Data-driven Probabilistic Constraint Elimination for Accelerated Optimal Power Flow

Constance Crozier, Kyri Baker
Department of Civil, Environmental and Architectural Engineering
University of Colorado Boulder
Email: {constance.crozier, kyri.baker}@colorado.edu

***Abstract*—Optimal power flow calculations are run by operators to determine the cheapest operation of electrical grids given a particular system loading. Considering the increasing uncertainty in demand and availability of renewables, there is a need to develop faster solutions which can be run quickly for a variety of loading scenarios. The non-linear constraints in the optimal power flow problem increase computational costs. However, in reality, many of the constraints are non-binding and do not affect the optimum of the problem. In this paper, we propose a data-driven method for determining constraints that may be excluded from the formulation. The framework combines a latin-hypercube sampling method for generating training scenarios, with maximum likelihood estimation of Gamma distributions. We demonstrate that the method can reduce the solution time of a 24k network from 85.4 seconds to 11.1 seconds.**

***Index Terms*—Computational speed, Constraints, Data-driven methods, Maximum likelihood estimation, Optimal power flow**

I. Introduction

Optimal power flow (OPF) seeks to find the lowest cost method of delivering electrical power to consumers, while meeting the physical constraints of the network. Although OPF is an established research area [1], increased intermittency from renewable generation requires faster and more efficient methods in order to economically balance supply and demand. Towards increased reliability, security constrained optimal power flow is often considered by operators as a stochastic extension to the traditional OPF problem. Each scenario considered adds another set of the power flow constraints to the optimization problem, so the computational complexity grows fast with the number of scenarios.

Acceleration of the traditional OPF problem is attractive to network operators because a larger number of scenarios can be considered in the feasible time horizon. Previous efforts to achieve speed ups have focused around approximations (e.g. [2]), and relaxations (e.g. [3]) to the problem formulation. However, altering the formulation can sacrifice optimality, and in some cases feasibility [4]. More recently, there has been a rise in the use of data-driven methods for accelerating OPF. In other words, methods which in some way learn from a set of OPF solutions under different loading conditions. For example, in [5] artificial neural networks are used to estimate a warm start point for the OPF, meaning the optimization starts closer to the optimal point. In [6] a data-driven convex relaxation for three-phase OPF is proposed. Some works attempt to directly predict a solution [7]–[9], although these may not be strictly feasible.

In order for inequality constraints to have an effect on an optimal solution, they must be "active" at the optimum. This means that the chosen solution must lie at the limit of the inequality. If the constraint is not active at the optimal point, then the inclusion of that constraint in the formulation was unnecessary for that particular set of problem inputs. In OPF there are typically thermal limits on each branch, while we know in practice only a small number of the constraints will be active [10]. Eliminating the constraints (particularly those which add non-convexities) can speed up the optimization algorithm. In security constrained optimal power flow, recent work attempts to identify a set of "umbrella constraints" – a set of constraints which are necessary and sufficient description of the feasible set of the problem. An optimization-based approach for identifying them has been proposed [11], and a deep learning method for predicting them [12].

In this paper, we consider whether data-driven methods can be used to identify constraints that are inactive in all feasible loading scenarios. This can perhaps be framed as the reverse of the umbrella constraint detection problem, where the constraints that do not affect the feasible set are removed. Rather than considering the network over a predefined set of scenarios, we try to identify constraints that will not be reached in any feasible scenario. In the security constrained optimal power flow framework, these constraints could be eliminated in each set of network constraints. The main advantage of the proposed approach over umbrella constraints, is that the computation can be done offline, while umbrella constraint detection relies on knowledge of the loading scenario.

Data-driven methods regarding OPF typically use solutions of various loading scenarios as training data. There are an infinite number of possible network loading scenarios, so an exhaustive loading set is not possible. For this application, it is important that the scenarios considered pushes the network to its feasible limits, otherwise it is impossible to make conclusions about which constraints will be hit. Additionally, given the impossibility of considering every loading scenario, we can not conclude that it is impossible that a constraint will be active just because that has not been observed. Therefore, generating a representative set of scenarios, and interpreting the results of these simulations, are paramount to the success of such a method.

