



Imperial College of Science, Technology and Medicine  
Department of Electrical and Electronics Engineering  
Control and Power Research Group

**Data generation and model usage for machine  
learning-based dynamic security assessment  
and control**

Al-Amin Bashir Bugaje

Submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy and the Diploma of Imperial College, March 2023

## Abstract

The global effort to decarbonise, decentralise and digitise electricity grids in response to climate change and evolving electricity markets with active consumers (prosumers) is gaining traction in countries around the world. This effort introduces new challenges to electricity grid operation. For instance, the introduction of variable renewable energy generation like wind and solar energy to replace conventional power generation like oil, gas, and coal increases the uncertainty in power systems operation. Additionally, the dynamics introduced by these renewable energy sources that are interfaced through converters are much faster than those in conventional system with thermal power plants.

This thesis investigates new operating tools for the system operator that are data-driven to help manage the increased operational uncertainty in this transition. The presented work aims to answer some open questions regarding the implementation of these machine learning approaches in real-time operation, primarily related to the quality of training data to train accurate machine-learned models for predicting dynamic behaviour, and the use of these machine-learned models in the control room for real-time operation.

To answer the first question, this thesis presents a novel sampling approach for generating 'rare' operating conditions that are physically feasible but have not been experienced by power systems before. In so doing, the aim is to move away from historical observations that are often limited in describing the full range of operating conditions. Then, the thesis presents a novel approach based on Wasserstein distance and entropy to efficiently combine both historical and 'rare' operating conditions to create an enriched database capable of training a high-performance classifier. To answer the second question, this thesis presents a scalable and rigorous workflow to trade-off multiple objective criteria when choosing decision tree models for real-time operation by system operators. Then, showcases a practical implementation for using a machine-learned model to optimise power system operation cost using topological control actions. Future research directions are underscored by the crucial role of machine learning in securing low inertia systems, and this thesis identifies research gaps covering physics-informed learning, machine learning-based network planning for secure operation, and robust training datasets are outlined.



## **Statement of originality and copyright declaration**

I hereby declare that this thesis is the result of my own work, and that any ideas, concepts or results from the work of other people, published or otherwise, are appropriately referenced.

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Al-Amin Bashir Bugaje

Department of Electrical and Electronic Engineering  
Imperial College of Science, Technology and Medicine, London, U.K.



## Acknowledgements

Alhamdulillah.

The PhD is a unique journey towards discovery: of the self and knowledge. It takes you to the frontiers, the brink, and just a little further. As such, guidance in charting out a good course, in navigating those muddy waters, and in recalibrating the course along the way if needed is no trivial task. To my supervision team, Professor Goran Strbac and Dr Jochen Cremer, I thank you for the opportunities, the guidance, the invaluable insights, time and energy to flesh out ideas, and most importantly, the chance to fail and learn.

Funding is the lifeblood of research; without it, the future work section disappears into oblivion, and well, that is just a shame. I want to thank my sponsors, the the Nigerian National Petroleum Corporation (NNPC) for this opportunity of a lifetime. I thank Ms Rakiya Sambo for her relentless efforts and Umar Abdulrahman for his kindness and help along the way. I thank Ndu Ughamadu for facilitating, and Dr. Maikanti Kachalla Baru for believing and investing in me; may your soul continue to rest in peace. I pray to return this kindness in many folds.

Emotional support is pivotal to a successful research effort. I have been blessed with an amazing family whose corny jokes across the atlantic nourished me in the depths of Covid lockdowns. I am grateful to my mum, Professor Mairo Adamu, and my dad, Alhaji Bashir Bugaje, for their unconditional and unwavering support. You made me whom I am, and everyday I discover new dimensions to how lucky I am to be yours. A son could not be more proud. May Allah bless your endeavours and crown your efforts. I thank my siblings for their comforting and thoughtful engagements, and the mothers who have taken me in and from whom I draw inspiration, Justice Sadiya Mayana, Dr Mairo Mandara, Dr Rabi Adamu, Dr Hadiza Nuhu, Dr Hauwa Bugaje, Dr Badiyya Bugaje, amongst others.

As much as the PhD is an individual journey, your colleagues and friends have a significant impact on oneself. I want to thank Dr Federica Bellizio and Alicia Blatiak for their friendship and the invaluable life lessons during our entrepreneurship stint. I am grateful to the EEE community for their socials, especially the EEE reps during my time organising events. I am also grateful for the climate entrepreneurship club, and the graduate school for their extracurricular engagements. These activities added some colour to otherwise dull days. I thank Dr

Mingyang Sun for the support, Yunhe Wei for collaboration. I thank Mohamed Aldarmon, Malek Alduhaymi, Saleh Komies and Faris Abualnaja; my days in college light up with your conversations.

Most importantly, to the "bane of my existence"; my rock, my true north. In life, and especially during the PhD, you will fail, and how you repeatedly get back up is pivotal to your continued growth and success. Iman, my love, my "moon on a starless night", you have made this endeavour considerably easy by providing me a safe haven, a warm cocoon of solitude, and for that, I will forever be grateful. Thank you for the space, the courage, and the warmth. "Baby PhD", Daddy can't wait to meet you. May Al-latif guide all your endeavours, ameen.

## **Dedication**

To all those who struggle in the pursuit of knowledge: stay strong, stay true.

To Nur.




# List of Tables

2.1	Confusion and Cost Matrix	46
3.1	Summary of relevant state-of-the-art works on database generation for ML-based DSA	52
3.2	Computational analysis of the GAPSPLIT algorithm where $ \Omega^N $ is a constant and $ \Omega^G $ corresponds to the number of iterations. The sorting step is the computational bottleneck of the algorithm.	62
3.3	$\mathcal{V}$ of 5000 samples computed for 10'000 random variable subsets $\{\hat{\Omega} \subset \Omega^N\}$ , where $ \hat{\Omega}  = 3$ for different sampling approaches	74
3.4	The $\mathcal{V}$ , $\mathcal{COV}$ , $h$ and $\lambda$ values of four different sets of 5000 samples computed for 10'000 random variable selections $\{\hat{\Omega} \subset \Omega^N\}$ , where $ \hat{\Omega}  = 3$	78
3.5	Computation time to generate 100'000 samples in IEEE 118-bus generator space	79
3.6	ANN trained for dynamic security on 1000 OCs from IEEE 68-bus system	80
3.7	DT trained for dynamic security on 1000 OCs from IEEE 68-bus system	81
3.8	$F1$ -score for 5 different ML models trained on 1000 OCs from the IEEE 68-bus system. Each type of model is trained 100 times.	81
4.1	Results for contingency representing a three-phase fault at bus 12	103
4.2	Results for contingency representing a three-phase fault at bus 15	103

4.3	Results for contingency representing a three-phase fault at bus 49 . . . . .	106
4.4	Results for contingency representing a three-phase fault at bus 80 . . . . .	106
4.5	Performance of USM (A) and USM (B) approaches for contingencies representing a three-phase fault at buses 12 and 15. . . . .	107
5.1	Actual normalised costs $N\Sigma$ (to the basis $10^{-2}$ ) considering errors of 20 % in $\tilde{\gamma}$ . The presented values are averages and standard deviations in the interval $0.8 \leq PCF \leq 1$ of Fig. 5.11. . . . .	131
5.2	Computation time for the offline and online workflows showing the proposed approach to compute cost curves for classifier selection scales well to large systems.	132
6.1	The relative cost and computation time and for different heuristics on 200 TS test instances . . . . .	148
6.2	Handling infeasible TS of the proposed RBFNN heuristic for $\mathcal{J} \leq 5$ . . . . .	150


# List of Figures

2.1	Power system operating states and control. The system responds to disturbances ( $\longleftrightarrow$ ) with control actions ( $\Longrightarrow$ ), while satisfying the equality (E) and inequality (I) constraints. Figure gotten from [1]. . . . .	23
2.2	Classification of power system stability [2]. . . . .	25
2.3	A synchronous generator showing the prime mover, governor and voltage excitation system. The internal state variables are denoted by red color. [3]. . . . .	29
2.4	Rotor angle trajectory for different transient faults [2]. Increase in rotor angles in cases 1 & 2 show loss of synchronism and transient instability. Case 3 is transient stable. . . . .	30
2.5	Different post fault trajectories from an operating condition (OC) (a), to an unstable OC (b), a static and dynamically secure OC (c), a static secure but dynamically insecure OC (d) [4]. . . . .	31
2.6	Extended power system stability definition covering resonance and converter driven instability. [5]. . . . .	32
2.7	A two-level VSC (top); block structure of the power converter model and its control scheme (bottom). [3]. . . . .	34
2.8	System-level control of grid-following operation mode (top) and grid-forming operation mode (bottom) [3]. . . . .	34

2.9	A visual illustration of the machine learning-based approach to DSA. In the offline stage, the model learn the security boundary (■ ■ ■ ■) from the training database containing secure (●) and insecure (●) operating conditions. In the online stage, the classifier predicts the security labels of new operating conditions (Ⓐ) (ⓧ) with varying degrees of confidence. . . . .	42
2.10	Machine learning workflow for dynamic security assessment. The orange coloured areas are investigated in this thesis . . . . .	43
3.1	(a) GAPSPLIT generates new samples by attempting to split maximal gaps. (b) GAPSPLIT converges to an infeasible sample when the primary target (★) is located in the infeasible region. Max gap is found in $X_2$  . . . . .	60
3.2	(a) GAPSPLIT* uses only secondary targets (★) to minimise the shortest distance to the feasible region. (b) GAPSPLIT* with only secondary targets (★) generates samples around the boundary of the feasible space (●) when $T_n$ is located in infeasible regions. (c) GAPSPLIT* re-directs sampling to other regions of the physical space by memorising infeasible targets (★) . . . . .	61
3.3	The iteration step of the proposed modification (ii) of GAPSPLIT* that introduces the set of infeasible samples $\Omega^{G'}$ . . . . .	64
3.4	The samples (blue circles) in (a) and (b) have the same $\mathcal{COV}$ but different $\mathcal{V}$ values. The proposed convex hull metric (dotted black line) is suitable for measuring multivariate coverage. . . . .	68
3.5	(a) The $\mathcal{V}$ of samples generated using the proposed modification (i) (■), the proposed modification (ii) (●), and $\mathcal{RS}$ (▲). . . . .	72
3.6	The value of $\beta = \frac{ \Omega^{G'} }{ \Omega^{G''} }$ for the proposed modification (ii) (●) reduces while regular GAPSPLIT's (●) increases as more samples are generated in both (a) small and (b) larger systems. . . . .	76

3.7	Modification (ii) better distributes samples across the feasible space than modification (i) in small systems, respectively in (b) and (a).	76
3.8	(a) GAPSPLIT* with only primary targets (shown for $P1(\bullet)$ and $Q3(\bullet)$ ) generates samples that are distributed along the axis of that variable and do not cover the entire feasible space. (b) Proposed use of GAPSPLIT* uses primary and secondary targets to cover the entire feasible space	78
3.9	Samples that have similar $\mathcal{COV}$ values are differentiated with the proposed $\mathcal{V}$	78
3.10	(a) The normalised $\mathcal{V}$ mean (—) and median (—) values of random subset selections are similar across different dimensions (b) There is some correlation between the $\mathcal{V}$ of a random subset and the $\mathcal{V}$ of the full dimension	80
4.1	Quality datasets balance historical relevance (—), coverage (□), and discriminative relevance (—) so that ML-models can be trained for accurately predicting secure (●) and insecure (●) OCs.	85
4.2	Proposed unified sampling considering all three properties of <i>quality</i> datasets. Phase A combines generic and copula-based historical sampling to generate historically relevant and diverse pre-fault OCs. Phase B directs generating labels to entropy-rich regions.	92
4.3	Wasserstein distance between two probability distributions $P(\bar{X})$ and $P(\underline{X})$ is analogous to the amount of 'work' to transform one distribution into another. 'Work' is the distance moved multiplied by the probability 'mass' at that distance.	94
4.4	Proposed phase B computes the entropy of clusters with OCs that are secure (●) and insecure (●). The cluster with the largest entropy initialises the bounds for resampling using the generic sampling modification (ii) of Section 3.2.3.	97
4.5	Single line diagram of the IEEE 118 bus test system showing fault locations at buses 12, 15, 49 and 80.	97

4.6	The relationship between relative volume $\hat{\mathcal{V}}$ and Wasserstein distance threshold $\tau$ . $\hat{\mathcal{V}}$ increases with the number of generated OCs and plateaus around $\tau \geq 10$ .	99
4.7	Boxplots showing mean (—) and median (—) values of 1,000 randomly selected OCs from different approaches corresponding to (a) the proposed $\mathcal{W}_p$ and (b) the normalised $\mathcal{V}$ .	100
4.8	CDF of p-values corresponding to historical data (—), CSM (—), GSM (—), and the proposed USM (A) (—) for (a) energy tests and (b) K-S tests.	101
4.9	Boxplots showing mean (—) and median (—) values for 100 models trained with different datasets to predict the post-fault status of a three-phase fault near buses (a) 12 (b) 15 (c) 49 (d) 80. Both training and testing data follow a similar distribution.	105
5.1	Overall workflow for ML-based security assessment [1].	110
5.2	The ROC curve for evaluating a classifier. Different combinations of $TPR$ and $FPR$ are obtained by varying the decision threshold $Z$ (shown with arrows). The perfect classifier would be in the top left corner having $TPR = 1$ and $FPR = 0$ .	113
5.3	Proposed cost-curve workflow to select the best DT classifier offline. The selection can be updated with real-time data.	119
5.4	Construction of cost-curves: In (a), each dashed linear curve corresponds to one blue point in the ROC curve Fig. 5.2. The combined minimal expected cost $N\tilde{\Sigma}$ of (a) is where the selected $Z$ equalled the cost-optimal decision threshold $\tilde{Z}^* = PCF$ and is shown in (b).	120
5.5	Proposed cost-curve based selection considering multiple criteria. The lime classifier has the best trade-off in terms of interpretability and expected cost in the relevant interval $0.8 \leq PCF \leq 1$ .	120

5.6	The predicting performance of classifiers with 7 different DT depths. The box plots in (a) show min, max, interquartiles, and median for $22 \times 7 = 154$ trees. Using (a), the 22 classifiers with the highest $F_1$ score are selected and the selection-frequencies are in (b).	123
5.7	Tree structure of a DT with DT depth of 3. P and Q are the active and reactive power-flow between buses, and V is the voltage at the bus.	125
5.8	Ratio between missed(■) and false(■) alarms for the 22 DTs without considering the impact of different costs. $\Pi_+$ is the class ratio of positive conditions.	126
5.9	Ratio between missed(■) and false(■) alarms for the 22 DTs by considering the impact of different costs.	127
5.10	Selection based on modified ROC curves for multiple criteria. The DT classifier in cyan has the best trade-off in terms of cost-optimality (x on the curve) and interpretability (the number of DT-nodes from 0 to 50, where 0 is on the left side of the color scheme  ).	129
5.11	Sensitivity study on errors in estimations of parameter $\tilde{\gamma}$ . The actual costs $N\Sigma$ under the error of +20 % (---) and -20 % (---) are presented against the estimated costs $N\tilde{\Sigma}$ that is shown as the baseline. In (a) and (b) are the lime and cyan DTs from Fig. 5.5.	130
6.1	ML workflow for real-time topology change	139
6.2	An example RBFNN architecture	143
6.3	Change in average dispatch cost ( $\Delta\bar{q}$ ) with variation in load distribution for the proposed RBFNN (▲) and (a) KNN (●) (b) a feed-forward ANN (★) heuristics.	146
6.4	The relative dispatch cost to best known solution ( $\frac{\hat{q}-q}{q}$ ) for 200 test instances using different heuristics	148

6.5	Computational time to obtain TS solutions for 200 test instances using different heuristics	149
6.6	Solution time for different MIP formulations	149

# List of Publications

- [1] Federica Bellizio, **Bugaje, Al-Amin B**, Jochen L Cremer, and Goran Strbac. Verifying machine learning conclusions for securing low inertia systems. *Sustainable Energy, Grids and Networks*, 30:100656, 2022.
- [2] **Bugaje, Al-Amin B**, Jochen L Cremer, and Goran Strbac. Split-based sequential sampling for realtime security assessment. *International Journal of Electrical Power & Energy Systems*, 146:108790, 2023.
- [3] **Bugaje, Al-Amin B**, Jochen L Cremer, Mingyang Sun, and Goran Strbac. Selecting decision trees for power system security assessment. *Energy and AI*, 6:100110, 2021.
- [4] **Bugaje, Al-Amin Bashir**, Jochen Lorenz Cremer, and Goran Strbac. Real-time transmission switching with neural networks. *IET Generation, Transmission & Distribution*, 2022.
- [5] **Bugaje, Al-Amin Bashir**, Jochen Lorenz Cremer, and Goran Strbac. Generating quality datasets for real-time security assessment: Balancing historically relevant and rare feasible operating conditions. *International Journal of Electrical Power & Energy Systems*, In Review.
- [6] Yunhe Wei, **Bugaje, Al-Amin B**, Federica Bellizio, and Goran Strbac. Reinforcement learning based optimal load shedding for transient stabilization. In *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–5. IEEE, 2022.



