Graph Convolutional Networks for probabilistic power system operational planning

Yasmin Bashir Sheikh-Mohamed
Sigurd Hofsmo Jakobsen
Espen Flo Bødal
Fredrik Marinius Haugseth
Erlend Sandø Kiel
SINTEF Energy Research
Trondheim, Norway

Signe Riemer-Sørensen
SINTEF Digital
Oslo, Norway

***Abstract*—Probabilistic operational planning of power systems usually requires computationally intensive and time consuming simulations. The method presented in this paper provides a time efficient alternative to predict the socio-economic cost of system operational strategies using graph convolutional networks. It is intended for fast screening of operational strategies for the purpose of operational planning. It can also be used as a proxy for operational planning that can be used in long term development studies. The performance of the model is demonstrated on a network inspired by the Nordic power system.**

***Index Terms*—probabilistic operational planning, power system reliability, contingency analysis, machine learning, graph neural networks**

I. Introduction

A. Motivation

Most power systems today are operated and planned according to deterministic criteria, which are often socio-economical sub-optimal. In order to keep adequate reliability of supply while also minimizing the socio-economic costs, moving towards probabilistic criteria is recommended by the European FP7 project, GARPUR [1]. ACER, the European Union's Agency for the Cooperation of Energy Regulators, adopted a decision for transmission system operators (TSOs) to develop a methodology on probabilistic risk assessment [2]. The move towards probabilistic criteria requires efficient and accurate methods for decision support. This study explores the possibility of using machine learning for modeling decision support according to probabilistic criteria in system operational planning. The idea is that the model can be used as a screening method for operational planning, or for including a more accurate and fast representation of operations in long-term planning studies. There is a multitude of considerations when implementing probabilistic operational planning. The method should not only consider the risk of violating physical constraints but also the cost of operation [3]. Methods that attempt to minimize the cost of power system operation are therefore considered in this work.

The research leading to these results has received funding from the Research Council of Norway through the project "Resilient and Probabilistic reliability management of the transmission grid" (RaPid) (Grant No. 294754), The Norwegian Water Resources and Energy Directorate, and Statnett.

A dynamic programming model was developed in [4] to quantify the socio-economic cost and detect the most favorable socio-economic operational strategy. Dynamic time-domain simulations were used to capture situations that normally go undetected by traditional static methods. However, this approach is computationally intensive. It is also necessary to predict consequences for a large sample space of possible contingencies, available corrective actions and other uncertainties, as steps toward identifying the optimal operational strategy.

As a response to these challenges, this paper proposes a two-step supervised learning model based on graph convolutional networks (GCNs) to rapidly predict the expected costs of simulated operational strategies, exemplified using data from [4].

B. Related works

Alternative methods to [4] for probabilistic operational planning have been presented in the literature [5]–[9]. In [5] a DC power flow was used to include power system response in a probabilistic operational planning model. Different cost-based criteria are compared in [6], where a transport model was used to model the power system response. An AC power flow and a linear approximation of frequency response were used in [7] to include frequency response in an operational planning model. More recent approaches use machine learning for generating proxy models of real-time operation to speed up probabilistic operational planning [8], [9]. In these papers, a machine learning model is trained to act as a DC-security constrained optimal power flow (SCOPF) and to predict the optimal corrective actions given a set of preventive actions. This is a promising approach, however, the use of a DC-SCOPF means that voltage, frequency and stability issues will not be captured. Moreover, the time domain characteristic of protection systems cannot be included. In the proposed approach, a GCN is trained to predict the result of a detailed time-domain simulation that calculates the cost of operating a power system given a set of preventive and corrective actions while considering the time-domain characteristics of protection systems [4].

