On the robustness of machine-learnt proxies for security constrained optimal power flow solvers

Nipun Popli* Elnaz Davoodi† Florin Capitanescu† Louis Wehenkel*

***Abstract*—In this paper, we focus on the robustness of machine learning based proxies used to speed up, alone or jointly with state-of-the-art mathematical optimization methods, optimal power flow and security-constrained optimal power flow calculations. On data sets for the Nordic32 alternative current security-constrained optimal power flow benchmark, we evaluate the robustness of proxies with respect to load distribution, power factors, on-line generators and network topology, and generator costs. We show that simplified random load sampling procedures that are used in most published academic studies, are insufficient to yield robust machine learnt proxies, and consequently limit their usefulness in the real world. Based on these results, we formulate recommendations for future research.**

***Index Terms*—artificial intelligence, machine learning, deep neural networks, random forests, security-constrained optimal power flow, reproducibility, proxies, robustness.**

I. Introduction

The optimal power flow (OPF) and the security-constrained optimal power flow (SCOPF), or the problems of computing a low-cost and secure operating state for an electric grid, are a widely investigated research area [1]. When using an alternating current (AC) physical model of the grid, OPF and more so SCOPF are high-dimensional optimization problems inherent with non-linearity and non-convexity. The so-called direct current (DC) versions, DC-OPF and DC-SCOPF, are based on a linearized physical model. These greatly reduce the computational complexity, but, at the expense of often unacceptable approximations.

In order to make these heavy computations more tractable without sacrificing accuracy, artificial intelligence (AI) and in particular various flavours of machine learning (ML) approaches are currently intensively investigated by the academic research community (see Refs. [2]–[17] discussed in section II). These researches propose methods to build proxies which could advantageously replace or complement analytical techniques based on classical non-linear and/or linear programming applied to physical power system (PS) models.

In a nutshell, the machine learning approach consists in generating a training sample of (SC)OPF problems for a given grid. The training sample is used to build (via machine learning) a 'proxy' (for instance a deep neural network, or a random forest). The proxy receives as input a description of the problem instance. Then it computes as output quantities that can be used in place of, or, in complement with the analytical solver in order to speed up the problem resolution. To evaluate the accuracy of the learnt proxy an independent test sample must be used, on which the machine learning based solver is 'statistically' compared with the analytical solvers in terms of accuracy and computing times.

The lessons that can be learnt from such studies obviously depend on the ranges and on the dimensions of variability covered by the training and test sample generation procedures. The main original contribution of this paper is to identify, raise awareness, and support quantitatively, the core relevant variability factors that should be covered when generating training and test samples for such studies. We first motivate the core variability factors that should be considered to safely assess machine learning based proxies. We next show how the choice of a too small subset of these variability factors when training and evaluating machine learning based proxies may lead to overly optimistic conclusions in terms of accuracy. Our investigation is based on an AC-SCOPF case study applied to a publicly available benchmark test system. In addition, we provide a set of recommendations that are crucial to precisely assess and document the robustness of machine learnt proxies, and to enable the reproducibility of results published by the research community. To the best of our knowledge, no such study has been previously published.

The remainder of the present paper is organized as follows. In Section II we outline the set of core variability factors that should be considered and we highlight the sparsity of the subsets of these factors that are actually covered in a representative sample of research works recently published. In Section III, we describe our empirical study in terms of test system, SCOPF formulation and solvers, and machine learning algorithms utilized. In Section IV, we provide the results of our robustness study highlighting the impact of non-extensive training datasets on poor generalization capabilities of the proxies for SCOPF calculations. In Section V, we conclude with recommendations and open problems for further research. To facilitate reproducibility of our study, and to make our paper self contained, we provide additional background material and some detailed results in the Appendices.

II. Core variability factors to consider

The increasing need for repeated SCOPF calculations to compute a sequence of optimal and secure network states is an inevitable consequence of a more and more dynamical

The authors acknowledge the funding from the Belgian Fonds de la Recherche Fondamentale Collective (F.R.S.-FNRS) for the project ML4SCOPF under grant T.0258.20, and the funding from Luxembourg National Research Fund (FNR) in the framework of the project ML4SCOPF (INTER/FNRS/19/14015062). Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the F.R.S.-FNRS under Grant No. 2.5020.11 and by the Walloon Region.

*Institut Montefiore, Dept of EE&CS, University of Liège, Liège, Belgium, (email: {n.popli, l.wehenkel}@uliege.be)

†Environmental Research and Innovation, Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg, (email: {elnaz.davoodi, florin.capitanescu}@list.lu)

TABLE I: Our study scope with respect to the literature of papers on ML based (DC/AC)-(SC)(O)PF proxies

Bib.	Targeted solvers			Pb.	PS models		ML algorithms		Variability factors considered to train and/or test the ML proxies				
Ref.	PF	OPF	SCOPF	Size	DC	AC	DNN	RF	Ld distr.	Ld pow. fact.	Gen. outages	Line outages	Gen. Cost
[2]			X	4M	X		X		X				
[3]			X	7K	X		X		X				
[4]		X		100K	X		X		X				
[5]		X		300	X		X		X				
[6]		X		600		X		X	X				
[7]		X		2K	X		X		X				
[8]		X		4K		X	X		X			X	
[9]	X			6K		X	X		X	X			
[10]		X		600		X	X		X				
[11]			X	2K		X	X		X				
[12]		X		6K	X	X	X		X	X			
[13]		X		20K		X	X		X			X	
[14]		X		250		X	X		X				
[15]		X		350		X	X		X	X			
[16]		X		250		X	X		X				
[17]		X		3K		X	X		X				
Our			X	6K		X	X	X	X	X	X	X	X

Problem size refers to the largest benchmark used in each paper:

- for DC-PF and DC-OPF, it is measured by the number of buses in the grid;
- for AC-PF and AC-OPF it is 2 times the number of buses in the grid;
- for DC or AC SCOPF these numbers are multiplied by (c+1) where c is the number of explicitly covered contingencies.

grid. The frequency and number of computations will be exacerbated by the following critical factors:

- *Variability and uncertainty of net demand*: proliferation of renewable generation powered by stochastic weather conditions, and of price-driven end-use consumptions.
- *Exogenous market disturbances*: coupling with external energy networks, for instance susceptibility of generator's production cost models to volatility in spot, day-ahead, or forward fuel prices (e.g. gas prices in 2022).
- *Variable system topology*: planned or unintended grid and generation outages, variable market-driven intra-day/hourly generation portfolios, and hence variable grid capacities and control resources.

In regard to these variability factors, Table I summarises the scope of various ML-based studies published in the last few years with the goal of speeding up power flow (PF), OPF, or SCOPF computations. The table also indicates the (computational) problem size of the empirical studies made in these papers, the type of physical PS model used (AC vs DC), and the general class of machine learning algorithms (DNN - deep neural network based methods, RF - random forest types of methods).

The last five columns of this table highlight whether and how the core variability factors are indeed covered by these studies in order to train and/or evaluate the robustness of the learnt proxies:

- variable level and distribution of active power consumed,
- variable power factors of (net) loads,
- variable generation outages, or unit commitments,
- variable grid topologies, e.g. in terms of line outages,
- variable objective function, e.g. costs of generators.

As it is shown in Table I, most of these core variability factors are unfortunately neglected by the machine learning studies published so far. The lack of attention accorded to these factors is a major drawback of state-of-the-art techniques, and, a huge barrier to the potential adoption of machine-learnt proxies in real-world applications.

The last line of Table I shows the characteristics of our AC-SCOPF based study made on the Nordic32 benchmark (see Appendix I for details). It is a medium sized benchmark, but the largest among the two AC-SCOPF studies of the table, and the second to largest among all studies using an AC model. Notice that our study covers all core variability factors.

III. Test Problem and Machine Learning Setup

A. Test problem physics

Our test system is a modified version [18] of the Swedish transmission network, i.e., the Nordic32 electric grid from [19] (refer to Fig. A1 and Fig. A2 in Appendix I). This 60-bus grid has 22 loads, 22 synchronous machines and a tie-line flow from Norway (represented as a single equivalent generator) as sources of real and reactive powers, and we consider a comprehensive set of 52 line and generator based contingencies (see Fig. A2 in Appendix I).

We rely on a physical model for the computation of PSCOPF solutions to generate multiple datasets (see Appendix II for a comprehensive statement of the PSCOPF problem). All simulations have been performed in the open-source Julia/JuMP programming language, resorting to IPOPT to solve all AC PSCOPF problems (see [20] for further details).

B. Overview of generated datasets

In section IV we report results about the robustness of proxies when they are learnt on a learning set covering a certain subset of core variability factors and tested on an independent test set, covering the same or possibly a different set of core variability factors.

For this purpose we have generated 9 different datasets, generically denoted by $\mathcal{S}_\gamma$, for $\gamma \in \{a, b, \dots, h, i\}$.

Each dataset $\mathcal{S}_\gamma$ has about 10,000 operating states of different real and reactive power demands described by 44 input features (corresponding to the vectors of active and reactive load demands), and 46 output features (corresponding to optimal real and reactive power generations from synchronous machines and the tie-line flow from Norway computed, as by the PSCOPF). By utilizing 80% of samples in each dataset $\mathcal{S}_\gamma$ for training, we build ML-based multi-input multi-output (MIMO) regression models: $i)$ $\hat{d}_\gamma : \mathbb{R}^{44} \to \mathbb{R}^{46}$ with feed-forward Deep Neural Networks (DNN) as a parametric learning-based approach, and, $ii)$ $\hat{e}_\gamma : \mathbb{R}^{44} \to \mathbb{R}^{46}$ using Extremely Randomized Trees, as a non-parametric supervised machine learning algorithm (see Section III-C).

The remainder of 20% samples in each dataset $\mathcal{S}_\gamma$ are utilized as test set $\mathcal{T}_\gamma$. These test sets are subsequently used to gauge robustness of the ML-based proxies. In Section IV we discuss only the results obtained by using DNN-based PSCOPF proxies (Tables A1 and A2 provided in Appendix IV show that our conclusions would be the same for proxies learnt by Extremely Randomized Trees).

C. ML Algorithms

1) Deep neural networks [21], [22]: In a preliminary cross-validation study we found a single fully-connected deep neural network architecture suitable for all scenarios presented in this paper. It is composed of four layers: an input layer with 44 units, an output layer with 46 units, and two hidden layers with 150 units each using ReLU activation functions. PyTorch-1.12 was used. In terms of data preprocessing, both inputs and outputs were standardized. Other hyper-parameters include: Adam as optimizer with a learning rate of 0.002, mean square error as a learning loss, and 300 epochs with a mini-batch size of 32.

2) Random forests: We use Extremely randomized trees (ERT), which is a tree ensemble-learning algorithm composed of multiple decision (regression) trees, similar to classical Random forests [23]. In the case of ERT, random cut-points are selected for the splits at each test node in the decision (regression) tree. On our machine learning problems, the ERT method was significantly faster in training than Random forests and yielded comparable accuracy. For this paper, we used the "Extra Tree" package of Scikit learn [24], with an ensemble size of 500 fully developed regression trees and default hyper-parameter settings.