# Contents

Abstract	i
Acknowledgements	v
List of Tables	ix
List of Figures	xi
List of Publication	xvii
List of Acronyms	2
1 Introduction	7
1.1 Motivation . . . . .	7
1.2 Machine learning approach to DSA . . . . .	13
1.3 The challenge of generating high quality training data . . . . .	15
1.4 The challenge of model selection . . . . .	18
1.5 The challenge of managing risks of model usage . . . . .	18
1.6 Objectives . . . . .	19

1.7 Contributions . . . . .	20
1.8 Structure of thesis . . . . .	21
<b>2 Background</b>	<b>22</b>
2.1 Introduction . . . . .	22
2.2 Power system stability . . . . .	25
2.2.1 Voltage stability . . . . .	26
2.2.2 Rotor angle stability . . . . .	27
2.2.3 Rotor angle stability . . . . .	30
2.2.4 Extended stability definition . . . . .	32
2.2.5 Power system reliability . . . . .	35
2.2.6 Security assessment . . . . .	36
2.2.7 Dynamic security assessment . . . . .	37
2.2.8 Machine learning-based approach to dynamic security assessment . . . . .	40
2.2.9 Machine learning-based DSA Workflow and contributions . . . . .	42
<b>3 Generation of Diverse Datasets</b>	<b>48</b>
3.1 Introduction . . . . .	49
3.1.1 Sampling approaches . . . . .	49
3.1.2 Measuring quality of sampling . . . . .	54
3.1.3 Contributions . . . . .	55
3.2 Split-based sampling . . . . .	57

3.2.1 Regular split-based approach . . . . .	57
3.2.2 Issues with GAPSPLIT . . . . .	60
3.2.3 Proposed split-based approach: GAPSPLIT* . . . . .	62
3.3 Measuring performance of samplers . . . . .	67
3.4 Case study . . . . .	70
3.4.1 Test System and assumption . . . . .	70
3.4.2 Effective sampling with GAPSPLIT* . . . . .	73
3.4.3 Addressing GAPSPLIT issues . . . . .	74
3.4.4 Measuring performance of samplers . . . . .	77
3.4.5 Computational Performance & Scalability . . . . .	79
3.4.6 Dynamic Security and Machine learning . . . . .	80
3.5 Conclusion . . . . .	81
<b>4 Trade-offs in Generating Balanced Datasets</b>	<b>83</b>
4.1 Introduction . . . . .	84
4.1.1 Contributions . . . . .	86
4.2 Single criterion-based sampling approaches . . . . .	86
4.2.1 State-of-the-art historical sampling . . . . .	87
4.2.2 State-of-the-art generic sampling approach . . . . .	89
4.2.3 State-of-the-art importance sampling approach . . . . .	90
4.3 Proposed unified sampling approach . . . . .	91
4.3.1 Knowledge discovery: trading off historical and rare OCs . . . . .	92

4.3.2 Dataset enrichment and labelling: importance sampling in the context of feasibility . . . . .	94
4.4 Case Study . . . . .	96
4.4.1 Test system and assumptions . . . . .	98
4.4.2 Trading-off historical relevance and coverage . . . . .	100
4.4.3 Historical relevance . . . . .	101
4.4.4 Coverage . . . . .	102
4.4.5 Security assessment for out-of-distribution OCs . . . . .	103
4.4.6 Security assessment for similar distribution OCs . . . . .	104
4.4.7 Balancing the distribution . . . . .	106
4.5 Conclusion . . . . .	107
<b>5 Model Selection</b>	<b>109</b>
5.0.1 Selection of the machine learning model . . . . .	110
5.0.2 Contributions . . . . .	112
5.1 Objectives in selecting DT classifiers for DSA . . . . .	113
5.1.1 Minimising the effects of inaccuracies . . . . .	114
5.1.2 Maximising the interpretability of DTs . . . . .	116
5.1.3 Maximising the robustness of classifiers . . . . .	116
5.2 Multi-criteria selection of classifiers . . . . .	117
5.2.1 Visual inspection using the ROC curve . . . . .	119
5.2.2 Proposed cost-curve approach for graphical inspection . . . . .	121

5.3 Case study . . . . .	123
5.3.1 Test system and assumption . . . . .	123
5.3.2 Selections based on single criteria . . . . .	125
5.3.3 Multi-criteria selection with modified ROC approach . . . . .	127
5.3.4 Multi-criteria selection with cost-curves . . . . .	128
5.3.5 Reduction of loss of load costs under uncertainties . . . . .	129
5.3.6 Computation time and scalability . . . . .	131
5.3.7 Discussion . . . . .	132
5.4 Conclusion . . . . .	133
<b>6 Machine Learning in Real-time Operation</b>	<b>135</b>
6.1 Introduction . . . . .	136
6.2 Heuristic approaches . . . . .	139
6.2.1 Proposed RBFNN Heuristic . . . . .	141
6.2.2 KNN Heuristic . . . . .	143
6.2.3 Sensitivity-based Greedy Search Heuristic . . . . .	144
6.2.4 Gurobic Heuristic . . . . .	144
6.2.5 Computational complexity . . . . .	144
6.3 Case Study . . . . .	145
6.3.1 Test System and assumption . . . . .	145
6.3.2 Performance of heuristics on out of distribution data . . . . .	146
6.3.3 Real-time computational performance . . . . .	148

6.3.4	Guaranteeing feasible topology near real-time	150
6.3.5	Offline computational improvements	151
6.3.6	Key Findings	151
6.4	Conclusion	152
<b>7</b>	<b>Conclusion</b>	<b>153</b>
7.1	Summary of Thesis Achievements	154
7.1.1	Quality of training data	155
7.1.2	Model selection and usage for real-time application	156
7.2	Future Research Directions	158
7.2.1	Machine learning in low inertia systems	158
7.2.2	Physics-informed learning	159
7.2.3	ML-based network planning for secure operations	161
7.2.4	Robust training dataset	162
	<b>Bibliography</b>	<b>163</b>



# List of Acronyms

**AC** Alternating current.

**ADASYN** Adaptive synthetic sampling.

**AI** Artificial intelligence.

**ANN** Artificial neural network.

**AVR** Automatic voltage regulator.

**CART** Classification and regression trees.

**CDF** Cumulative distribution function.

**CIG** Converter-interfaced generator.

**CSM** Copula-based sampling model.

**DC** Direct current.

**DER** Distributed energy resource.

**DL** Deep learning.

**DSA** Dynamic security assessment.

**DTs** Decision trees.

**ED** Economic dispatch.

**EV** Electric vehicle.

**FN** False negative.

**FP** False positive.

**FPR** False positive rate.

**GNN** Graph neural network.

**GSM** Generic sampling model.

**HI** High inertia.

**HM** Historical model.

**HR** Hit and run.

**HVDC** High-voltage direct current.

**ICT** Information and communication technology.

**IPOPT** Interior point optimiser.

**K-S** Kolmogorov-Smirnov.

**KNN** K nearest neighbour.

**LHS** Latin hypercube sampling.

**LI** Low inertia.

**LSLR** Least squares linear regression.

**MC** Monte Carlo.

**MG** Multivariate gaussian.

**MIP** Mixed integer programming.

**ML** Machine learning.

**OCs** Operating conditions.

**ODE** Ordinary differential equation.

**OPF** Optimal power flow.

**OTS** Optimal transmission switching.

**PCF** Probability cost function.

**PINN** Physics informed neural network.

**PLL** Phase-locked loop.

**PMUs** Phasor measurement units.

**PR** Precision-recall.

**PSS** Power system stabiliser.

**PTDF** Power transfer distribution factor.

**RBFNN** Radial basis function neural network.

**RLC** Resistor inductor capacitor.

**RMSprop** Root mean squared propagation.

**ROC** Receiver operating characteristic.

**SMOTE** Synthetic minority oversampling technique.

**SO** System operator.

**SVMs** Support vector machines.

**TN** True negative.

**TP** True positive.

**TPR** True positive rate.

**TS** Transmission switching.

**USM** Unified sampling model.

**V2G** Vehicle-to-grid.

**VSC** Voltage source converter.



# Chapter 1

## Introduction

إِنَّ مَبَادِ كُلِّ فَنٍّ عَشْرُهُ      الْحَدُّ وَالْمَوْدُوعُ ثُمَّ الثَّمَرَةُ  
وَفَضْلُهُ وَنِسْبَةُ وَالْوَاضِعُ      وَالْأَسْمُ لَا سِتِمْدَادَ مُحْكَمُ الْأَشَارِعِ  
مَسَائِلُ وَالْبَعْضُ بِالْبَعْضِ اكْتَفَى      وَمَنْ دَرَى الْجَمِيعَ حَازَ الشَّرَفَ

*The tenets of every discipline are ten: the definition, the content, and then the fruit; its merit, interdisciplinarity, and its pioneer; its name, its sources, the legal ruling; the issues discussed, while some have sufficed with a little bit of it; the one who knows the field completely becomes honoured.*

—Shaykh Muhammad Al-Yaqoubi

### 1.1 Motivation

The increasing concern for climate and environmental sustainability has led to the evolution of power systems by replacing traditional fossil fuel-based energy sources with renewable energy technologies. This coordinated shift towards renewable energy is a crucial step in decarbonising the power sector, which follows the adoption of the internationally binding climate treaty at the UN Climate Change conference in Paris in 2015. This treaty aimed

at limiting the global temperature increase to 1.5°C above pre-industrial levels, and 196 countries have agreed to this goal. The transition to net-zero carbon emissions has accelerated the adoption of renewable technologies, such as wind, solar, and battery energy storage systems. It has also paved the way for the electrification of the transport and heating sectors. Most of the power sector decarbonisation is currently taking place in highly polluted countries, such as Europe and America, where conventional power systems are dominated by large synchronous generators operating on fossil fuels. At the same time, the replacement of gas boilers and combustion engines with electric heat pumps and electric vehicles, respectively, in the heating and transport sectors, offers a unique opportunity to deeply decarbonise the entire economy. The integration of renewable energy sources, and the electrification of heating and transport sectors, present new operational challenges and opportunities for power systems. The shift towards renewable energy sources and the electrification of heating and transport has revolutionised the energy mix, requiring power systems to adapt and evolve to meet the demands of a low-carbon future.

Conventional power systems transport electrical energy in one direction, from mega sources (generators) to large sinks (loads). The power system architecture is typically centralised, with the generated power flowing through an interconnected network of lines and cables, also known as the transmission system. The transmitted power is at high voltage levels and is ultimately made available to the end user at low voltage levels through the distribution system. To maintain the equilibrium of the power system, the aggregate generator output must always equal the total load demand. In legacy power systems, this equilibrium was easily maintained by regulating the amount of fuel burned to match demand, as the generated power came from synchronous machines and consist of prime movers that automatically and in real-time adjust the generator output to match the demand. Consequently, the power system can, on most occasions, accommodate the uncertainty in energy demand via increasing or reducing the generator output, as the case may be. The system operator sits at the centre of this architecture and plays a critical role in maintaining this equilibrium by reliably operating the power system infrastructure and ensuring that generated electricity is available to the end users. However, achieving net-zero carbon emissions

targets presents significant challenges to maintaining this equilibrium as the power system architecture and infrastructure undergo crucial changes. Crucially, the introduction of large amounts of renewable energy sources in the generation mix is accompanied by power electronics devices that make real-time organic control of the generator output difficult. Two key challenges arise in this context.

The first challenge in maintaining power systems' equilibrium concerns handling uncertainties arising from new faults encountered during operations [6]. The substitution of synchronous machines with power electronics-interfaced generation at high voltage levels and the integration of renewable energy sources in the form of distributed energy resources (DERs) in medium and low voltage levels introduce uncertainty in energy availability tied to weather conditions. The dependence on weather makes controlling energy generation to match demand more difficult, as both demand and generation become highly unpredictable. Furthermore, seasonal variations in demand contrast with renewable energy outputs. For instance, in winter the energy demand is high, but solar irradiance is low, while the opposite is the case in summer. During normal operation, the system operator's objective is to maintain system variables close to their desired values and ensure the system can survive disturbances. By maintaining variables close to their desired values, the system operator ensures that the equipment is not overloaded, and the physical constraints are satisfied around the clock. The system operator must ensure that the system can survive small or large disturbances. For small disturbances such as variations in demand levels, the system constantly adjusts the generation output through automatic controllers. For large disturbances, such as equipment failure or line overloads, the system operator may trigger preventive or corrective actions to restore normal operations. Large disturbances are of major concern to system operators as they can lead to power blackouts. Conventional power system design follows a redundancy principle, where large safety margins on underutilised assets and preventive control actions are incorporated to ensure reliable system operation. This design ensures that an asset failure does not disrupt operations, as most assets are operating at below capacity. However, this approach is highly inefficient and expensive, and for future power systems, it is unscalable as operational uncertainty increases. To

overcome these challenges, a scalable and efficient approach would utilise existing assets close to their capacity with smaller safety margins and deployment of 'smart' corrective control to reliably operate the system [7]. Investigating new operating tools that allow the efficient utilisation of existing assets is the focus of this thesis.

The second challenge is the evolution of power flows in power systems from unidirectional to bidirectional (from low and medium voltage levels to high voltage levels), as prosumers may wish to trade their electricity as a form of flexibility. The rise of the prosumer follows legislative and economic incentives that facilitate the proliferation of DERs, such as battery energy storage systems and embedded solar generation. In such an architecture, power generation is no longer centralised but decentralised, and the demand side allows for some flexibility as local generation can offset some demand. There, an opportunity exists where DERs can provide additional value, such as providing flexibility services to mitigate the uncertainty from renewable sources, eliminating network congestion and safeguarding against cyber-attacks [8]. Local energy markets may serve as a coordinating mechanism for end-users to trade excess energy locally and as aggregated flexibility services to the grid, thereby minimising their electricity costs and maximising their revenues. Future energy systems can have local users that can organise themselves into 'smart' microgrids, enabling them to become fully energy independent by relying solely on local generation and storage, thereby operating off-grid. Additional DERs will result from electrifying the transport and heating sectors, where electric vehicles and electric heat pumps can act as energy storage devices that can offer flexibility services. Albeit, the electrification of transport and heat can increase the peak electricity demand levels, which traditionally would require upgrading infrastructure, as mentioned in the first challenge. The bi-directional flow of energy as a result of DERs and the increase of peak demand represent a challenge requiring 'smarter' controls to efficiently utilise available assets [8, 9].

