Gaussian process learning-based probabilistic optimal power flow," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 541–544, 2020.

[20] A. S. Zamzam and K. Baker, "Learning optimal solutions for extremely fast ac optimal power flow," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2020, pp. 1–6.

[21] X. Pan, T. Zhao, M. Chen, and S. Zhang, "Deepopf: A deep neural network approach for security-constrained dc optimal power flow," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 1725–1735, 2020.

[22] M. K. Singh, V. Kekatos, and G. B. Giannakis, "Learning to solve the ac-opf using sensitivity-informed deep neural networks," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2833–2846, 2021.

[23] M. Mitrovic, A. Lukashevich, P. Vorobev, V. Terzija, S. Budenny, Y. Maximov, and D. Deka, "Data-driven stochastic ac-opf using gaussian processes," 2022. [Online]. Available: https://arxiv.org/abs/2207.10781

[24] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer school on machine learning*. Springer, 2003, pp. 63–71.

[25] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[26] D. Kraft, "A software package for sequential quadratic programming," *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*, 1988.

[27] T. Koller, F. Berkenkamp, M. Turchetta, and A. Krause, "Learning-based model predictive control for safe exploration," in *2018 IEEE conference on decision and control (CDC)*. IEEE, 2018, pp. 6059–6066.

[28] A. Girard, C. E. Rasmussen, J. Quinonero-Candela, and R. Murray-Smith, "Gaussian process priors with uncertain inputs? application to multiple-step ahead time series forecasting," in *Advances in neural information processing systems*, vol. 15, 2003.

[29] M. P. Deisenroth, *Efficient reinforcement learning using Gaussian processes*. KIT Scientific Publishing, 2010, vol. 9.

[30] M. Titsias, "Variational learning of inducing variables in sparse gaussian processes," in *Artificial intelligence and statistics*. PMLR, 2009, pp. 567–574.

[31] L. Hewing, J. Kabzan, and M. N. Zeilinger, "Cautious model predictive control using gaussian process regression," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 6, pp. 2736–2743, 2019.

[32] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "Casadi: a software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 1–36, 2019.

[33] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, vol. 106, no. 1, pp. 25–57, 2006.

[34] L. Thurner, A. Scheidler, F. Schäfer, J.-H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, "pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, 2018.

[35] F. Abbaspourtorbati and M. Zima, "The swiss reserve market: Stochastic programming in practice," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1188–1194, 2015.

[36] I. Mezghani, S. Misra, and D. Deka, "Stochastic ac optimal power flow: A data-driven approach," *Electric Power Systems Research*, vol. 189, p. 106567, 2020.