Graph neural networks (GNNs) is a collective term describing neural networks that process data structured as graphs. Use

Author Accepted Manuscript version of the paper by Yasmin Bashir Sheikh-Mohamed et al.
In 2023 IEEE Belgrade PowerTech (2023), DOI: http://dx.doi.org/10.1109/PowerTech55446.2023.10202799

of GNNs in power system applications has recently gained popularity [10]. The graph structure allows the model to capture specific entity attributes and their relations. This is useful for representing and enforcing power system topology in deep learning models. In [11] and [12] GNNs are used to model power flows, predicting optimal power flow and load shedding, respectively. Other applications include transformer fault diagnosis [13], weather related power outages [14] and complex energy systems [15]. Multiple studies show that GNNs outperforms other machine learning models for graph-structured data [11]–[13].

C. Structure of paper

The probabilistic operational planning approach applied in this paper and some fundamentals of GCNs is explained in Section II. The data generation method, model training and how to do the predictions is presented in Section III. Results from a case study are given in Section IV, and conclusions and further work are given in Section V.

II. Theory and definitions

A. Probabilistic Operational Planning

Probabilistic operational planning often aim to find a strategy that minimizes expected operational costs given numerous fault scenarios, their probabilities, and preventive and corrective measures. This is different from traditional power system operational schemes that are mainly based on the N-1 criterion. The criterion states that the system should be able to withstand any credible contingency at all times in such a way that the system is capable of accommodating the new operational situation without violating operational security limits [16].

One way of doing this is through the use of preventive and corrective measures during operations. Preventive and corrective measures are defined in this paper as an adaptation of the understanding of remedial actions and preventive remedial actions applied to a deterministic reliability criterion found in [17]: A *preventive measure* is understood as *an action taken to ensure that the system will adhere to a reliability criterion in the future, based on a predicted future system state and threat exposure*. Examples of preventive measures are the allocation of reserves, transmission switching or reducing transmission capacities between market areas. A *preventive strategy* in this work is subsequently understood as *a set of measures taken prior to the operational hour that change the system operating state such that the system will adhere to a reliability criterion subject to a contingency*. Similarly, a *corrective measure* is understood as *an action to ensure the system's compliance with a reliability criterion based on an observed and estimated present system state and threat exposure*.

The main objective of the method presented in this paper is to identify the preventive operational strategy with the lowest expected socio-economic cost according to the framework presented in [4]. The socio-economic costs are given by:

$$\Pi = \min_{p \in \mathcal{P}} S(p \mid \mathcal{C},\ \mathcal{U}). \tag{1}$$

Where $S(p \mid \mathcal{C},\ \mathcal{U})$ is the system operation cost for a preventive strategy p, given a set of uncertainties, $\mathcal{U}$ and corrective actions, $\mathcal{C}$. The set of uncertainties can include variation in load, generation price, contingency, etc. Uncertainties related to contingencies based on line outages is the focus of this paper. The cost of an operational strategy is given by:

$$S(p \mid \mathcal{C},\mathcal{U}) = D(p,\mathcal{U}) + \sum_{u \in \mathcal{U}} \rho_u \min_{c \in \mathcal{C}} C(p,u,c). \tag{2}$$

Where D is the dispatch cost and C is the corrective operation cost for each realization of the uncertainty u. The expected operational cost is found by using the probability ρ_u for the realization of the uncertainties and assuming that the least-cost corrective measure is chosen for each of these realizations.

The operational cost includes system protection schemes (SPS) failures by weighing the cost of the system response with and without successful SPS activation using the probability of successful SPS γ_c as shown in (3), where no corrective action is indicated by c^0, and a contingency, f, is the realization of the uncertainty:

$$C(p,f,c) = \gamma_c R(p,f,c) + (1-\gamma_c) R(p,f,c^0). \tag{3}$$

$$R(p,f,c) = FCR(p,f,c) + SPS(c) + CENS(p,f,c) + BAL(p,f,c). \tag{4}$$

The system response cost includes the cost of frequency containment reserves (FCR), SPS activation, cost of energy not supplied (CENS) and balancing *BAL* (4). Here, FCR and SPS are obtained from time-domain simulations, while CENS and *BAL* are obtained using two variations of static DC Optimal Power Flow (OPF) models as explained in detail in [4].