The integration of renewable energy sources at high voltage levels and local DERs with bi-directional power flows at low and medium voltage levels necessitates implementing an extensive information and communication technology (ICT) infrastructure to provide emerging control tools with necessary real-time data. For instance, controlling local DERs

in both centralised and decentralised systems would require a communication infrastructure to instantly transmit optimal set-points to the local DERs. Albeit, a decentralised system would require less infrastructure and provide safeguards against cyber-attacks at the expense of sub-optimal operation, in comparison to a centralised system. The vision of digitalisation in power systems and the adoption of 'smart' control to replace traditional approaches to power system security analysis and control requires instrumentation at various voltage levels across power systems to collect large quantities of real-time data from phasor measurement units (PMUs), coupled with advances in wireless communication and cloud computing [10]. The digitalisation of power systems can act as a major catalyst for efficiently utilising existing assets thereby ushering in a 'smart' grid era. Such a 'smart' grid is characterised by an increased penetration of renewable energy sources, where the system architecture is decentralised and allows for bi-directional power flows from DERs, with the proliferation of local energy markets. This transformation of the generation, transmission, distribution, consumption and trading of electricity can enhance power system reliability, flexibility, and efficiency [11]. Furthermore, advancements in artificial intelligence for sequential decision-making [12] can support the autonomous operation of power systems and system planning [13], while game theory [14, 13] supports the development of more agile energy markets [15]. From the transmission system perspective, this digitalisation vision encompasses the integration of three functional 'smart' blocks: control centres, transmission networks and substations, integrated through a common digital platform that relies on the ICT infrastructure [16]. The implementation of a central digital platform for integrating these blocks can facilitate more secure and reliable grid operations using real-time data from PMUs [17]. Albeit, there are data privacy and security considerations when handling these massive datasets. Recent advancements in distributed ledgers used in blockchain technologies can support privacy-centred and secure data distribution [18, 19]. The development of such digital platforms in tandem with novel control tools and strategies is essential to fully exploit the benefits of this digitalisation paradigm [19].

Following this digitalisation paradigm, the objective of this thesis is to investigate novel approaches for operating and controlling power systems such that existing assets are utilised

more effectively as opposed to the worst-case guarantee principle. Specifically, the focus is on the secure operation of power systems, as these novel approaches can quickly evaluate the system response to a fault and adjust pre-fault and post-fault operating set-points of controllable assets. These new tools will improve the situational awareness of the system operator to maintain adequate security levels without relying on traditional approaches such as asset redundancy to prevent worst-case scenarios of power blackouts [6]. The use of preventive and corrective control actions to lower safety margins on equipment can help manage insecurity in power systems operations. Security assessments measure a power system's vulnerabilities to faults and equipment failure. The system operator carries out contingency studies ahead of the operation horizon to evaluate the response of the system to probable faults. In the past, the system operator utilised numerical integration methods to simulate the system's response to a fault. These simulations provided insight into the post-fault security status according to specified criteria. Once the dynamic security assessment (DSA) was completed, operators determined the optimal cost-effective preventive control actions to activate to ensure adequate security levels [20]. The expansion of modern power systems coupled with the integration of renewable energy sources introduces a high degree of uncertainty to power system operations that challenges the conventional approach to DSA. The conventional approaches of simulating the system's response to faults and assessing post-fault dynamic security using standard criteria become intractable in near or real-time settings due to the exponentially increasing number of possible operating conditions (OCs) and probable contingencies. For instance, simulating a single contingency on the French transmission system takes 56 s [21], and assessing tens of thousands of OCs and thousands of contingencies would be required. Therefore, the computational complexity of these detailed time-domain dynamic simulations becomes a significant bottleneck for real-time DSA.

In recent years, the digitalisation of power systems has been proposed as a potential solution to the computational bottleneck of DSA [1]. This approach involves processing and analysing historical data to infer patterns to speed up the assessment and control framework, thus making them tractable in real-time operation. Machine learning (ML) has

shown promise in efficiently inferring underlying patterns from complex data and providing instantaneous solutions. This technique has been previously investigated to address the challenges surrounding power system DSA and control [22]. However, there are several barriers to the adoption of ML methods in the power systems and broader energy sector. These barriers include the need to develop trust among system operators, concerns regarding data privacy and security, data quality issues, generalisation concerns, data ownership, and regulatory challenges. In contrast, the use of ML in other domains, such as finance [23] and language models [24], has quickly become commonplace.

## 1.2 Machine learning approach to DSA

The machine learning approach to DSA was proposed in the 1980s [25] to enable a faster assessment of power system security [22]. Since then, ML techniques have garnered considerable interest as a scalable approach for managing uncertainty and improving the reliability of the power system, particularly for transmission system operators [1]. In this new approach, the system operator prepares a classifier offline with data samples that mimic expected OCs a few days before the operation horizon. These data samples can include features such as voltages, power flows, power injections and phase angles. The aim is to train this classifier to accurately predict the security status of, current or future OCs, whether secure or insecure [1]. The offline training phase is designed to minimise residual uncertainties due to different network configurations or anticipated contingencies between when the classifier is trained and when it is used. These methods are promising for predicting various power system stability phenomena ranging from small to large disturbances, such as short-term [26] and dynamic [27] voltage stability [28], and transient stability [29][30]. Several classification models have been explored, including support vector machines (SVMs) [31, 32], random forest (RF) [33], k-nearest neighbours [34], and artificial neural networks (ANNs) [35], for transient stability analysis. In power systems security assessment, the choice of model is an important consideration. For instance, while ANN models offer higher accuracy, system operators prefer decision trees (DTs) and DT

ensembles [33, 36, 37, 38] as a balance between accuracy, computational complexity, and interpretability [39, 21, 38]. For ML-based DSA, there is a preference for so-called 'white box' models as they offer high interpretability. Interpretability in ML-based DSA is crucial since it enables system operators to comprehend the decision-making process of a classifier with a minimal examination, thereby maintaining their involvement in the control loop.

Although ML-based approaches to DSA have generated some traction in recent years and show significant promise for real-time operations [21], there are existing challenges that hinder their adoption for online DSA. The applicability of ML-based models for day-to-day operations faces several challenges that need to be addressed. This section outlines three primary challenges: (i) improving the quality of training data, (ii) trading off multiple objectives in the training and selection of the ML-based model, and (iii) managing risks when applying ML-based models in real-time operation.

Firstly, the quality of training data has a significant impact on the accuracy of ML-based models for DSA. Therefore, to improve the prediction performance of ML-based models for DSA, the training data must be of high quality. Typically, historical data is an initial data source for training ML models. However, historical data which characterises normal power systems operation is highly imbalanced with a distribution that is skewed towards the secure class [28, 40]. In addition to the class imbalance, the imbalance in the cost of wrong predictions needs to be considered, as missed alarms can lead to power blackouts and must be avoided while false alarms may only require re-dispatching generation assets. The training data should also take into account the sensitivity to parameters that can affect system security, such as fault clearance time, as they are assumed to be fixed when obtaining the security label of an OC.

Secondly, the system operator must balance multiple and often contrasting objectives in the training and selection process of the ML model. For instance, while high predictive accuracy is a prerequisite for these models, interpretability and efficient use of data are also important. In security assessments, interpretability is crucial for understanding the decision-making mechanism of a classifier, which is essential for building trust and adapting

to changing assumptions such as topological reconfiguration or data distribution. Additionally, minimising the amount of data required to achieve high accuracy is important due to limited computational resources. The selection of appropriate performance metrics is essential, and the objective function should be designed to take into account various operational objectives such as power system stability, operational efficiency, and economic costs.

Thirdly, the system operator must develop a mechanism to manage and balance risks when applying an ML-based model in real-time operations. As previously mentioned, the impacts of wrong predictions differ and thus carry different penalties. A wrong prediction can either be a missed alarm (predicted secure while the system is insecure) or a false alarm (predicted insecure while the system is secure). Missed alarms can lead to system collapse and are generally way more 'costly' than false alarms, as false alarms require deploying unnecessary resources at worst. In real-time operation, managing the residual risk of these classifiers is crucial as they can act as 'black boxes' for prediction or control. Approaches exist to quantify worst-case guarantees of these models, verify near real-time, or adopt a probabilistic outlook. It is necessary to develop models that can deal with a range of uncertainties, including unforeseen events and data quality issues. Additionally, there is a need to establish a framework for monitoring the performance of ML-based models to identify and mitigate any unexpected outcomes.

### **1.3 The challenge of generating high quality training data**

The quality of training data is an important factor that affects the prediction performance of ML models, which relies on several criteria such as coverage, variability, and balance [41]. However, as mentioned earlier, using historical data for training ML-based DSA models has limitations. One significant limitation is that historical observations mostly correspond to secure OCs, while a good training database must also include insecure ones

[1]. Additionally, extreme operating scenarios are seldom present in historical observations, and hence synthetic OCs are generated using sampling approaches. To generate synthetic samples, an OC is first sampled and then evaluated using time-domain simulations for the contingency considered.

The generation of synthetic data for ML-based DSA is crucial for accurate predictions. Three sampling approaches have been proposed in the literature to optimise data quality: *historical sampling*, *generic sampling* and *importance sampling*. Each sampling approach aims to optimise a specific property of quality datasets, respectively, *historical relevance*, *coverage*, and *discriminative relevance* [42, 1]. The first approach, *historic sampling*, uses historical records [43] and fits a probability distribution, such as vine-copulas [21, 40], to capture dependencies between variables such as loads and wind power outputs, or to identify 'relevant' buses for sparse PMU measurements [44, 45]. Subsequently, Monte-Carlo (MC) sampling can generate synthetic OCs [46, 47]. *Historic sampling* is suitable to sample OCs following the same distribution. However, future OCs can be different from historical OCs, and sampling from distributions is unsuitable for creating extreme OCs typically found at the tails of distributions. The second approach, *importance sampling*, aims to maximise the information content of the final training dataset [39, 37, 48, 49, 50]. The classifier quantifies the 'importance' of possible OCs generated [51, 52], and security assessment is carried out only on OCs with low confidence. The third approach, *generic sampling*, generates OCs uniformly distributed in the feasible space to explore all possible OCs. However, generating large amounts of OCs is computationally expensive, and many OCs add little knowledge to the final dataset. Several techniques have been proposed to overcome this challenge, including the use of Latin hypercube sampling (LHS) [53], sampling within the feasible neighbourhood of OCs [54], and outer approximation to convexify the original non-convex feasible space [55]. Infeasibility certificates based on separating hyperplanes [56] can be used to discard infeasible portions of the input space. [57] proposed generating representative OCs that span the AC OPF feasible space by uniformly sampling loads from a convex input space and using infeasibility certificates to reduce the search space. However, each of these approaches has limitations. The *historical*

and importance sampling approaches may neglect some feasible OCs, while *generic sampling* faces the challenge of sampling in high dimensions. Therefore, an efficient *generic sampling* approach that scales to larger systems and can generate extreme synthetic OCs beyond historical records is essential for accurate ML-based DSA in power systems with high shares of intermittent renewables. This thesis identifies two existing gaps in the generation of quality training data for ML-based DSA.

The first gap concerns the assumption in most state-of-the-art works that generator outputs are scheduled based on optimal power flow (OPF) solutions that minimise generation costs [21, 39, 57, 37, 48, 45, 58, 50]. However, since faults can occur at non-optimal set points, it is necessary to develop methods that explore likely operating conditions (OCs) beyond the OPF solution [51]. While some methods in the literature, such as Latin hypercube sampling (LHS) [51, 52, 53, 55, 56], explore the entire feasible space using *generic sampling* approaches, current methods attempt to discard sections of the search space through rapid rejection sampling [52, 57, 55, 56, 52]. There is an opportunity for iterative methods that optimally explore the feasible space to address this research gap.

The second research gap pertains to the need for methods that balance different properties of quality datasets, as state-of-the-art approaches focus on maximising a specific property. Approaches that prioritise maximising *coverage* of the feasible space [55, 56, 57, 59] neglect variable dependency structures and therefore sacrifice *historical relevance*. Conversely, approaches that prioritise *historical relevance*, for instance, copulas [42, 21], autoencoders, and conditional variational autoencoders [44], fail to generate feasible but 'rare' OCs, thus sacrificing *coverage*. Additionally, approaches that focus on *historical relevance* do not generalise well to previously unobserved OCs. Finally, approaches that prioritise *discriminative relevance* neglect variable dependency structures that represent typical power system OCs and do not generalise to feasible but 'rare' OCs. Hence, there exists a research gap in developing methods that effectively balance these contrasting properties.

## 1.4 The challenge of model selection

Another challenge regarding adopting ML-based models includes the selection procedure of these models to build trust by the SO. One facet of such a selection process involves trading off multiple objectives of accuracy, interpretability and 'cost' sensitivity. The selection strategy must consider the primary objective of the system operator to ensure secure operations where the training data may be imbalanced while also being modular and scalable. Several performance metrics have been proposed in the literature to compare model performance in a class imbalance context. The limitations of using test accuracy, which is fast, scalable, and has traditionally been used as the performance evaluation metric, have been previously discussed [60][61][62]. Other fast and scalable metrics [63] include precision and recall score (for example [29]), AUC (for example, [64][65]) F-Measure, G-Mean (for example, [62][28][30]), brier score, still fail to generalise performance of a classifier over a range of sample distribution. One popular alternative metric is the  $F_1$ -score. However, it equally weights precision and recall and does not consider the different costs of the various inaccuracies [66]. That gap was filled using advanced metrics and graphical tools firstly by receiver operating characteristic (ROC) curves (ushered into the machine learning community by [61], and later extended the idea of convex hulls in [67]) and subsequently by precision-recall curves [68]. The state-of-the-art graphical tool is the cost curve [69], and it allows for comparing classifiers over a range of class or cost distributions and provides confidence intervals for a classifier's performance. Still, there is a gap in a scalable workflow to compare many classifiers for usage in real-time operation.

## 1.5 The challenge of managing risks of model usage

The challenge of managing risks when applying ML-based models in real-time operation can be underscored from a practical viewpoint. The control strategies proposed by these ML-based models may be ineffectual or infeasible to implement. For instance, grid topolog-

ical reconfiguration has been proposed as a cheap control mechanism to mitigate against voltage excursions, and line overloads [70], improve system security [71], and can even reduce operational costs and line losses [72]. Topological reconfiguration through transmission switching has previously been considered as an NP-hard mixed-integer problem that minimises the operational costs while meeting security requirements [73, 74, 75]. As the problem becomes multidimensional with the expansion of power systems, solving this linear and non-linear combinatorial problem tractably is an open challenge for solvers. Heuristics aiming to reduce the computational requirements are promising [76]. Even more promising is the use of ML-based models to instantaneously propose transmission switching solutions, and there, the SO needs to verify the feasibility of the proposed solutions as a way to manage the risks of using these ML-based models.

## 1.6 Objectives

The digitalisation of power systems has opened new avenues for using ML-based models to predict the uncertainty of power system operations and assessing the impact of dynamic phenomena on electricity grids in real-time. These models hold promise in proposing preventive and corrective control actions that can mitigate power blackouts. However, questions remain regarding the quality of training data which affects the the models' training procedures and quality of predictions. Furthermore, questions persist concerning the practical implementation of these models in real-time operations, including model selection and model usage. As such, this thesis aims to address these outstanding issues and provide answers to the following questions:

- How can a representative database be generated for training machine-learned models to accurately predict dynamic behaviour?
- How can machine-learned models be utilised for real-time operation, including model selection and modal usage in the control room?

The contributions of this thesis is to answer these questions with novel methods and prac-

tical examples in the following.

## 1.7 Contributions

The first contribution addresses the first question of generating a representative database. Consequently, this thesis develops a novel method that sequentially generates physically feasible but 'rare' OCs that have not been previously observed in power systems operations. The developed method aims to move beyond the limitations of historical observations to explore the full range of possible OCs, by efficiently analysing previously generated OCs within the feasible space of power systems operations. Further, this thesis develops an approach to efficiently combine the three properties of quality datasets, viz coverage, discriminative relevance, and historical relevance, trading-off historically relevant with feasible but 'rare' OCs using the Wasserstein distance, and considering discriminative relevance using entropy-informed sampling to create an enriched database capable of training a high-performing classifier.

The second contribution addresses the second question of using machine-learned models in the control room for real-time operation. This thesis proposes a scalable workflow to trade-off between multiple objective criteria of accuracy, interpretability and cost sensitivity using graphical methods when selecting machine-learned models for real-time operation by system operators. Additionally, this thesis showcases a practical implementation for using a machine-learned model to optimise the cost of power system operations using topological control actions. The proposed heuristic is more robust to deviations in the training data and provides significant computational savings with similar accuracy performance as state-of-the-art heuristics for transmission switching.