Data-driven methods involving scenario generation and OPF have been used previously in studies with other objectives. In [13] 6000 loading scenarios on a real network were studied to investigate the optimality gap of various OPF relaxations. While [14] uses a scenario-based approach in order to assess future transmission expansion. Additionally, it is worth mentioning that data-driven constraint optimization has been investigated in other optimization problems (e.g. supply-chain management [15]). However, to the authors' knowledge, this is the first paper investigating whether OPF constraints can be removed using data-driven methods.

Therefore, the contributions of the paper can be summarized as follows. First, that we propose a method for generating loading scenarios which push the network to its feasibility limits. Second, that we propose a Bayesian method for estimating the probability that each constraint is active, meaning non-zero probabilities may occur when an active constraint has not been observed. Third, that we propose a framework for removing constraints, and demonstrate the computational speed up that can be achieved for two different networks.

II. Background

This section provides the necessary background to the OPF problem and gamma probability distributions, which are exploited in the proposed method.

A. AC Optimal Power Flow

Optimal power flow aims to minimize the cost of meeting consumer demand on an electricity network, while adhering to the physical constraints of the network. At a given time the network state is described by: the voltage magnitude and angle at each bus, v, θ, the real and reactive power output of each generator, p_g, q_g, and the real and reactive power of each load, p_l, q_l.

We assume that the load is uncontrollable, so the objective is to find the generator values which minimize cost. Here we assume that each generator has a quadratic cost function, such that:

$$\min_{v,\theta,p,q} \sum_g c_{g_0} + c_{g_1} p_g + c_{g_2} p_g^2\,, \tag{1}$$

where $c_{g_0}, c_{g_1}, c_{g_2}$ represent the constant, linear, and quadratic cost coefficients of generator g. The constraints of the optimal power flow problem can be written as:

$$v_{min} \le v_i \le v_{max} \quad \forall i \tag{2}$$
$$0 \le \theta_i \le \pi \quad \forall i \tag{3}$$
$$p_{g_{min}} \le p_g \le p_{g_{max}} \quad \forall g \tag{4}$$
$$q_{g_{min}} \le q_g \le q_{g_{max}} \quad \forall g \tag{5}$$
$$-p_{e_{max}} \le p_e \le p_{e_{max}} \quad \forall e \tag{6}$$
$$-p_{f_{max}} \le p_f \le p_{f_{max}} \quad \forall f \tag{7}$$
$$\sum_{g\in G_i} p_g + \sum_{j\in J_i} p_j + \sum_{e\in E_i} p_e + \sum_{f\in F_i} p_f = 0 \quad \forall i \tag{8}$$
$$\sum_{g\in G_i} q_g + \sum_{j\in J_i} q_j + \sum_{e\in E_i} q_e + \sum_{f\in F_i} q_f = 0 \quad \forall i\,, \tag{9}$$

where i represents the bus index, g the generator index, e the line index, and f the transformer index. Constraints (2) and (3) are bounds on the voltage magnitude and angle at each bus, (4) and (5) are bounds on the outputs of each generator, (6) and (7) are thermal limits of each line and transformer, and (8) and (9) ensures real and reactive power balance at each bus.

Depending on the cost coefficients c_g, the generator bounds (5)-(7) may be frequently active (e.g. if some generators are significantly cheaper than others) and constraints (3) are necessary for ensuring a physical solution. However, it is unlikely that many of the voltage magnitude bounds (2) and thermal constraints (6)-(7) will all be active at once (due to network physics). Inactive constraints increase the computational complexity of the problem, but do not affect the optimal solution.

The line and transformer flows (p_e, q_e, p_f, q_f) are non-linear functions of the voltage magnitudes and angles at either end of the line, for example:

$$p_e = g\ v_o^2 - \Big(g\ cos(\theta_o - \theta_d) + b\ sin(\theta_o - \theta_d)\Big) v_o v_d\,, \tag{10}$$

where g is the line conductance, b is the line susceptance, and the subscripts $_{o,d}$ denote the line's origin and destination buses respectively. Similar expressions are available for q_e, p_f, q_f. The cross-multiplied decision variables and the sinusoidal functions means that the bounds for these terms, constraints (6)-(9), are non-convex. This makes the thermal constraints an especially costly addition to the optimization. This paper explores whether simulation over a large range of loading scenarios can identify constraints that may always be inactive.

B. Gamma Distribution

With thousands of continuous loads, there are an infinite number of possible loading scenarios on the network. Therefore, it would be inappropriate to assume that a constraint that is inactive in all observed scenarios will never be active. Instead we fit a probability distribution function to the observed data, meaning that a probability of the constraint being active can be non-zero even with zero observations.