IV. Robustness study

For large-scale electric grids, data samples corresponding to SCOPF solutions are typically high-dimensional in nature. To build ML-based proxies, high-dimensional data samples are utilized to approximate lower-dimensional spaces or manifolds. This, in theory, is known as manifold hypothesis. It states that samples of high-dimensional data form low-dimensional non-linear manifolds, i.e., datasets lie in spaces of arbitrary dimensions embedded within high-dimensional space [25]. This is attributed to constraints arising from underlying physical laws of the real-world phenomena, for which data samples are collected [26]. Suppose we aim to build a MIMO regression model as a high-accuracy SCOPF estimator. A prerequisite to achieve this objective involves fitting low-dimensional nonlinear manifolds with samples mapping a wide range of operating conditions. This, however, begets a combinatorial approach for an extensive dataset generation. We illustrate inadequacies of datasets towards high-fidelity proxy construction, if factors influencing SCOPF solutions are unaccounted for. We cover a subset of factors, particularly *Variability and Uncertainty*, *Exogenous Disturbances*, *Variable System Structure*, as enlisted in Section I.

A. Variability and uncertainty of net demand

The cardinal source of variability and uncertainty are load scenarios. Now consider following mathematical formulations to generate random real and reactive power demands:

$$\mathbf{P}_D = \left(\alpha \cdot \mathbf{P}_D^{\text{high}} + (1-\alpha) \cdot \mathbf{P}_D^{\text{low}}\right) \odot (\mathbf{1} + \beta_P \cdot \mathbf{w}_P), \tag{1a}$$

$$\mathbf{pf} = \mathbf{pf}^{\text{high}} \odot (\mathbf{1} + \beta_Q \cdot \mathbf{w}_Q), \tag{1b}$$

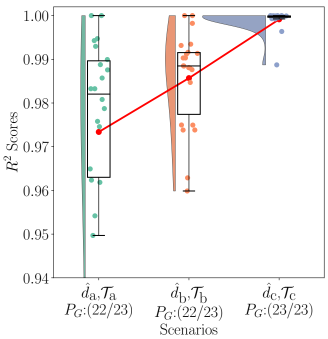
where $\mathbf{P}_D^{\text{low}}, \mathbf{P}_D^{\text{high}} \in \mathbb{R}_+^{22}$ vectorize minimum and peak values for real power demands, respectively, with $\mathbf{P}_D^{\text{low}} = 0.6 \cdot \mathbf{P}_D^{\text{high}}$.

The real power demands are varied homothetically between their extremes with the scalar α drawn from a uniform distribution $\mathcal{U}(0, 1)$. To randomize further, we take an element-wise (aka Hadamard) product ($\odot$) with Gaussian noises $\mathbf{w}_P \in \mathbb{R}^{22}$, generated independently with normal distributions $\mathcal{N}(0, 1)$ and scaled by $\beta_P \in \mathbb{R}_+$.

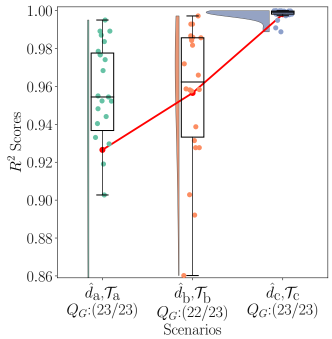
The corresponding reactive power profiles are generated by randomizing power factors $\mathbf{pf}$ in Eq. (1b) with Gaussian noises $\mathbf{w}_Q \in \mathbb{R}^{22}$, once again generated independently with normal distributions $\mathcal{N}(0, 1)$ and scaled by $\beta_Q \in \mathbb{R}_+$, where $\mathbf{pf}^{\text{high}} \in \mathbb{R}^{22}$ contains power factors for peak demand scenario.

We leverage the pair $\{\beta_P, \beta_Q\}$ in Eq. (1) to generate demand samples or load scenarios with varying degree of homotheticity and noise characteristics. These load scenarios serve as inputs to our physics-based simulator used to compute PSCOPF solutions. For each assumption of the pair $\{\beta_P, \beta_Q\}$, we generate around $10,000$ feasible PSCOPF solutions with random load scenarios. These PSCOPF solutions are utilized to construct datasets $\mathcal{S}_\gamma$, as described in Section III.

1) Assumptions about the load distribution: We begin by considering $\{\beta_P = 0.07, \beta_Q = 0.02\}$. These non-zero values result in non-homothetic load scenarios for generation of dataset $\mathcal{S}_\text{a}$. We refer to PSCOPF solutions in $\mathcal{S}_\text{a}$ as base-case scenario. Next, we generate dataset $\mathcal{S}_\text{b}$ with $\{\beta_P = 0.07, \beta_Q = 0\}$. By setting $\beta_Q = 0$ we ensure that power factors are constant. We also generate a dataset $\mathcal{S}_\text{c}$ by considering homothetic real power demands and constant power factors, i.e., $\{\beta_P = 0, \beta_Q = 0\}$. Now we consider DNN-based MIMO regression models $\hat{d}_\text{a}, \hat{d}_\text{b}$, and $\hat{d}_\text{c}$, trained with 8000 samples from $\mathcal{S}_\text{a}, \mathcal{S}_\text{b}$, and $\mathcal{S}_\text{c}$, respectively. We assess predictive accuracies of these models using goodness-of-fit test. For each of the 46 output features, we calculate the coefficient of determination or R^2 score (see Appendix III for further explanations about this criterion). If predicted values of an output feature very well match ground truths in a test set, then that output feature's coefficient-of-determination is close to unity, or, $R^2 \approx 1$ signifies a high-accuracy predictor. However, if $R^2 \notin [0, 1]$, then the predictor is of extremely low-accuracy. From Section III, recall that we left 20% of the samples in test sets. We now assess how accurately MIMO models $\hat{d}_\text{a}, \hat{d}_\text{b}$, and $\hat{d}_\text{c}$ predict test samples in sets $\mathcal{T}_\text{a}, \mathcal{T}_\text{b}$, and $\mathcal{T}_\text{c}$, respectively. In Fig. 1, we visualize R^2 scores as *raincloud* plots [27], [28].



(a) $\boldsymbol{P}_G$: *Real Power Generations*



(b) $\boldsymbol{Q}_G$: *Reactive Power Generations*

Fig. 1: R^2 scores for output features predicted under scenarios: $\{\hat{d}_a, \mathcal{T}_a\}$, $\{\hat{d}_b, \mathcal{T}_b\}$, $\{\hat{d}_c, \mathcal{T}_c\}$

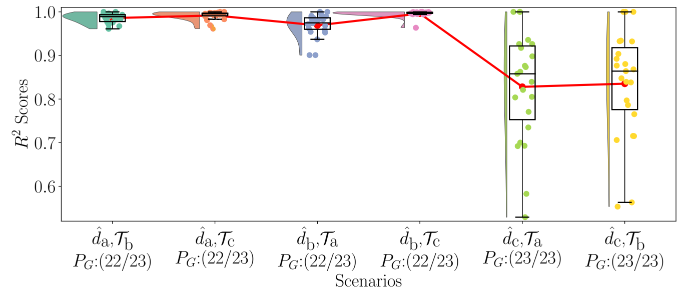
TABLE II: Prediction errors for tie-line flows

Deep Neural Network: MIMO Regressors			
Error	$\hat{d}_c$, $\mathcal{T}_c$	$\hat{d}_c$, $\mathcal{T}_a$	$\hat{d}_c$, $\mathcal{T}_b$
Root Mean Square	ΔP_{tie}=2.02 MW ΔQ_{tie}=0.66 MVar	ΔP_{tie}=113 MW ΔQ_{tie}=35.22 MVar	ΔP_{tie}=113.9 MW ΔQ_{tie}=34.18 MVar
Mean Absolute	ΔP_{tie}=1.47 MW ΔQ_{tie}=0.46 MVar	ΔP_{tie}=89.76 MW ΔQ_{tie}=27.77 MVar	ΔP_{tie}=89.78 MW ΔQ_{tie}=26.73 MVar

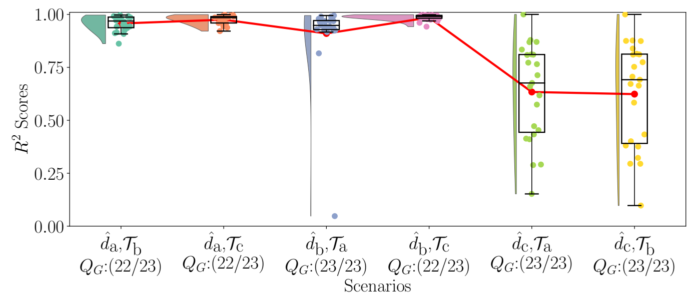
It depicts distributions and box-plots for R^2 scores for three testing scenarios, plotted separately for real and reactive power generations. Only output features with R^2 scores in range $[0, 1]$ are considered, and their counts are numbered along x-axes. The red curves delineate averages of R^2 scores under all scenarios, i.e., R^2_{avg}. These curves lie above 0.92, signifying high predictive accuracies for real and reactive power demands. Notice that average goodness-of-fit for output features (R^2_{avg}) increases with higher degree of homotheticity. For scenario $\{\hat{d}_c, \mathcal{T}_c\}$ in Fig. 1, distributions of R^2 scores have low variance and their averages are very close to unity. Effectively, PSCOPF manifolds for homothetic load scenarios are simple to learn. In Fig. 2, we highlight the impact of homotheticity on generalization capabilities. It depicts raincloud plots for cross-testing scenarios: $\{\hat{d}_a, \mathcal{T}_b\}$, $\{\hat{d}_a, \mathcal{T}_c\}$, $\{\hat{d}_b, \mathcal{T}_a\}$, $\{\hat{d}_b, \mathcal{T}_c\}$, $\{\hat{d}_c, \mathcal{T}_a\}$, $\{\hat{d}_c, \mathcal{T}_b\}$.

The predictor $\hat{d}_a$ was constructed by training a DNN with PSCOPF solutions for non-homothetic load scenarios with variable power factors. It transfers well to constant power-factor ($\mathcal{T}_b$) and homothetic ($\mathcal{T}_c$) load scenarios with $R^2_{\text{avg}} \geq 0.91$ for both real and reactive power generations. The average goodness-of-fit for real ($R^2_{\text{avg}} \leq 0.83$) and reactive ($R^2_{\text{avg}} \leq 0.61$) power generations are lowest with predictor $\hat{d}_c$. As a point-in-case, in Fig. 3 we contrast tie-line flow estimates for scenarios $\{\hat{d}_c, \mathcal{T}_c\}$, $\{\hat{d}_c, \mathcal{T}_a\}$, $\{\hat{d}_c, \mathcal{T}_b\}$.