## 1.8 Structure of thesis

Chapter 2 provides an overview of power system stability and security assessment, as well as ML-based approaches to DSA. The first contribution of this thesis is presented in Chapter 3, which introduces a novel generic sampling method that generates representative and feasible OCs. In Chapter 4, a novel approach to generating quality datasets capable of training high-performing classifiers is presented, which trades off three properties of the quality datasets. Chapter 5 presents the second contribution of this thesis, a scalable workflow to select machine-learned models in real-time while considering multiple objectives. Further, as part of the second contribution, Chapter 6 provides a practical implementation for ML-based approaches in real-time operation, proposing topological reconfiguration to optimise dispatch costs. Finally, Chapter 7 concludes this thesis by summarising the thesis achievements, underscoring the role of ML in securing low inertia systems, and providing an outlook for future research.

# Chapter 2

## Background

*“The two most important days in your life are the day you are born and the day  
you find out why.”*

— Mark Twain

### 2.1 Introduction

The electric power system is an indispensable infrastructure that sustains modern life in the 21st century. It is responsible for transporting electricity across a network of lines and cables to power homes, industries, and critical sectors, such as transportation and health, thereby supporting economic activities. The effective operation of power systems requires the continuous balancing of power demand with generated energy. Hence, conventional power systems are powered by generators with prime movers that can accelerate or decelerate to match the variation of power demand. The system operator (SO) sits at the epicentre of this infrastructure, ensuring rigorous standards related to power quality are satisfied and overall reliability, including maintaining a constant frequency and operating within acceptable voltage levels.

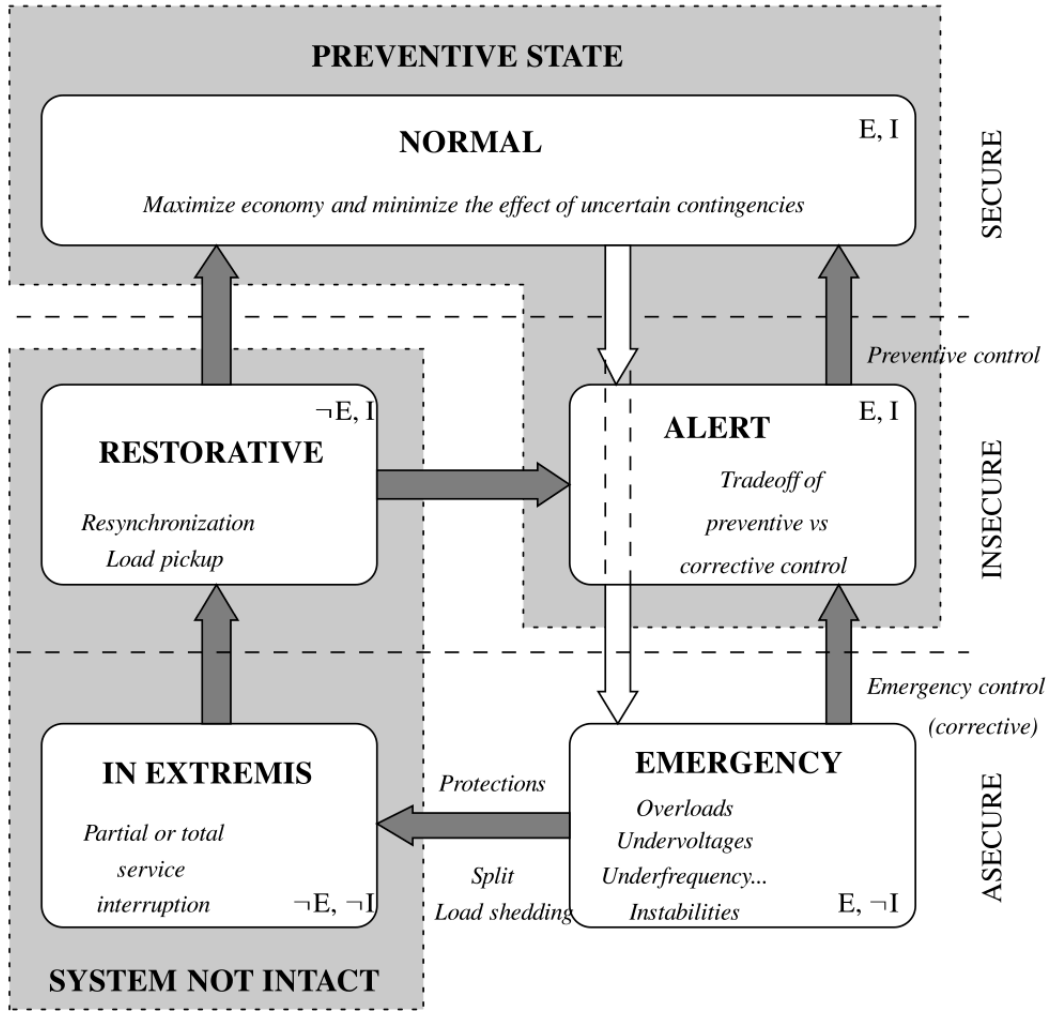


Figure 2.1: Power system operating states and control. The system responds to disturbances ( $\longleftrightarrow$ ) with control actions ( $\longrightarrow$ ), while satisfying the equality (E) and inequality (I) constraints. Figure gotten from [1].

The reliability of the power system measures the preparedness and ability of the system to maintain electricity supply without disruption despite foreseen or unforeseen disturbances [1]. From a practical perspective, power system reliability covers adequacy and security. Adequacy is the ability to maintain electricity supply in the event of equipment failure, for instance, generator or transmission line tripping. Security measures the elasticity of the power system to survive sudden disturbances without failing as it transition between different operational states.

On a conceptual level, there are five operational states of the power system, each with its

associated control objective to ensure reliable operation of the power system [2]. These operational states are presented in Figure 2.1. The *normal state* is the default operation state where all system variables are within acceptable operational limits. The control objective is to maintain system variables close to their desired values and be able to survive sudden disturbances. The system triggers the *alert state* if the security level drops beyond a specified threshold or if unfavourable weather increases the likelihood of disturbances on the system. In the *alert state*, while the system operates within acceptable limits and all constraints are met, the system strength has dwindled such that equipment overload from a contingency event is highly likely, leading to an *emergency state*. The control objective is preventative action such as generation re-dispatch to restore the system to the *normal state*. The system remains in the *alert state* if such restorative efforts fail to the system to the *normal state*. In the event of a disturbance of sufficient severity while the system is in the *alert state*, the system triggers the *emergency state*, where equipment overloads beyond the short-term emergency ratings. The control objective is active emergency corrective actions such as excitation control or generation tripping to return the system to the *alert state*. Otherwise, the system reaches a critical state and becomes *in extremis*, leading to cascading outages and potentially causing partial blackouts. Load shedding and controlled islanding can be used to contain the spread of the blackout from crippling the entire system. After a partial or full blackout, the SO restores the system with synchronised load pick-ups and re-connection of generators and aims to transition the system operation to the *alert state* or *normal state* [2].

This chapter presents some background on power system stability and security. Subsequently, this chapter discusses security assessment, with a focus on machine learning-based approaches to power system security assessment.

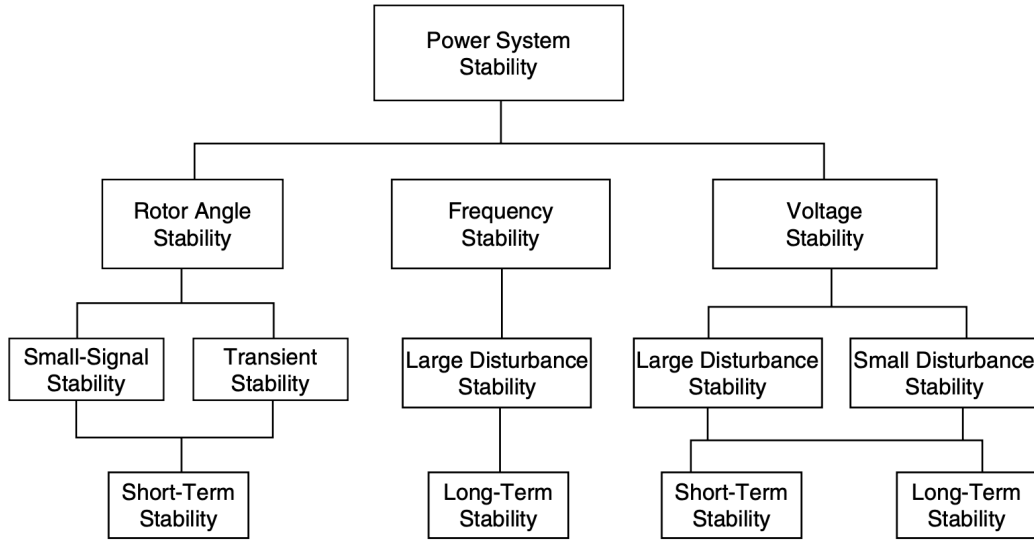


Figure 2.2: Classification of power system stability [2].

## 2.2 Power system stability

Power system stability is an inherent characteristic of a power system that enables it to response to small or large disturbances [77]. It is the ability of a power system to maintain an equilibrium state under normal operating conditions and to recover to an acceptable state of equilibrium after being subjected to a physical disturbance. Fig. 2.2 shows a classification of power system stability of conventional power systems. The stability of conventional power systems with large synchronous generators is primarily focused on ensuring synchronism between all generators, where the dynamics of the generator rotor angle provide insight into the stability of the system. Power systems are highly non-linear high-dimensional multi-variate systems whose dynamic performance is determined by a range of factors, including the wide array of devices with different response rates and inherent characteristics, system topology, initial operating condition and the type of disturbance. These systems use their spinning reserve to continuously modify generator output to match small load variations. For large disturbances, such as loss of generation or three-phase faults, the system response is influenced by the equipment, as the protective devices for individual equipment may respond to variations in system variables and influence system performance. For instance, a fault on a critical component is followed by protective relays isolating the faulted item to

protect the power system. This isolation can cause variations in bus voltages and machine rotor speed. Voltage variations can in turn trigger voltage regulators on the transmission lines and generators. The generator speed variations can actuate prime mover governors. Further, variations in voltage and frequency can trigger different behaviours on individual protective equipment, which ultimately affects system performance. Albeit, only a handful of equipment has a substantial impact on system stability.

There are various forms of instability the power system may experience depending on the physical nature of the phenomena that cause the instability, the size of the disturbance and the underlying equipment. The following sections explore some of these stability categories relating to voltage and rotor angles.

### **2.2.1 Voltage stability**

Voltage stability refers to a power system's capacity to maintain steady and acceptable voltages at all system buses under normal operating conditions and after being subjected to a disturbance [77]. When a disturbance such as an increase in load demand or change in system conditions leads to a progressively worsening and uncontrollable drop in voltage, the system enters a state of voltage instability. The primary cause of voltage instability is the power system's failure to provide the necessary reactive power to meet demand. In the short term, voltage stability is related to the response of voltage controllers such as automatic voltage regulators on generators or power electronic converters in flexible AC transmission systems. Voltage instability is usually contained within a limited region of the power system. However, its effects can have a far-reaching impact. A simple criterion for voltage stability is inferred from the V-Q sensitivity, such that the system is unstable if the V-Q relationship is negative for at least one bus of the system. Otherwise, the system is stable. Voltage instability can also occur as a result of generator rotor angles going out of step. Hence, voltage instability often coexists with rotor angle instability, and one may lead to the other.

### 2.2.2 Rotor angle stability

Rotor angle stability refers to the ability of a power system's synchronous machines to continue operating in sync [77]. The maintenance of synchronism is crucial as the electric power output of synchronous machines varies with respect to their rotor movements. A typical synchronous machine has two fundamental parts, the field, which is mounted on the rotor, and the armature, found on the stator. The field winding is energised by a direct current source. When the rotor rotates due to a prime mover such as a turbine, the rotating magnetic field from the field winding induces alternating voltages in the stator's three-phase armature windings. The frequency of the induced voltages and the resulting currents that flow when a load is connected are dependent on the rotor's speed.

To maintain synchronism when multiple synchronous machines are linked, the stator voltages and currents of all the machines must share the same frequency, and the mechanical speed of each rotor must be synchronised with this frequency. Hence, all rotors of the interconnected synchronous machines must be in sync. When the power system is operating under normal conditions, a three-phase current flows in the stator armature windings, inducing a magnetic field whose frequency is in sync with the rotor speed. The interaction between the rotor and stator fields results in an electromagnetic torque. For a synchronous generator, this electromagnetic torque opposes the rotor movement, and a mechanical torque must be consistently applied through the prime mover to sustain the rotor movement. Thus, the mechanical torque from the prime mover changes the position of the rotor relative to the revolving stator magnetic field and controls the electrical power output of the generator. The amount of electrical power generated depends on the angular separation between the rotor field and the revolving stator field rotating at the same speed. The converse is true for a synchronous motor, where the electromagnetic torque induces rotation, and the mechanical loads retards rotation.

The amount of power transmitted from a generator to a motor (infinite bus or load) is dependent on the angular separation between the rotors of the two machines, characteristic

by a nonlinear sinusoidal relationship.

$$P = \frac{V_g V_m}{X_T} \sin \delta \quad (2.1)$$

where  $P$  is the transmitted power,  $\delta$  is the angular separation between the two machines,  $X_T = X_g + X_m + X_l$  is the sum of generator, motor, and transmission line reactance, and  $V_g, V_m$  are the generator and motor voltages, respectively. The power transfer  $P$  increases linearly with an increase in angle until the maximum transfer limit at  $90^\circ$ , after which the power transfer decreases with an increase in angle. When multiple machines are present, the exchange of power is similarly influenced by their relative angular positions. However, the limiting values of power transfers are a more complex mechanism of generation and load distribution [2].

In a system with multiple machines, a steady-state operating condition is when the balance between the input mechanical torque and output electrical torque of each machine is maintained with their speeds remaining constant. Such a system operates in equilibrium. This equilibrium is disturbed when a perturbation results in the acceleration or deceleration of the rotors of the machines. If one generator rotates faster than another, the relative position of its rotor compared to that of the slower machine will change. This angular separation results in load transfer from the slower to the faster machine, which narrows the speed difference and, in turn, the angular separation. However, when the angular separation exceeds a certain threshold, the power transfer decreases as the angular separation increases, leading to instability. The stability of the power system is determined by the balance of forces that are either restoring or increasing the angular separation of the rotors. The loss of synchronism can occur either between one machine and the rest of the system or between groups of machines. In the former, the discrepancy between the stator field, which is in sync with the system frequency, and the rotor field results in significant variations in the machine's power output, which triggers the protection system to disconnect the unstable machine or out of step protection. In the latter, synchronism can be maintained between the different machine groups.

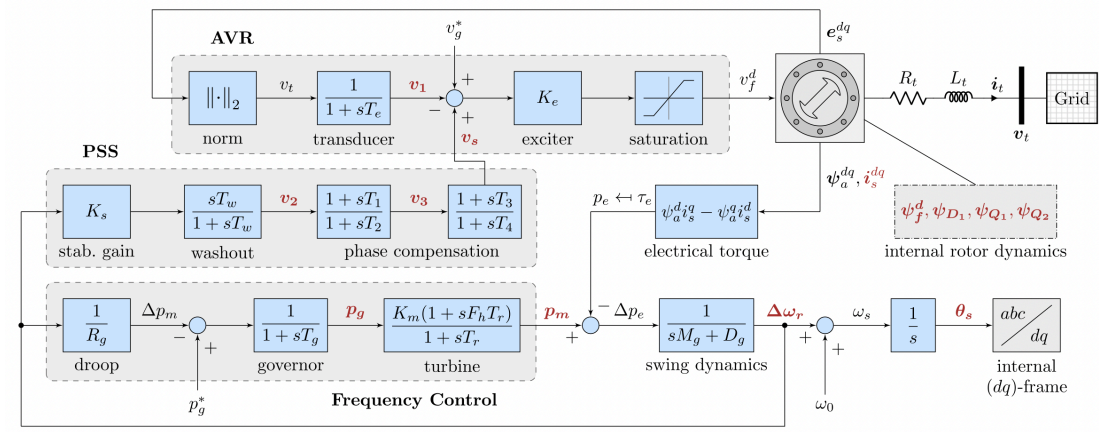


Figure 2.3: A synchronous generator showing the prime mover, governor and voltage excitation system. The internal state variables are denoted by red color. [3].