Here we use the gamma distribution. The gamma distribution of variable x is defined by the probability distribution function:

$$f(x; k, \phi) = \frac{\phi^{-k} x^{k-1} e^{-\frac{x}{\phi}}}{\Gamma(k)} \tag{11}$$

where the distribution is defined by the shape parameter k and the scale parameter ϕ. The distribution is only defined for variables $x > 0$ and parameters $k, \theta > 0$. Figure 1 shows gamma distribution with various parameters. As the distribution is constrained to $x > 0$ we can express x as the distance of each variable from its constraint, given that in OPF the constraints may never be violated. The scale parameter is also particularly useful for this application, because some constraint values will be fairly insensitive to load while some are sensitive. By learning a constraint specific scale parameter

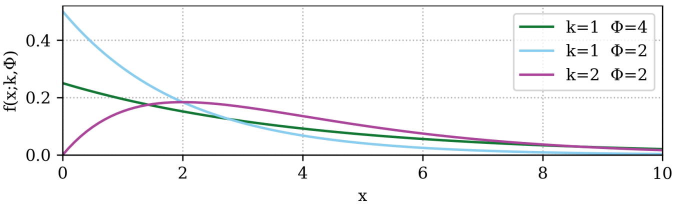


Fig. 1: Three gamma probability distribution functions.

we can factor in the spread of values across the simulations into our predictions.

The probability that x is less than some positive value α is given by the gamma cumulative distribution function, which has the form:

$$P(x \leq \alpha) = \frac{\gamma(k, \frac{\alpha}{\phi})}{\Gamma(k)}\,, \tag{12}$$

where γ is the lower incomplete gamma function. Gamma distributions are undefined at $x = 0$, so for the purposes of this paper we assume that the probability that the constraint is active is equivalent to $P(x \leq 10^{-4})$. In this case we choose 10^{-4} because it is the numeric precision of the constraints in our optimization solver, meaning constraints within this distance are considered active by the solver.

III. Proposed Framework

This section details the proposed methodology to identify constraints that may be eliminated from an optimal power flow problem. The method for generating scenarios is discussed, then the distribution fitting procedure is explained, and finally a criteria for excluding constraints is provided.

A. Scenario Generation

In order to capture a large range of feasible loading conditions of the network, we need to generate a range of loading scenarios. To meet the objective of this study, these scenarios must include cases that push the network to the limits of its feasible region [16]. It is also important that the loading scenarios cover as wide a range of the functional space as possible, given limited computational resources.

Therefore, we propose an iterative latin hypercube sampling process. Latin hypercube sampling is a method which, unlike random sampling, purposefully chooses dissimilar points [17]. We use an iterative process with increasing load perturbations, because it is impossible to know ahead of time how much load the network can tolerate without violating its operating limits. Therefore, we propose beginning with a small randomized perturbation to each of the loads, and gradually increasing the perturbation until none of the sampled loading points are feasible. This algorithm is described in Algorithm 1. The function LHS provides n_b latin hypercube samples of the network loads p_l, q_l using bounds of $p_l^0(1 \pm s_f), q_l^0(1 \pm s_f)$. The OPF function runs n_b optimal power flow problems and returns n_{feas} sets of bus voltages v and branch overheads b_o, where $n_{feas} \leq n_b$ is the number of loading points which had a feasible solution. The while loop terminates once a batch of loading conditions yields zero feasible solutions. It is advisable to keep the batch number n_b large such that the likelihood of all loadings being infeasible is small. Here we specifically use perturbation step sizes of 10%, but other values are also possible. This means that in the first batch all loads are between 90% and 100% of their initial value.

```
n_feas = 1
s_f = 0.1
while n_feas ≥ 0 do
    p_l, q_l ← LHS(p_l^0, q_l^0, s_f, n_b)
    v^sf, b_o^sf, n_feas ← OPF(p_l, q_l)
    s_f += 0.1
end
return v^0.1, b_o^0.1, v^0.2, b_o^0.2 ...
```

Algorithm 1: Algorithm for generating voltage and branch overhead data with varied loading and starting points.