Notice that tie-line active and reactive power flow estimates under scenario $\{\hat{d}_c, \mathcal{T}_c\}$ near-perfectly mirror ground truths. In contrast, prediction inaccuracies and interquartile ranges are large in cross-testing scenarios $\{\hat{d}_c, \mathcal{T}_a\}$, $\{\hat{d}_c, \mathcal{T}_b\}$. Thereby, mean-squared and mean-absolute errors for tie-line predictions are larger in scenarios $\{\hat{d}_c, \mathcal{T}_a\}$, $\{\hat{d}_c, \mathcal{T}_b\}$, as shown in Table II. To summarize, an overall drop in generalization capability is observed with increasing degree of homotheticity used for the training sample generation.

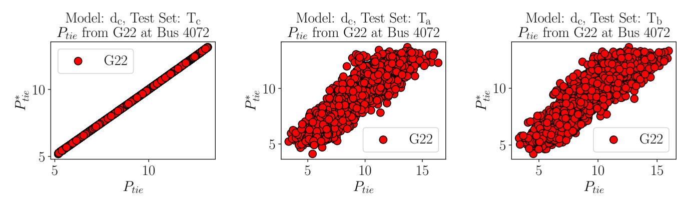


(a) $\boldsymbol{P}_G$: *Real Power Generations*

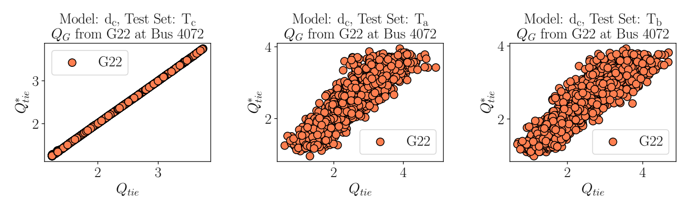


(b) $\boldsymbol{Q}_G$: *Reactive Power Generations*

Fig. 2: R^2 scores for output features predicted under scenarios: $\{\hat{d}_a, \mathcal{T}_b\}$ $\{\hat{d}_a, \mathcal{T}_c\}$, $\{\hat{d}_b, \mathcal{T}_a\}$ $\{\hat{d}_b, \mathcal{T}_c\}$, $\{\hat{d}_c, \mathcal{T}_a\}$ $\{\hat{d}_c, \mathcal{T}_b\}$



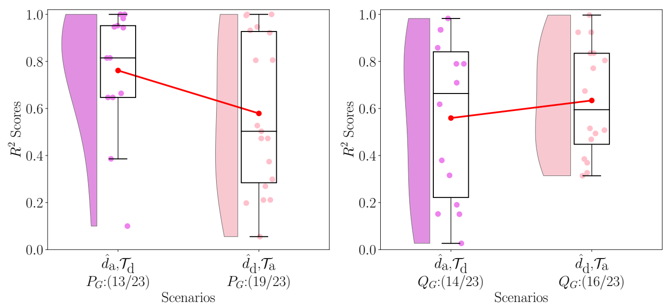
(a) P_{tie}: *Real Power Flow via Tie-Line* (marked as g22 in Fig. A2)



(b) Q_{tie}: *Reactive Power Flow via Tie-Line* (marked as g22 in Fig. A2)

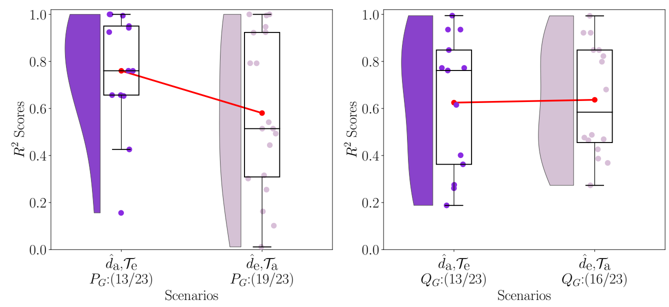
Fig. 3: Tie-line flow (i.e. g22 production) predictions $(P^*_{\text{tie}}, Q^*_{\text{tie}})$ vs. ground truths $(P_{\text{tie}}, Q_{\text{tie}})$ under scenarios: $\{\hat{d}_c, \mathcal{T}_c\}$ $\{\hat{d}_c, \mathcal{T}_a\}$, $\{\hat{d}_c, \mathcal{T}_b\}$. Flows are expressed in *per unit of a 100 MVA base*

2) Assumptions about power factors: The parameter β_Q in Eq. (1) characterizes the range over which the power factor of net demands is expected to fluctuate. As non-synchronous technologies proliferate and replace synchronous machines, at both transmission and distribution level, a significant impact is forecasted on reactive power demands and generations [29]. Reactive power demands may substantially alter generation profiles, and hence, PSCOPF solutions. To illustrate this, we generate a dataset $\mathcal{S}_d$ similar to base-case dataset $\mathcal{S}_a$ in terms of active power variability, but with amplified variability of power factors, i.e, $\{\beta_P = 0.07, \beta_Q = 0.05\}$.



(a) $\boldsymbol{P}_G$: *Real Power Generations* (b) $\boldsymbol{Q}_G$: *Reactive Power Generations*

Fig. 4: R^2 scores for output features predicted under scenarios: $\{\hat{d}_{\mathrm{a}}, \mathcal{T}_{\mathrm{d}}\}$, $\{\hat{d}_{\mathrm{d}}, \mathcal{T}_{\mathrm{a}}\}$



(a) $\boldsymbol{P}_G$: *Real Power Generations* (b) $\boldsymbol{Q}_G$: *Reactive Power Generations*

Fig. 5: R^2 scores for output features predicted under scenarios: $\{\hat{d}_{\mathrm{a}}, \mathcal{T}_{\mathrm{e}}\}$, $\{\hat{d}_{\mathrm{e}}, \mathcal{T}_{\mathrm{a}}\}$

Once again, we train a DNN by using 8000 training samples from $\mathcal{S}_{\mathrm{d}}$. The resulting MIMO regressor $\hat{d}_{\mathrm{d}}$ exhibits an average goodness-of-fit $R^2_{\mathrm{avg}} \geq 0.923$ on the test set $\mathcal{T}_{\mathrm{d}}$. But, consider in Fig. 4, box-plots and distributions for R^2 scores under cross-testing scenarios: $\{\hat{d}_{\mathrm{a}}, \mathcal{T}_{\mathrm{d}}\}$, $\{\hat{d}_{\mathrm{d}}, \mathcal{T}_{\mathrm{a}}\}$. The goodness-of-fit averages (R^2_{avg}) less than 0.8, for both real and reactive power generations. The total number of output features with $R^2 \in [0,1]$ also strongly drops. Note that $\hat{d}_a$ was obtained with 8000 PSCOPF solutions in $\mathcal{S}_{\mathrm{a}}$, computed for non-homothetic load scenarios. Yet, it does not well transfer to $\mathcal{T}_{\mathrm{d}}$. We conclude that the randomization of reactive power demands profoundly impacts a proxy's generalization capability.

To summarize, random demand scenarios for dataset generation must reflect potential variabilities and uncertainty bounds of both active and reactive power. Furthermore, parameter α in Eq. (1a) can be vectorized for finer modeling, as large industrial and commercial loads may exhibit non-conventional diurnal variations. Similarly, $\{\beta_P, \beta_Q\}$ could be modeled as a pair of time-varying vectors. This would enable temporal and spatial disaggregation of uncertainty bounds.

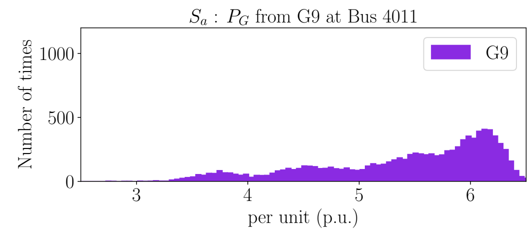
B. Exogenous market disturbances

The dependencies of electric grids on external energy networks are exogenous sources of disturbances. A prime example is of infrastructural couplings with gas and oil pipelines. For optimal power flow solutions, such externalities alter generation profiles via the objective function. Consider a dataset $\mathcal{S}_{\mathrm{e}}$ generated with parameters used to construct base-case dataset $\mathcal{S}_{\mathrm{a}}$, but with different production costs for gas-fired generators $\{g2, g11, g14, g18\}$, and for tie-line flows (marked as $g22$ in Fig. A2). These costs were scaled-up to 2-3 times of the values assumed to construct $\mathcal{S}_{\mathrm{a}}$. In Fig. 5, we contrast reduced predictive accuracies for cross-testing scenarios: $\{\hat{d}_{\mathrm{a}}, \mathcal{T}_{\mathrm{e}}\}$, $\{\hat{d}_{\mathrm{e}}, \mathcal{T}_{\mathrm{a}}\}$.

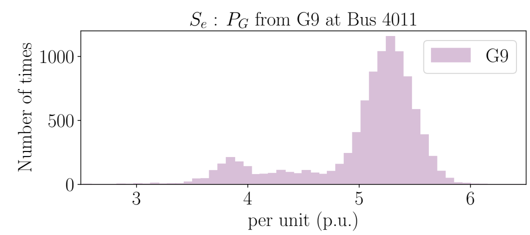
The abundances of (or the interruptions in) fuel supplies manifest as cost variations and may result in new manifolds of PSCOPF solutions, for which the model is untrained. As an example, in Fig. 6 we include histogram plots for real power outputs of generator $g9$ in datasets $\mathcal{S}_{\mathrm{a}}$ and $\mathcal{S}_{\mathrm{e}}$.

C. Variable system configuration and topology

A power system's topology is perpetually in a state of flux. Some common causes include scheduled maintenances of lines/generators/substations, market-driven unit commitments, and inadvertent tripping or failures of power system components. A rich dataset therefore must also contain PSCOPF solutions for topologies resulting from planned network configurations and plausible outages. As we illustrate next, PSCOPF solutions under certain system topologies may lie in a qualitatively different low-dimensional manifold, while PSCOPF solutions for some system topologies are generalisable with a smaller sample set.