Rotor angle stability can broadly be categorised into small-signal and transient stability [2]. Small-signal stability is the ability of a power system to survive small disturbances usually due to a slight load of generation variations without losing synchronism. The analysis of small-signal stability involves linearization of the system equations as the disturbances are considered to be sufficiently small. The system response to these small disturbances depends on the initial operating condition, the system strength and the generator excitation mechanism. Instability arises due to either a lack of synchronisation or damping torque, which results in an increase in the rotor angle and amplitude of rotor oscillation, respectively.

The dynamics of a traditional two-pole synchronous generator can be described by the swing equation, given as follows:

$$M_g \dot{\omega}_r = \Delta p_e - D_g (\omega_r - \omega_0) \quad (2.2)$$

where  $M_g$  and  $D_g$  represent the inertia and damping constants, respectively, and  $\Delta p_e$  denotes the difference between the mechanical and electrical power at the output of the generator [3]. Fig. 2.3 presents the block diagram and control structure of the considered synchronous generator connected to the grid. The model also shows a voltage excitation system consisting an automatic voltage regulator (AVR) and a power system stabiliser

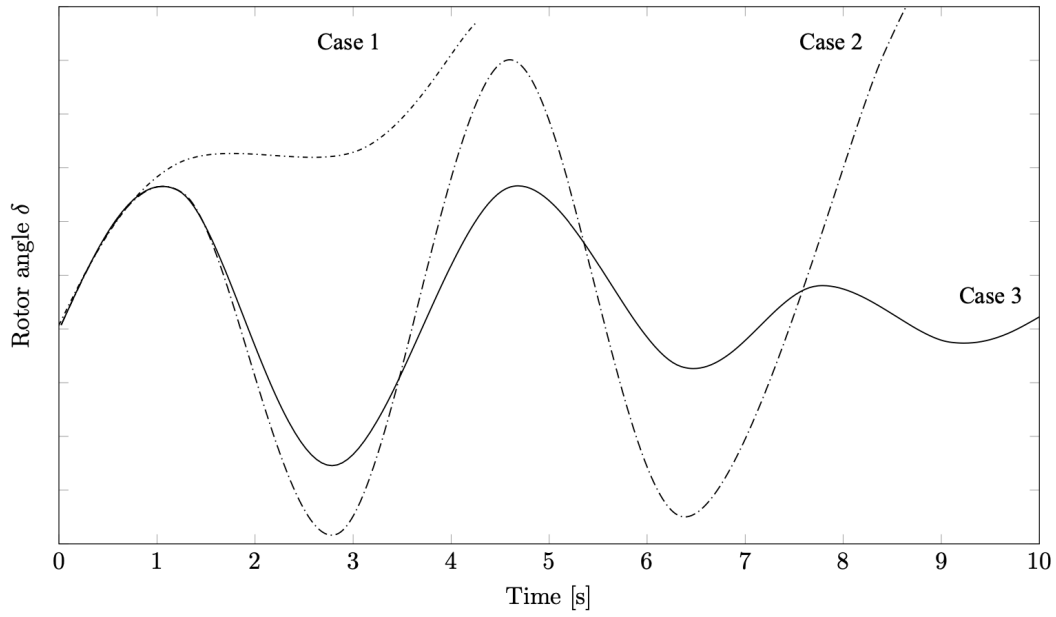


Figure 2.4: Rotor angle trajectory for different transient faults [2]. Increase in rotor angles in cases 1 & 2 show loss of synchronism and transient instability. Case 3 is transient stable.

(PSS), a prime mover and a governor.

Transient stability is discussed in the following.

### 2.2.3 Rotor angle stability

Transient stability is a type of rotor angle stability which refers to the ability of a power system to survive sudden large disturbances while maintaining synchronism [2]. The response of the system to such disturbances involves significant movements in the generator rotor angles and is shaped by the non-linear power-angle relationship. The stability of the system is dependent on the initial operating conditions and the intensity of the disturbance, often leading to a change in the post-disturbance steady-state operation compared to the pre-disturbance state. The system design considers a list of probable disruptions (contingencies) such that the operation is not disrupted when they occur. These contingencies are typically phase-to-ground, phase-to-phase-to-ground or three-phase short-circuits on the transmission lines or transformers, which trigger fault breakers to isolate the faulted item. An illustration of transient stability is presented in Fig. 2.4, which shows the behaviour of

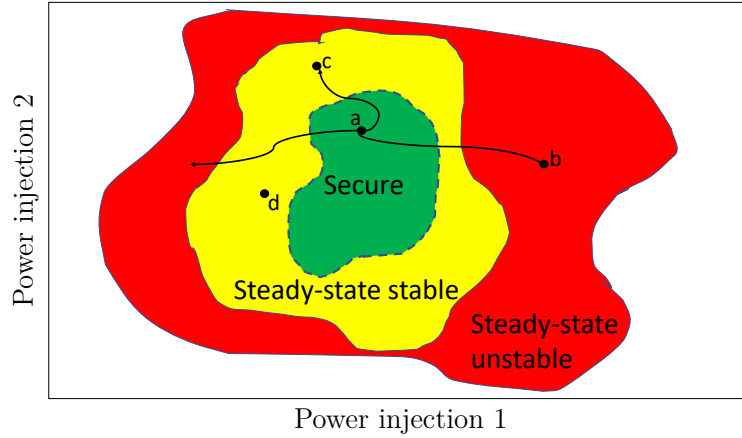


Figure 2.5: Different post fault trajectories from an operating condition (OC) (a), to an unstable OC (b), a static and dynamically secure OC (c), a static secure but dynamically insecure OC (d) [4].

the rotor angle of a synchronous machine under unstable and stable conditions. In case 3, the system is transient stable, as the rotor peaks, then decreases and swings with smaller amplitudes until it reaches a steady state. In both cases 1 and 2, the system is unstable. The rotor angle increases until it slides into instability, losing synchronism. Such a case (case 1) is termed first swing instability and can be mitigated with sufficient synchronising torque. In case 2, the system is stable during the first few swings but transitions into instability due to increasing oscillations in subsequent swings. Such a case of instability usually manifests when the system is also small-signal unstable. Hence, transient stability needs to be analysed for periods between 3 and 5 seconds for small power systems and up to 10 seconds for larger systems.

Transient stability differs from steady-state stability (small-signal stability) in that the former analyses if the system can reach a new steady-state condition after a contingency, while the latter only examines the existence of the new steady-state condition, not whether the system can reach it. The existence of a steady-state condition refers to maintaining energy balance without violating any physical network constraints. Fig. 2.5 illustrates this difference between transient and steady-state stability, where the axes show power injections at two buses of a power system that describe the subspace of feasible operational states. The full physical feasible space is a multi-dimensional hyperspace described by the voltages, active and reactive power generations, active and reactive loads and phase

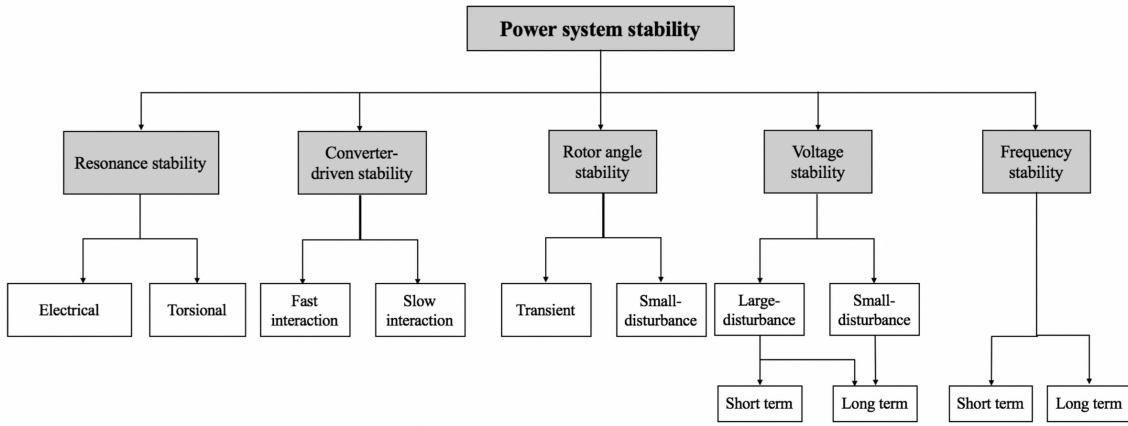


Figure 2.6: Extended power system stability definition covering resonance and converter driven instability. [5].

angles at each bus. The red, yellow and green areas show unstable, stable and secure regions, respectively. The initial operating condition (a) is at a steady-steady equilibrium and in the secure region. The trajectory evolution of the system from this initial operating condition, (a), following a contingency has three possibilities, shown as (b), (c) and (d). (b) is an unstable operation, as both the transient and steady-state conditions are not satisfied. (c) is both transient and steady-state stable, while (d) is steady-state stable but transient unstable at the trajectory does not settle to the post-fault equilibrium.

## 2.2.4 Extended stability definition

The increasing penetration of converter-interfaced generators (CIGs) into bulk power systems requires new considerations regarding the classification and definition of power system stability phenomena [5]. Compared to synchronous generators, CIGs exhibit different dynamic behaviour, resulting in the emergence of two new stability classes that must be considered: converter-driven stability and resonance stability (Fig. 2.6). Resonance instability arises from the resonance between series compensation and the mechanical or electrical characteristics of the generator, while converter-driven instabilities result from the cross-couplings of CIG control loops with both the electromechanical dynamics of machines and the electromagnetic transients of the network. Despite these new phenomena,

the traditional definition of power system security still applies if a fault affecting only the CIG does not lead to cascading instability in the main system. The conventional definition refers to the ability of a power system to withstand sudden disturbances without major service interruptions in real time. However, the integration of power electronics devices scales down the timescale of interest to electromagnetic transients, necessitating the consideration of the two new stability classes mentioned above. Regarding the existing stability classes, namely rotor angle, voltage, and frequency stability, the presence of CIGs does not affect their definitions. However, the fast control of CIG converters during and after faults can significantly influence the transient rotor stability. The integration of CIGs with fast-acting load components may also introduce new potential causes of voltage instability. Finally, CIGs are typically associated with renewable sources and cannot provide fast primary frequency response, which is crucial in low-inertia systems requiring fast controllers to prevent frequency drops and maintain system stability. Despite the lack of inertial response, the fast response provided by CIGs during faults can contribute to the stability of the system.

Fig. 2.7 presents the common control architecture for power converters, which is a two-level voltage source converter (VSC) where an outer system-level control provides a reference for the converter's terminal voltage that is subsequently tracked by a cascaded device-level controller. The block structure consists of the power converter model, system-level control, and device-level control. The power converter model consists of a DC-link capacitor, a lossless switching block, which transforms the DC voltage  $v_{dc}$  into a three-phase AC voltage  $v_{sw}$ , and an output RLC filter ( $r_f, l_f, c_f$ ).  $r_t$  and  $l_t$  are the transformer's per-unit resistance and inductance,  $v_t$  is the voltage at the connection terminal, and  $\omega_r$  is the normalised reference for the angular velocity of the dq-frame. From a system-level control perspective, the dynamic behaviour of the power converter is described by a dynamic controller. A three-phase power converter model can be operated in grid-forming and grid-following operation modes [3]. The two operation modes are shown in Fig. 2.8.

The detailed description of the dynamical model of CIGs, including a filter, transformer, AC-side controller and Phase-Locked Loop (PLL), can be found in [3].

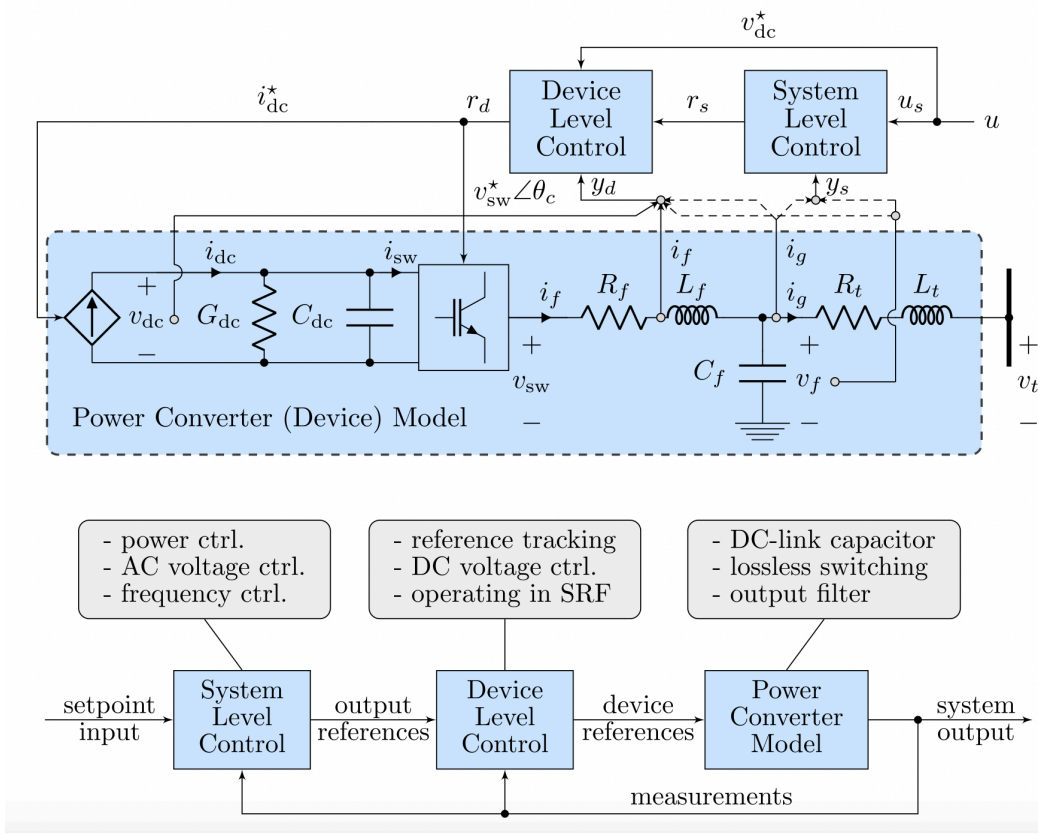


Figure 2.7: A two-level VSC (top); block structure of the power converter model and its control scheme (bottom). [3].

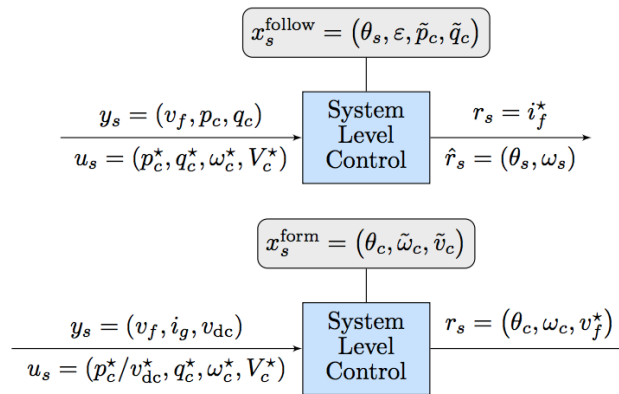


Figure 2.8: System-level control of grid-following operation mode (top) and grid-forming operation mode (bottom) [3].