B. Fitting Probability Distribution

In order to determine the probability of a constraint being active, we fit a gamma distribution to the distance of the variable from its bound. Here we are specifically concerned with the voltage bounds and thermal limits. Therefore, for each bus or branch, gamma distributions must be fit based on the observations. Given a set of N observations x_i, the values of k and ϕ which best describe the data are the ones that maximize the likelihood $p(x_1, ..X_n \mid k, \phi)$. Given the exponential form of the probability distribution function, it is numerically easier to maximize the log-likelihood [18]:

$$l(k, \phi) = (k-1)\sum_{i=1}^{N} \ln(x_i) - \sum_{i=1}^{N} \frac{x_i}{\phi} - Nk\ln(\phi) - N\ln(\Gamma(k))\,. \tag{13}$$

Maximizing this function remains non-trivial, particularly when it must be done for more than 10^5 sets. Therefore, we consider reduced ranges of the parameters $k < 2, \phi < 0.05$ for voltages, and $k < 5, \phi < 100$ for thermal constraints. These ranges were deemed reasonable based on the qualities of the targets; voltages can only vary between 0.95 and 1.05 so the scale parameter must be small, while thermal constraints can be orders of magnitude below their limit.

C. Criteria for Constraint Exclusion

Once probability distributions have been fitted based on observations, a probability that the constraint will be violated can be calculated using (12). Here we use a rule-based method to determine whether the constraint should be included. We define a threshold probability ϵ at which the constraint should be included if $P(x = 0) \geq \epsilon$. The value chosen will depend on the requirements for speed versus robustness, given that a value $\epsilon = 10^{-2}$ is likely to eliminate many more constraints than $\epsilon = 10^{-4}$.

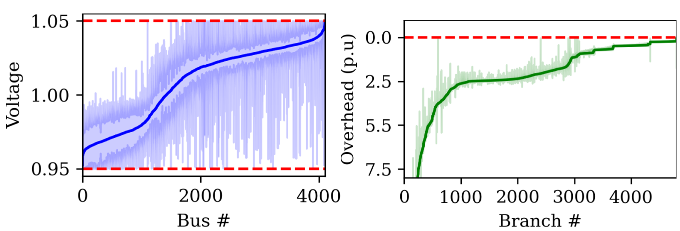
IV. Results & Discussion

Here we demonstrate the proposed method and its effectiveness using two large and realistic networks.

A. Test System and Computational Setup

Here we demonstrate the use of the method on one of the networks from the ARPA-E Grid Optimization Competition [19]. These networks were designed to be representative of real transmission networks in the US. Specifically, we use *C2S7N04102* a network with 4102 buses, 1673 loads, and 399 generators.

Scenarios were generated using the method defined in Section III-A. We used a batch size of 1000 and $s_f = 0.1$, resulting in 4102 feasible loading scenarios after 13 iterations. To run the optimal power flow problems, we used a Python code that directly interfaces with the *IPOPT* package. In order to speed up the code, analytic forms of the Jacobian and Hessian matrices were provided, calculated in vector form using numpy for computational efficiency. These efficiencies meant that we could run a large number of problems using a laptop in a relatively small timeframe.



Bus voltages　　Branch overhead

Fig. 2: The median and range of each of the bus voltages and branch flows across all simulations.

Figure 2 show the variation in bus voltage and branch overhead across all the simulations. On the figure on the left, the buses have been ordered by their median voltage across the scenarios. The median is shown with a solid line, while the shading shows the variation between maximum and minimum voltage experienced across the scenarios. In this network, all voltages are constrained between 0.95 and 1.05 per unit. It can be seen that many of the buses' voltages are very sensitive to the loading scenarios. There were a few buses which had both upper or lower bounds active for different scenarios. However, for the majority of the buses, only one limit was hit.

The figure on the right shows the branch overhead. That is, for each branch, the distance to the flow limit of that branch, in per unit power. Branches are sorted by their median overhead, and the variation is shown with shading. In this case, the branch flows were much less sensitive to loading scenario. Overall, it was rare that any of the branch limits were hit in this network, suggesting that the voltage bounds are more constraining in this case.

B. Probability distributions

For each bus, gamma distributions were fitted to the distance from both voltage limits. Figure 3 shows the cumulative probability distributions for each bus being above a certain voltage for the upper bounds, and below a certain voltage for the lower bound. The buses are ordered as in the previous figure, by their median voltage, so it is unsurprising that the buses at risk of violating the upper bound are at the far right, and vice versa.