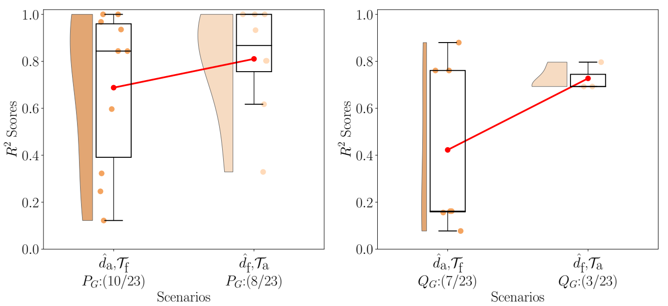
(a) P_G for $g9$ in Dataset $\mathcal{S}_{\mathrm{a}}$



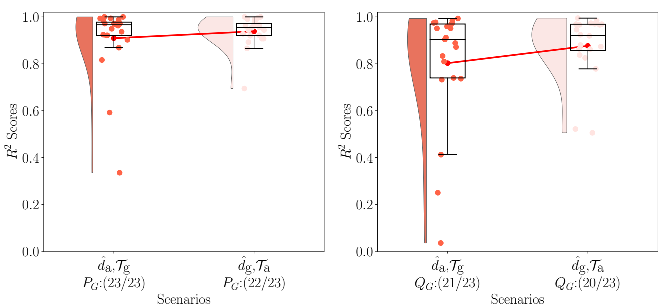
(b) P_G for $g9$ in Dataset $\mathcal{S}_{\mathrm{e}}$

Fig. 6: Histogram plots for real power outputs of generator $g9$ in PSCOPF-solution datasets: a) $\mathcal{S}_{\mathrm{a}}$, b) $\mathcal{S}_{\mathrm{e}}$. Generation powers are expressed in *per unit of a 100 MVA base*

1) Grid-based topology variations: We now consider datasets $\mathcal{S}_{\mathrm{f}}$ and $\mathcal{S}_{\mathrm{g}}$, containing PSCOPF solutions for two grid topologies. In each topology, we assume an out-of-service line at sub-transmission level (130 kV). In Fig. A2, these lines connect buses $\{1043{,}1044\}$ ($\mathcal{S}_{\mathrm{f}}$) and $\{1011{,}1013\}$ ($\mathcal{S}_{\mathrm{g}}$). Each line has a flow limit of 175 MVA. To generate $\mathcal{S}_{\mathrm{f}}$ and $\mathcal{S}_{\mathrm{g}}$, load scenarios were randomized with $\{\beta_P = 0.07, \beta_Q = 0.02\}$ in Eq. (1), similar to values assumed to construct $\mathcal{S}_{\mathrm{a}}$. The corresponding MIMO regressors, $\hat{d}_{\mathrm{f}}$ and $\hat{d}_{\mathrm{g}}$, obtained by training DNNs with samples from $\mathcal{S}_{\mathrm{f}}$ and $\mathcal{S}_{\mathrm{g}}$, respectively, provide $R^2_{\mathrm{avg}} \geq 0.907$ for scenarios $\{\hat{d}_{\mathrm{f}}, \mathcal{T}_{\mathrm{f}}\}$, $\{\hat{d}_{\mathrm{g}}, \mathcal{T}_{\mathrm{g}}\}$. Fig. 7(a) and Fig. 7(b) depict rain-

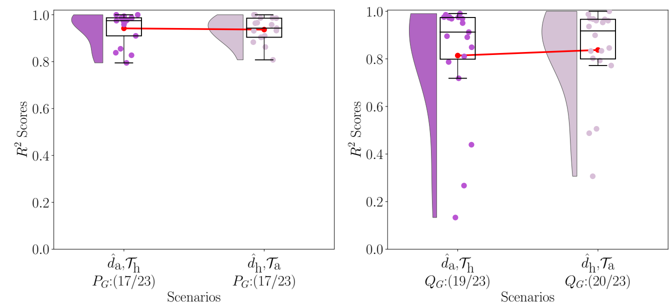


(a) $\boldsymbol{P}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{f}\}$, $\{\hat{d}_\mathrm{f}, \mathcal{T}_\mathrm{a}\}$ (b) $\boldsymbol{Q}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{f}\}$, $\{\hat{d}_\mathrm{f}, \mathcal{T}_\mathrm{a}\}$

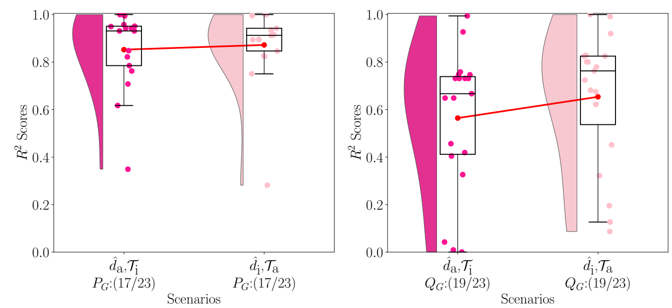


(c) $\boldsymbol{P}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{g}\}$, $\{\hat{d}_\mathrm{g}, \mathcal{T}_\mathrm{a}\}$ (d) $\boldsymbol{Q}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{g}\}$, $\{\hat{d}_\mathrm{g}, \mathcal{T}_\mathrm{a}\}$

Fig. 7: R^2 scores for real and reactive power generations predicted under scenarios: a), b) $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{f}\}$, $\{\hat{d}_\mathrm{f}, \mathcal{T}_\mathrm{a}\}$ c), d) $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{g}\}$, $\{\hat{d}_\mathrm{g}, \mathcal{T}_\mathrm{a}\}$



(a) $\boldsymbol{P}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{h}\}$, $\{\hat{d}_\mathrm{h}, \mathcal{T}_\mathrm{a}\}$ (b) $\boldsymbol{Q}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{h}\}$, $\{\hat{d}_\mathrm{h}, \mathcal{T}_\mathrm{a}\}$



(c) $\boldsymbol{P}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{i}\}$, $\{\hat{d}_\mathrm{i}, \mathcal{T}_\mathrm{a}\}$ (d) $\boldsymbol{Q}_G$: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{i}\}$, $\{\hat{d}_\mathrm{i}, \mathcal{T}_\mathrm{a}\}$

Fig. 8: R^2 scores for real and reactive power generations predicted under scenarios: a), b) $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{h}\}$, $\{\hat{d}_\mathrm{h}, \mathcal{T}_\mathrm{a}\}$ c), d) $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{i}\}$, $\{\hat{d}_\mathrm{i}, \mathcal{T}_\mathrm{a}\}$

cloud plots for R^2 scores under scenarios: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{f}\}$, $\{\hat{d}_\mathrm{f}, \mathcal{T}_\mathrm{a}\}$. Mean goodness-of-fit for real power outputs are 0.688 and 0.81, averaged for 10 and 8 sources out of 23, respectively. The prediction accuracies are lower for reactive power outputs with R^2_avg values 0.72 and 0.42, averaged for 7 and 3 sources, respectively. Now, contrast these performance indices with those obtained for cross-testing scenarios: $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{g}\}$, $\{\hat{d}_\mathrm{g}, \mathcal{T}_\mathrm{a}\}$. As shown in Fig. 7(c) and Fig. 7(d), R^2_avg scores are greater than 0.803 for real and reactive power output estimates, and are averaged for at least 20 out of 23 features. Even with identical capacity ratings of 175 MW for out-of-service lines, generalizabilities of PSCOPF solutions in datasets $\mathcal{S}_\mathrm{f}$ and $\mathcal{S}_\mathrm{g}$ differ significantly. One must note that in the Nordic32 system, bulk of electricity is transmitted from north to south via five high-capacity interconnections. By opening a single branch between $\{1043,1044\}$ in south, which is relatively closer to bulk transmission lines in centre, qualitatively different PSCOPF solutions are obtained. In comparison, the out-of-service status of line connecting buses $\{1011,1013\}$ in northern most region, has a minimal impact on north-to-south flows. To conclude, certain topologies necessitate an exclusive or a dedicated dataset construction. For instance, sampling low-load conditions where in a few lines are opened to preclude voltage transients/overshoots.

2) Generation-based topology variations: The generation portfolio varies with unit commitments. Thus, irrespective of planned or unforeseen grid outages, system topology changes due to start-up and shutdown of generators within an operating hour. Let us construct PSCOPF-solution datasets $\mathcal{S}_\mathrm{h}$ and $\mathcal{S}_\mathrm{i}$ by assuming $g1$ and $g8$ in offline mode, respectively. The rated nameplate capacity for generator $g1$ is 720 MW, and 965 MW for generator $g8$. In each case, load scenarios were randomized with parameters $\{\beta_P = 0.07, \beta_Q = 0.02\}$ in Eq. (1). We once again construct DNN-based MIMO regressors $\hat{d}_\mathrm{h}$ and $\hat{d}_\mathrm{i}$ by using 8000 samples from $\mathcal{S}_\mathrm{h}$ and $\mathcal{S}_\mathrm{i}$, respectively. An average goodness-of-fit is greater than 0.942 under self-testing scenarios $\{\hat{d}_\mathrm{h}, \mathcal{T}_\mathrm{h}\}$, $\{\hat{d}_\mathrm{i}, \mathcal{T}_\mathrm{i}\}$. Now consider R^2 scores for cross-testing scenarios $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{h}\}$ $\{\hat{d}_\mathrm{h}, \mathcal{T}_\mathrm{a}\}$ in Fig. 8(a) and Fig. 8(b). The distribution of R^2 scores for real power generations exhibits a low variance. A plausible reason is the fact that cost objective is a function of real power outputs. So, different reactive power generations in dataset $\mathcal{S}_\mathrm{h}$ enable real power outputs near-similar to those in base-case set $\mathcal{S}_\mathrm{a}$. This is indicated by $R^2_\mathrm{avg} > 0.937$ for 17 out of 23 sources in Fig. 8(a). We notice a similar observation under scenarios $\{\hat{d}_\mathrm{a}, \mathcal{T}_\mathrm{i}\}$ $\{\hat{d}_\mathrm{i}, \mathcal{T}_\mathrm{a}\}$ in Fig. 8(c) and Fig. 8(d), albeit with lower goodness-of-fit averages for real power outputs ($R^2_\mathrm{avg} < 0.87$). For illustrative purposes, we considered modest variations in generation portfolio. But, in practice, multiple generators often disconnect from, or, reconnect to the network within an operating hour. In the future, inter- and intra-hour variations in generation portfolios are expected to be more pronounced. For example, frequent unit commitments are required to balance duck-shaped diurnal variations caused by solar parks.

V. Recommendations and Research Directions

In this paper we have studied the impact of relevant dimensions on the data generation process used to build and evaluate ML-based (SC)(O)PF proxies. We raise awareness that a set of multiple lower-order manifolds results from topological variations, random load scenarios, fluctuations in production costs, to cite a few influencing factors. These factors will also impact predictive accuracies of ML-based proxies proposed for optimal power flow, power flow, direct current power flow, and other possible variants [5]–[15]. Our extensive case study was based on 9 different dataset sampling assumptions, each one covered by 10,000 AC-SCOPF computations for a 60-bus system and 52 contingencies. Using both DNN and RF, two very different state-of-the-art ML methods, we generated 2×9 different machine learnt proxies, and evaluated them in 2×34 train/test combinations. To the best of our knowledge this is by far the most comprehensive empirical robustness study of machine-learnt proxies for (SC)(O)PF computations.

We show quantitatively and systematically that an abundance of training samples is useless if they are not representative of typical operating conditions witnessed in real-world utilities. Yet, the process of constructing an extensive dataset is rather combinatorial in nature. Possible solutions to do sound academic research pass by the explicit consideration of our 5 main variability factors for dataset generation.