B. Graph Convolutional Neural Networks

Data is often represented in the form of graphs which conveys relations between entities, from social networks to power systems. Neural networks that exploit and adapt to these relations have increased in popularity, and generalizations of methods such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to the non-Euclidean space have seen rapid advancement [18]. A GCN is a type of GNN which generalize the convolutional operation in a CNNs to work on graph data [19]. Thus a GCN benefit from the convolutional operation in a CNN applied to data such that relations are preserved, e.g. to incorporate information about the relationship between units in a power system topology.

Entities and relations are in this paper referred to as nodes and edges in a graph, respectively. In the Graph Network (GN) framework developed in [20] a graph, G, consists of global attributes, ϑ, and a set of nodes, $V = \{v_1, v_2, ..., v_n\}$, which are connected by edges, $E = \{e_1, e_2, ..., e_m\}$, as illustrated in Fig. 1. An edge is represented as a connection between a pair of nodes, $e = \{v_i, v_j\}$. If multiple edges are assigned to the same nodes, the graph is referred to as a multi-graph, and contains a mapping $\phi : E \rightarrow \{\{i,j\} : i,j \in V\}$, as shown in (5).

$$G = \{\vartheta, V, E, \phi\}. \tag{5}$$

Author Accepted Manuscript version of the paper by Yasmin Bashir Sheikh-Mohamed et al.
In 2023 IEEE Belgrade PowerTech (2023), DOI: http://dx.doi.org/10.1109/PowerTech55446.2023.10202799

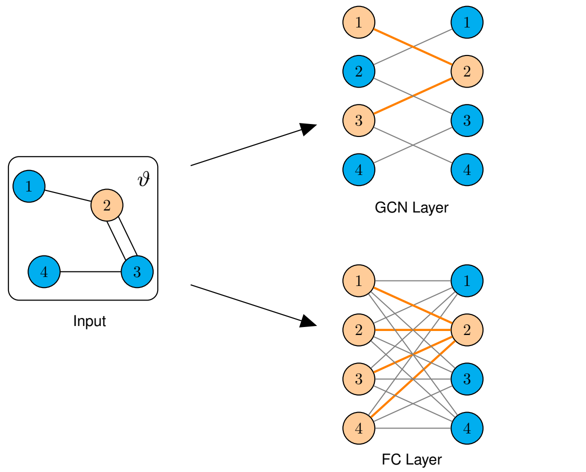

Fig. 1. Left: Illustration of a graph input (left) with nodes, V, connecting edges, E, subject to a global attribute, ϑ. Right: Demonstration of propagation in a graph convolutional network (GCN) layer (upper) and a fully connected (FC) layer (lower).

The foundation of GCN is that nodes in a new layer are represented by aggregating information from neighboring nodes in the previous layer. This differs from fully connected networks with all-to-all relations where each node gathers information from all nodes in the previous layer, as illustrated in Fig. 1. The restriction in information gathering per convolutional layer isolates the most relevant information for the task [20].

III. Method

The main purpose of the method proposed is to use machine learning to identify near socio-economic optimal preventive strategies. Two GCNs models are trained to achieve this. A classification method is trained to identify operational strategies with expected low costs, to separate out operational strategies which have very high costs from those with more moderate costs. A regression model is then trained to predict the dispatch and operation costs for the low cost strategies. The operation cost for high cost scenarios is set to a fixed representative value since it is more important to know the value of the low cost scenarios accurately. This two-step approach avoids that the regression model used to predict the dispatch and operation costs must take into account the more extreme values found in the prohibitively high costs outcomes, and is expected to yield more accurate estimates for the low cost scenarios.

