Selecting Critical Scenarios of DER Adoption in Distribution Grids Using Bayesian Optimization

Olivier Mulkin, Miguel Heleno, *Senior Member, IEEE*, and Mike Ludkovski

***Abstract*—We develop a new methodology to select scenarios of DER adoption most critical for distribution grids. Anticipating risks of future voltage and line flow violations due to additional PV adopters is central for utility investment planning but continues to rely on deterministic or ad hoc scenario selection. We propose a highly efficient search framework based on multi-objective Bayesian Optimization. We treat underlying grid stress metrics as computationally expensive black-box functions, approximated via Gaussian Process surrogates and design an acquisition function based on probability of scenarios being Pareto-critical across a collection of line- and bus-based violation objectives. Our approach provides a statistical guarantee and offers an order of magnitude speed-up relative to a conservative exhaustive search. Case studies on realistic feeders with 200-400 buses demonstrate the effectiveness and accuracy of our approach.**

***Index Terms*—Distribution planning, DER adoption, scenario selection, Bayesian optimization, Gaussian process surrogates.**

I. Introduction

INVESTMENTS in new distribution grid capacity account for a significant share of utility capital expenditures [1], requiring utilities to periodically justify those investments to regulators through detailed assessments of distribution grid needs [2], [3]. These assessments rely heavily on long-term forecasts (typically spanning 5 to 10 years) of distribution system demand and Distributed Energy Resource (DER) penetration [4], which drive distribution capacity needs.

However, while there are methodologies to predict future DER diffusion at the city, substation, or neighborhood level [5], understanding "where" exactly these DERs will appear within a distribution feeder remains a challenge. In fact, for feeders with hundreds or even thousands of consumers, a single DER diffusion trajectory can translate into a high-dimensional set of scenarios of potential individual DER adoption at the nodes of the distribution grid. Planning localized feeder upgrades, such as reconductoring or voltage regulation, requires identifying, in this high-dimensional space, the specific combinations of nodal adoption patterns that drive new capacity needs.

To address this issue, this paper proposes a new methodology, based on Bayesian Optimization (BO), to forecast and identify scenarios of behind-the-meter DER adoption that are critical to distribution grid planning.

A. Literature Review

Long-term forecasts of loads and DERs start at the system level to provide a broad picture of load and DER growth across the utility's territory while ensuring compatibility with other planning processes [3], [6]. Utilities capture data on historical consumption, demographic and economic indicators, weather trends, and DER policy adoption rates [2], [7] and use various statistical-based methodologies to generate such aggregate-level forecasts. For example, [8] introduces a probabilistic forecasting model to predict the magnitude and timing of peak electricity demand, [9] applies spatial forecasting to assess the impact of load growth in power distribution networks, [10] proposes a density forecasting approach to address the uncertainty associated with long-term planning, and [11] uses hourly data to model long-term trends in probabilistic forecasts. A methodology that uses an eXtreme Gradient Boosting algorithm with different sequential configurations is proposed in [12] to combine energy consumption and peak power demand forecasting. In a different direction, the work of [13] proposes a hybrid approach that combines top-down features (economic and weather) with bottom-up features (individual consumer loads and DER adoption propensity) to provide long-term forecasts. More recently, [14] introduce a method based on system dynamics to forecast long-term electricity consumption by dividing the driving factors of energy use into population, per capita GDP, energy intensity, and energy structure. In [15] a hybrid convolutional neural network and long short-term memory method is proposed to forecast monthly energy consumption for three year planning horizons.

For distribution system planning, utilities disaggregate these system-level forecasts into smaller geographic units using tools like LoadSEER [2], [6]. These tools rely on statistical spatial regressions [16] or agent-based models [3], [17] to generate future scenarios of load growth and DER diffusion. In the literature, earlier works proposed land-use pattern recognition techniques to anticipate new demand using cellular data [18], [19], while recent studies focus on forecasting a combination of loads and DERs, including their temporal characteristics [20], [21]. However, these models are often based on geographical

Received 18 January 2025; revised 20 September 2025; accepted 9 January 2026. Date of publication 19 January 2026; date of current version 22 June 2026. This work was supported by the U.S. Department of Energy under Grant DE-AC02-05CH11231. Paper no. TPWRS-00120-2025. *(Corresponding author: Olivier Mulkin.)*

Olivier Mulkin and Mike Ludkovski are with the Department of Statistics & Applied Probability, UC Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: omulkin@ucsb.edu).

Miguel Heleno is with Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA (e-mail: MiguelHeleno@lbl.gov).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2026.3655693.

Digital Object Identifier 10.1109/TPWRS.2026.3655693

representations (e.g., neighborhoods, zip codes), which do not necessarily correspond to specific feeder locations. To forecast household-level consumption, [22] proposes an ensemble of forecasts based on the historical median distribution among similar households. The forecast of DERs at the consumer site can be based on rules (e.g., assuming DER power proportional to base load [23]) or on individual consumer models [24]. For example, [25] develops a feeder-specific agent-based model integrating customer GIS data, EV adoption criteria (e.g., neighbor influence, car age), and charging station placement. This feeder-level adoption is highly granular, down to individual parcels and specific customer distribution transformers. In the utility space, Portland General Electric (PGE) developed its own AdopDER tool, a statistical model that captures consumer-level propensity to adopt DERs [26].

Although these DER forecasting methods provide feeder-level spatial granularity, they may not be suitable for planning purposes. This is because they primarily identify load and DER patterns that are more likely to occur, rather than those that would impose greater stress on the system. To address this limitation, recent works have proposed methods to select load growth and DER penetration scenarios that could trigger capacity upgrades by mapping these scenarios to specific equipment capacity violations. For instance, the authors of [27] presented a method to forecast substation annual maximum demand by analyzing extreme load events and their relationship with external factors, leveraging extreme value theory. Similarly, [28] introduced a data-driven approach to estimate transformer ratings under varying temperature and load conditions. Regarding DER adoption, the authors of [29] proposed a method based on a binomial distribution to predict the probability of transformer overloading caused by electric vehicle (EV) adoption.

Nonetheless, a major limitation of these approaches is that they identify the most critical scenarios in relation to a single equipment (e.g., substations or service transformers). Expanding to the entire distribution system would require identifying all critical scenarios corresponding to all potential system-level violations. In distribution grids with hundreds of nodes and millions of potential scenarios of DER adoption, this becomes a search in a high-dimensional space making exhaustive evaluation infeasible. We propose a guided "smart" search via Bayesian Optimization, designed to address this challenge by substantially reducing the number of scenarios that must be evaluated, as demonstrated in feeders with several hundred nodes. While we focus on these realistic test cases, the proposed approach offers a scalable foundation that could be extended to larger networks in future work.

B. Contributions

This paper introduces a novel methodology for forecasting long-term DER adoption patterns for the purpose of distribution system planning. Unlike existing high-resolution locational DER forecast approaches [23], [24], [25], [26], which focus on identifying the most probable scenarios, our methodology prioritizes finding the scenarios that are most critical to the grid and, therefore, more relevant for planning exercises. Previous forecasting approaches following this logic have been limited to mapping critical scenarios based on the capacity of single distribution grid assets [27], [28], [29]. Our work extends this concept to a system-wide perspective, leveraging BO to navigate the high-dimensional space and identify critical scenarios relevant for planning upgrades in the entire grid.

Our main contributions are threefold:

- We link long-term DER adoption scenarios to future violations of distribution grid limits (voltage and line capacity), in order to identify the set of critical scenarios that are most relevant for distribution grid planning. We map such critical scenarios to the Pareto frontier of the above grid stress objectives vector.
- We propose an algorithm that efficiently performs this search, provides statistical guarantees, and achieves an order-of-magnitude speed-up compared to a conservative exhaustive search.
- Using two realistic case studies, we demonstrate that the DER adoption scenarios critical for distribution grid planning are not necessarily those with extreme aggregated adoption patterns. This highlights the importance of conducting a thorough and systematic locational search, such as the one presented in this paper.

Statistically, our methodology leverages Gaussian Process (GP) surrogates to efficiently forecast line and bus voltage violations. GPs have been previously used in several power system applications, for instance, to describe the uncertainty of optimal power flows [30], to perform transient stability analysis of large-scale power systems [31], [32], to simulate renewable energy power generation [33], [34], etc. However, this paper is, to the best of our knowledge, the first application of GPs in long-term forecasts of distribution grid load and stress for the purpose of planning.

C. Paper Organization

The rest of the paper is organized as follows. Sections II and III introduce the problem statement and the methodology, respectively. Numerical results on two realistic feeders are presented in Section IV; Section V discusses the performance of our model and Section VI concludes.

II. Problem Statement

A distribution planning process implies a grid needs assessment [2], which is conducted in three main steps: i) define scenarios for load and DER penetration; ii) run a feeder-level power flow evaluation for those scenarios and assess potential line/transformer capacity and voltage violations; iii) propose a planning solution based on distribution capital projects to upgrade the system and solve those violations. The selection of scenarios in the first step is critical, as it directly influences the power flow analysis and ultimately shapes the planning solutions. In this section, we describe the problem of identifying the set of *critical scenarios*, i.e. those whose evaluation and resolution of violations ensures the grid remains feasible for all other potential scenarios. Throughout the paper, we use the behind-the-meter adoption of PV to better illustrate the problem and the methodology proposed. The extension to other types of DERs or to load is straightforward.

A. Scenarios of Adoption

We consider a radial distribution grid with L lines and B buses. Let $\mathcal{A}$ be the set of $A = |\mathcal{A}|$ potential PV adopters, i.e. network nodes where a PV system can be installed. We simulate scenarios of PV adoption with an agent-based extension of the dGen diffusion model [35], introduced in [36]. The probability of an agent adopting at time $t + \Delta t$ given the network state at time t is

$$\mathbb{P}(a_{t+\Delta t,j} = 1 | a_{t,j} = 0) = \left(p + \frac{q}{A} \sum_{k \in \mathcal{A}} a_{t,k} \right), \quad (1)$$

where p and q are the coefficients of innovation and adoption respectively, and $a_{t,j} = 1$ if agent j has adopted PV at time t, and $a_{t,j} = 0$ otherwise. The initial adoption status $a_{0,j}$ is uniformly sampled with a probability matching currently observed adoption rates. For a given time horizon T, scenarios describing potential future PV adoption over the network are obtained by independent Monte Carlo simulations with transition probabilities (1).