On the other hand, the availability of non-simulated datasets, provided from historians of TSO SCADA platforms would be most useful, e.g. in order to help designing representative simulated datasets. In our study, we found that the range of (net) demand patterns is an important subject, and both active and reactive demand combinations need to be well covered. Also, changing the cost function or the system configuration may jeopardize the validity of machine-learnt proxies.

Reinforcement learning [16], training multiple proxies for different topological variations [4], or use of proxies to reduce iterations in optimization processes [17], [30] are relevant research directions. However, alternative solutions must be explored wherein proxies are shown to be able to learn new abstractions that indeed span the desired range of conditions targeted by their practical application scenarios in planning and operation. As an example, graph-theoretic data representations can potentially generalize (SC)(O)PF solutions under topological variations [31], [32].

Finding solutions to these open problems is crucial to inspire trust and confidence for the deployment of AI/ML-driven (SC)(O)PF calculators in the real world.

Appendix I
Nordic32 System

The Nordic32 network, prepared for voltage stability and security assessment [19], is a synchronous interconnection of Swedish network, and parts of Norwegian and Finnish transmission-level networks along with Zealand, the eastern part of the Danish network, as depicted in Fig. A1. The one-line diagram of the modified Nordic32 system is shown in Fig. A2. Here, the original 74-bus system in [19] is modified to a 60-bus network. Structurally, there are two main modifications. First of all, for the purpose of preventive security-constrained optimal power flow calculations, 22 distribution-level transformers aren't modeled and 22 loads are connected directly to high-voltage buses.

Secondly, interconnected transmission-level (400, 220 kV) and sub-transmission-level (130 kV) networks are bifurcated into two regions, ***NORTH*** and ***SOUTH***, in contrast to four

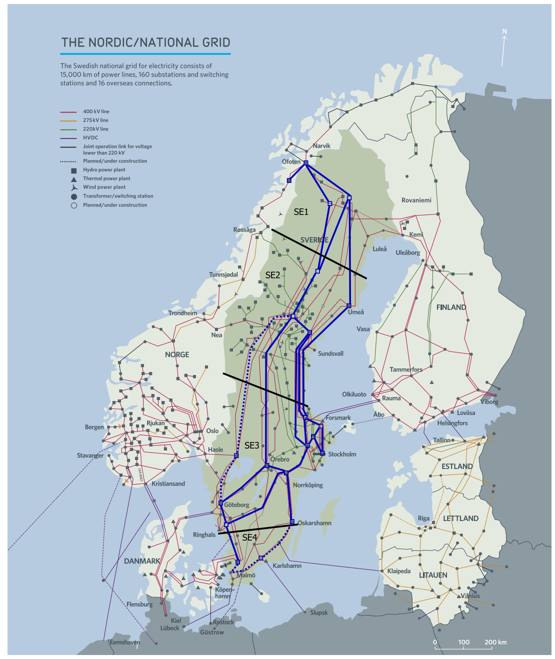


Fig. A1: Nordic32 network location within Swedish transmission-level system (image courtesy: [33])

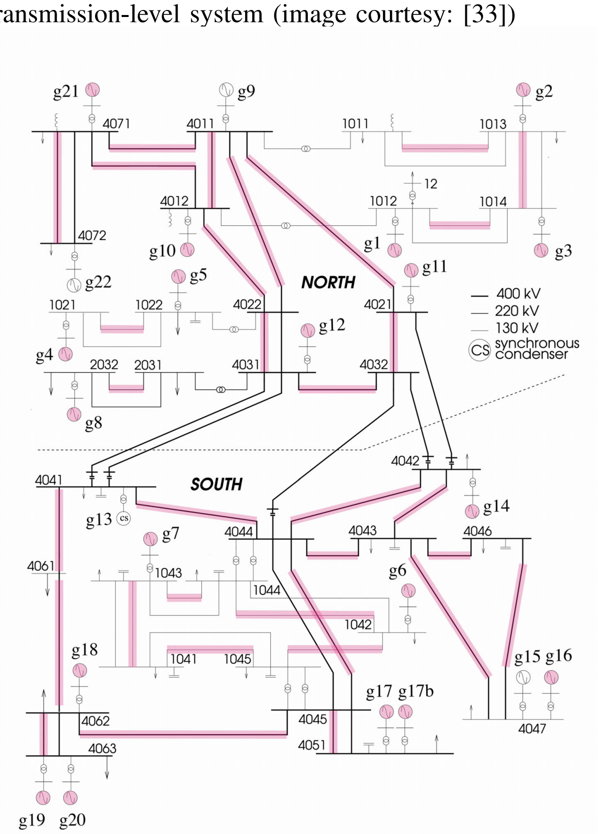


Fig. A2: One-line diagram for modified Nordic32 system: highlighted generator and line contingencies for reliability assessments

areas in [19]. The control area in ***NORTH*** is generation dominant with moderate load level, while the area in ***SOUTH*** has significant loading level and low generation capacity. Then, interconnecting the generation-heavy ***NORTH*** to the load-heavy ***SOUTH*** are long series compensated 400 kV transmission lines, two between buses {4031,4041}, and one each between buses $\{4032{,}4044\}$, $\{4032{,}4042\}$, $\{4021{,}4042\}$. For reactive power management, there are three inductive and one capacitive shunts in ***NORTH***, and eight capacitive shunts in ***SOUTH***. The electric grid consists of 21 generators, a synchronous condenser (g13 at bus 4041) and a tie-line flow from Norway (equivalenced as generator g22 at bus 4072), coupled via 23 step-up transformers. Lastly, there are 52 non-transformer and non–series-compensator branches, i.e. true transmission lines.

For grid reliability, security-constrained optimal power flow solutions account for 33 transmission line-based contingencies $\{\{1011{,}1013\}, \{1012{,}1014\}, \{1013{,}1014\}, \{1021{,}1022\}, \{1041{,}1043\}, \{1041{,}1045\}, \{1042{,}1044\}, \{1042{,}1045\}, \{1043{,}1044\}, \{2031{,}2032\}, \{4011{,}4012\}, \{4011{,}4021\}, \{4011{,}4022\}, \{4011{,}4071\}, \{4012{,}4022\}, \{4012{,}4071\}, \{4021{,}4032\}, \{4022{,}4031\}, \{4031{,}4032\}, \{4041{,}4044\}, \{4041{,}4061\}, \{4042{,}4043\}, \{4042{,}4044\}, \{4043{,}4044\}, \{4043{,}4046\}, \{4043{,}4047\}, \{4044{,}4045\}, \{4045{,}4051\}, \{4045{,}4062\}, \{4046{,}4047\}, \{4061{,}4062\}, \{4062{,}4063\}, \{4071{,}4072\}\}$, and 19 generator-based contingencies {g1, g2, g3, g4, g5, g6, g7, g8, g10, g11, g12, g14, g16, g17, g17b, g18, g19, g20, g21}.

Appendix II
Preventive AC-SCOPF formulation

The present section provides fundamentals about the standard AC-SCOPF formulation as we have implemented it in our study.

There are two variants of SCOPF calculations [1]: preventive SCOPF [34], and corrective SCOPF [35]. In this paper we only consider the preventive one (denoted by PSCOPF).

An electric grid can be represented as a graph $\Gamma = (\mathcal{N}, \mathcal{B})$ with generators and loads connected to nodes or buses $n \in \mathcal{N}$ and branches $b \in \mathcal{B}$, where $\mathcal{B} = \mathcal{B}_l \cup \mathcal{B}_t \cup \mathcal{B}_p$ with multiple sets for transmission lines ($\mathcal{B}_l$), transformers ($\mathcal{B}_t$), and phase-shift transformers ($\mathcal{B}_p$). A branch k from node i^k to node j^k is defined as $b = ((i^k, j^k),\ (\boldsymbol{Y}_x^k, y_x^k, I_x^k) |\ i^k, j^k \in \mathcal{N},\ i^k \neq j^k,\ \boldsymbol{Y}_x^k \in \mathbb{C}^{2\times 2},\ y_x^k \in \mathbb{C},\ I_x^k \in \mathbb{R}_+,\ k \in \{1, 2, \ldots, |\mathcal{B}|\},\ x \in \{l, t, p\})$. For a branch indexed as k, $\boldsymbol{Y}_x^k$ and y_x^k are complex nodal admittance matrix and branch admittance in rectangular forms, respectively, I_x^k denotes its rated maximum current flow, and $x \in \{l, t, p\}$ indicates the branch type.

The subset $\mathcal{G} \subset \mathcal{N}$ contains the indices of the nodes with generators. The subset $\mathcal{G}_\mathrm{f} \subset \mathcal{G}$ contains indices of the nodes postulated to experience single generator failure. The subset $\mathcal{L}_\mathrm{f} \subset \mathcal{B}_l$ contains transmission-line branches for expected single-line-failures.

For an operating scenario indexed as c, $\boldsymbol{P}_\mathrm{G}^c$, $\boldsymbol{Q}_\mathrm{G}^c \in \mathbb{R}^{|\mathcal{N}|}$ represent real and reactive power generations at all nodes, respectively. The base case or the normal operating scenario is labeled as $c = 0$, while line and generator contingencies are indexed with labels in sets $\{1, 2, \ldots, |\mathcal{L}_\mathrm{f}|\}$ and $\{|\mathcal{L}_\mathrm{f}|+1, |\mathcal{L}_\mathrm{f}|+2, \ldots, |\mathcal{L}_\mathrm{f}|+|\mathcal{G}_\mathrm{f}|\}$, respectively. Out of $|\mathcal{G}|$ nodes, one node with a generator is modeled as a reference node or a slack bus. We label this node with index 's'. The problem formulation relies on rectangular coordinates of complex voltage $\boldsymbol{V}^c \in \mathbb{C}^{|\mathcal{N}|}$.