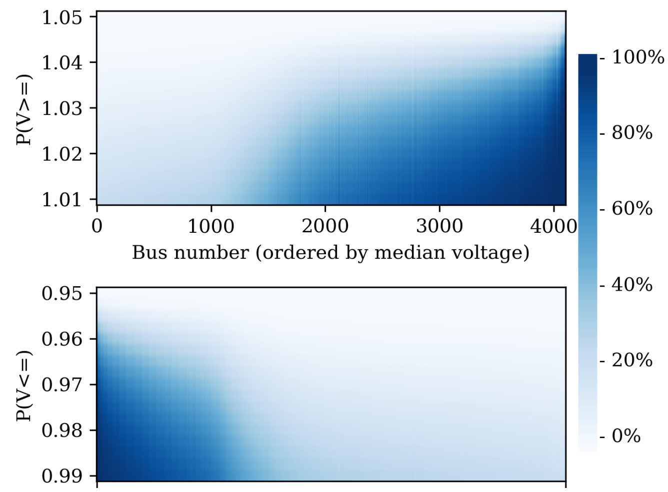


Fig. 3: The cumulative probability distribution of the voltage at each bus with respect to the lower and upper bounds.

Likewise, for each branch a gamma distribution is fitted to the overhead data. Figure 4 shows the cumulative probability of the branch being within a certain distance of its limit. As expected, the probability of a branch constraint being active is smaller than of a voltage constraint. Unlike the voltage distributions, the branches tend to group (see the *blocks* on the right hand side of the distribution). This is likely partly due to the fact that the branch rating varies between branches, so those with similar ratings are likely to have similar overheads.

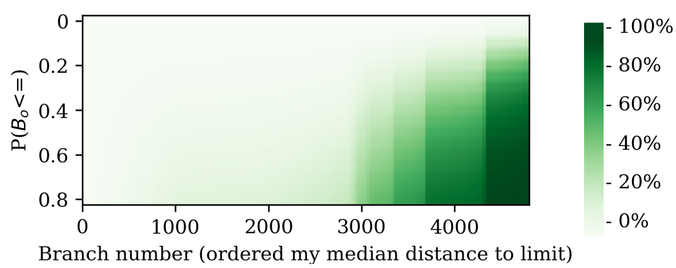


Fig. 4: The cumulative probability distribution of the overhead at each branch.

In order to determine which constraints should be included in the reduced formulation, we calculate the probability that the distance to the limit is $\leq 10^{-4}$ and we include all constraints where this value exceeds 1%. Figure 5 shows a scatter plot with this probability for each constraint, and the cut-off point. It can be seen that several of the bus voltages have more than a 1% chance of being active, but none of the branches do.

C. Computational Savings

In order to assess the computation time savings of the proposed method, 100 optimal power flow simulations were

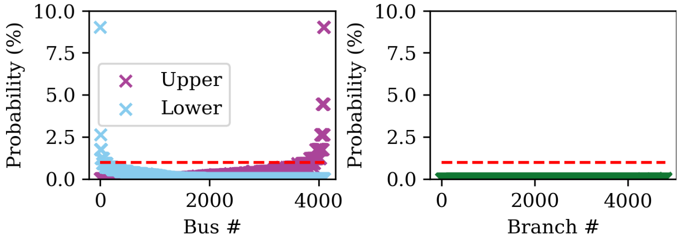


Fig. 5: The probability that a bus voltage constraint or a branch thermal limits is active.

TABLE I: The calculation time (in seconds) for an optimal power flow with and without constraint reduction.

Network	All Constraints	Red. Thermal	Red. Voltage	Both
C2S7N04102	3.609 s	1.438 s	5.649 s	1.355 s
C2S7N24465	85.40 s	11.59 s	87.34 s	11.13 s

performed with loading and generator starting points chosen randomly as described in Section III-A. Table I shows the average number of seconds taken to solve the problem in a variety of cases: (1) with all constraints in the problem, (2) with the chosen thermal constraints removed, (3) with the chosen voltage constraints removed, and (4) with both removed. The process was also repeated for a larger network, C2S7N24465, which has more than 24k buses. It was found that for this network, only two of the thermal constraints fell above the 1% threshold, while many of the bus voltages did.

It can be seen removing the thermal constraints reduces the problem time by 60% for the 4k network, and 86% for the 24k network. Removing the inactive bus constraints in addition to the thermal limits in both cases achieved a small further reduction. However, removing voltage limits without thermal limits increased the computation time. This may be the bounds are linear (so not computationally complex) and including them can speed up convergence.