We denote by $\mathcal{X}$ the space of all scenarios based on model (1). Formally, $\mathcal{X} = \{0, 1\}^A$ as any scenario has a strictly positive probability, but many of those scenarios are extremely unlikely. Thus, a scenario $\mathbf{x} \in \mathcal{X}$ is a A-dimensional binary vector describing the state of adoption for every agent in the network at the end of the simulation period T, with j^{th} entry given by $x_j = a_{T,j}$.

B. Measuring Distribution Grid Impact

Assessing the impact of different PV adoption scenarios requires a measure of grid stress. Under a given adoption pattern $\mathbf{x} \in \mathcal{X}$ solving a power flow determines the voltage at each bus and the power flow through each line in the distribution network. For distributed PV planning studies, for example, the power flow evaluation is performed at the minimum daytime load, when solar radiation and reverse power flows are highest.

We represent this grid stress as a vector-valued function $\boldsymbol{f}$ of the scenario $\mathbf{x}$ and evaluate it at both buses and lines as function of potential planning solutions. For buses, we group them into P clusters $\mathcal{B}_1, \ldots, \mathcal{B}_P$ that partition $\mathcal{B}$, representing voltage control zones within the distribution grid. The voltage violation in each cluster (or voltage zone) is measured by the maximum deviation of the computed bus voltage—measured per unit (p.u.) relative to the nominal system voltage—from its acceptable operating range (often $\pm 5\%$). This measure quantifies the magnitude of the voltage control problem in each zone, which can be addressed by planning solutions, such as placement of capacitor banks (for undervoltage problems) or voltage regulators (for both over and undervoltage problems). The voltage zones (and number of clusters) can be defined *a priori* by the utility or automatically obtained by clustering techniques. The left panel of Fig. 1 shows an illustrative partition of buses into 12 clusters created with the Louvain community algorithm [37] for a representative feeder (p4rhs8).

For distribution lines, stress is defined by the deviation of the power flow (measured in p.u.) from the line capacity.

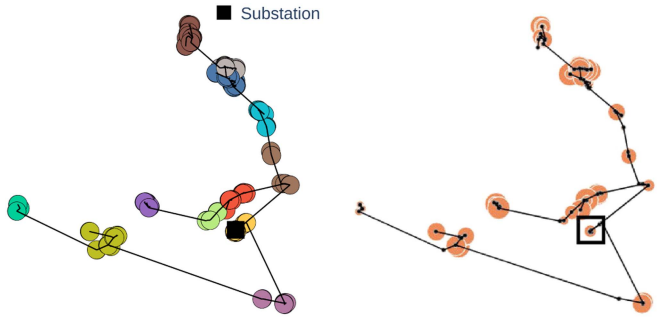

Fig. 1. Feeder p4rhs8 with 248 buses, 202 lines and 159 potential PV adopters, cf. Table I. *Left:* partition of buses into 12 bus objectives (different colors) via the Louvain community algorithm. *Right:* A representative critical scenario of PV adoption; circle size is proportional to installed PV capacity.

Thus, for a given scenario $\mathbf{x}$, let $volt_b(\mathbf{x})$ be the computed voltage in p.u. for bus $b \in \mathcal{B}$, and let v_b^u, v_b^b be its upper and lower voltage tolerance bounds. Furthermore, for a line $k \in \mathcal{L} := \{P+1, \ldots, P+L\}$, let $flow_k(\mathbf{x})$ be the computed power flow and $flow_k^M$ be its capacity, i.e. the maximum power flow line k can accommodate. Then, the stress function (using common notation for buses and lines) is defined by

$$f_k(\mathbf{x}) = \begin{cases} \max_{b \in \mathcal{B}_k} \left\{ \max \left(volt_b(\mathbf{x}) - v_b^u, v_b^l - volt_b(\mathbf{x}) \right) \right\} \\ \qquad \text{if } k \in \{1, \ldots, P\} \\ flow_k(\mathbf{x}) - flow_k^M \\ \qquad \text{if } k \in \{P+1, \ldots, P+L\}. \end{cases} \quad (2)$$

Positive stress values $f_k(\mathbf{x}) > 0$ quantify the severity of the bus or line violation. Otherwise, for a bus group, $f_k(\mathbf{x}) \leq 0$ measures how close the most stressed bus in $\mathcal{B}_k$ is to be violated. Similarly, for a line, $f_k(\mathbf{x}) \leq 0$ measures how close line k flow is to being overloaded. We collect all objectives in $\boldsymbol{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_P(\mathbf{x}), f_{P+1}(\mathbf{x}), \ldots, f_{P+L}(\mathbf{x})]^T$.

C. Defining Critical Scenarios

To define critical scenarios, we introduce the mapping

$$V_k(y) = \begin{cases} y^+ & k \leq P \\ \sum_{j=0}^{\infty} j \times \mathbb{I}_{[c_j, c_{j+1})}(y^+) & k > P \end{cases} \quad (3)$$

where $y^+ = \max(y, 0)$. V maps stress values to violation objectives. For bus objectives $k \leq P$, it simply takes the magnitude of the violation in each control zone, which will determine the voltage control solutions in the investment planning. For line objectives $k > P$, V_k bins excess power flows by levels of severity with respect to pre-specified thresholds c_j. This mirrors real-world practices, where utility planners often flag lines for reinforcement or upgrades when certain severity thresholds are exceeded (corresponding to ampacity upgrade options in reconductoring). Thus, for a given line, overloads that fall under the same c_j bracket require the same reconductoring solution and are equivalent from the grid needs assessment perspective.

To rank scenarios with violations, we introduce the dominance relation $\mathbf{x}' \succ \mathbf{x}$, such that $\mathbf{x}'$ dominates $\mathbf{x}$ if for all objectives

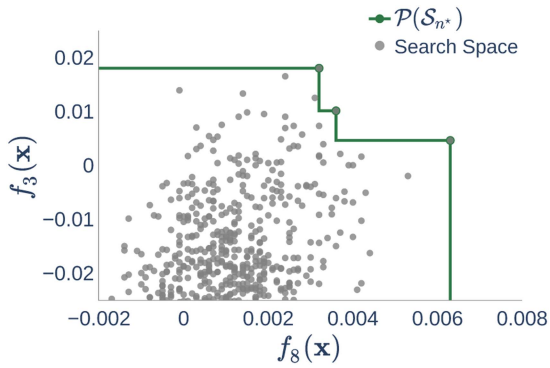

Fig. 2. Pareto frontier $\mathcal{P}(\mathcal{S}_{n^*})$ projected on two representative bus objectives f_3 and f_8. Each point represents the stress objective values $(f_3(\mathbf{x}), f_8(\mathbf{x}))$ of a scenario $\mathbf{x}$. The three green points on the staircase are the critical non-dominated scenarios.

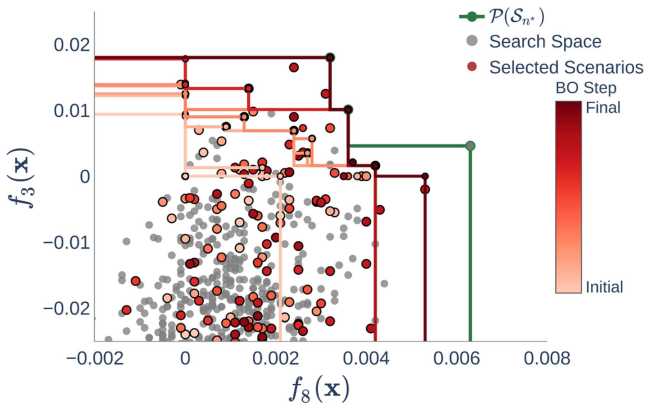

Fig. 3. Progression of the Pareto frontier projected onto two representative bus objectives. Each point represents the objective values $(f_3(\mathbf{x}), f_8(\mathbf{x}))$ of a scenario $\mathbf{x} \in \mathcal{S}_{n^*}$. Evaluated scenarios $\mathbf{x}_n$ (and respective Pareto fronts $\mathcal{P}_n$) are color-coded in terms of the step n of the BO algorithm. Non-evaluated scenarios are shown in gray.

k, $V_k(f_k(\mathbf{x})) \leq V_k(f_k(\mathbf{x}'))$ and $V_k(f_k(\mathbf{x})) < V_k(f_k(\mathbf{x}'))$ for at least one k. We call a scenario $\mathbf{x}$ *critical* if it is in the Pareto set with respect to $V(\boldsymbol{f})$, where the $(P+L)$-vector $V(\boldsymbol{f}(\mathbf{x}))$ stacks all the $V_k(f_k(\mathbf{x}))$. The Pareto set $\mathcal{P}(\mathcal{X}) = \{\mathbf{x} \in \mathcal{X} : \nexists\ \mathbf{x}' \in \mathcal{X}\ s.t.\ \mathbf{x}' \succ \mathbf{x}\}$ is the set of all non-dominated scenarios. The image of $\mathcal{P}(\mathcal{X})$ under $V(\boldsymbol{f})$ is called the *Pareto front*.

Using $\succ$ we map the search for critical scenarios, i.e. $\mathbf{x}$'s such that there are no other scenarios with a more serious violation in any bus group or line, to the task of identifying the Pareto set in the multi-objective optimization problem

$$\max_{\mathbf{x}\in\mathcal{X}} V(\boldsymbol{f}(\mathbf{x})). \tag{4}$$

We call an objective $k \in \{1, \ldots, P+L\}$ critical if it can be violated due to PV adoption: $\exists \mathbf{x} \in \mathcal{X} : V_k(f_k(\mathbf{x})) > 0$. Note that many objectives are typically non-critical, i.e. never bind. The right panel of Fig. 1 illustrates one critical scenario for the representative feeder p4rhs8 and Fig. 2 illustrates the Pareto frontier with respect to two bus objectives. Each point represents a different adoption scenario $\mathbf{x}$ with those lying on the staircase line being critical scenarios.

D. Use in Distribution Grid Planning

Given the definitions above, the set of critical DER adoption scenarios in the Pareto frontier captures the worst-case violation patterns across all possible adoption decisions. For distribution planning, this small set can replace the high-dimensional space of all adoption realizations. From a utility perspective, this greatly reduces planning complexity regardless of the methodology used. In less sophisticated environments, where planning is still manual, engineers can easily evaluate this small set of critical scenarios. In more advanced utilities, where optimal expansion planning is applied, these scenarios can serve as inputs to more robust or scenario-based optimization.