### 2.2.5 Power system reliability

Stability is an inherent characteristic of a power system. Power system reliability, on the other hand, refers to the degree of assurance in delivering a constant supply of electricity to the end-users (adequacy) and surviving to withstand sudden and unforeseen disruptions or disconnections of some of its components (security) [1]. Adequacy focuses on the sizing of the system to handle the unpredictable and fluctuating demand from end-users, while also considering the occasional unavailability of system components. Adequacy planning considers a long time horizon that can range from several months to several years, and it is measured by metrics such as the loss of load probability or the expected energy not supplied. According to the North American Electricity Reliability Council, security is defined as 'the ability of the bulk power electric system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system components' [78]. Power system security focuses on the functioning of the system as it transitions between operational states when an unforeseen external disruption acts on the system. These disturbances commonly referred to as contingencies, cause large changes in operating conditions and are more often followed by relay operations to protect the system from abnormal conditions. The relay operations can result in the disconnection of the load, transmission line, generator or transformer. Power system security has stricter restrictions than stability (green and yellow regions in Fig. 2.5), as security requires that the system remains stable during the transient response and that all operating constraints are satisfied before and after a fault occurs. The security level of a power system is tied to a defined list of probable contingencies and is dependent on the limits set by the system operator to establish security standards. For instance, the green region in Fig. 2.5 contains the operating conditions that are secure against a list of contingencies. The yellow region contains operating conditions that are steady-state stable but insecure. The red region contains operating conditions that are both insecure and steady-state unstable. As stability is inherent to a power system, the demarcation between stable and unstable regions is immutable. However, the separation between secure and stable regions can change relative to the defined security level. The list of probable contingencies varies from one control area to another, which can include

$N-1$  security and more complex contingencies that include cascading faults.  $N-1$  security is typically standard practice, which secures the system against loss of a single component such as transformers, transmission lines or generators, where  $N$  is the total number of a component type.

Power system security can be categorised into static and dynamic security [1]. Static security refers to the system meeting all physical constraints in the steady state after a contingency. However, it does not address whether the system can make the transition from pre-fault to post-fault. Dynamic security is focused on whether the system can make that transition. Static security can be used as a first criterion to determine a system's response to a contingency, after which a more detailed dynamic simulation assesses the full dynamics of the contingency.

### 2.2.6 Security assessment

Security assessment is a prerogative of system operators to reliably operate power systems. It enhances the situational awareness by determining the security level of present and future operating conditions for a list of probable contingencies. The security criteria for normal operation consists of the following [78]:

- Transmission line flows within normal ratings pre and post fault.
- Bus voltage within 5 % (this can change between system operators) of nominal values pre and post fault
- Transient flows do not exceed transient limits (e.g. equipment ratings, system constraints from offline simulations in addition to safety margins).

These criteria entail investigating whether the system can withstand the transient response and reach a steady-state post-fault condition where all assets stay within their operating limits. Thus security assessment represents a mathematical analysis of the system response after a fault and the new equilibrium position. Security assessment has two components,

static and dynamic security assessment. Static security assesses whether the system arrives at a steady-state equilibrium (within 60 minutes) after a contingency [1]. The primary physical aspect of concern is whether the system components do not exceed their capacities, such as conductor current limits or electric insulation capabilities. In static analysis, the focus is on whether the system operating condition (OC) fulfills all physical limits in the post-fault state (OC c and d in Fig. 2.5). In practice, static security assessment involves solving a Power Flow problem to determine the post-contingency state for each component for a list of probable contingencies. Otherwise, a security-constrained Optimal Power Flow, for instance, [79], is solved to determine the minimum generation cost solution to guarantee  $N-1$  security subject to satisfying all network constraints such as power balance at each node (Kirchhoff's laws), line flow limits and generator physical limits. However, the system may be steady-state secure while the transient response of the system is insecure, as illustrated by OC c and d of Fig. 2.5.

Dynamic security assessment (DSA), which analyses the transients, is discussed in the following.

### 2.2.7 Dynamic security assessment

DSA measures the ability of a power system to successfully transition from a pre-fault operating state to a stable post-fault operating state for a list of probable contingencies. The formal definition by the Institute of Electrical and Electronics Engineers, Power Engineering Society's working group on DSA is, 'an evaluation of the ability of a certain power system to withstand a defined set of contingencies and to survive the transition to an acceptable steady-state condition' [77]. Dynamic security analysis focuses on whether the system survives the transition from the pre-disturbance to the post-disturbance condition (OC c in Fig. 2.5). The system's dynamic response is dependent on the initial OC, and hence, simulating many OCs for a wide range of contingencies can be computationally challenging, as the power system is highly nonlinear and these assessments require numerical analysis including detailed time-domain simulations [80, 81]. Numerical analy-

sis via time-domain simulations is challenging in real-time operations as the simulations require significant computational capacity [82], as each possible disturbance needs consideration as event-type perturbations, and many possible OCs need assessing, requiring a separate time-domain simulation. These simulations output the post-fault security status of the system, as secure or insecure, after the simulated fault. The computational burden is proportional to the size of the power system, the number of OCs and the number of contingencies. As dynamic security analysis is more difficult than static analysis, consequently, in practice, system operators often rely on static security assessments for a short list of severe faults and place large safety margins on their operating equipment to ensure a secure dynamic response of the system [83, 84]. These conservative actions underutilise the power system assets and can lead to economic loss, leading to inefficiently investing in asset redundancy to secure the system against future uncertainties and demand growth. A more efficient strategy would be to consider dynamic security as well as more contingencies to improve situational awareness such that the system operator can operate the grid closer to operational limits. Considering dynamic security enhances situational awareness and enables the closer operation to operating limits, thereby reducing the need for new investments. To assess dynamic security, all dynamic phenomena, such as rotor angle, voltage, and frequency stability, must be individually studied [81].

The challenge of DSA lies in the need to consider a large number of possible contingency scenarios, as a high number of considered contingencies correlate with a higher level of security. However, the contingency list goes beyond just single faults. Multiple assets failing simultaneously and cascading failures that are less likely but have severe consequences need to be analysed. Accordingly, there are different methods at the disposal of the system operator to decide which contingencies to analyse. These methods are broadly categorised under deterministic, probabilistic, heuristic, and data-driven methods [85].

Deterministic methods can be simulation-based, such as numerical integration or direct / Lyapunov methods. Numerical integration methods solve first-order differential equations that describe the dynamics of the system to obtain the solution of systems stability after a fault. However, such methods require significant computation power and thus occur

in the offline settings. Lyapunov methods, otherwise known as transient energy function methods, introduce stability criteria to replace the time-consuming numerical integration. The post-fault stability of the system is calculated by comparing the Lyapunov function at the instant to the critical Lyapunov value. However, estimating the Lyapunov function that accurately represents the internal characteristics of a power system is not trivial, and the method is limited to only inferring the stability of operating conditions within the estimated stability region. An example of a deterministic method currently used by system operators to ensure secure operation for each fault is the  $N-1$  criterion. However, the pitfall with deterministic security assessment is that it may not account for those contingencies that have a low probability of occurring but carry a high level of risk. Additionally, the risk and impact of a contingency can change over time due to factors such as asset degradation, weather changes, and changes in the system's network topology. Despite this, most operators continue to use a deterministic approach, which means they accept the potential impact of those contingencies not included on the list. However, probabilistic security assessment methods can be used to address the challenge of the vast number of contingency scenarios considering impact probabilities and risk.

Probabilistic security assessment evaluates the security of the system by taking into account the risk associated with each contingency, based on both their likelihood and impact. The contingencies with the highest impact are given priority in this evaluation. However, estimating the risk, which is time-varying, is a challenging task. For instance, the impact may be evaluated based on the violation of operational limits, while the probabilities are estimated based on actual weather forecasts. However, the actual impact is the cost of expected energy not supplied to the end-user for a given loss of load probability. Estimating these costs is more difficult than calculating the violation of operational limits. Although the probabilistic approach has been extensively researched for static security assessment, it is still not common practice for DSA. As the dynamics become more important in future power systems, new methods are required to study the dynamical phenomena.

Expert heuristics is another approach to determining the stability of the system, where a knowledge base is accumulated from the extensive expertise of system operators to synthe-

size a set of inferred rules for system operation. Even though these heuristics can be useful for normal power systems operations, the limitation of this approach is that the limits of the system can not be known, and therefore the system assets are not fully utilised to efficiently operate the grid.

This thesis focuses specifically on transient stability, and not all aspects of system dynamics are extensively examined or checked against. However, the developed methods in this thesis can still be used to investigate broader aspects of system dynamics

The limitations of the above approaches can be addressed with the data-driven methods. Data-driven methods, such as a machine learning-based approach to DSA, are discussed in the following.

### **2.2.8 Machine learning-based approach to dynamic security assessment**

There is a need to scalably assess the dynamic security of power systems to efficiently utilise grid assessment. However, the common practice amongst system operators is to use static security assessments with large safety margins. Current software for dynamic security assessment is unscalable as they have high computational demands from solving numerical integration [81, 86], making online dynamic security assessment with conventional methods such as Forward Euler or Runge-Kutta methods impractical [87]. Machine learning approaches have gained some traction in the last couple of years and may provide a scalable alternative to improve the situational awareness of the system operator to reliably operate the grid [1].

The machine learning-based approach to DSA is to predict the outcome of the stability analysis [22], and consists of two stages, offline and online. In the offline stage, the goal is to curate a database of many possible operating conditions with the corresponding information on whether the condition is secure or not for one or multiple contingencies. This database is used to train a model as a binary classifier for use in the online stage.

The online stage represents near to real-time operational setting, where the learned model takes the current operating condition as an input and outputs a prediction of the security level of the system. Although the model may sometimes be inaccurate, the main benefit is that operating conditions that were not part of the database can be taken as inputs, and most importantly, no computational time (for simulations) is needed in real-time operations. Whereas the DSA for large systems via simulation can take several seconds, for instance, 56 s for a single contingency and operating condition on French system [42], a machine learning model does it using considerably less time  $\ll 0.1$  s. As the prediction is instantaneously available, it allows the system operator (SO) to consider a large number of possible operating conditions and disturbances in real time and has the potential to increase the situational awareness. As an illustrative example, Fig. 2.9 depicts a simple case of the machine learning-based approach to DSA. All the circles (without colour consideration) represent pre-fault operating conditions, which form initial conditions to simulate the system's dynamic response to a specified contingency. Each operating condition is then classified into one of two classes, secure (black) or insecure (red), based on the system response to the simulated fault. A classifier then aims to learn the security boundary of the system (green line) from these pre-fault operating conditions and their security labels. Then, in a real-time setting, the classifier can instantaneously predict the security labels of new operating conditions (in blue circles). In this case, the classifier in Fig. 2.9 can predict the security labels of operating conditions labelled (A) with more confidence than operating conditions (X).

In light of this development, machine-learned models including decision trees (DTs)[88][28], support vector machines (SVMs)[89], and more recently deep learning models[90] have shown promise to assess dynamic stability problems ranging from voltage stability[28][91], transient stability[88][90] and frequency stability [92][93]. Recent works [94, 95, 96, 97] show real promise for real-time probabilistic DSA including considering topological changes [98, 99]. There, using ML, estimating the dynamic security boundary [100] particularly works well for future low inertia grids [101] and may become possible through the increasing availability of a large amount of phasor measurement units (PMUs) data and monitoring

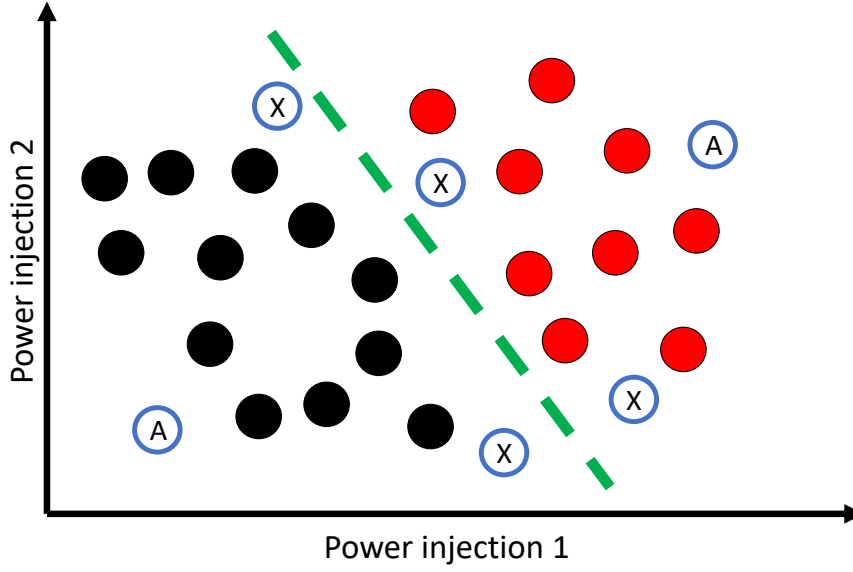


Figure 2.9: A visual illustration of the machine learning-based approach to DSA. In the offline stage, the model learn the security boundary (■ ■ ■ ■) from the training database containing secure (●) and insecure (●) operating conditions. In the online stage, the classifier predicts the security labels of new operating conditions (A) (X) with varying degrees of confidence.

tools in control centers[102].

### 2.2.9 Machine learning-based DSA Workflow and contributions

Machine learning-based security assessment  $X \rightarrow \mathcal{Y}$  takes as an input the power system operating conditions  $X$  for a given network topology  $\Psi_{TOP}$  and outputs security labels  $\mathcal{Y}$  for a set of probable contingencies  $\Omega^J$ . The power system topology  $\Psi_{TOP,a,b} \in \{0,1\}$  represents the interconnection of lines connecting all assets in the network, where  $a,b \in \Omega^N$  are buses and  $a \neq b$ . The security labels  $\mathcal{Y}_i^j \in \{0,1\} \forall j \in \Omega^J$  represent secure and insecure operating conditions, respectively, where  $i$  is the index of operating condition  $X_i$ . Typically, the input variables are the static pre-fault set-point of all generators and loads. These input variables define the operating condition  $X_i$  and are bounded by the power system's physical limits, such as generator limits, line limits, and complex network

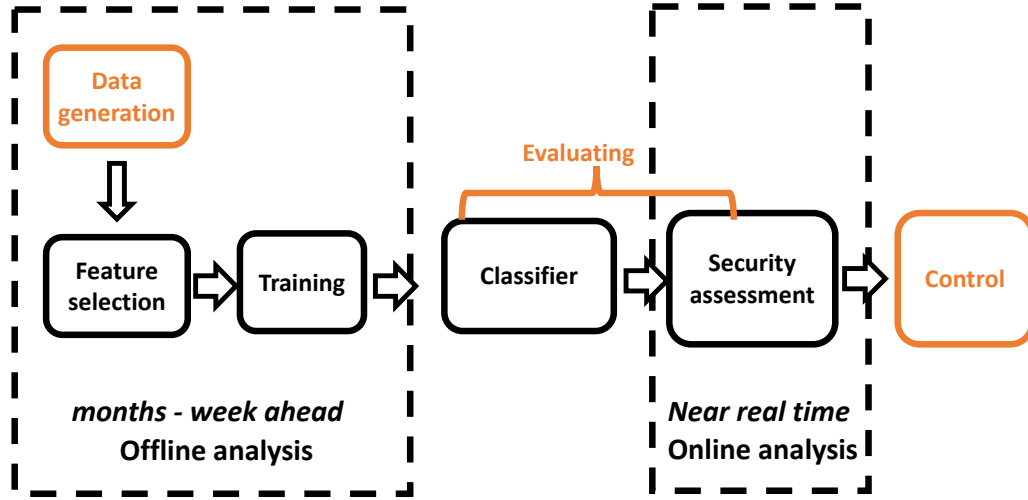


Figure 2.10: Machine learning workflow for dynamic security assessment. The orange coloured areas are investigated in this thesis

constraints.

For a more concrete representation, Fig. 2.10 presents the machine learning-based DSA workflow showcasing the following four major steps: data generation, data pre-processing (feature selection), training the machine learning model, and evaluating the trained model. The contributions of this thesis are in the first and fourth steps.