The decision variables are real and reactive power schedules for generators, and complex voltages at generator nodes (except voltage angle for the slack generator at node 's'), defined for pre- and post-contingency states of operations. These are formalized as $\boldsymbol{E}\boldsymbol{P}_\mathrm{G}^c$, $\boldsymbol{E}\boldsymbol{Q}_\mathrm{G}^c$ and $\boldsymbol{E}_\mathrm{r}\boldsymbol{V}^c$, where $\boldsymbol{E} \in \{0,1\}^{|\mathcal{G}|\times|\mathcal{N}|}$ and $\boldsymbol{E}_\mathrm{r} \in \{0,1\}^{(|\mathcal{G}|-1)\times|\mathcal{N}|}$ contain subsets of rows from $|\mathcal{N}|$-dimensional identity matrix $\boldsymbol{I}_{|\mathcal{N}|}$ corresponding to indices of nodes in sets $\mathcal{G}$ and $\mathcal{G}\backslash\{\mathrm{s}\}$, respectively. The reference node voltage is $\boldsymbol{e}_\mathrm{s}\boldsymbol{V}^c$, where $\boldsymbol{e}_\mathrm{s}$ is the s$^\mathrm{th}$ row in $\boldsymbol{I}_{|\mathcal{N}|}$. Its imaginary component is assumed to be zero for all contingency scenarios. The vectors $\boldsymbol{P}_\mathrm{D}$, $\boldsymbol{Q}_\mathrm{D} \in \mathbb{R}^{|\mathcal{N}|}$ denote nodal real and reactive power demands, respectively. For line-based contingencies, the nodal or bus admittance matrices are obtained as $\boldsymbol{Y}^c = \boldsymbol{Y}^0 + \boldsymbol{Y}_\mathrm{sh} - \Delta\boldsymbol{Y}^c$, where $\boldsymbol{Y}^c, \boldsymbol{Y}^0, \boldsymbol{Y}_\mathrm{sh}, \Delta\boldsymbol{Y}^c \in \mathbb{C}^{|\mathcal{N}|\times|\mathcal{N}|}$. The base case admittance matrix $\left(\boldsymbol{Y}^0 + \boldsymbol{Y}_\mathrm{sh}\right)$, where $\boldsymbol{Y}_\mathrm{sh}$ accounts for shunt admittances, is adjusted for each line contingency. A sparse matrix $\Delta\boldsymbol{Y}^c$ is constructed using nodal admittance matrix $\boldsymbol{Y}_x^k$ of failed branch between nodes $\left(i^k, j^k\right)$ with index k. It consists of four non-zero elements: $\Delta\boldsymbol{Y}^c\left(i^k, i^k\right) = \boldsymbol{Y}_x^k(1,1)$, $\Delta\boldsymbol{Y}^c\left(j^k, j^k\right) = \boldsymbol{Y}_x^k(2,2)$, $\Delta\boldsymbol{Y}^c\left(i^k, j^k\right) = \boldsymbol{Y}_x^k(1,2)$, $\Delta\boldsymbol{Y}^c\left(j^k, i^k\right) = \boldsymbol{Y}_x^k(2,1)$. The branch-to-node adjacency matrix for base case ($c = 0$) is $\boldsymbol{A} \in \{0, 1, -1\}^{|\mathcal{B}|\times|\mathcal{N}|}$. Notice that we do not equivalence or aggregate admittances of multiple branches between same pair of nodes, as their rated maximum current flows may differ numerically. Next, we provide the detailed PSCOPF formulation.

$$\min_{\substack{\boldsymbol{E}\boldsymbol{P}_\mathrm{G}^c,\\ \boldsymbol{E}\boldsymbol{Q}_\mathrm{G}^c, \boldsymbol{E}_\mathrm{r}\boldsymbol{V}^c}} \quad (\boldsymbol{E}\boldsymbol{P}_\mathrm{G}^0)^\top \mathrm{diag}(\boldsymbol{a})(\boldsymbol{E}\boldsymbol{P}_\mathrm{G}^0) \ + \ \boldsymbol{b}^\top\boldsymbol{E}\boldsymbol{P}_\mathrm{G}^0 \ + \ \boldsymbol{c}^\top\boldsymbol{1} \tag{A1}$$

subject to

Nodal or Bus Constraints

Power Balancing - Base Case and Contingency Scenarios:

$$\boldsymbol{P}_\mathrm{G}^c - \boldsymbol{P}_\mathrm{D} = \mathrm{Re}\left[\bar{\boldsymbol{V}}^c \odot (\boldsymbol{Y}^c\boldsymbol{V}^c)\right], \tag{A2a}$$

$$\boldsymbol{Q}_\mathrm{G}^c - \boldsymbol{Q}_\mathrm{D} = -\mathrm{Im}\left[\bar{\boldsymbol{V}}^c \odot (\boldsymbol{Y}^c\boldsymbol{V}^c)\right], \tag{A2b}$$

Voltage Limits - Base Case and Contingency Scenarios:

$$\boldsymbol{V}_\mathrm{min} \odot \boldsymbol{V}_\mathrm{min} \le \bar{\boldsymbol{V}}^c \odot \boldsymbol{V}^c \le \boldsymbol{V}_\mathrm{max} \odot \boldsymbol{V}_\mathrm{max}, \tag{A2c}$$

Reference Voltage - Base Case and Contingency Scenarios:

$$\mathrm{Im}\left(\boldsymbol{e}_\mathrm{s}\bar{\boldsymbol{V}}^c\right) = 0, \tag{A2d}$$

for all

$c \in \{0\} \cup \{1, 2, \ldots, |\mathcal{L}_\mathrm{f}|\} \cup \{|\mathcal{L}_\mathrm{f}|+1, |\mathcal{L}_\mathrm{f}|+2, \ldots, |\mathcal{L}_\mathrm{f}|+|\mathcal{G}_\mathrm{f}|\}$.

Generator-based Constraints

Base Case:

$$\boldsymbol{P}_\mathrm{min} \le \boldsymbol{E}\boldsymbol{P}_\mathrm{G}^0 \le \boldsymbol{P}_\mathrm{max}, \tag{A3a}$$

$$-\boldsymbol{Q}_\mathrm{min} \le \boldsymbol{E}\boldsymbol{Q}_\mathrm{G}^0 \le \boldsymbol{Q}_\mathrm{max}. \tag{A3b}$$

Line Contingencies:

$$\boldsymbol{E}\boldsymbol{P}_\mathrm{G}^c = \left[\boldsymbol{e}_\mathrm{s}\boldsymbol{P}_\mathrm{G}^c \ \left(\boldsymbol{E}_\mathrm{r}\boldsymbol{P}_\mathrm{G}^0\right)^\top\right]^\top, \tag{A4a}$$

$$\boldsymbol{e}_\mathrm{s}\boldsymbol{P}_\mathrm{min} \le \boldsymbol{e}_\mathrm{s}\boldsymbol{P}_\mathrm{G}^c \le \boldsymbol{e}_\mathrm{s}\boldsymbol{P}_\mathrm{max}, \tag{A4b}$$

$$-\boldsymbol{Q}_\mathrm{min} \le \boldsymbol{E}\boldsymbol{Q}_\mathrm{G}^c \le \boldsymbol{Q}_\mathrm{max}, \tag{A4c}$$

$$(\boldsymbol{E}_\mathrm{r}\bar{\boldsymbol{V}}^c) \odot (\boldsymbol{E}_\mathrm{r}\boldsymbol{V}^c) = (\boldsymbol{E}_\mathrm{r}\bar{\boldsymbol{V}}^0) \odot (\boldsymbol{E}_\mathrm{r}\boldsymbol{V}^0), \tag{A4d}$$

for all $c \in \{1, 2, \ldots, |\mathcal{L}_\mathrm{f}|\}$.

Generator Contingencies:

$$\boldsymbol{E}^c\boldsymbol{P}_{\mathrm{G}}^c = \left[\boldsymbol{e}_{\mathrm{s}}\boldsymbol{P}_{\mathrm{G}}^c \ \left(\boldsymbol{E}_{\mathrm{r}}^c\boldsymbol{P}_{\mathrm{G}}^0\right)^\top\right]^\top, \tag{A5a}$$

$$\boldsymbol{e}_{\mathrm{s}}\boldsymbol{P}_{\min} \leq \boldsymbol{e}_{\mathrm{s}}\boldsymbol{P}_{\mathrm{G}}^c \leq \boldsymbol{e}_{\mathrm{s}}\boldsymbol{P}_{\max}, \tag{A5b}$$

$$-\boldsymbol{Q}_{\min}^c \leq \boldsymbol{E}^c\boldsymbol{Q}_{\mathrm{G}}^c \leq \boldsymbol{Q}_{\max}^c, \tag{A5c}$$

$$\left(\boldsymbol{E}_{\mathrm{r}}^c\bar{\boldsymbol{V}}^c\right)\odot\left(\boldsymbol{E}_{\mathrm{r}}^c\boldsymbol{V}^c\right) = \left(\boldsymbol{E}_{\mathrm{r}}^c\bar{\boldsymbol{V}}^0\right)\odot\left(\boldsymbol{E}_{\mathrm{r}}^c\boldsymbol{V}^0\right), \tag{A5d}$$

$$\text{for all } c \in \{|\mathcal{L}_{\mathrm{f}}|+1, |\mathcal{L}_{\mathrm{f}}|+2, \ldots, |\mathcal{L}_{\mathrm{f}}|+|\mathcal{G}_{\mathrm{f}}|\}.$$

Line Flow Limits

Base Case and Line Contingencies:

$$\left(\bar{\boldsymbol{y}}_{\mathrm{B}}\boldsymbol{A}\bar{\boldsymbol{V}}^c\right)\odot\left(\boldsymbol{y}_{\mathrm{B}}\boldsymbol{A}\boldsymbol{V}^c\right) \leq \boldsymbol{I}_{\max}\odot\boldsymbol{I}_{\max}, \tag{A6a}$$

$$\text{for all } c \in \{0\}\cup\{1,2,\ldots,|\mathcal{L}_{\mathrm{f}}|\}.$$

Generator Contingencies:

$$\left(\bar{\boldsymbol{y}}_{\mathrm{B}}^c\boldsymbol{A}\bar{\boldsymbol{V}}^c\right)\odot\left(\boldsymbol{y}_{\mathrm{B}}^c\boldsymbol{A}\boldsymbol{V}^c\right) \leq \boldsymbol{I}_{\max}\odot\boldsymbol{I}_{\max}, \tag{A6b}$$

$$\text{for all } c \in \{|\mathcal{L}_{\mathrm{f}}|+1, |\mathcal{L}_{\mathrm{f}}|+2, \ldots, |\mathcal{L}_{\mathrm{f}}|+|\mathcal{G}_{\mathrm{f}}|\}.$$

In Eq. (A1), the objective function is the cost of generation schedules where $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c} \in \mathbb{R}^{|\mathcal{G}|}$. The conservation of real and reactive powers at all nodes are formalized in Eq. (A2a) and in Eq. (A2b), respectively, where $\bar{\boldsymbol{V}}^c$ is the conjugate of $\boldsymbol{V}^c$ and '$\odot$' represents element-wise multiplication of matrices or their Hadamard product. In Eq. (A2c), the voltage magnitudes at all nodes are constrained within their rated minimum and maximum values, where $\boldsymbol{V}_{\min}, \boldsymbol{V}_{\max} \in \mathbb{R}_+^{|\mathcal{N}|}$ and '$\leq$' denotes element-wise inequality.