III. Proposed Methodology

To search for critical scenarios, we propose an algorithm based on BO [38]. The main spirit of BO is to guide the search for critical scenarios using a surrogate-based acquisition function that balances exploration and exploitation. Rather than directly maximizing the f_k's which requires extensive power flow evaluations, BO constructs a statistical model that uses previously evaluated scenarios to learn the relationship between scenarios and stress objectives. It then carries out a targeted search that only evaluates scenarios that iteratively maximize the acquisition function $\alpha : \mathcal{X} \to \mathbb{R}$, approximating the probability of a scenario being critical.

The mechanics of this search are illustrated in Fig. 3. The BO algorithm sequentially evaluates chosen scenarios $\mathbf{x}_n$ (shown by colored points) among the vast search space $\mathcal{X}$. The staircase lines represent the Pareto fronts (i.e. stresses of critical scenarios) at different steps n of the search. Observe how the frontier is progressively "pushed out" towards the upper-right, corresponding to finding scenarios with stronger and stronger violation combos.

A. Gaussian Process Surrogates for Grid Stresses

The statistical surrogate that we use to predict the likelihood of unevaluated scenarios being critical is a GP. GPs are analytically tractable to integrate with many acquisition functions, have good predictive performance and can naturally quantify the uncertainty of predictions. Roughly speaking, the GP interpolates the stresses of evaluated scenarios to predict the stress of unevaluated ones. A separate GP is built for each objective f_k. To this end, the GP employs a probabilistic framework, constructing a covariance kernel $\kappa_k(\cdot,\cdot)$ that represents dependence among scenarios and imposing a Gaussian distribution on the predicted stress of the un-evaluated scenarios. An important challenge is handling our representation of PV adoption scenarios as binary vectors $\mathbf{x}$.

Related ideas for BO over combinatorial input spaces include choosing an alternative surrogate that can handle high-dim./combinatorial inputs [39], [40], performing the search in a lower-dimensional and/or continuous embedding [41], [42] or modifying the GP kernel to encode the structure of the objective function with respect to the input space [43], [44].

Here, we adapt the method of Wan et al. [44] who propose a separable categorical GP kernel based on the Hamming distance between inputs, where relevant "active" dimensions are automatically given higher importance in predicting the output. The categorical kernel is a product of A univariate terms

$$\kappa_k(\mathbf{x}_1, \mathbf{x}_2) := \eta_k \exp\left(\frac{-1}{A}\sum_{j=1}^{A}\theta_{k,j}\mathbf{1}\left(x_{1,j} \neq x_{2,j}\right)\right). \quad (5)$$

Specifically, suppose we have evaluated n scenarios $\mathbf{X}_n = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ with corresponding stress $f_k(\mathbf{X}_n) = [f_k(\mathbf{x}_1), \ldots, f_k(\mathbf{x}_n)]$, $k = 1, \ldots, P+L$, summarized as $\mathcal{D}_n = \{(\mathbf{x}_i, \boldsymbol{f}(\mathbf{x}_i)), i = 1, \ldots, n\}$. Then, the stress of new, unevaluated scenarios $\mathbf{X}' = [\mathbf{x}'_1, \ldots, \mathbf{x}'_M]$ has the predictive normal distribution

$$f_k(\mathbf{X}')|\mathcal{D}_n \sim \mathcal{N}\left(\mu_k^{(n)}(\mathbf{X}'), \Sigma_k^{(n)}(\mathbf{X}')\right) \quad (6)$$

where letting $\mathbf{K}_k^{(n)} := K_k(\mathbf{X}_n, \mathbf{X}_n)$ with $\mathbf{K}_{k,i,j} := \kappa_k(\mathbf{x}_i, \mathbf{x}_j)$ be the GP design matrix, we have

$$\begin{aligned}\mu_k^{(n)}(\mathbf{X}') &= K_k(\mathbf{X}', \mathbf{X}_n)[\mathbf{K}_k^{(n)} + \sigma_k^2\mathbf{I}_n]^{-1} f_k(\mathbf{X}_n);\\ \Sigma_k^{(n)}(\mathbf{X}') &= K_k(\mathbf{X}', \mathbf{X}')\\ &\quad - K_k(\mathbf{X}', \mathbf{X}_n)[\mathbf{K}_k^{(n)} + \sigma_k^2\mathbf{I}_n]^{-1} K_k(\mathbf{X}_n, \mathbf{X}'). \quad (7)\end{aligned}$$

In (7) $K_k(\mathbf{X}_n, \mathbf{X}')$ is the *Gram* matrix containing pairwise correlations between considered scenarios, and $\mathbf{I}_n$ the $n \times n$ identity matrix. The observation variance σ_k^2 is a nugget term, which ensures numerical stability during the modeling process. The hyperparameters $\theta_{k,j}$ in (5) are inferred using maximum likelihood estimation based on $\mathcal{D}_n$, which allows the GP to learn the adopters j that most affect each objective f_k, a process referred to as Automatic Relevance Determination (ARD). The posterior mean $\mu_k^{(n)}(\mathbf{x}')$ represents the predicted stress of objective k at scenario $\mathbf{x}'$ and the posterior standard deviation $\Sigma_k^{(n)}(\mathbf{x}')$ is used for uncertainty quantification to measure the confidence in the above prediction, employed below to enforce exploration in the scenario space.

B. Acquisition Function

In order to guide the search for scenarios on the Pareto front, we construct an acquisition function $\alpha_{ND}(\cdot)$ (where "ND" stands for non-dominated) that proxies the probability of scenario $\mathbf{x}$ being non-dominated with respect to $V(\boldsymbol{f})$. Thus, the idea is to have $\alpha_{ND}(\mathbf{x}; \mathcal{D}_n) \approx \mathbb{P}(\mathbf{x} \in \mathcal{P}(\mathcal{X})|\mathcal{D}_n)$, so that maximizing α_{ND} corresponds to proposing scenarios that are most likely to improve the current Pareto front $\mathcal{P}(\mathbf{X}_n)$.

Because it is important to incorporate the covariance structure among unobserved scenarios and rank their potential violations consistently, we evaluate α_{ND} in parallel over M candidate scenarios $\mathbf{X}'_n = [\mathbf{x}'_{1,n}, \ldots, \mathbf{x}'_{M,n}]$. To do so, we first draw N samples of $\boldsymbol{f}(\mathbf{X}'_n)|\mathcal{D}_n$ using the latest GPs, denoted as $\widetilde{\boldsymbol{f}}_1(\mathbf{X}'_n), \ldots, \widetilde{\boldsymbol{f}}_N(\mathbf{X}'_n)$. These capture the distribution of potential violations forecasted for $\mathbf{X}'_n$. We then compute the subset of critical scenarios $\widetilde{\mathcal{P}}_i \subseteq \mathbf{X}'_n$ in each of these N samples.

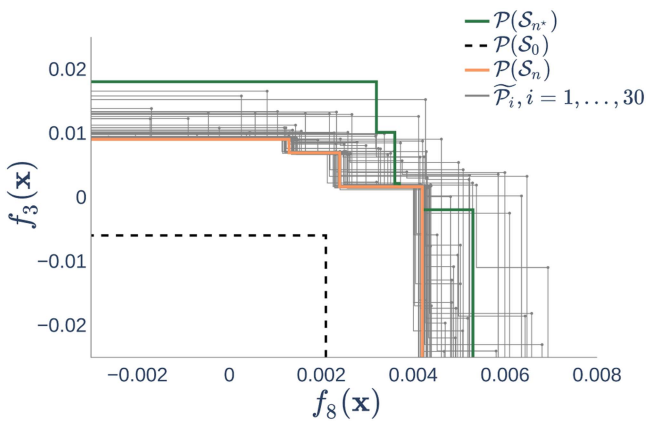

Fig. 4. Initial Pareto front $\mathcal{P}(\mathcal{S}_0)$ (dashed gray), Pareto front $\mathcal{P}(\mathcal{S}_n)$ at step n = 55 (orange) and true front $\mathcal{P}(\mathcal{S}_{n^\star})$ (green) across two representative objectives $f_3(\mathbf{x})$, $f_8(\mathbf{x})$. We also plot 30 simulated fronts $\widetilde{\mathcal{P}}_i$ (solid gray curves) obtained by sampling the GPs across $M = 500$ non-evaluated scenarios.

Then for a candidate $\mathbf{x}'_{m,n} \in \mathbf{X}'_n$, we set

$$\alpha_{ND}(\mathbf{x}'_{m,n}; \mathcal{D}_n) := \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}_{\widetilde{\mathcal{P}}_i}(\mathbf{x}'_{m,n}). \quad (8)$$

Thus, $\alpha_{ND}(\mathbf{x}'_{m,n})$ is the frequency that $\mathbf{x}'_{m,n}$ is critical among $\mathbf{X}'_n \cup \mathbf{X}_n$.

The conditional simulation of Pareto fronts closely follows the approach of [45], although here we map simulated samples to violations. This construction is visualized in Fig. 4 which shows 30 simulated Pareto fronts $\widetilde{\mathcal{P}}_i$ (gray "staircases") for two bus objectives f_1 and f_8. Then, $\alpha(\mathbf{x}'_{m,n}; \mathcal{D}_n)$ is proportional to the number of times the objectives sampled at $\mathbf{x}'_{m,n}$ lie on the staircase. Note that $\mathbf{x}'_{m,n}$ itself is not being plotted, so visually the same scenario would yield N different corners. Although a larger number of samples N in (8) better accounts for the covariance structure $\Sigma^{(n)}(\mathbf{X}'_n)$, leading to more accurate α_{ND} (and hence more accurate search) it comes at the cost of a higher computational burden. Similarly, jointly sampling from a GP at M non-evaluated scenarios $\mathbf{X}'_n$ to compute α_{ND} and τ_n becomes prohibitive as M grows.