V. Conclusion

In this paper, we have proposed a data-driven method for excluding constraints from an optimal power flow problem. First, we generate loading scenarios using latin-hypercube sampling, with increasing load variance until the feasibility limits of the network are reached. Then we fit a Gamma distribution to each constraint using maximum likelihood estimation, such that the probability of the constraint being active can be predicted. Finally we remove all constraints whose active probability is below a threshold.

Using two large realistic networks (up to 24,465 buses) we show that removing inactive thermal limits using this method can reduce computation time by 86%. However, we found that removing only voltage bounds could slow down the convergence of the solver. This is likely because the voltage bounds are linear, so the additional cost of including the individual constraint is smaller compared to the thermal limits (which are non-linear).

In order to further develop this method, there are several promising directions for future work. Firstly, a scenario generation method that looks directly scenarios at the feasibility limit could increase the robustness of the method. Secondly, variations in generator costs could also be considered. Finally, the analysis should be repeated for a larger number of networks, to better understand the network dependence of these results.

References

[1] S. Frank, I. Steponavice, and S. Rebennack, "Optimal power flow: a bibliographic survey I," *Energy Systems*, vol. 3, no. 4, 2012.

[2] D. Shchetinin, T. De Rubira, and G. Hug, "On the construction of linear approximations of line flow constraints for AC optimal power flow," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1182–1192, 2018.

[3] 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.

[4] K. Baker, "Solutions of DC OPF are never AC feasible," in *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, ser. e-Energy '21. Association for Computing Machinery, 2021, p. 264–268.

[5] 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)*, 2020, pp. 1–6.

[6] Q. Li, "Uncertainty-aware three-phase optimal power flow based on data-driven convexification," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1645–1648, 2021.

[7] X. Lei, Z. Yang, J. Yu, J. Zhao, Q. Gao, and H. Yu, "Data-driven optimal power flow: A physics-informed machine learning approach," *IEEE Transactions on Power Systems*, vol. 36, no. 1, pp. 346–354, 2021.

[8] Y. Zhou, B. Zhang, C. Xu, T. Lan, R. Diao, D. Shi, Z. Wang, and W.-J. Lee, "A Data-driven Method for Fast AC Optimal Power Flow Solutions via Deep Reinforcement Learning," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1128–1139, 2020.

[9] W. Sun, M. Zamani, M. R. Hesamzadeh, and H.-T. Zhang, "Data-driven probabilistic optimal power flow with nonparametric bayesian modeling and inference," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1077–1090, 2020.

[10] F. Bouffard, F. Galiana, and A. Conejo, "Market-clearing with stochastic security-part II: case studies," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1827–1835, 2005.

[11] A. J. Ardakani and F. Bouffard, "Identification of umbrella constraints in dc-based security-constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3924–3934, 2013.

[12] A. Jahanbani Ardakani and F. Bouffard, "Prediction of umbrella constraints," in *2018 Power Systems Computation Conference (PSCC)*, 2018, pp. 1–7.

[13] T. W. K. Mak, L. Shi, and P. V. Hentenryck, "Phase transitions for optimality gaps in optimal power flows a study on the french transmission network," 2018, arXiv preprint arXiv:1807.05460.

[14] M. Sun, J. Cremer, and G. Strbac, "A novel data-driven scenario generation framework for transmission expansion planning with high renewable energy penetration," *Applied Energy*, vol. 228, pp. 546–555, 2018.

[15] J. Wikarek and P. Sitek, "A data-driven approach to constraint optimization," in *Automation 2019*, R. Szewczyk, C. Zieliński, and M. Kaliczyńska, Eds. Cham: Springer International Publishing, 2020, pp. 135–144.

[16] T. Joswig-Jones, K. Baker, and A. S. Zamzam, "OPF-Learn: An Open-Source Framework for Creating Representative AC Optimal Power Flow Datasets," 2021, arXiv preprint arXiv:2111.01228.

[17] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, 1979.

[18] T. P. Minka, "Estimating a gamma distribution," https://tminka.github.io/papers/minka-gamma.pdf, 2002.

[19] ARPA-E, "Grid optimization competition: Challenge 2 datasets," https://gocompetition.energy.gov/challenges/23/datasets.