A. Data generation

The power system used in this paper is depicted in Fig. 2. The training data is generated by running the dynamic model described in [4] on a power system based on [21], which is inspired by three market areas in the Nordic power system[1]. The preventive strategies, $\mathcal{P}$, consists of 1401 transmission capacities between the market areas, the set $\mathcal{F}$ consists of single line contingencies of the lines between the market areas, and generators and loads depicted are included as possible SPS measures.

[1]A fourth area is connected by a single HVDC link. This is considered part of one of the other areas in this study.

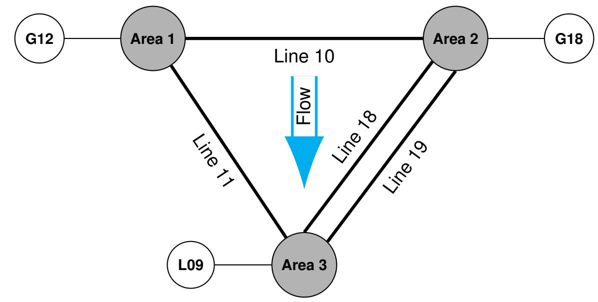

Fig. 2. Simple representation of the three area network. The figure illustrates the three market areas, inter-line connections, and the three generators and loads that are disconnected as an SPS measure in this study. Figure from [4].

Data was generated with the following procedure:

- Choose preventive strategy, $p \in \mathcal{P}$.
- Solve OPF based on preventive strategy, p.
- Simulate selected contingencies, $f \in \mathcal{F}$, using time-domain simulations.
- If the system response does not violate the set conditions:
  1) Report system operational costs.
- If the system response violates the set conditions:
  1) Simulate each SPS measure in $\mathcal{C}$ using time-domain simulation, and report system operation costs.

Each scenario, with a given preventive strategy, contingency, and corrective action, is an instance represented by a graph, (5), where V contains area-specific attributes, E contains information on the inter-area lines, and $\mathbf{u}$ contains the system operational cost for the scenario. All costs are given in per-unit (p.u.) values which are relative to the OPF solution without preventive measures.

Our dataset consists in per-unit (p.u.) values which are of 7953 scenarios, with scenarios based on 1401 preventive strategies, 3 corrective actions, and 4 possible line contingencies.

B. Model predictions

The aim is to predict the total system operation cost given a preventive strategy, $p \in \mathcal{P}$, a contingency, $f \in \mathcal{F}$, and corrective action $c \in \mathcal{C}$ as seen in (2). This is done by using the machine learning model to estimate the predictors $\hat{D}$ and $\hat{C}$ for $D(p, \mathcal{U})$ and $C(p, u, c)$ respectively.

Only the initial OPF solution is required for prediction with the GCN model. The idea is for the model to replace the time-domain simulations during screening so that only the initial OPF is required at prediction time.

C. Model set up and training

The model is trained on data from the time-domain simulation but there is no guarantee that it will adhere to all physical constraints captured by the time-domain simulations.

TABLE I
NODE ATTRIBUTES

Attribute	Description
P_i^G	Production in area i prior to contingency.
P_{sys}	Inter area flow for the whole system, prior to contingency.
$isSPS_i$	1 if SPS measure activated in area. Else, 0.
$isF_{1,i}$	1 if fault in line 10, 11 or 18 and area connected to line. Else, 0.
$isF_{2,i}$	1 if fault in line 19 and area connected to line. Else, 0.

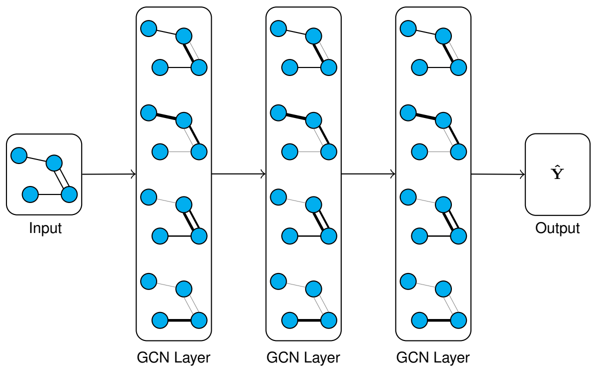

Fig. 3. Network architecture for both classification and regression.