C. Search Space and Stopping Criterion

To initialize the GPs and the search we begin by simulating and evaluating n_0 scenarios. Thereafter, in order to reflect the likelihood of scenarios under the agent-based diffusion model (1) we do not maximize the acquisition function over the entire $\mathcal{X}$ but over smaller search spaces $\mathcal{S}_n$ that grow in n. To this end, we first simulate N_{init} scenarios to construct $\mathcal{S}_1$ and then augment with N_{expand} additional new scenarios at each step, so that $|\mathcal{S}_n| = n_0 + N_{\text{init}} + (n-1) \cdot N_{\text{expand}}$.

Evaluation of $\alpha_{ND}(\cdot)$ on every scenario in the search space $\mathcal{S}_n$ becomes expensive when $|\mathcal{S}_n| \gg 10^3$. Therefore, we further restrict to a set of M candidate scenarios $\mathbf{X}'_n$, where M is set to the size of the initial search space $\mathcal{S}_1$. As $\mathcal{S}_n$ grows, we mitigate the risk of promising scenarios being overlooked by sampling them with probability inversely proportional to the frequency with which they have already has been sampled. Thus, scenarios

Algorithm 1: Critical Scenario Search With Bayesian Optimization.

1: Initialize $\mathcal{D}_0 = \{(\mathbf{X}_0, \boldsymbol{f}(\mathbf{X}_0))\}$ w/ n_0 evaluated scenarios
2: Set $n \leftarrow 0$ and $\tau_n \leftarrow \infty$
3: **while** $\tau_n > \bar{\tau}$ **do**
// Search until stopping criterion is satisfied, see (10)
4: $\mathcal{C}_n \leftarrow \{k : \exists \mathbf{x} \in \mathcal{D}_n, f_k(\mathbf{x}) > 0\}$ // Critical objectives
5: $\mathcal{E}_n \leftarrow \{k : \max_{\mathbf{x}\in\mathcal{S}_n} f_k(\mathbf{x}) > \bar{s}\}$
6: $\mathcal{O}_n \leftarrow \mathcal{C}_n \cup \mathcal{E}_n$ // Active objectives
7: Update/fit the surrogates $\mathcal{GP}(\mu_k^{(n)}, \sigma_k^{(n)})\ \forall k \in \mathcal{O}_n$ based on $\mathcal{D}_n$
8: $\mathbf{x}^{(n)} \leftarrow \arg\max_{\mathbf{x}\in\mathcal{S}_n} \alpha_{ND}(\mathbf{x}; \mathcal{D}_n)$ // see (8)
// Evaluate and augment to training dataset
9: Evaluate objectives $\boldsymbol{f}(\mathbf{x}^{(n)})$
10: $\mathcal{D}_n \leftarrow \mathcal{D}_{n-1} \cup \{(\mathbf{x}^{(n)}, \boldsymbol{f}(\mathbf{x}^{(n)})\}$
// New scenarios to test
11: $\mathcal{S}_n \leftarrow \mathcal{S}_n \cup \{N_{\text{expand}}$ fresh scenarios from $\mathcal{X}\}$
12: $n \leftarrow n + 1$
13: **end while**
14: Return violating scenarios $\{\mathbf{x} \in \mathbf{X}_{n^\star} : \exists k\ f_k(\mathbf{x}) > 0\}$, found critical objectives $\mathcal{C}_{n^\star}$ and critical scenarios $\mathcal{P}(\mathcal{S}_{n^\star})$

that have never appeared in $\{\mathbf{X}'_k, k \le n\}$ have higher probability of being sampled into $\mathbf{X}'_{n+1}$.

The acquisition function α_{ND} is further used as a stopping criterion to terminate the search, interpreted as a statistical guarantee on the set of critical scenarios discovered, which is of particular interest to utility planners. Indeed, the probability of having missed a critical scenario can be approximated as

$$\mathbb{P}(\exists\, \mathbf{x} \in \mathcal{P}(\mathcal{X}) \setminus \mathbf{X}_n) \lessapprox \sum_{\mathbf{x}\in\mathcal{S}_n\setminus\mathbf{X}_n} \mathbb{P}(\mathbf{x} \in \mathcal{P}(\mathcal{X})) \approx \textstyle\sum_{\mathbf{x}\in\mathcal{S}_n\setminus\mathbf{X}_n} \alpha_{ND}(\mathbf{x}; \mathcal{D}_n). \quad (9)$$

Therefore, the search is stopped when this upper bound has crossed a specified tolerance $\bar{\tau}$, set by the user. Since $\mathcal{S}_n$ grows at every step n, we compute (9) on a random subsample of M unevaluated scenarios $\mathbf{X}' \in \mathcal{S}_n \setminus \mathbf{X}'_n$

$$\tau_n := \textstyle\sum_{\mathbf{x}'\in\mathbf{X}'} \alpha_{ND}(\mathbf{x}'; \mathcal{D}_n). \quad (10)$$

Our methodology is summarized in Algorithm 1.

In order to speed up the search, we make a few adjustments to Algorithm 1. Firstly, there is no need to model objectives that have no violations as they do not affect the Pareto front, i.e. it is sufficient to focus on critical objectives only. Since the latter are not known until we have evaluated a scenario that leads them to be violated, we compute α_{ND} based on a set $\mathcal{O}_n = \mathcal{C}_n \cup \mathcal{E}_n$ (line 7) where $\mathcal{C}_n$ is the set of objectives that have already been violated, and $\mathcal{E}_n$ is the set of "stressed" objectives, i.e. those that are close to being violated. We take $\mathcal{E}_n = \{k : \max_{\mathbf{x}\in\mathcal{S}_n} f_k(\mathbf{x}) > \bar{s}\}$ where $\bar{s} < 0$ is a threshold decided by the user. The idea is to look for buses and lines that are operating close to their tolerance/capacity, and hence might actually be violated in yet-to-be evaluated scenarios. Thus, including such objectives in the acquisition function will guide the search towards scenarios that are likely to push the objective towards a violation, thereby "detecting" a new critical objective. Thus, when computing α_{ND}, the Pareto fronts $\widetilde{\mathcal{P}}_1, \ldots, \widetilde{\mathcal{P}}_N$ are taken with respect to objectives $V_k(\boldsymbol{f}_k)$, for $k \in \mathcal{O}_n$. Furthermore, we alternately let active objectives $\mathcal{O}_n$ be either bus or line objectives, but not both simultaneously. To be consistent with this choice, we track a stopping criterion for both buses and lines, denoted by τ_n^b and τ_n^l, and terminate the search when $\tau_n := \max(\tau_n^b, \tau_n^l) < \bar{\tau}$.

Secondly, fitting GP hyperparameters at every iteration via minimization of the log-likelihood (line 8) unnecessarily increases the computational burden without yielding significant improvements in model performance. GP hyperparameters tend to stabilize after a few iterations and minor changes to them have minimal impact on the BO search. Therefore, we refit the GP hyperparameters (while still updating the posterior means/standard deviations as new evaluations are added) only every N_0 steps. Finally, to reduce overhead, we select the top $B = 4$ scenarios (rather than just the top-1in line 8) that maximize α_{ND}, evaluating this batch in parallel.

IV. Numerical Results

We implement our approach in Python, using the GP implementation from the GPyTorch[1] and BoTorch[2] open source libraries [46]. We built our own acquisition functions and kernels on top of the modules available in BoTorch.

We proceed to a case study applying Algorithm 1 to identify critical grid stress scenarios on the real-life feeder, named *p4rhs8* with 12 bus groups shown in Fig. 1. Next we will analyze a larger feeder, *p3uhs27*, to show consistency of the methodology proposed. The networks are obtained from National Renewable Energy Laboratory's SMART-DS open dataset[3] with the same names as in the paper. PV capacities of adopting agents are computed using the PySAM[4] library. Bus voltage and line flows are computed using an optimal power flow solver with the pandapower[5] open source library for Python.

A. Algorithm Performance

The settings and main results are summarized in Table I. All scenarios are generated from the agent diffusion model (1) with $p = 0.01$ and $q = 0.164$. We first simulate $n_0 = 75$ scenarios to evaluate power flows and initialize the GPs. Algorithm 1 then iteratively expands the scenario space, simulating $N_{\text{expand}} = 200$ scenarios per iteration, except at the first step where we simulate an extra 2800 scenarios (thus $\mathcal{S}_1 = 3000$). The stopping criterion (10) with $\bar{\tau} = 0.1$ is used to automatically terminate the search. The shown run terminates after evaluating a total of 333 (75 initial + 258 selected) scenarios.

Fig. 3 shows the progression of the Pareto frontier projected on two of the 5 critical bus objectives. For better visibility, we

[1] https://gpytorch.ai
[2] https://botorch.org
[3] https://data.openei.org/submissions/2981
[4] https://sam.nrel.gov/software-development-kit-sdk/pysam.html
[5] https://www.pandapower.org

TABLE I
SUMMARY OF SETTINGS AND SEARCH RESULTS FOR FEEDERS P4RHS8 AND P3UHS27. THE NUMBER OF VIOLATING SCENARIOS DISCOVERED IS THE SIZE OF THE SET $\{\mathbf{x} \in \mathbf{X}_{n^\star} : f_k(\mathbf{x}) > 0$ FOR SOME $k\}$.

	p4rhs8		p3uhs27	
Feeder Characteristics				
Number of buses	202		439	
Number of lines	202 (29 miles)		339 (13 miles)	
Peak load demand (MW)	0.26		1.08	
Energy Consumption (GWh)	6.37		15.5	
Number of agents (A)	159		329	
Voltage tolerance (v_l^b, v_b^b)	(0.95, 1.05)		(0.965, 1.035)	
Scenario / Algorithm Settings				
Initial search space $\lvert\mathcal{S}_1\rvert$	3000		3000	
Initial sample size n_0	75		100	
Simulations per step N_{expand}	200		300	
Simulated scenarios $\lvert\mathcal{S}_{n^\star}\rvert$	9200		21800	
Evaluated scenarios $\lvert\mathbf{X}_{n^\star}\rvert$	333		452	
Stopping threshold $\bar{\tau}$	0.1		0.2	
Evaluation batch size B	4		5	
Results				
	Buses	Lines	Buses	Lines
Objectives	12	202	9	339
Violating scenarios found	213	255	337	281
Percentage found	82.56%	98.84%	95.74%	79.83%
Crit. objectives $\lvert\mathcal{C}^*_{n^\star}\rvert$	5	10	4	6
Crit. objectives found $\lvert\mathcal{C}_{n^\star}\rvert$	5	10	4	6
Crit. scenarios $\lvert\mathcal{P}(\mathcal{S}_{n^\star})\rvert$	15	32	36	20
Crit. scenarios found $\lvert\mathcal{P}(\mathcal{S}_{n^\star}) \cap X_{n^\star}\rvert$	13	31	34	19

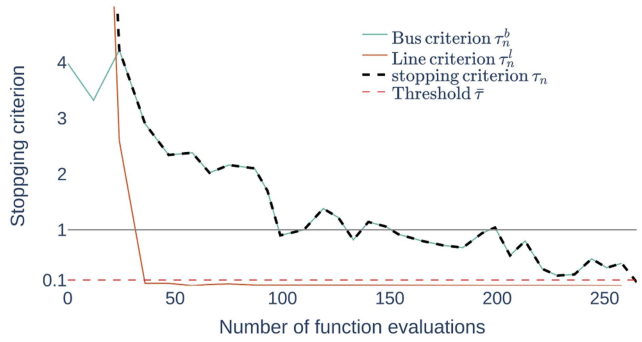

Fig. 5. Evolution of statistical stopping criteria τ_n^b for buses and τ_n^l for lines as a function of number of evaluations n. The overall stopping criterion $\tau_n = \max(\tau_n^b, \tau_n^l)$ is triggered at the threshold $\bar{\tau} = 0.1$.