Then, base case real and reactive power generation schedules are constrained within their rated limits in Eq. (A3a) and in Eq. (A3b), respectively, where $\boldsymbol{P}_{\min}, \boldsymbol{P}_{\max}, \boldsymbol{Q}_{\min}, \boldsymbol{Q}_{\max} \in \mathbb{R}_+^{|\mathcal{G}|}$. For post line-contingency state, Eq. (A4a) formalizes real power generation schedules. Notice that the term $\boldsymbol{E}_{\mathrm{r}}\boldsymbol{P}_{\mathrm{G}}^0$ restricts the real power schedules for generators at non-slack nodes, i.e., at nodes with labels in $\mathcal{G}\backslash\{\mathrm{s}\}$, to their base case schedules. The slack generator's post line-contingency real power schedules $\boldsymbol{e}_{\mathrm{s}}\boldsymbol{P}_{\mathrm{G}}^c$ are adjusted to match variations in line losses, which are manifestations of redistributed power flows. These adjusted schedules must conform to the rated limits, as formalized in Eq. (A4b). A generator's reactive power, unlike its real power, can be regulated near instantaneously. So post line-contingency reactive power schedules for all generators in Eq. (A4c) are adjusted to accommodate changes in power flows. In Eq. (A4d), pre and post line-contingency voltage magnitudes for generators at nodes $\mathcal{G}\backslash\{\mathrm{s}\}$ are constrained to be identical. The constraints for post generator-contingency scenarios in Eq. (A5a) - Eq. (A5d) are qualitatively similar to those in Eq. (A4a) - Eq. (A4d), respectively. The matrices $\boldsymbol{E}^c \in \{0,1\}^{(|\mathcal{G}|-1)\times|\mathcal{N}|}$ and $\boldsymbol{E}_{\mathrm{r}}^c \in \{0,1\}^{(|\mathcal{G}|-2)\times|\mathcal{N}|}$ contain subset of rows from $\boldsymbol{I}_{|\mathcal{N}|}$ corresponding to indices of the vertices in sets $\mathcal{G}\backslash\{\mathrm{g}^c\}$ and $\mathcal{G}\backslash\{\mathrm{s},\mathrm{g}^c\}$, respectively, where $\mathrm{g}^c \in \mathcal{G}_{\mathrm{f}}$ is the node for generator contingency c. The vectors $\boldsymbol{Q}_{\min}^c, \boldsymbol{Q}_{\max}^c \in \mathbb{R}_+^{(|\mathcal{G}|-1)}$ for all $c \in \{|\mathcal{L}_{\mathrm{f}}|+1, |\mathcal{L}_{\mathrm{f}}|+2, \ldots, |\mathcal{L}_{\mathrm{f}}|+|\mathcal{G}_{\mathrm{f}}|\}$. Finally, Eq. (A6a) and Eq. (A6b) characterize line flow constraints for pre- and post-contingency states, where $\boldsymbol{I}_{\max} \in \mathbb{R}_+^{|\mathcal{B}|}$. For base case and generator contingencies, $\boldsymbol{y}_{\mathrm{B}} \in \mathbb{C}^{|\mathcal{B}|}$ is constructed with branch admittances as $\boldsymbol{y}_{\mathrm{B}} = \left[y_x^1, y_x^2, \ldots, y_x^{|\mathcal{B}|}\right]^\top$. Lastly, $\boldsymbol{y}_{\mathrm{B}}^c = \boldsymbol{y}_{\mathrm{B}} - \Delta\boldsymbol{y}_{\mathrm{B}}$ where $\Delta\boldsymbol{y}_{\mathrm{B}}$ contains exactly one non-zero element corresponding to failed line with index k, i.e., $\Delta\boldsymbol{y}_{\mathrm{B}}(k,1) = y_x^k$.

Appendix III
Machine Learning Fundamentals

The concepts defined below are classical ones; the reader already familiar with machine learning can safely skip this subsection.

Definition 1 (Homogenous function): A function $f : \mathbb{R}^{n\times1} \to \mathbb{R}$ is a homogenous function of degree α if $f(\lambda\boldsymbol{x}) = \lambda^\alpha f(\boldsymbol{x})$ for all $\lambda > 0$ where $\boldsymbol{x} \in \mathbb{R}^{n\times1}$. ◇

Definition 2 (Monotonic transformation): A monotonic transformation of a function $f : \mathbb{R}^{n\times1} \to \mathbb{R}$ is defined as a composite function $g \circ f : \mathbb{R}^{n\times1} \to \mathbb{R}$ where $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing. ◇

Definition 3 (Homothetic function): A function $h : \mathbb{R}^{n\times1} \to \mathbb{R}$ is a homothetic function if it is a monotonic transformation of a homogenous function as $h = g \circ f$ where $h : \mathbb{R}^{n\times1} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function, and $f : \mathbb{R}^{n\times1} \to \mathbb{R}$ is a homogenous function. ◇

All homogenous functions are homothetic functions. The level set of a function $f : \mathbb{R}^{n\times1} \to \mathbb{R}$ is a set where it takes a constant value, i.e., $l_c(f) = \{\boldsymbol{x}|f(\boldsymbol{x}) = c, c \in \mathbb{R}\}$. If a function $f : \mathbb{R}^{n\times1} \to \mathbb{R}$ is also homogenous, and hence homothetic by default, then level sets of $f : \mathbb{R}^{n\times1} \to \mathbb{R}$ are radial expansions of one another, i.e., if $f(\boldsymbol{x}) = f(\boldsymbol{y})$ then $f(\lambda\boldsymbol{x}) = f(\lambda\boldsymbol{y})$ for $\lambda > 0$ based on *Definition* 1. In other words, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are on same level set, then their positively scaled values $\lambda\boldsymbol{x}$ and $\lambda\boldsymbol{y}$ are on same level set as well. As a consequence, gradients of tangent hyperplanes to level sets along rays from the origin are constant, i.e., $\frac{\partial f(\lambda\boldsymbol{x})}{\partial x_i}/\frac{\partial f(\lambda\boldsymbol{x})}{\partial x_j}$ equals $\frac{\partial f(\boldsymbol{x})}{\partial x_i}/\frac{\partial f(\boldsymbol{x})}{\partial x_j}$ for all i, j and $\lambda > 0$.

Definition 4 (Coefficient of Determination [36]): Given ground truths $y_i \in \mathbb{R}$ and their predicted values $\hat{y}_i \in \mathbb{R}$ for $i \in \{1,2,3,\ldots,N\}$, the coefficient of determination R^2 between them is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2},$$

where $\bar{y} = (\sum_{i=1}^N y_i)/N$. ◇

The R^2 metric is the standard metric to assess regression methods in machine learning studies in a dimension-less way. A value close to 1 gotten on a large enough test sample is an indicator of good generalization to the distribution used to generate that test sample. A value close to 0 reflects the (disappointing) fact that the predictor is only as accurate as the estimation via the best constant model (in the least-squares sense this is the sample mean $\bar{y}$). Significantly negative values indicate an even more pathological situation, which may be due to strong overfitting on the training sample or bad transfer from a learning sample data-generating distribution towards a different test-sample data-generating distribution.

We strongly advocate to use the R^2 metric for robustness assessment of SCOPF proxies. For the sake of comparison, we will nevertheless provide some error metrics expressed in MW/MVAr as often used in the power system literature.

Appendix IV
Detailed results

TABLE A1: R^2_{avg} Scores for Deep Neural Networks

Deep Neural Network: MIMO Regressors			
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{b}}$	**Test Set**: $\mathcal{T}_{\mathrm{c}}$
$\hat{d}_{\mathrm{a}}$	R^2_{avg}=0.973 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.927 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.986 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.957 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.992 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.975 $\boldsymbol{Q}_G$:(22/23)
$\hat{d}_{\mathrm{b}}$	R^2_{avg}=0.969 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.910 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.986 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.957 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.996 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.985 $\boldsymbol{Q}_G$:(22/23)
$\hat{d}_{\mathrm{c}}$	R^2_{avg}=0.828 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.634 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.835 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.624 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.999 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.998 $\boldsymbol{Q}_G$:(23/23)
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{d}}$	**Test Set**: $\mathcal{T}_{\mathrm{e}}$
$\hat{d}_{\mathrm{a}}$	R^2_{avg}=0.973 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.927 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.761 $\boldsymbol{P}_G$:(13/23) R^2_{avg}=0.559 $\boldsymbol{Q}_G$:(14/23)	R^2_{avg}=0.760 $\boldsymbol{P}_G$:(13/23) R^2_{avg}=0.625 $\boldsymbol{Q}_G$:(13/23)
$\hat{d}_{\mathrm{d}}$	R^2_{avg}=0.579 $\boldsymbol{P}_G$:(19/23) R^2_{avg}=0.634 $\boldsymbol{Q}_G$:(16/23)	R^2_{avg}=0.956 $\boldsymbol{P}_G$:(20/23) R^2_{avg}=0.923 $\boldsymbol{Q}_G$:(22/23)	NA
$\hat{d}_{\mathrm{e}}$	R^2_{avg}=0.580 $\boldsymbol{P}_G$:(19/23) R^2_{avg}=0.637 $\boldsymbol{Q}_G$:(16/23)	NA	R^2_{avg}=0.942 $\boldsymbol{P}_G$:(20/23) R^2_{avg}=0.909 $\boldsymbol{Q}_G$:(20/23)
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{f}}$	**Test Set**: $\mathcal{T}_{\mathrm{g}}$
$\hat{d}_{\mathrm{a}}$	R^2_{avg}=0.973 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.927 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.688 $\boldsymbol{P}_G$:(10/23) R^2_{avg}=0.423 $\boldsymbol{Q}_G$:(07/23)	R^2_{avg}=0.909 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.803 $\boldsymbol{Q}_G$:(21/23)
$\hat{d}_{\mathrm{f}}$	R^2_{avg}=0.810 $\boldsymbol{P}_G$:(08/23) R^2_{avg}=0.727 $\boldsymbol{Q}_G$:(03/23)	R^2_{avg}=0.918 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.905 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.822 $\boldsymbol{P}_G$:(08/23) R^2_{avg}=0.491 $\boldsymbol{Q}_G$:(05/23)
$\hat{d}_{\mathrm{g}}$	R^2_{avg}=0.938 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.876 $\boldsymbol{Q}_G$:(20/23)	R^2_{avg}=0.666 $\boldsymbol{P}_G$:(10/23) R^2_{avg}=0.401 $\boldsymbol{Q}_G$:(07/23)	R^2_{avg}=0.978 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.917 $\boldsymbol{Q}_G$:(23/23)
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{h}}$	**Test Set**: $\mathcal{T}_{\mathrm{i}}$
$\hat{d}_{\mathrm{a}}$	R^2_{avg}=0.973 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.927 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.942 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.814 $\boldsymbol{Q}_G$:(19/23)	R^2_{avg}=0.852 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.564 $\boldsymbol{Q}_G$:(19/23)
$\hat{d}_{\mathrm{h}}$	R^2_{avg}=0.937 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.837 $\boldsymbol{Q}_G$:(20/23)	R^2_{avg}=0.986 $\boldsymbol{P}_G$:(21/23) R^2_{avg}=0.949 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.896 $\boldsymbol{P}_G$:(14/23) R^2_{avg}=0.687 $\boldsymbol{Q}_G$:(17/23)
$\hat{d}_{\mathrm{i}}$	R^2_{avg}=0.872 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.654 $\boldsymbol{Q}_G$:(19/23)	R^2_{avg}=0.927 $\boldsymbol{P}_G$:(14/23) R^2_{avg}=0.715 $\boldsymbol{Q}_G$:(16/23)	R^2_{avg}=0.982 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.942 $\boldsymbol{Q}_G$:(22/23)