The power system is represented by graphs with $X \in \mathbb{R}^{mxn}$ containing node features $X = (X_1, X_2, ..., X_n)$ as listed in Table I. One-hot encodings are used to model SPS measures and line contingencies. Two features are added to differentiate between faults on Line 18 and Line 19, since they are connected to the same nodes.

The edge weights are set based on the undirected power flow on the relevant line prior to contingency:

$$\phi_{ijk} = \begin{cases} p_{ijk}, & (i,j,k) \in \mathbb{L} \\ 0, & otherwise \end{cases} \tag{6}$$

where ϕ_{ijk} is the k*th* line between area i and j. $\mathbb{L}$ is the set of lines in the system. The adjacency matrix is given by:

$$A_{ij} = \begin{cases} \sum_k \phi_{ijk}, & (i,j) \in \mathbb{L} \\ 0, & otherwise \end{cases} \tag{7}$$

The propagation rule described in [22] is used in this study, which can be expressed in a simplified form without self-connections as:

$$H^{l+1} = ReLU(\tilde{D}^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)}). \tag{8}$$

X is the node feature matrix, $H^0 = X$, A is the adjacency matrix, l is the layer number, D is the diagonal matrix, where $D_{ii} = \sum_j A_{ij}$. The standard Rectified Linear Unit activation function, $ReLU(x) = max(0, x)$, is applied. The same architecture is applied for both the classification and regression models which are shown in Fig. 3 with three hidden layers each applying the convolution in (8) followed by a linear transformation for the regression model and a Sigmoid transformation for the classification. Hyperparameters for the classification and regression models is given in Table II.

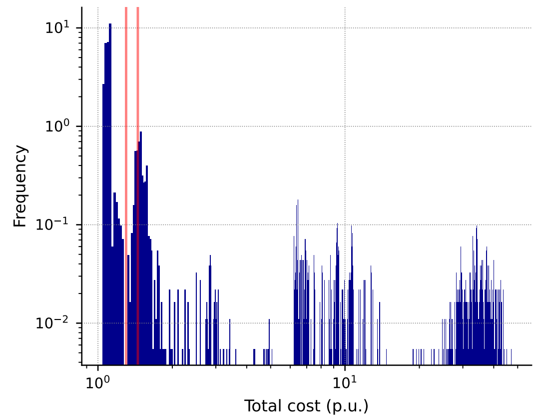

Fig. 4. Histogram of the cost distribution in the data. All costs are given relative to the OPF solution without preventive measures. The vertical lines indicate the low-high cost threshold, $thr = 1.3$, and the maximum cost value at which the regression data are fitted, $S_{\max} = 1.45$.

1) Classification model: The classification model is trained with a binary logistic loss function. The low and high cost classes are defined by a chosen threshold cost value, thr. There is a gap in the simulated costs used in the training process seen in Fig. 4 around 1.3 p.u. which is used to distinguish between classes. This split leads to imbalanced data, as the majority of scenarios are categorized as low cost. To prevent biasing the model, class weights are introduced. This increases the penalty for misclassification of the minority class. The class weights are set based on the ratio of observations in each class.

Classification models predict the probability of each instance belonging to a class. For two-class models, it is standard to use a threshold of 0.5 between the classes. However, in many cases, better performance can be obtained by adjusting the threshold used for predictions after training. This may help to unbias a model in addition to training with class weights. The post-training probability threshold is chosen such that it maximizes the F2-score, giving more weight to the recall [23]:

$$F_2 = \frac{TP}{TP + 0.2FP + 0.8FN} \tag{9}$$

where TP, FP, and FN are the true positive, false positive, and false negative predictions respectively.