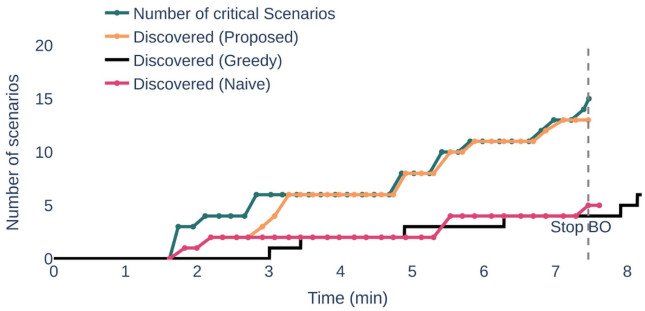

Fig. 6. Number of bus-critical scenarios in the search space $\mathcal{S}_n$(green line) and how many of those scenarios are identified by Algorithm 1 (orange line) and a greedy comparator (in black) as a function of time. The number of critical scenarios found by a naïve comparator that evaluates the $N_{top} = 25$ scenarios by total PV is the magenta curve. When Algorithm 1 stops after 7.5 minutes, the greedy approach has only identified 4 critical scenarios.

crop part of the search space where the objectives are negative. The ultimate goal of our methodology is to recover the critical scenarios lying on the outermost staircase line. Critical scenarios are discovered in the later stages of the BO, suggesting that it does not stop prematurely. Consistently with this fact, a few later-stage evaluations are relatively far from the frontier, probably because the algorithm focuses on the corners of the Pareto front for other critical objectives as well (where the two objectives shown might not be simultaneously violated). The picture looks similar for other objectives, with all but one scenario on the Pareto frontier being discovered.

In all, we find 13 out of 15 bus-critical and 31 out of 32 line-critical scenarios among a vast set of 9200 simulated scenarios, with only 333 evaluations. The algorithm also discovers all critical objectives (5 for the buses and 10 for the lines). To get a sense of how targeted the search is, among the 333 scenarios that the BO evaluates, the vast majority cause violations during PF: there is some bus violation in over 83% and a line violation in over 98% of evaluated scenarios (Table I). Fig. 5 shows the stopping criterion τ_n against n. The stopping criterion bounds from above the likelihood of missing a critical scenario, effectively stating that discovered critical scenarios account for no less than $(1 - \bar{\tau}) \times 100\%$ of all simulated critical scenarios. One can set $\bar{\tau}$ higher to quicken the search at the cost of lower confidence in omitting critical scenarios.

B. Comparators

One baseline competitor is a brute force approach that sequentially evaluates every simulated scenario to identify the critical ones. In the above case study, this would require 27.6× more evaluations (9200/333) and would take 7.65× longer runtime after accounting for the overhead of our algorithm including the computation of the acquisition function and fitting GPs on all the bus and line objectives. Furthermore, the brute force approach lacks statistical analysis of criticality.

Fig. 6 illustrates how our methodology picks up bus-critical scenarios. The green curve counts the total number of critical scenarios in the search space $\mathcal{S}_n$, which increases in n. The orange curve shows the size of the subset of these that have been selected and evaluated by the BO, i.e. scenarios that are correctly identified as critical. Our algorithm is able to quickly identify critical scenarios shortly after they are added to the search space, cf. the fact that the orange curve closely tracks the green one. In contrast, the black staircase representing the number of critical scenarios found by the brute force approach lags significantly behind, unable to keep up with the growing size of $\mathcal{S}_n$, leading to a widening gap between the number of critical scenarios selected by our algorithm and those discovered through brute force.

A second comparator approach is to rank simulated scenarios by total PV adoption and compare the critical scenarios obtained by the proposed search algorithm with the existing utility practice, which consists of selecting the extreme cases of aggregated capacity ("high" and "low" load or PV scenarios) as worst-cases to explore. This naïve approach evaluates the N_{top} scenarios that have the highest adoption capacity. Fig. 6 shows that this naïve approach (with $N_{top} = 25$) misses the majority of critical scenarios that our algorithm finds and barely improves upon

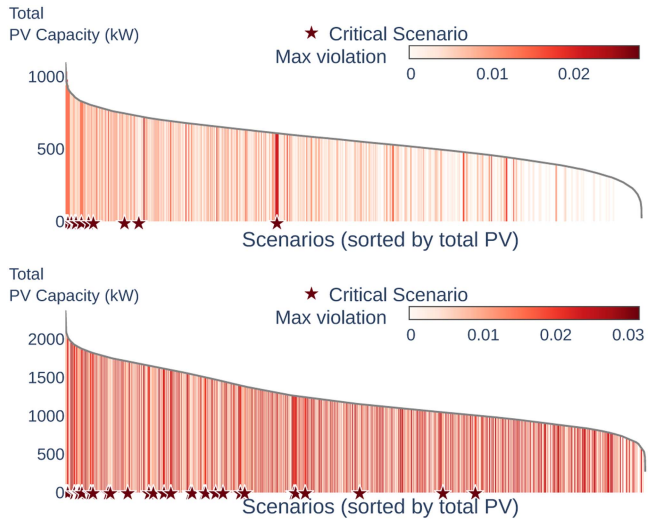

Fig. 7. Bus violations within the simulated PV adoption scenarios for feeders p4rhs8 (top, 9200 scenarios) and p3uhs27 (bottom, 21800 scenarios), sorted left-to-right by their total installed PV capacity. Each vertical bar corresponds to a scenario, with color indicating the maximum bus voltage violation observed under that scenario. Adoption scenarios identified as critical by Algorithm 1 (i.e., belonging to the Pareto set of violations) are marked with a star.

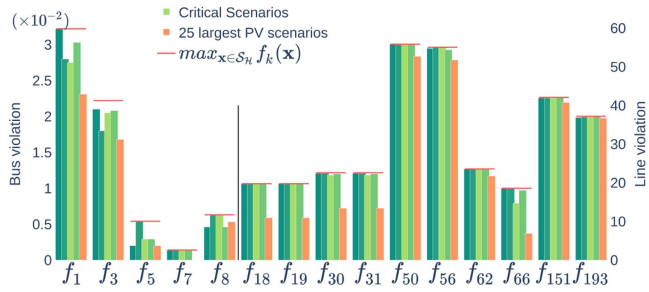

Fig. 8. Critical objective values found across 4 runs (green bars) of Algorithm 1 and a fixed database $\mathcal{S}_H$ with $H = 25000$ scenarios. Orange bars show the maximum violation attained in the top $N_{top} = 25$ scenarios by total PV. The horizontal lines show the maximum violation in this database. The critical bus objectives are f_k (left y-axis) $k \in \{1, 3, 5, 7, 8\}$, and the critical line objectives (right y-axis) are $k \in \{18, 19, 30, 31, 50, 56, 66, 151, 193\}$.

greedily evaluating every simulated scenario. The top panel of Fig. 7 further shows the distribution of the 9200 simulated scenarios sorted by their aggregated adopted PV, and colored by the maximum violation over the buses. Notably, most scenarios with the largest PV adoption are not critical, while other scenarios with relatively low PV capacity are critical (indicated with a star) due to specific locational adoption patterns. For example, high PV penetration throughout a feeder can lead to overvoltages at nodes distant from the substation while alleviating overloading in lines closer to it. However, a moderate PV penetration at the end of the feeder can still cause the same overvoltages but may not be sufficient to alleviate overloading, resulting in a more challenging planning scenario. This highlights the importance of systematic locational search algorithms and explains how, with our method, we successfully find critical scenarios that the greedy and naive approaches miss.

Fig. 8 visualizes the critical violations found versus the ground truth (orange bars) and versus the violations within the top-25 scenarios ranked by total PV capacity. We see that critical scenarios of PV adoption lead to more serious violations in nearly every critical objective than the largest PV scenarios. For the bus objective f_7, simply looking at the largest PV scenarios would entirely miss the violation.

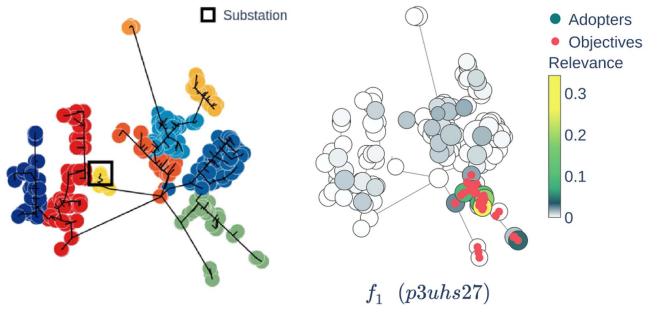

Fig. 9. *Left:* Feeder p3uhs27 and bus groupings obtained using the Louvain community algorithm. *Right:* Spatial map of adopter relevance $\tilde{\theta}_{1,j}$ for a representative bus objective f_1. Adopters j marked with an orange dot are buses that are part of objective f_1. Larger values (brighter colors) indicate stronger influence of the adopter on the respective objective. Circle size is proportional to the adopter's PV system capacity.