In the first step, the training database can be extrapolated from historical observations that may be biased toward secure operating conditions [21] or can be generically created in a balancing way as done in [55, 59]). An operating condition is described by the phase angles, voltage magnitudes, active and reactive power injections at every bus and power flows on the transmission lines. This first step is crucial as the prediction performance of machine learning models is generally a reflection of the quality (coverage, variability, and balance) of the data used in training [41]. In Machine learning-based DSA, the choice of training data differs from other machine learning applications where historical observations is sufficient. Often, the majority of historical observations which characterises normal power system operation are secure, however, an effective training database needs to consider both, secure and insecure conditions [1]. Moreover, historical operating conditions

rarely involve extreme operating scenarios. Hence, sampling approaches are used to synthetically generate operating conditions. Subsequently, the pre-fault operating conditions (for instance load scenarios) are assessed with the stability analysis for a single or set of disturbances, and a post-fault metric for stability/security is adopted (e.g., as in [86]). The set of contingencies that need to be taken into account is determined by the operator and typically includes the most significant equipment failures, such as major transmission lines or large generators. As previously mentioned, the system's response to each contingency corresponds to binary labels, with 1 representing an insecure response to a contingency and 0 representing a secure response. For instance, operating conditions (A) and (X) in Fig. 2.9 can be synthetically generated and verified to fulfill the non-convex AC model of the optimal power flow. Subsequently the system response to various contingencies is recorded. The combination of pre-fault operating conditions and the post-fault metric makes up the training database. Generating synthetic operating conditions that cover a large section of the feasible space is a first pre-step that ensures that a more accurate model is trained. As discussed in the introduction, the state-of-the-art approaches that move beyond historical observations to generate synthetic operating conditions fall under one of three approaches: *historical sampling*, *generic sampling* and *importance sampling*, each typically focusing on maximising a specific property of what constitutes *quality* datasets, respectively, *historical relevance*, *coverage* and *discriminative relevance* [42, 1]. *Historical relevance* shows how much the generated operating conditions represent historical power systems operations through variable dependency structure. *Coverage* measures how much of the power system feasible region the generated operating conditions span. *Discriminative* relevance depicts how much new information the generated operating conditions add to training a model. However, there are pending unaddressed challenges to curating *quality* datasets for security assessment. Consequently, this thesis investigates methods to generate representative and quality datasets. Additionally, this thesis investigates trade-offs between historically relevant operating conditions with 'rare' but feasible operating conditions.

In the second step, the data is pre-processed [99] for the concurrent training of the model. The database is analysed to select or extract relevant features for reducing the number

of dimensions [103] and representing the information in a lower dimension [40]. This pre-processing, for example, the Synthetic Minority Oversampling Technique (SMOTE) [104], can also balance the database as often more secure than insecure data is available. This step is not the focus of the thesis and we adopt the approaches in the literature in our studies.

In the third step, the machine learning model is trained. Decision trees (DTs) are often used because of their high interpretability, which is crucial for such a critical task as security assessment, where operators require a manual inspection to understand and rely on these machine learning models. Typically one DT model is trained per disturbance as the stability/security is different for each disturbance [22, 42]. This step is not the focus of the thesis and we adopt the approaches in the literature in our studies.

In the fourth step, the models are evaluated, selected, and eventually updated. The selection and evaluation of the model involve finding the model with the highest performance out of a large set of trained models. Typically, the performance is measured by testing how the model performs on data that is not part of the training. For instance, a testing set is used to compute the testing error (the ratio of inaccurate predictions). The case of a binary classifier, for an operating condition  $i$  described by input features such as active and reactive power injections  $X_i$ , the prediction  $f(X_i) = \{0, 1\}$  can either assess the system security as secure  $f(X_i) = 0$  or insecure  $f(X_i) = 1$ . Whereas the true label for each operating condition for a contingency  $j$  is denoted as  $\mathcal{Y}_i^j = \{0, 1\}$ , respectively for secure and insecure labels. Therefore, there are two types of accurate and inaccurate predictions. The accurate predictions can be true positives (TP) or true negatives (TN), which denote accurately predicting secure and insecure labels, respectively. On the other hand, there are two types of incorrect predictions - false positives (FP), where an insecure operating condition is wrongly predicted as secure, and false negatives (FN), where a secure condition is wrongly predicted as insecure. These two errors are more commonly referred to missed (FP) and false alarms (FN). Table 2.1 presents the different accuracies and inaccuracies and their associated costs. Using this notation, the error on the testing set is calculated as the ratio of the incorrect predictions over all the predicted values

$\frac{FP+FN}{TP+FP+TN+FN}$ . However, in the context of DSA the testing error may be inadequate to quantify the performance of the model as class and cost imbalances are present [105]. The class imbalance results from having more historically secure than insecure OCs, while the cost imbalance results from the skewed cost disparity, where missed alarms are much more "costly" than false alarms. This thesis investigates scalable approaches to evaluate and select tree-based models considering multiple objectives of the system operator for the task of DSA.

	$\mathcal{Y}_i^j = 0$	$\mathcal{Y}_i^j = 1$
$f(X_i) = 0$	$TP, C_{TP}$	$FP, C_{FP}$
$f(X_i) = 1$	$FN, C_{FN}$	$TN, C_{TN}$

Table 2.1: Confusion and Cost Matrix

Other performance measures like the  $F_1$  score [66]

$$F_1 = \frac{TP}{TP + 0.5 \times (FP + FN)} \quad (2.3)$$

allows for a harmonic balance of the precision and recall for different errors [106] as used in [29], or the G-mean score that computes the geometric mean as used in [28]. Also, graphical approaches can be used to select models, such as the precision-recall (PR) curve [107] or receiver operating characteristic (ROC) curve [108] as applied to DSA in [31, 109]. However, selecting a model based on a single criterion may be sub-optimal. In DSA, there is need to trade-off the following three metrics: accuracy, interpretability and cost sensitivity. Firstly, predicting errors for the different classes can have various impacts. A missed alarm is much more severe than a false alarm ( $C_{FP} \gg C_{FN}$ ). Missing an alarm can result in power blackouts and load shedding that have high expected costs, however, a false alarm may require only preventive and corrective control measures (e.g., generation re-dispatch) to be taken that are significantly cheaper. Secondly, it is crucial to consider the interpretability (and complexity) of the selected model. Models that are high in their complexity are not interpretable for operators [1] and this renders them unsuitable for the application to DSA. Operators responsible for the critical task of security assessments

may prefer interpretable models in their decision-support tools such that manual inspection remains possible and errors can be identified [42]. Recent work on SHAP explanations [110] are promising to improve the interpretability of complex models. Thirdly, it is also crucial to consider frequent changes in the system. Frequent changes in system parameters may require to change the selection of models. For instance, the weather changes frequently, and with that, the likelihood of contingencies [111, 112]. If the probability is high, an operator may select more conservative models than at times with low likelihood. In practice, 1000s of models may be used in real-time [42] and a fast, adaptable selection process is needed.

This thesis also studies a real-time application of machine learning models for power system control, specifically real-time topological reconfiguration. There, machine learning can improve the solution times to propose feasible topology  $\Psi_{TOP}$  much faster than Fisher's [73] canonical NP-hard mixed integer problem to transmission switching.

This chapter provided an overview of power system stability and security assessment, and introduced ML-based approaches to DSA in power systems. ML-based approaches can predict the security level of the system based on current OCs and address the challenges of scalability and real-time operability of traditional time domain simulations or the practice of using static security assessments with large safety margins. The chapter also discussed the steps involved in the ML-based DSA workflow, including data generation, data pre-processing, model training, and model performance evaluation. The next chapter discusses the first contribution of this thesis on data generation.

## Chapter 3

# Generation of Diverse Datasets

*“Not all those who wander are lost.”*

— J.R.R. Tolkien, *The Fellowship of the Ring*

Machine learning (ML) for real-time security assessment requires a diverse training database to be accurate for scenarios beyond historical records. Generating diverse operating conditions is highly relevant for the uncertain future of emerging power systems that are completely different to historical power systems. In response, for the first time, this work proposes a novel split-based sequential sampling approach based on optimisation that generates more diverse operation scenarios for training ML models than state-of-the-art approaches. This work also proposes a volume-based coverage metric, the convex hull volume ( $\mathcal{V}$ ), to quantify the quality of samplers based on the coverage of generated data. This metric accounts for the distribution of samples across multidimensional space to measure coverage within the physical network limits. Studies on IEEE test cases with 6, 68 and 118 buses demonstrate the efficiency of the approach. Samples generated using the proposed split-based sampling cover 37.5% more volume than random sampling in the IEEE 68-bus system. The proposed  $\mathcal{V}$  metric can assess the quality of generated samples (standard deviation of 0.74) better than a distance-based coverage metric which outputs the same value (standard deviation of  $< 0.001$ ) for very different data distributions in the IEEE

68-bus system. As this thesis demonstrates, the proposed split-based sampling is relevant as a pre-step for training ML models for critical tasks such as security assessment.

## 3.1 Introduction

As discussed in Chapters 1 & 2, the integration of renewable energy introduces uncertainty in power systems operations, challenging reliability management. Conventional approaches with large safety margins are inefficient. The availability of real-time operation data, for instance, from PMUs [113] allows for carrying out state estimation [114] and the subsequent dynamic security analysis (DSA) of the system. ML is particularly promising as it allows for predictions in real-time with little computational time [1]. A functioning ML-based DSA tool has the potential to increase situational awareness, support reliability management, improve uncertainty handling, and efficiently integrate more renewable energy. However, as these ML-based approaches to DSA are data-driven these approaches can only be as good as their training database. Using only historical observations as training data is insufficient [51, 43]. Therefore, the generation (sampling) of data is highly relevant for the success of all aforementioned approaches [52].

### 3.1.1 Sampling approaches

The prediction performance of ML models is generally a reflection of the quality (coverage, variability, and balance) of the data used in training [41]. The choice of training databases in DSA application differs from other ML applications that use (recorded) observations. Using recorded, historical data for DSA has limitations as highlighted in Chapter 2. A good training database needs to consider both secure and insecure conditions [1] while the majority of historical observations are often secure. Also, historical OCs rarely involve extreme operating scenarios. Hence, sampling approaches are used to generate synthetic OCs. When generating synthetic samples, firstly, an OC is sampled, then it is assessed with a time-domain simulation for the considered contingency.

The generation of data for ML-based DSA is highly relevant which is why many contributions were made along three types of approaches: the first type of approach, *historic sampling* uses historical records [43], fits a probability distribution to it (e.g., vine-copulas in [21][40] to capture the dependencies between loads and wind power outputs), then generates OCs using Monte-Carlo (MC) type samplings [46][47]. This type of approach is suitable to sample OCs following the same distribution as historical observations. Another variant of *historic sampling* determines the 'relevant' buses to obtain sparse PMU measurements. By selecting subsets of these 'relevant' buses for sampling, the issue of high dimensionality can be mitigated as only a smaller ('relevant') dimension of variables need to be sampled as shown in [44, 45]. However, future OCs may be different than historical OCs, and sampling from distributions is unsuitable for creating extreme OCs typically found at the tails of distributions. The second type of approach, *importance sampling* is where the sequence of sampling and classifier training iteratively repeats to maximise high information content [39, 37, 48, 49, 50]. In each iteration, the sampling (e.g., with MC-sampler) generates possible OCs. Then, the classifier quantifies the importance of these OCs based on the predicting confidence. Subsequently, the security assessment is used only on samples with low confidence. For instance, Yan [51] uses entropy as a metric to generate 'relevant' samples closer to the decision boundary. [52] uses 'directed walk' methods to samples around the decision boundary. The third type of approach, *generic sampling*, generates points uniformly distributed in the feasible space to explore all possible OCs. However, large systems require large amounts of generated data, and most data adds little knowledge to the database. For instance, Jafar [53] uses the Latin hypercube sampling (LHS) approach to uniformly sample the entire search space, and researchers in [54] sample within the feasible neighbourhood of OCs, while researchers in [55] proposed an outer approximation to convexify the original non-convex feasible space, then sample from the convex region to generate samples close to the security boundary. Venzke [56] uses infeasibility certificates based on separating hyperplanes to discard large portions of the input space as infeasible. More recently, the authors in [57] developed a framework to generate representative samples that span the AC OPF feasible space by uniformly sampling loads

from a convex input space and using infeasibility certificates to reduce the search space. The drawback of the first and second type of sampling approaches is that they neglect some feasible OCs. The first approach is biased towards historical observations, and the second towards the importance of learning the security boundary. Hence, sampling extreme OCs with those approaches is rare. However, studying extreme OCs beyond historical records and probabilistic approaches that approximate historical distributions is crucial as these can be dangerous for system operations. The challenge of the third type of approach is that sampling in high-dimensions is not trivial. Therefore, a current research gap and need is an efficient *generic sampling* approach that scales to larger systems and can generate extreme synthetic OCs as the introduction of intermittent renewables into the energy mix means that the power system will experience new OCs that were historically not covered and cannot be generated by statistical methods.

Other fields faced with similar sampling challenges from large solution spaces have proposed novel approaches. In particular, bio-engineering employs random sampling  $\mathcal{RS}$  techniques to investigate constraint-based metabolic reactions that have a large solution space. A popular sampling technique often employed is the family of "hit-and-run"(HR) samplers (ACHR([115]), CHRR([116])) that randomly choose directions to traverse a model's solution space based on warm start positions. This approach relies on the convexity of the solution space and requires relaxation of non-convex models as found in the II-ACHR sampler [117]. A new approach called GAPSPLIT was introduced to sample models directly [118]. The sampler generates points by jumping to unexplored regions of the space in contrast with the random walk approach employed by HR samplers. GAPSPLIT is a competitive alternative to HR samplers that scales relative to the size of the model and can sample directly from non-convex models. Samples generated with GAPSPLIT also have better coverage than ACHR and CHRR on unbounded model variables. The approach was used in [119] to sample a highly constrained solution space. The ACHR and CHRR are not tested as they require convex models or convex approximations.