2) Regression model: The regression model has a dual output and predicts both dispatch and correction costs individually. The regression model is used to predict the cost of scenarios already classified as low cost and is only trained on scenarios with a true cost lower than $S_{\max}$. However, some scenarios may be mis-classified, and hence based on Fig. 4 scenarios with a total cost up to $S_{\max} = 1.45$ is included. Apart from this, the training data are the same as for the classification model.

3) Training: The data is split into training (70%), validation (10%), and test samples (20%). All scenarios for a given preventive strategy are needed to compute the expected cost in (2), thus the data is split into preventive strategies rather than individual scenarios. Hence the test set contains all possible

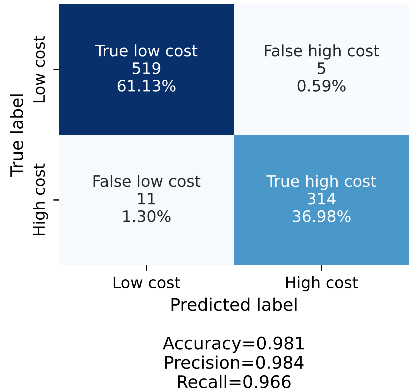

Fig. 5. The confusion matrix for the classification model.

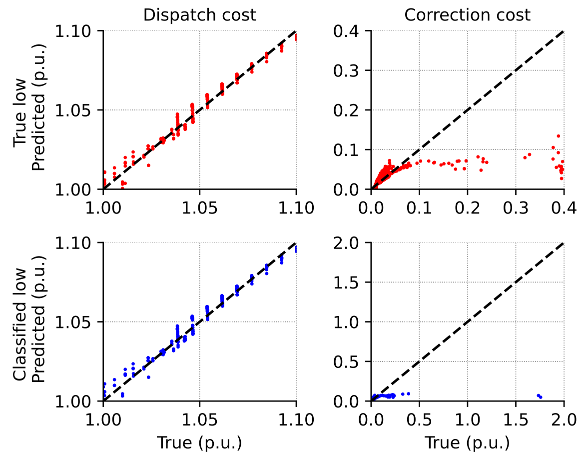

Fig. 6. The predicted versus true values for the regression model on the training dataset (upper panels) and test data-set (lower panels). Extreme outliers are removed but show the same trend.

scenarios for the included preventive strategies. The same split is used for both models but the regression model is only trained and evaluated on scenarios with a total cost less than $S_{\max}$.

Early stopping, where the training is stopped if the validation performance does not increase over 10 training epochs, is applied to prevent overfitting. Additionally, learning rate decay is set to 95%, which reduces the learning rate on the fly during training. Other hyper-parameters include initial learning rate, hidden channels, and dropout rate. These parameters are tuned with a search over 50 randomly sampled combinations, in order to find optimal parameters for each model. Table II provides a description of all model parameters as well as the chosen values for the two models.

IV. Results

A. Classification model

For the trained model, the post-training classification probability leading to the maximal F_2 score is $\gamma = 0.56$. With this threshold, the classification model obtains an accuracy of 98.1% on the test data and the confusion matrix given in Fig. 5. It can be seen that the accuracy, precision, and recall are high.

B. Regression model

Fig. 6 shows the true and predicted output values for the regression model on the test data. Evaluated on the true low cost scenarios, the normalized root mean square error (RMSE) for the model is 0.00015 for the dispatch cost, and 0.00336 for the corrective cost. The model has best performance on the dispatch cost, whereas the corrective cost is accurately predicted for low cost scenarios, but under-estimated for scenarios with high corrective costs. The corrective cost for scenarios classified as high cost is replaced with the median of corrective costs for the true high cost scenarios. This value will in many cases over-estimate the corrective cost, but only for scenarios already classified as less interesting.