C. Second Feeder

To validate the performance of our algorithm, we consider a larger 439-bus network (feeder p3uhs27) with 9 bus groups and 300+ adopters (shown in left panel of Fig. 9), see the right column of Table I. For this feeder, a stopping threshold of $\bar{\tau} = 0.2$ is set and we use batch evaluations with $B = 5$ scenarios. Results are similar to the first feeder p4rhs8. The search strongly targets scenarios leading to violations (over 95% of evaluated scenarios result in bus voltage violations, and over 79% in line flow ones). We recover 34 out of 36 bus-critical and 19 out of 20 line-critical scenarios. This shows the robustness of our methodology. The lower panel of Fig. 7 shows the distribution of the total adopted PV of the 21800 simulated scenarios and color-coded by the maximum bus violation. The Figure highlights that violations are common (even for scenarios with lower total adopted PV capacity), but only a few scenarios are critical, and those are not necessarily the ones with the largest PV. Even if we evaluated the top 452 scenarios by total PV, we would miss a majority of the critical scenarios.

D. Algorithm Stability

To evaluate the stability of our search, we first fix a large database $\mathcal{S}_H$ of $H = 25000$ scenarios and run Algorithm 1 four times on feeder p4rhs8 where for each run $i = 1, \ldots, 4$ the search space $\mathcal{S}_n^{(i)}$ is constructed by randomly sampling without replacement from $\mathcal{S}_H$, with a different seed for every run, while keeping algorithm parameters fixed. For each run, we record in Table II the number of function evaluations $n^\star$, the search space size at termination ($\mathcal{S}_n^{(i)}$), and the number of found bus- and line-critical scenarios. The maximum violation achieved by evaluated scenarios (i.e. the marginal Pareto front) for each of the 5+10 critical objectives is shown in Fig. 8.

TABLE II
STABILITY ANALYSIS OF ALGORITHM 1. EACH ROW IS A DIFFERENT RUN WITH THE SEARCH SPACE CONSTRUCTED BY INDEPENDENT SAMPLING FROM A FIXED DATABASE OF $H = 25000$ SCENARIOS.

Run	Simulated	Evaluated	Critical Bus scens	Critical Line scens
1	10600	392	16	33
2	14950	524	15	31
3	11500	398	11	29
4	13000	440	15	29

TABLE III
SENSITIVITY ANALYSIS OF ALGORITHM 1. EACH ROW CORRESPONDS TO THE AVERAGE OF 10 RUNS ON A FIXED DATABASE $\mathcal{S}_H$ OF $H = 25000$ SCENARIOS, WITH $\pm$ INDICATING THE RESPECTIVE STANDARD DEVIATION ACROSS RUNS. BASELINE PARAMETERS CORRESPONDING TO THE RESULTS IN TABLE I ARE IN BOLD.

	Simulated $\lvert\mathcal{S}_{n^\star}\rvert$	Evaluated $\lvert\mathbf{X}_{n^\star}\rvert$	Critical Bus scens	Critical Line scens
$N = 10$	9540 ± 3590	362 ± 138	14.3 ± 4.0	23.3 ± 3.1
$N = 25$	9000 ± 219	338 ± 69	15.2 ± 3.8	22.9 ± 2.4
$\mathbf{N = 50}$	9480 ± 1450	368 ± 52	14.4 ± 1.7	22.4 ± 2.3
$N = 100$	8740 ± 148	330 ± 54	14.0 ± 1.9	21.7 ± 2.4
$M = 500$	5763 ± 898	216 ± 37	14.8 ± 2.7	22.4 ± 4.2
$M = 1000$	7900 ± 1310	312 ± 50	12.6 ± 2.8	23.3 ± 2.1
$M = 2000$	8727 ± 1150	342 ± 48	15.0 ± 3.4	23.4 ± 2.8
$\mathbf{M = 3000}$	9480 ± 1450	368 ± 52	14.4 ± 1.7	22.4 ± 2.3
$M = 4000$	9620 ± 2200	361 ± 72	14.4 ± 3.5	23.2 ± 2.8
$N_e = 50$	3695 ± 247	210 ± 38	11.0 ± 3.7	20.2 ± 3.4
$N_e = 100$	5845 ± 540	322 ± 39	12.6 ± 3.0	22.4 ± 3.6
$\mathbf{N_e = 200}$	9480 ± 1450	368 ± 52	14.4 ± 1.7	22.4 ± 2.3
$N_e = 300$	12570 ± 2360	369 ± 68	17.2 ± 3.9	23.6 ± 2.7
$N_e = 400$	16200 ± 4590	378 ± 93	17.7 ± 3.4	24.6 ± 4.0
$\bar{\tau} = 0.05$	10360 ± 2040	393 ± 68	17.0 ± 2.5	23.6 ± 1.8
$\bar{\tau} = \mathbf{0.1}$	9480 ± 1450	368 ± 52	14.4 ± 1.7	22.4 ± 2.3
$\bar{\tau} = 0.2$	8491 ± 950	328 ± 38	15.2 ± 2.1	23.6 ± 3.2
$\bar{\tau} = 0.5$	6818 ± 1387	265 ± 55	13.3 ± 2.3	22.4 ± 3.6

The whole database has $|\mathcal{P}(\mathcal{S}_H)| = 26$ bus-critical and 54 line-critical scenarios. While not all critical scenarios are recovered since Algorithm 1 only gets exposed to less than half of the entire search space when it decides to stop, the overall number of critical scenarios is comparable and the number of function evaluations before stopping is stable across the runs. Furthermore, and most importantly, in each run, the magnitude of the identified max violations is almost identical, showing that while we do not identify exactly the same Pareto set, the overall Pareto front is nearly the same, and we find scenarios very close to the maximum violation in every run. This is a key take-away since, while we may not find the exact layout of PV adoption defined to be critical, the scenarios found lead to the same issues in grid equipment, making our methodology a reliable tool for upgrade planning.

Next, we perform a sensitivity analysis for four key parameters in Algorithm 1: the number of candidate scenarios M at each BO step, the number of samples N to compute α_{ND} in (8), the stopping threshold $\bar{\tau}$, and the number of new scenarios simulated at each BO step N_{expand}. To control for the intrinsic randomness due to scenario simulation, all the reported values are averages over 10 runs of Algorithm 1; we moreover list the respective standard deviations. Table III reports the total number of simulated scenarios, the number of evaluations (equivalent to computational time), and the number of found critical scenarios for the bus and line objectives. we report the average number of simulated, evaluated, bus-critical and line-critical scenarios. The discussion below is relative to the baseline parameters from Table I.

While the number of scenarios simulated by our algorithm is by design sensitive to above tuning parameters, robustness is reflected in the stable number of *evaluated* scenarios, as well as a consistent number of found critical scenarios. Performance is little impacted by N, although setting it too low leads to a larger variance in all key metrics, i.e., a less stable performance, as the acquisition function becomes too noisy. The number of candidates in the stopping criterion, M, is also benign, though it must be sufficiently large to prevent the algorithm from terminating prematurely due to a lack of promising scenarios considered when computing the stopping criterion. The algorithm is most sensitive to $N_e = N_{\text{expand}}$ that controls the growth of the search space. Setting N_{expand} too low causes the algorithm to exhaust the search space and stop prematurely. Taking it too large, causes the algorithm to be overwhelmed by the number of potential candidates with the result that the stopping criterion decreases very slowly. Finally, the threshold $\bar{\tau}$ directly controls the stopping rule, and a large τ causes early termination, missing some critical scenarios.

In terms of violation magnitudes, we ran a statistical test that confirmed that the distribution of maximum found violations is statistically the same between the baseline and the variants in Table III, so the algorithm consistently finds the worst violations of each objective. Overall, the stable number of evaluations ($|\mathbf{X}_{n^\star}| \in [300, 370]$) and of the critical scenarios ($14 - 16$ for buses $22 - 24$ for lines) discovered, indicates the robustness of our approach; the exception being the rows with $N = 10$ (very large variance across runs) and $M = 500$, $N_{\text{expand}} = 50$ and $\bar{\tau} = 0.5$ (insufficient number of evaluations leading to some critical scenarios missed).

Another algorithm parameter that ought to be mentioned is the number of bus groups P that determines the number of bus objectives. This parameter represents the number of voltage control areas in the network, which are often a small number and very specific of each feeder. Similarly, the violation threshold $\bar{s}$ (Line 5 of Algorithm 1) is another user-defined parameter that is network-dependent. One way to pick $\bar{s}$ is to look at the distribution of objective values on the set of initial evaluated scenarios.

V. DISCUSSION

A. Surrogate Performance

It is instructive to illustrate the evolution of the GP surrogates $\widehat{f}_k(\cdot)$ over the BO iterations $n = n_0, \ldots$. Because the search space $\mathcal{X}$ is high-dimensional, we cannot directly visualize $\hat{f}_k^{(n)}(\cdot)$; instead we show in Fig. 10 a scatterplot of the predicted mean (and standard deviations) $\mu_k^{(n)}(\cdot)$ objective values against the true values $f_k(\cdot)$ at the initial step $n = n_0$ and at the last step $n = n^\star$ when the algorithm terminated. The predictions

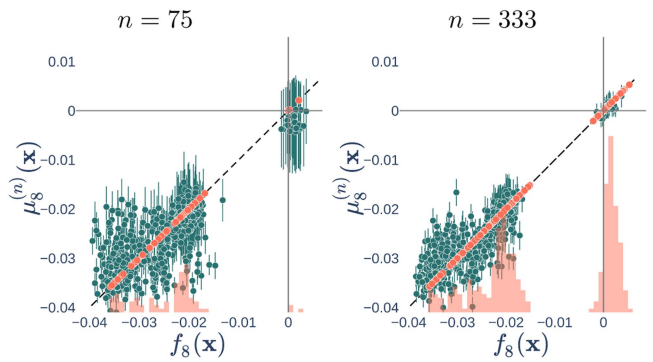

Fig. 10. True stress objective (x-axis) vs. GP prediction, i.e. posterior mean (y-axis) for the critical bus objective $f_8(\mathbf{x})$, after evaluating $n = 75$ (*left*) and $n = 333$ scenarios (*right*) respectively. Vertical error bars indicate ± 2 standard deviations around the mean. Orange points are in-sample scenarios and green points are out-of-sample. The rug plots indicate the distribution of the training outputs $\{f(\mathbf{x}), \mathbf{x} \in \mathcal{D}_n\}$.