TABLE A2: R^2_{avg} Scores for Random Forests

Extremely Randomized Trees: MIMO Regressors			
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{b}}$	**Test Set**: $\mathcal{T}_{\mathrm{c}}$
$\hat{e}_{\mathrm{a}}$	R^2_{avg}=0.878 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.852 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.882 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.818 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.989 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.935 $\boldsymbol{Q}_G$:(22/23)
$\hat{e}_{\mathrm{b}}$	R^2_{avg}=0.884 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.855 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.893 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.821 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.987 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.941 $\boldsymbol{Q}_G$:(22/23)
$\hat{e}_{\mathrm{c}}$	R^2_{avg}=0.819 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.654 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.826 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.655 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=1.0 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=1.0 $\boldsymbol{Q}_G$:(23/23)
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{d}}$	**Test Set**: $\mathcal{T}_{\mathrm{e}}$
$\hat{e}_{\mathrm{a}}$	R^2_{avg}=0.878 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.852 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.666 $\boldsymbol{P}_G$:(14/23) R^2_{avg}=0.543 $\boldsymbol{Q}_G$:(16/23)	R^2_{avg}=0.635 $\boldsymbol{P}_G$:(15/23) R^2_{avg}=0.594 $\boldsymbol{Q}_G$:(15/23)
$\hat{e}_{\mathrm{d}}$	R^2_{avg}=0.536 $\boldsymbol{P}_G$:(20/23) R^2_{avg}=0.605 $\boldsymbol{Q}_G$:(17/23)	R^2_{avg}=0.845 $\boldsymbol{P}_G$:(21/23) R^2_{avg}=0.743 $\boldsymbol{Q}_G$:(23/23)	NA
$\hat{e}_{\mathrm{e}}$	R^2_{avg}=0.542 $\boldsymbol{P}_G$:(20/23) R^2_{avg}=0.617 $\boldsymbol{Q}_G$:(17/23)	NA	R^2_{avg}=0.805 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.770 $\boldsymbol{Q}_G$:(23/23)
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{f}}$	**Test Set**: $\mathcal{T}_{\mathrm{g}}$
$\hat{e}_{\mathrm{a}}$	R^2_{avg}=0.878 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.852 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.613 $\boldsymbol{P}_G$:(10/23) R^2_{avg}=0.521 $\boldsymbol{Q}_G$:(07/23)	R^2_{avg}=0.881 $\boldsymbol{P}_G$:(22/23) R^2_{avg}=0.748 $\boldsymbol{Q}_G$:(21/23)
$\hat{e}_{\mathrm{f}}$	R^2_{avg}=0.638 $\boldsymbol{P}_G$:(10/23) R^2_{avg}=0.519 $\boldsymbol{Q}_G$:(08/23)	R^2_{avg}=0.744 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.721 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.634 $\boldsymbol{P}_G$:(10/23) R^2_{avg}=0.533 $\boldsymbol{Q}_G$:(08/23)
$\hat{e}_{\mathrm{g}}$	R^2_{avg}=0.842 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.780 $\boldsymbol{Q}_G$:(20/23)	R^2_{avg}=0.609 $\boldsymbol{P}_G$:(10/23) R^2_{avg}=0.520 $\boldsymbol{Q}_G$:(07/23)	R^2_{avg}=0.877 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.804 $\boldsymbol{Q}_G$:(23/23)
Model	**Test Set**: $\mathcal{T}_{\mathrm{a}}$	**Test Set**: $\mathcal{T}_{\mathrm{h}}$	**Test Set**: $\mathcal{T}_{\mathrm{i}}$
$\hat{e}_{\mathrm{a}}$	R^2_{avg}=0.878 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.852 $\boldsymbol{Q}_G$:(22/23)	R^2_{avg}=0.907 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.707 $\boldsymbol{Q}_G$:(20/23)	R^2_{avg}=0.812 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.600 $\boldsymbol{Q}_G$:(19/23)
$\hat{e}_{\mathrm{h}}$	R^2_{avg}=0.833 $\boldsymbol{P}_G$:(17/23) R^2_{avg}=0.713 $\boldsymbol{Q}_G$:(20/23)	R^2_{avg}=0.935 $\boldsymbol{P}_G$:(20/23) R^2_{avg}=0.813 $\boldsymbol{Q}_G$:(23/23)	R^2_{avg}=0.833 $\boldsymbol{P}_G$:(13/23) R^2_{avg}=0.613 $\boldsymbol{Q}_G$:(16/23)
$\hat{e}_{\mathrm{i}}$	R^2_{avg}=0.826 $\boldsymbol{P}_G$:(16/23) R^2_{avg}=0.594 $\boldsymbol{Q}_G$:(19/23)	R^2_{avg}=0.872 $\boldsymbol{P}_G$:(13/23) R^2_{avg}=0.629 $\boldsymbol{Q}_G$:(16/23)	R^2_{avg}=0.910 $\boldsymbol{P}_G$:(23/23) R^2_{avg}=0.794 $\boldsymbol{Q}_G$:(23/23)

References

[1] F. Capitanescu, J. Martinez Ramos, P. Panciatici, D. Kirschen, A. Marano Marcolini, L. Platbrood, and L. Wehenkel, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731–1741, 2011.

[2] A. Velloso and P. Van Hentenryck, "Combining Deep Learning and Optimization for Preventive Security-Constrained DC Optimal Power Flow," *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3618–3628, 2021.

[3] W. Chen, S. Park, M. Tanneau, and P. Van Hentenryck, "Learning optimization proxies for large-scale Security-Constrained Economic Dispatch," *Electric Power Systems Research*, vol. 213, p. 108566, 2022.

[4] S. Liu, Y. Guo, W. Tang, H. Sun, W. Huang, and J. Hou, "Varying Condition SCOPF Optimization Based on Deep Learning and Knowledge Graph," *IEEE Transactions on Power Systems*, pp. 1–12, 2022.

[5] X. Pan, T. Zhao, and M. Chen, "DeepOPF: Deep Neural Network for DC Optimal Power Flow," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2019, pp. 1–6.

[6] K. Baker, "Learning Warm-Start Points For Ac Optimal Power Flow," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6.

[7] S. Misra, L. Roald, and Y. Ng, "Learning for Constrained Optimization: Identifying Optimal Active Constraint Sets," *INFORMS Journal on Computing*, vol. 34, pp. 463–480, 2022.

[8] M. Zhou, M. Chen, and S. H. Low, "DeepOPF-FT: One Deep Neural Network for Multiple AC-OPF Problems with Flexible Topology," *IEEE Transactions on Power Systems*, pp. 1–4, 2022.

[9] L. Chen and J. E. Tate, "Hot-Starting the AC Power Flow with Convolutional Neural Networks." arXiv, 2020.

[10] W. Dong, Z. Xie, G. Kestor, and D. Li, "Smart-PGSim: Using Neural Network to Accelerate AC-OPF Power Grid Simulation," in *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, 2020, pp. 1–15.

[11] Z. Yan and Y. Xu, "A Hybrid Data-Driven Method for Fast Solution of Security-Constrained Optimal Power Flow," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4365–4374, 2022.

[12] A. Robson, M. Jamei, C. Ududec, and L. Mones, "Learning an Optimally Reduced Formulation of OPF through Meta-optimization." arXiv, 2019.

[13] M. Chatzos, T. W. K. Mak, and P. V. Hentenryck, "Spatial Network Decomposition for Fast and Scalable AC-OPF Learning," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2601–2612, 2022.

[14] N. Guha, Z. Wang, M. Wytock, and A. Majumdar, "Machine Learning for AC Optimal Power Flow," 2019.

[15] R. Nellikkath and S. Chatzivasileiadis, "Physics-Informed Neural Networks for AC Optimal Power Flow," *Electric Power Systems Research*, vol. 212, p. 108412, 2022.

[16] Z. Yan and Y. Xu, "Real-Time Optimal Power Flow: A Lagrangian Based Deep Reinforcement Learning Approach," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3270–3273, 2020.

[17] K. Baker, "Emulating AC OPF Solvers With Neural Networks," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4950–4953, 2022.

[18] F. Capitanescu, "Suppressing ineffective control actions in optimal power flow problems," *IET Generation, Transmission & Distribution*, vol. 14, no. 13, pp. 2520–2527, 2020.

[19] T. Van Cutsem *et al.*, "Test Systems for Voltage Stability Studies," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 4078–4087, 2020.

[20] E. Davoodi, F. Capitanescu, and L. Wehenkel, "A methodology to evaluate reactive power reserves scarcity during the energy transition," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4355–4368, 2023.

[21] G. Louppe, "Lectures for INFO8010 - Deep Learning, Université de Liège," 2022, accessed: Spring 2022.

[22] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press.

[23] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[25] L. Cayton, "Algorithms for manifold learning." University of California - San Diego, 2005, accessed: December 2022.

[26] C. Fefferman, S. Mitter, and H. Narayanan, "Testing the Manifold Hypothesis," *Journal of the American Mathematical Society*, vol. 29, no. 4, pp. 983–1049, 2016.

[27] M. Allen, D. Poggiali, K. Whitaker, T. Marshall, J. van Langen, and R. Kievit, "Raincloud plots: a multi-platform tool for robust data visualization," *Wellcome Open Research*, vol. 4, no. 63, 2021.

[28] "RainCloud Plot Script," 2020.

[29] F. Capitanescu, "Evaluating reactive power reserves scarcity during the energy transition toward 100% renewable supply," *Electric Power Systems Research*, vol. 190, p. 106672, 2021.

[30] N. Popli, E. Davoodi, F. Capitanescu, and L. Wehenkel, "Machine learning based binding contingency pre-selection for AC-PSCOPF calculations," *IEEE Transactions on Power Systems*, pp. 1–4, doi: 10.1109/TPWRS.2023.3338971, 2023.

[31] B. Donon, B. Donnot, I. Guyon, and A. Marot, "Graph Neural Solver for Power Systems," in *International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8.

[32] B. Donon, "Deep statistical solvers & power systems applications," PhD Dissertation, Université Paris-Saclay, 2022.

[33] S. Müller, "Development of Nordic 32 system model and performance analysis based on real operation statistics," *KTH, School of Electrical Engineering and Computer Science (EECS)*, 2019.

[34] O. Alsac and B. Stott, "Optimal Load Flow with Steady-State Security," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 3, pp. 745–751, 1974.

[35] A. Monticelli, M. V. F. Pereira, and S. Granville, "Security-Constrained Optimal Power Flow with Post-Contingency Corrective Rescheduling," *IEEE Transactions on Power Systems*, vol. 2, no. 1, pp. 175–180, 1987.

[36] T. O. Kvalseth, "Cautionary Note about R2," *The American Statistician*, vol. 39, no. 4, pp. 279–285, 1985.