By using (2) the expected cost of preventive strategies is calculated using output from the regression model (input with the median corrective cost values for the high cost scenarios). In Fig. 7 the total costs is plotted against the power flow on Line 10 and Line 11, with $\rho_f = 0.01$. The true values are labeled as N-1 safe or not N-1 safe based on whether the contingencies lead to interruptions and associated CENS. There is a reduction of expected cost with increasing power flow on the line up until 60 MW on Line 10 and 150 MW on Line 11. Above those values the risk of energy not supplied increases, leading to some scenarios and preventive strategies being classified as high costs with consequent high predictions of total system operation cost.

The model shows the ability to identify the range of preventive strategies close to the least-cost strategy shown in Fig. 7, as well as provide information on which contingencies contribute to the highest costs.

V. Conclusion and Further Work

A graph convolutional network (GCN) based method for power system operational planning is implemented in this work. The methodology consists of one GCN classifier model for labeling low and high cost scenarios and a GCN regression model for predicting the cost of low cost scenarios. Training data for the models are obtained from a dynamic programming model with time-domain simulations that capture situations that normally go undetected by traditional static methods. The model is used in an illustrative case study on a test network inspired by the Nordic power system.

High cost scenarios affect the regression model's ability to detect low cost scenarios. Therefore, the classification model is first used to filter out high cost scenarios before the regression model predicts the dispatch and corrective cost for each scenario. While the predictions are not perfect, the framework can be applied for the identification of near-optimal preventive strategies and thereby significantly reduce the sample space where one has to run time-domain simulations. Reducing the amount of time-domain simulations is important to make that type of detailed simulation applicable in practical operations. The effect of additional load scenarios with more variation should be explored in further work.

TABLE II
MODEL PARAMETERS

Parameter	Description	Classification	Regression
`Layers`	Number hidden layers	3	3
`Hidden channels`	Number of neurons in each hidden layer	29	41
`Loss`	Loss function for gradient descent	`BCEWithLogits`	`MSE`
`Epoch`	Number of times a model is trained on a data set	80	100
`Learning rate`	Initial learning rate determining the step size in updates	0.0009	0.0035
`Learning rate decay`	Rate at which learning rate decreases per epoch	0.95	0.95
`Dropout rate`	Rate that decides number of neurons randomly nullified in each training epoch	0.1	0.25

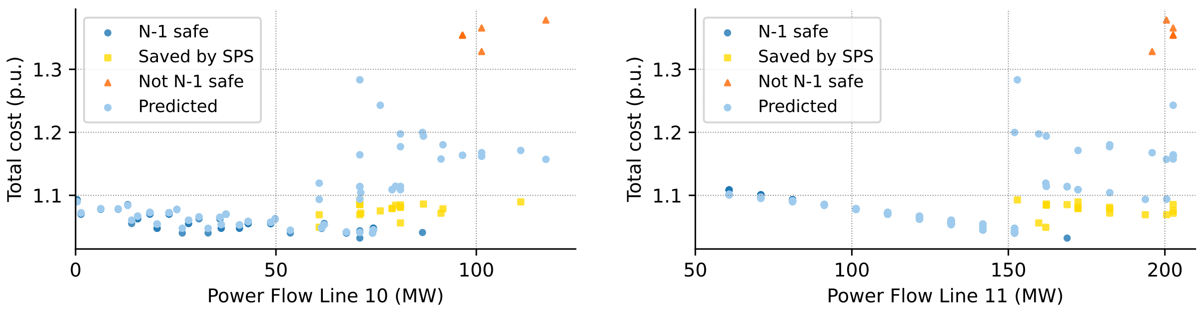

Fig. 7. The total system operation cost against power flow on Line 10 (left) and Line 11 (right) for $\rho_f = 0.01$.

ACKNOWLEDGMENT

The authors would like to thank the RaPid project participants, Gerd Kjølle and Oddbjørn Gjerde, for their valuable input and discussion of the work.