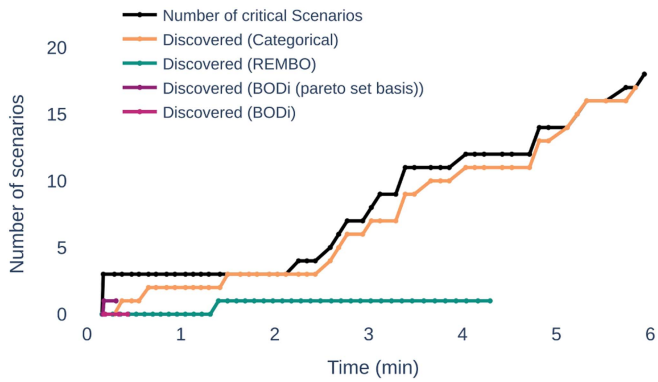

Fig. 11. Number of bus-critical scenarios in the search space $\mathcal{S}_n$ (black curve) and how many of those scenarios are identified by Algorithm 1 using four different GP kernels, cf. Table IV. Each run is on the same search space but has its own stopping criterion, leading to different termination times.

are shown for the already evaluated scenarios $\mathcal{D}_n$, as well as a (subset) of the current search set $\mathcal{S}_n$. The uncertainty of the predictions is indicated by the vertical bars that span $\mu_k^{(n)}(\mathbf{x}) \pm 2\sigma_k^{(n)}(\mathbf{x})$. Note that for $\mathbf{x} \in \mathcal{D}_n$, there is almost no uncertainty due to the small σ_k^2.

Accurate predictions would make $\mu_k^{(n)}(\mathbf{x}) = f_k(\mathbf{x})$ lie on the diagonal. As n increases, two effects occur. First, the algorithm focuses on scenarios that are close to critical, i.e. with $f_k(\mathbf{x}) > 0$. Therefore, it shifts learning to the "right", cf. the rug plots in the left and right panels. Second, for $n = 333$ the GPs have more training data and therefore yield more accurate (smaller prediction errors) and more precise (smaller $\sigma_k^{(n)}(\mathbf{x})$) fits. In conjunction, these two effects shift the training observations and asymmetrically improve accuracy, as witnessed in the right panel of Fig. 10. The much improved quality of $\hat{f}_k(\cdot)$ in the top-right quadrant at $n = 333$ compared to initial fit at $n = 75$ shows that the surrogates are adept at learning individual f_k's in their critical regions as the BO progresses.

B. Kernel Selection

The performance of the GP surrogate in Algorithm 1 hinges on the choice of its kernel. Standard kernels, such as the Matèrn or radial basis function (RBF), assume continuous inputs and are ill-suited for binary vectors modeling our adoption scenarios. The Hamming distance we employ is the natural choice to measure distance between scenarios. An alternative is to employ embedding techniques to project the high-dimensional discrete $\mathbf{x}$ into a lower-dimensional continuous space, where an RBF or Matèrn kernel can then be applied. Embedding can be done via random projection, for example the REMBO algorithm [47]; via dictionary embedding based on Hamming distance to an exogenously specified basis set, or via variational autoencoders.

Embedding is conceptually promising for our problem since only a subset of adopters is expected to influence each bus or line objective. Thus, the effective dimensionality of our problem is likely much lower than the input dimension. Motivated by this, we compare our categorical kernel to two prominent embedding approaches. First, we implemented REMBO [47] and vary the embedding dimension and the resampling frequency of the random projection matrix, and test both RBF and Matèrn kernels. Second, we applied the dictionary embedding method (BODi) [48], evaluating both the binary-wavelet dictionary and a variant of our own, where at each step the scenarios in the Pareto set are used as basis vectors. For compactness, we report only the best-performing runs. Importantly, we emphasize that our use of these kernels is limited to surrogate modeling; we still evaluate our acquisition function on a subset of the search space.

TABLE IV
PREDICTIVE ACCURACY OF THE FIVE CRITICAL BUS OBJECTIVES FOR DIFFERENT CHOICES OF GP KERNELS. WE REPORT PREDICTIVE RMSE ($\times 10^{-2}$) OVER 500 UNEVALUATED SCENARIOS RANDOMLY SELECTED FROM THE SEARCH SPACE AFTER EVALUATING $n = 333$ SCENARIOS SELECTED BY ALGORITHM 1.

	REMBO	BODi (wavelet)	BODi (Pareto set)	Cat. (Eqn (5))
f_1	1.62	1.12	1.17	0.43
f_3	1.63	0.99	0.95	0.41
f_5	0.94	0.6	0.58	0.36
f_7	0.8	0.69	0.65	0.15
f_8	1.57	1.01	1.12	0.33

Fig. 11 reports the number of critical scenarios identified when running Algorithm 1 with different kernels. REMBO was manually terminated after 500 evaluations because its stopping criterion remained far from $\bar{\tau}$ and showed no indication of halting. BODi failed to recognize any scenarios as potentially critical and stopped prematurely after fewer than 50 evaluations, despite requiring a larger number of initial training points. For both REMBO and BODi, we fixed the embedding dimension at 20 (compared to $A = 159$ for the categorical kernel), which substantially reduced GP fitting costs and which explains the shorter runtime for REMBO, despite having more function evaluations. The poor performance of REMBO and BODi stems from their apparent inability to capture the dependence of bus and line voltages on adoption scenarios. Table IV summarizes out-of-sample prediction errors across 500 scenarios, comparing our categorical kernel (5) the embedding-based alternatives. For

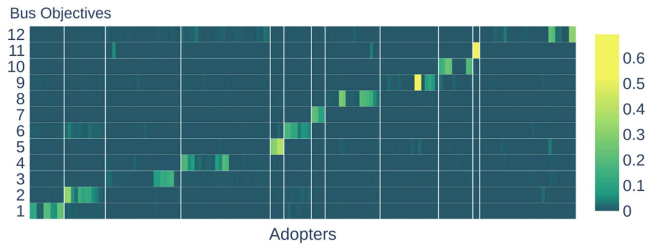

Fig. 12. Heatmap of adopter relevance $\tilde{\theta}_{k,j}$ (brighter is higher) in feeder p4rhs8. Rows represent bus objectives f_k, $k = 1, \ldots, 12$, columns are the adopters $j = 1, \ldots, 159$. Vertical white lines separate adopters according to the bus groups they belong to.

every bus objective, the κ_k in (5) achieves consistently lower RMSE and outperforms both REMBO and BODi. Similar results hold for line objectives, underscoring the advantage of directly modeling the binary structure of the input space.

C. Most Relevant Adopters

Our choice of the categorical GP kernel furnishes a tool to interpret adopter relevance, i.e. to get information about which specific adopter buses are driving different bus and line stresses. Specifically, leveraging the separable nature of (5) the fitted relevance parameters $\boldsymbol{\theta}_k$ indicate how much changes in the adoption pattern $\mathbf{x}$ affect the GP predictions. High value for $\theta_{k,j}$ indicates that adoption by agent j significantly impacts the stress in objective k. In that way, fitting the GP kernels as part of the BO provides the association between adopters and future grid stresses.

We plot the relevance of each adopter $\tilde{\theta}_{k,j} := 1 - e^{-\frac{\theta_{k,j}}{A}}$ in Fig. 12 for feeder p4rhs8, with the bus objectives on the y-axis and the adopters on the x-axis, sorted by the partitions $\mathcal{B}_p$ to which they belong. The ARD feature of our GPs endogenously identifies the relevant adopters for each objective, shown by the brighter colors. The sparsity of the heatmap implies that for each bus objective, only a few adopters are relevant and there is a low-dimensional subspace of $\mathcal{X}$ with "active" dimensions. This sparsity helps the GPs to focus on these dimensions to model the stress values. The right panel of Fig. 9 visualizes $\tilde{\theta}_{1,j}$ for objective f_1 (based on buses indicated by the orange points) for feeder p3uhs27 spatially overlaid on the grid network. We find that the resulting most-relevant "active" adopters are consistent with grid topology, corresponding to buses that are in the cluster $\mathcal{B}_1$ defining f_1, or to nearby adopters with large PV capacities.

The identification of these relevant adopters provides utilities with a group of consumers that are likely to trigger grid investments through their adoption decisions. This information can be used to schedule grid upgrades in response to PV interconnection requests from this group. It can also offer useful insights on where to incentivize load growth or procure local flexibility services to defer such upgrades.

VI. Conclusions and Future Work

In this work, we propose a new methodology to efficiently search through thousands of scenarios of PV adoption to identify those that lead to worst violations of bus voltages and line flows. Our Bayesian Optimization-driven search accurately navigates the high-dimensional scenario space, dramatically improving ($27 - 52\times$ fewer evaluations) over a brute force approach that evaluates all simulated scenarios. We also demonstrate that the search is highly nontrivial and that naive heuristics like sorting by aggregate adopted PV capacity will miss most critical scenarios.

Several aspects of our methodology should be investigated further. Further analysis of the best GP surrogates could offer faster initialization and re-fitting of the f_k's; for example one could try to leverage the latent lower-dimensional structure hinted in Section V-C. It might be worthwhile to incorporate correlation among GP surrogates to fuse learning of related stress objectives. Another idea is to fit GPs directly for individual bus voltages $volt_b(\mathbf{x})$ in (2), rather than on groups of buses $\mathcal{B}_k$, and then group predictions. Some preliminary experiments showed promising gains in accuracy, although at a higher computational cost. Alternative acquisition functions might be warranted for very large feeders (over 1000 nodes) where α_{ND} becomes slow.

The present work focused on distribution planning for radial grids. Further analysis is warranted for larger and/or meshed grids where more lines are likely to be critical. Finally, one could extend the methodology to consider multi-horizon planning where DER adoption evolves over multiple time steps and criticality is defined both locationally (what is violated) and temporally (when).