REFERENCES

[1] GARPUR Consortium, “D11.1a: Results and recommendations towards stakeholders,” 2017. Available at https://www.sintef.no/globalassets/project/garpur/deliverables/.
[2] Agency for the Cooperation of Energy Regulators, “Decision no 07/2019 of the agency for the cooperation of energy regulators,” tech. rep., Agency for the Cooperation of Energy Regulators, 06 2019.
[3] D. S. Kirschen and D. Jayaweera, “Comparison of risk-based and deterministic security assessments,” *IET Generation, Transmission and Distribution*, vol. 1, no. 4, pp. 527–533, 2007.
[4] E. F. Bødal, S. H. Jakobsen, and O. Gjerde, “Probabilistic operational planning using dynamic programming with time-domain simulations,” *Electric Power Systems Research*, vol. 211, p. 108379, Oct. 2022.
[5] E. Karangelos and L. Wehenkel, “Probabilistic reliability management approach and criteria for power system short-term operational planning,” in *Bulk Power Systems Dynamics and Control-X The Power System of the Future: Global Dynamics arising from Distributed Actions (IREP), 2017 IREP Symposium*, 2017.
[6] G. Warland, A. T. Holen, G. Solem, K. Uhlen, and I. Gimmestad, “Decision support for network operation in an open power market,” *European transactions on electrical power*, vol. 17, no. 4, pp. 319–332, 2007.
[7] M. Negnevitsky, D. H. Nguyen, and M. Piekutowski, “Risk assessment for power system operation planning with high wind power penetration,” *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1359–1368, 2015.
[8] L. Duchesne, E. Karangelos, and L. Wehenkel, “Using machine learning to enable probabilistic reliability assessment in operation planning,” in *2018 Power Systems Computation Conference (PSCC)*, pp. 1–8, 2018.
[9] L. Duchesne, E. Karangelos, A. Sutera, and L. Wehenkel, “Machine learning for ranking day-ahead decisions in the context of short-term operation planning,” *Electric Power Systems Research*, vol. 189, p. 106548, 2020. Publisher: Elsevier Ltd.
[10] W. Liao, B. Bak-Jensen, J. Radhakrishna Pillai, Y. Wang, and Y. Wang, “A Review of Graph Neural Networks and Their Applications in Power Systems,” *arXiv e-prints*, p. arXiv:2101.10025, Jan. 2021.
[11] D. Owerko, F. Gama, and A. Ribeiro, “Optimal power flow using graph neural networks,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5930–5934, IEEE, 2020.
[12] C. Kim, K. Kim, P. Balaprakash, and M. Anitescu, “Graph convolutional neural networks for optimal load shedding under line contingency,” in *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2019.
[13] W. Liao, D. Yang, Y. Wang, and X. Ren, “Fault diagnosis of power transformers using graph convolutional network,” *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 241–249, 2020.
[14] D. Owerko, F. Gama, and A. Ribeiro, “Predicting power outages using graph neural networks,” in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 743–747, IEEE, 2018.
[15] F. Fusco, B. Eck, R. Gormally, M. Purcell, and S. Tirupathi, “Knowledge-and data-driven services for energy systems using graph neural networks,” in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 1301–1308, IEEE, 2020.
[16] GARPUR Consortium, “D1.1: State of the art on reliability assessment in power systems,” 2014. Available at https://www.sintef.no/globalassets/project/garpur/deliverables/.
[17] UCTE, “UCTE Operations Handbook,” 2004.
[18] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, “A comprehensive survey on graph neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 32, no. 1, pp. 4–24, 2020.
[19] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, “Geometric deep learning: Going beyond euclidean data,” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.
[20] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, *et al.*, “Relational inductive biases, deep learning, and graph networks,” *arXiv preprint arXiv:1806.01261*, 2018.
[21] I. B. Sperstad, E. H. Solvang, S. H. Jakobsen, and O. Gjerde, “Data set for power system reliability analysis using a four-area test network,” *Data in Brief*, vol. 33, p. 106495, 2020.
[22] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
[23] C. J. van Rijsbergen, *Information Retrieval*. London: Butterworths, 1975.