References

[1] Edison Electric Institute, "Industry capital expenditures," Tech. Rep., Sep. 2025. [Online]. Available: https://web.archive.org/web/20210125021710/https://www.eei.org/issuesandpolicy/Finance\%20and\%20Tax/EEI_Industry_Capex_Functional_2019.10.16.pdf

[2] Pacific Gas and Electric Company, "Grid needs assessment report"Tech. Rep., 2023. [Online]. Available: https://www.pge.com/content/dam/pge/docs/about/doing-business-with-pge/GNA.pdf

[3] Southern California Edison, "Distribution system plan," Tech. Rep., 2021. [Online]. Available: https://efiling.energy.ca.gov/GetDocument.aspx?tn=237453

[4] M. Olearczyk et al., "Load forecasting for modern distribution systems," Electric Power Res. Inst.,Tech. Rep. 1024377, 2013. [Online]. Available: https://restservice.epri.com/publicdownload/000000000001024377/0/Product

[5] F. Heymann, F. vom Scheidt, F. J. Soares, P. Duenas, and V. Miranda, "Forecasting energy technology diffusion in space and time: Model design, parameter choice and calibration," *IEEE Trans. Sustain. Energy*, vol. 12, no. 2, pp. 802–809, Apr. 2021.

[6] Hawaiian Electric, "Distribution DER hosting capacity grid needs report," Tech. Rep., 2021. [Online]. Available: https://www.hawaiianelectric.com/documents/clean_energy_hawaii/integrated_grid_planning/20211108_distribution_der_hosting_capacity_grid_needs_report.pdf

[7] PacifiCorp, "Distribution system planning," Tech. Rep., 2021. [Online]. Available: https://www.pacificorp.com/energy/oregon-distribution-system-planning.html

[8] P. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1166–1172, May 2005.

[9] V. A. Evangelopoulos and P. S. Georgilakis, "Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks," *Electric Power Syst. Res.*, vol. 207, 2022, Art. no. 107847. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779622000773

[10] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1142–1153, May 2010.

[11] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with Honurly information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 456–462, Jan. 2014.
[12] T. Zhang et al., "Long-term energy and peak power demand forecasting based on sequential-XGBoost," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 3088–3104, Mar. 2024.
[13] M. Dong and L. Grumbach, "A hybrid distribution feeder long-term load forecasting method based on sequence prediction," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 470–482, Jan. 2020.
[14] J. Li, S. Wei, Y. Lei, and Y. Luo, "Long-term electricity consumption forecasting for future power systems combining system dynamics and impact equation," *IEEE Trans. Ind. Appl.*, vol. 58, no. 5, pp. 5955–5965, Sep./Oct. 2022.
[15] O. Rubasinghe et al., "A novel sequence to sequence data modelling based CNN-LSTM algorithm for three years ahead monthly peak load forecasting," *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 1932–1947, Jan. 2024.
[16] R. Bernards, J. Morren, and H. Slootweg, "Development and implementation of statistical models for estimating diversified adoption of energy transition technologies," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1540–1554, Oct. 2018.
[17] S. A. Robinson and V. Rai, "Determinants of spatio-temporal patterns of energy technology adoption: An agent-based modeling approach," *Appl. Energy*, vol. 151, pp. 273–284, 2015.
[18] H.-C. Wu and C.-N. Lu, "A data mining approach for spatial modeling in small area load forecast," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 516–521, May 2002.
[19] E. M. Carreno, R. M. Rocha, and A. Padilha-Feltrin, "A cellular automaton approach to spatial electric load forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 532–540, May 2011.
[20] C. Ye, Y. Ding, P. Wang, and Z. Lin, "A data-driven bottom-up approach for spatial and temporal electric load forecasting," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1966–1979, May 2019.
[21] F. Heymann, V. Miranda, F. J. Soares, P. Duenas, I. Perez Arriaga, and R. Prata, "Orchestrating incentive designs to reduce adverse system-level effects of large-scale EV/PV adoption–the case of Portugal," *Appl. Energy*, vol. 256, 2019, Art. no. 113931.
[22] L. Botman et al., "A scalable ensemble approach to forecast the electricity consumption of households," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 757–768, Jan. 2023.
[23] A. Rodriguez-Calvo, R. Cossent, and P. Frías, "Integration of PV and EVs in unbalanced residential LV networks and implications for the smart grid and advanced metering infrastructure deployment," *Int. J. Elect. Power Energy Syst.*, vol. 91, pp. 121–134, 2017.
[24] M. Heleno, D. Sehloff, A. Coelho, and A. Valenzuela, "Probabilistic impact of electricity tariffs on distribution grids considering adoption of solar and storage technologies," *Appl. Energy*, vol. 279, 2020, Art. no. 115826.
[25] L. Sun and D. Lubkeman, "Agent-based modeling of feeder-level electric vehicle diffusion for distribution planning," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 751–760, Jan. 2021.
[26] Portland General Electric, "Distribution system plan," Tech. Rep., 2021. [Online]. Available: https://portlandgeneral.com/about/who-we-are/resource-planning/distribution-system-planning
[27] Y. Li and B. Jones, "The use of extreme value theory for forecasting long-term substation maximum electricity demand," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 128–139, Jan. 2020.
[28] M. Dong, "A data-driven long-term dynamic rating estimating method for power transformers," *IEEE Trans. Power Del.*, vol. 36, no. 2, pp. 686–697, Apr. 2021.
[29] J. M. Sexauer, K. D. McBee, and K. A. Bloch, "Applications of probability model to analyze the effects of electric vehicle chargers on distribution transformers," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 847–854, May 2013.
[30] P. Pareek and H. D. Nguyen, "Gaussian process learning-based probabilistic optimal power flow," *IEEE Trans. Power Syst.*, vol. 36, no. 1, pp. 541–544, Jan. 2021.
[31] K. Ye, J. Zhao, N. Duan, and Y. Zhang, "Physics-informed sparse Gaussian process for probabilistic stability analysis of large-scale power system with dynamic PVs and loads," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2868–2879, May 2023.
[32] K. Ye, J. Zhao, H. Li, and M. Gu, "A high computationally efficient parallel partial gaussian process for large-scale power system probabilistic transient stability assessment," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 4650–4660, Mar. 2024.
[33] M. Ludkovski, G. Swindle, and E. Grannan, "Large scale probabilistic simulation of renewables production," 2022, *arXiv:2205.04736*.
[34] Q. Li and M. Ludkovski, "Probabilistic spatiotemporal modeling of day-ahead wind power generation with input-warped Gaussian processes," *Spatial Statist.*, vol. 2025, 2025, Art. no. 100906.
[35] B. Sigrin, et al., "Distributed generation market demand model (dGen): Documentation," Natl. Renewable Energy Lab., Tech. Rep., Feb. 2016, doi: 10.2172/1239054.
[36] Y. Li, S. Cordova, A. Moreira, and M. Heleno, "Netload range cost curves for coordinated transmission-distribution planning under DER growth uncertainty," *IEEE Trans. Energy Markets Policy Regulation*, vol. PP, no. 99, pp. 1–15, 2025, doi: 10.1109/TEMPR.2025.3638837.
[37] V. A. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: Guaranteeing well-connected communities," *Sci. Rep.*, vol. 9, no. 1, Art. no. 5233.
[38] P. I. Frazier, "A tutorial on Bayesian optimization," 2018, *arXiv:1807.02811*.
[39] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proc. Int. Conf. Learn. Intell. Optim.*, Springer, 2011, pp. 507–523.
[40] J. T. Springenberg, A. Klein, S. Falkner, and F. Hutter, "Bayesian optimization with robust Bayesian neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4141–4149.
[41] M. Binois, D. Ginsbourger, and O. Roustant, "A warped kernel improving robustness in Bayesian optimization via random embeddings," in *Learning and Intelligent Optimization*, C. Dhaenens, L. Jourdan, and M.-E. Marmion, Eds. Berlin, Germany: Springer, 2015, pp. 281–286.
[42] M. Binois, D. Ginsbourger, and O. Roustant, "On the choice of the low-dimensional domain for global optimization via random embeddings," *J. Glob. Optim.*, vol. 76, no. 1, pp. 69–90, 2020.
[43] C. Oh, J. Tomczak, E. Gavves, and M. Welling, "Combinatorial Bayesian optimization using the graph Cartesian product," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 2914–2924.
[44] X. Wan, V. Nguyen, H. Ha, B. Ru, C. Lu, and M. A. Osborne, "Think global and act local: Bayesian optimisation over high-dimensional categorical and mixed search spaces," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 10663–10674.
[45] M. Binois, D. Ginsbourger, and O. Roustant, "Quantifying uncertainty on Pareto fronts with Gaussian process conditional simulations," *Eur. J. Oper. Res.*, vol. 243, no. 2, pp. 386–394, 2015.
[46] M. Balandat et al., "BoTorch: A framework for efficient Monte Carlo Bayesian optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 21524–21538.
[47] Z. Wang, F. Hutter, M. Zoghi, D. Matheson, and N. De Freitas, "Bayesian optimization in a billion dimensions via random embeddings," *J. Artif. Int. Res.*, vol. 55, no. 1, pp. 361–387, 2016.
[48] A. Deshwal, S. Ament, M. Balandat, E. Bakshy, J. R. Doppa, and D. Eriksson, "Bayesian optimization over high-dimensional combinatorial spaces via dictionary-based embeddings," in *Proc. 26th Int. Conf. Artif. Intel. Stat.*, 2023, pp. 7021–7039.

Olivier Mulkin received the bachelor's degree in econometrics from Maastricht University, Maastricht, Netherlands, and the master's degree in econometrics from Erasmus Rotterdam University, Rotterdam, Netherlands. He is currently working toward the Ph.D. degree in statistics and applied probability with UC Santa Barbara, Santa Barbara, CA, USA. He was a Intern with the Erasmus School of Economics, ABN AMRO Bank, Lawrence Berkeley National Laboratory and Plaid.

Miguel Heleno (Senior Member, IEEE) received the M.Sc. degree in electrical engineering and computer science from the University of Porto, Porto, Portugal, and the Ph.D. degree in sustainable energy systems from MIT Portugal Program, University of Porto. He is currently a Staff Scientist with the Lawrence Berkeley National Laboratory, Santa Barbara, CA, USA, leading research and innovation projects in the areas of power systems planning and economics.

Mike Ludkovski received the Ph.D. degree in operations research & financial engineering from Princeton University, Princeton, NJ, USA. He is currently a Professor of statistics & applied probability with UC Santa Barbara, Santa Barbara, CA, USA. His research interests include stochastic control, stochastic modeling and Gaussian process surrogates. His research focuses on the statistical analysis of renewable generation and its short-term uncertainty quantification.