The state-of-the-art methods in the literature focusing on generating data for a training database of ML-based DSA is presented in Table 3.1. The proposed approach is fundamen-

Table 3.1: Summary of relevant state-of-the-art works on database generation for ML-based DSA

Reference	Type	Sampling of initial OCs	Advantages	Shortcomings
[57]	Generic sampling	Solving OPF to minimise generation cost	Uses convex relaxations and hyperplanes to discard large sections of the input space. Explores load space via MC sampling	Only considers generator outputs obtained from solving OPF to represent conventional operation, which is a small subset of the feasible space.
[56]	Generic sampling	LHS or uniform sampling.	Systematically covers the search space with uniform sampling while discarding large hyperplanes of infeasible regions	Fitting a multivariate distribution around secure OCs only generates OCs of similar distribution and not other possibilities.
[55]	Generic sampling	LHS	Systematically covers the search space with uniform sampling while discarding hyperspheres of infeasible regions	Discarding hyperspheres of many initialisation points in high-dimensions is not computationally trivial
[37]	Importance sampling	Solving OPF to minimise generation cost	Sampling close to the decision boundary is computationally efficient	Biassing the sampling towards the security boundary ignores rare OCs. Only considers OPF solutions
[21]	Historical sampling	Solving OPF to minimise generation cost	High density sampling of OCs from historical records.	Neglects unseen or rare OCs that are critical to be analysed.
[52]	Generic + Importance sampling	Grid search, uniform sampling in each dimension or LHS	Focuses on sampling close to the decision boundary using enhancement methods such as directed walks, the prediction model as a pre-selection tool for relevant samples and performance guarantee of entire regions.	Relies on resampling techniques to bias sampling to narrow regions of the space. Using performance guarantees significantly reduces the search space and can affect model performance.
[120]	Generic sampling	Sequentially generated to explore the feasible space	Sequentially explores the feasible space to maximise distance from generated samples.	Performance on larger systems (bus $\geq 68$ ) is not tested.
[48]	Importance sampling	Solving OPF to minimise generation cost	Computes quadratic approximation of the security boundary and use importance sampling to generate OCs	Dataset represents only a small portion of the feasible space.
[39]	Importance sampling	Solving OPF to minimise generation cost	Identifies the decision boundary and fits a polynomial function so as to sample OCs close to the boundary	Dataset represents only a small portion of the feasible space
[51]	Importance sampling	LHS	Uses a transient stability index to sample in a high-information region formulated as an optimisation problem.	Focuses only on generating datasets for identifying the transient stability boundary.
[44]	Historical sampling	Solving OPF to minimise generation cost	Dimensionality reduction using neural networks reduces the search space thereby improving computational time.	Part of the search space is ignored. Rare OCs are not considered.
[53]	Importance sampling	LHS	Considers rare cases by fitting a generalised pareto distribution to the tail-region.	Dataset represents only a small portion of the feasible space as only OPF solutions are considered.
[45]	Historical sampling	Solving OPF to minimise generation cost	Exploits GANs to mitigate against missing PMU data when implementing ML-based DSA.	Dataset represents only a small portion of the feasible space as only OPF solutions are considered. Method cannot generate arbitrarily new OCs.
[49]	Importance sampling	Solving OPF to minimise generation cost	Interpolating between secure and insecure cases to sample new OCs ensures the creation of relevant samples	Dataset represents only a small portion of the feasible space
[58]	Historical sampling	Solving OPF to minimise generation cost	Adopts a feature selection strategy to optimise PMU data collection for fast and robust prediction.	Dataset represents only a small portion of the feasible space as only OPF solutions are considered. Method cannot generate arbitrarily new OCs.
[43]	Historical sampling	Historical records.	Use of a cycleGAN model to refine simulated data such that it mimics actual transients from historical data and improve synthetic data quality.	Dataset represents only a small portion of the feasible space as only historical records are considered.
[50]	Importance sampling	Solving OPF to minimise generation cost	Improves computation time needed to build a transient stability assessment database using a semi-supervised ensemble learning approach.	Dataset represents only a small portion of the feasible space
Proposed approach	Generic sampling	Sequentially generated to explore the physical feasible space	Sequentially explores the feasible space to maximise distance from previously generated samples.	Method does not currently consider class imbalance in the formulation

tally novel to other peer-reviewed works we have investigated. Specifically, the proposed approach is novel in the way the initial OCs are being sampled. The proposed approach conceptually outperforms other state-of-the-methods in its practicality to sequentially generate all (feasible) OCs. Typically, most state-of-the-art works consider generator outputs to be scheduled to represent conventional systems operations, often as a result of solving the optimal power flow problem that minimises generation cost [21][39][57][37][48][45][58][50]. However, as it is likely that the initial OC where a fault occurs is different from the optimal set-points, it is necessary to develop methods that explore these likely OCs [51]. Thus, a first point of comparison to generate pre-fault OCs is with methods that consider the OPF to generate initial OCs. The other approaches in the literature explore the entire feasible space in a generic way via random sampling, often using the LHS to generate initial OCs [53][55][56][52][51]. As a consequence, a second and more pivotal comparison is with those methods that aim to uniformly cover the search space using techniques like the LHS. As highlighted in Table 3.1, additionally, a major shortcoming of *historical* and some *importance sampling* approaches is that the resulting database of OCs represents only a small portion of the feasible space. Consequently *generic sampling* allows the exploration of the full physical feasible space. While existing generic sampling approaches currently in the literature attempt to discard sections of the search space via rapid rejection sampling [57][55][56][52], the proposed method differs fundamentally from state-of-the-art approaches by optimally exploring the feasible space in an iterative fashion. This exploration is done by varying the objective function and active constraints while respecting all the physical feasible constraints. The novelty of this work stems from presenting for the first time a *generic sampling* method that categorically explores the feasible space in an optimal manner. Finally, the proposed approach is versatile and could be further developed towards the combination with other database generation approaches like importance sampling and together with historical records. In parallel to this work, the *generic sampling* approach [120] investigates multiple objective functions to explore the feasible space.

### 3.1.2 Measuring quality of sampling

The quality of a training database is a measure of coverage of the feasible space and data usefulness, representing the pre- and post- fault data, respectively, in ML DSA application. For the coverage, typically, a set of points is said to uniformly cover a region when the points satisfy the following characteristics: (1) placed equidistant relative to one another (2) cover the entire region/volume of interest (3) distributed equally along all directions [121]. Point-to-point coverage measures focus on the first characteristic and aim to quantify how well the points are placed relative to one another. Examples of such metrics include the coverage metric ( $\mathcal{COV}$ ) used in [118], the coefficient of variation ( $\lambda$ ) and mesh ratio ( $\gamma$ ). Volumetric coverage measures, however, combine the first characteristic with one or both of the other two. Examples of volumetric measures based on Voronoi tessellation include point norm distribution ( $h$ ), point distribution ratio ( $\mu$ ), regularity metric ( $\chi$ ), etc. [121]. In high dimensional space, proximity measures (point-to-point coverage measures) used in two or three-dimensional space do not carry the same intuitive descriptive information quality [122]. The intuition of Euclidean distance falls apart, and a skin-effect-like tendency is observed such that the volume is concentrated around the skin of a high dimensional hyper-sphere instead of the centre [123]. Due to this concentration effect, the relative contrast between far and near points diminishes as the dimensionality increases, making it difficult to discriminate between far and near points [124]. Therefore, a current research gap and need is a metric that can quantify the quality of a training database for *generic sampling* approaches that have the objective to generate diverse OCs in the physical feasible space for power system DSA application.

For data usefulness, typically the issue of imbalanced datasets in the post-fault label is in focus. In DSA application, the distribution of secure/insecure OCs represents an important consideration for training ML models. There, the state-of-the-art methods in the literature use a preprocessing step such as synthetic minority oversampling (SMOTE) [64] and adaptive synthetic sampling (ADASYN) [125] to achieve this balance, usually to supplement with insecure OCs. The second way to address this imbalance is by combining historical

records with OCs generated using a *generic sampling* approach. In power system security assessment, accurately predicting insecure OCs is more important than predicting secure ones [95], and as historical OCs are disproportionately biased with more secure OCs, it motivates the creation of synthetic datasets.

This work focuses on the first quality measure of coverage and aims to generate pre-fault data so that the datasets can have diverse OCs. The motivation of this work is to fully concentrate on the issue of variability in the pre-fault database.

### 3.1.3 Contributions

This work proposes a novel split-based *generic sampling* approach, **GAPSPLIT\***. This novel split-based approach is a modification of the **GAPSPLIT** approach [118]. The novel split-based sampling approach aims to systematically generate diverse pre-fault operating conditions. The proposed approach covers previously unexplored OCs that are physical feasible but have not occurred in the past. With the proposed approach, high-quality databases (of pre-fault OCs) can be generated for training ML models used in real-time DSA. The proposed algorithm's crucial advantage over other statistical, distribution-based approaches that require fitting to a pre-existing database is the ability to consider the full physical search space defined by the AC PF and requires no historical data to work. In this chapter, the contribution is threefold: first, for the first time, this work investigates the **GAPSPLIT** approach for power system application. Second, this work modifies the **GAPSPLIT** approach to make it suitable for power system application. Third, this work investigates metrics to assess the quality of a *generic sampling* approach.

In the first contribution, the **GAPSPLIT** approach uses mathematical optimisation for sampling [118]. In this proposed work, sampling feasible OCs considers all power system constraints, such as power flow equations, line flow constraints, and node balances from the Alternating Current (AC) model. Then, an optimisation is solved sequentially for each new sample. Each sequence considers previously generated samples and determines

the maximal gap in the entire feasible region, then uses optimisation to add physical constraints at the maximal gap, which is the sampling target. The approach considers primary and secondary targets, where the primary target is a hard-constraint on the maximal gap in the optimisation, and the secondary targets are in the objective function to minimise the Euclidean distance to the target.

In the second contribution, the proposed modification from GAPSPLIT to the proposed GAPSPLIT\* approach has two pivotal advancements: to avoid converging to infeasible samples that do not satisfy the power flow equations and to efficiently analyse previously generated data to boost scalability to larger systems. The first advancement to avoid infeasibility is achieved by one of two proposed approaches: (i) relaxing the hard constraint on the primary target and activating only the constraints on secondary targets and (ii) storing infeasible samples and considering them as closed gaps to prevent the sampling from diverging. The second advancement to support scalability to larger systems is achieved by improving the sorting of sets containing the OCs.

In the third contribution, this work proposes a new volumetric coverage assessment metric, the convex hull volume ( $\mathcal{V}$ ) to assess the quality of a *generic sampling* approach. The convex hull is the union of all simplices with vertices in a set, i.e., the smallest convex polygon that surrounds a set of points. The  $\mathcal{V}$  of this envelope serves as a metric to represent the coverage of points. In the case studies studies, the benefits of  $\mathcal{V}$  as a better coverage metric to distance-based coverage metrics is shown.

The rest of the chapter is structured as follows: Section 3.2 discusses the regular split-based sampling GAPSPLIT and the proposed modified split-based approach GAPSPLIT\*. Section 3.3 introduces performance measuring metrics, including coverage using the proposed  $\mathcal{V}$  metric. Section 3.4 outlines case studies to compare the performance of the proposed modified sampling approach, the proposed performance metric and the computational performance. Tests are carried out on the IEEE 6-bus, the IEEE 68-bus, and the IEEE 118-bus systems. Section 3.5 concludes the chapter.

## 3.2 Split-based sampling

The proposed split-based approach follows the idea of *generic sampling* that aims to uniformly cover the full physical feasible space with all possible OCs.

---

**Algorithm 1** GAPSPLIT algorithm

---

- 1: Define samples-set  $\Omega^G = \{\}$  with each sample  $X^u \in \mathcal{R}^{|\Omega^N|}$
- 2: Define range  $r_n = X_n^{UB} - X_n^{LB} \quad \forall n \in \Omega^N$
- 3: Define normalisation parameter  $w_n = \frac{1}{r_n^2} \quad \forall n \in \Omega^N$
- 4: **while** true **do**
- 5:   Sort  $\Omega^G \quad \forall n \in \Omega^N$
- 6:   Compute  $\Delta X_n^{(u)} = \frac{(X_n^{(u+1)} - X_n^{(u)})}{r_n} \quad \forall n \in \Omega^N$
- 7:   Search  $\forall n \in \Omega^N \quad \Delta X_n^{(max)} = \max\{\Delta X_n^{(k)} \mid \forall k = 1, 2, \dots, (|\Omega^G| + 1)\}$
- 8:   Select  $\tilde{n}$  s.t  $\Delta X_{\tilde{n}}^{(max)} = \max\{\Delta X_n^{(max)} \mid \forall n \in \Omega^N\}$
- 9:   Compute  $T_n = \frac{X_n^{(d+1)} - X_n^{(d)}}{2} + X_n^{(d)} \quad \forall n \in \Omega^N$
- 10:   Select  $\Omega^{Q^*} = \{q \mid q = f(v), f : [N] \mapsto \Omega^Q, N \leq |\Omega^Q|\}$
- 11:   Solve

$$\min_{X^*} \sum_{n \in \Omega^{Q^*}} w_n (X_n^* - T_n)^2$$

$$g(X^*) \leq 0$$

$$(1 - \zeta)T_{\tilde{n}} \leq X_{\tilde{n}}^* \leq (1 + \zeta)T_{\tilde{n}}$$

- 12:   Update  $\Omega^G \leftarrow X^*$
  - 13:   Recalculate  $\Delta X_n^{(max)} \quad \forall n \in \Omega^N$
  - 14:   **if**  $|\Omega^G| \leq \mathcal{S}^G$  **then** return  $\Omega^G$
  - 15:   **end if**
  - 16: **end while**
- 

### 3.2.1 Regular split-based approach

This section describes the GAPSPLIT sampling algorithm [118] that formulates the sampling as an optimisation problem with physical model-based constraints. *Algorithm (1)* illustrates this sampling strategy that comprises an initialisation step, an iteration step that performs analysis and optimisation, and criteria to stop the iterating algorithm.

The algorithm initialises with an empty set of samples  $|\Omega^G| = 0$ . The symbol  $|\cdot|$  denotes the cardinality of a set. The subsequently generated samples  $\Omega^G$  have a sample vector  $X^u \in \mathcal{R}^{|\Omega^N|}$  that describes the OC of the power system in  $|\Omega^N|$  dimensions and satisfies

the constraints of the physical model.  $\Omega^N$  is a finite set of variables defining an operating condition, including active and reactive power set points, bus voltages and angles, and line flows. The lower and upper bounds of the  $n$ -th variable are denoted as  $X_n^{LB}$  and  $X_n^{UB}$ , respectively, and the ranges are  $r_n = X_n^{UB} - X_n^{LB}$ .

In each iteration of **GAPSPLIT**, the algorithm generates a single optimised sample  $X^*$ , starting with an analysis of previous samples  $\Omega^G$ . The analysis begins with sorting the samples  $\Omega^G$  for each variable  $n \in \Omega^N$  resulting in  $|\Omega^N|$  ordered sets of the same samples  $\Omega^G$ . **GAPSPLIT** sorts these sets according to the values in variable  $n$

$$\Omega_n^{Go} = \{X_n^{(u)} \mid \forall u = 1, 2, \dots, |\Omega^G|, X_n^{(u+1)} \geq X_n^{(u)}\} \quad (3.1)$$

where  $X_n^{(u)}$  corresponds to the  $u$ -th largest sample in the  $n$ -th variable. Subsequently, the algorithm computes the gaps of the samples next to each other

$$\Delta X_n^{(u)} = \frac{(X_n^{(u+1)} - X_n^{(u)})}{r_n} \quad (3.2)$$

$r_n$  is a normalising parameter. The algorithm then identifies the maximal gap in each  $n$ -th variable

$$\Delta X_n^{(max)} = \max\{\Delta X_n^{(k)} \mid \forall k = 1, 2, \dots, (|\Omega^G|)\}, \quad (3.3)$$

where the algorithm denotes the two samples next to the maximal gap  $\Delta X_n^{(max)} = X_n^{(d+1)} - X_n^{(d)}$  with  $(d+1)$  and  $(d)$ . The maximal gap among all variables is

$$\Delta X_{\tilde{n}}^{(max)} = \max\{\Delta X_n^{(max)} \mid \forall n = 1, 2, \dots, |\Omega^N|\}, \quad (3.4)$$

where  $\tilde{n}$  denotes the variable with the maximal gap called the primary variable. All other variables are called secondary variables  $\Omega^Q = \Omega^N \setminus \tilde{n}$ . Subsequently, the algorithm computes targets for all primary and secondary variables at the centre of their respective maximal

gaps

$$T_n = \frac{X_n^{(d+1)} - X_n^{(n)}}{2} + X_n^{(d)}; \quad (3.5)$$

these are accordingly called primary and secondary targets, e.g., the primary target is  $T_{\tilde{n}}$ . Subsequently, the algorithm considers a subset of secondary variables  $\Omega^{Q^*} \subset \Omega^Q$  as not all secondary variables are further needed. There are multiple user-specific ways to select the subset of secondary variables  $\Omega^{Q^*}$ . One way is to consider a random selection

$$\Omega^{Q^*} = \{q \mid q = f(v), f : [N] \mapsto \Omega^Q, |f| = N, N \leq |\Omega^Q|\} \quad (3.6)$$

of a subset of  $N$  elements from  $\Omega^Q$ , where  $N$  is fixed and defined by the user (e.g.,  $N = 0.05 |\Omega^Q|$ ). Then, the algorithm uses random selection  $f : [N] \mapsto \Omega^Q$  in each iteration. Other ways to select the secondary variables are in [118], including to select  $|\Omega^{Q^*}| = 0$  as empty. In the remainder of the text, referring to secondary variables and targets corresponds to the subset of secondary variables  $\Omega^{Q^*}$ .

After the above analysis, the **GAPSPLIT** algorithm generates a single, new sample with the mathematical optimisation

$$\begin{aligned} & \underset{X^*}{\text{minimise}} && \sum_{n \in \Omega^{Q^*}} w_n (X_n^* - T_n)^2 \\ & \text{subject to} && g(X^*) \leq 0 \\ & && (1 - \zeta)T_{\tilde{n}} \leq X_{\tilde{n}}^* \leq (1 + \zeta)T_{\tilde{n}}, \end{aligned} \quad (3.7)$$

where the optimisation considers a constraint on the value  $X_{\tilde{n}}^*$  of the primary variable  $\tilde{n}$  at the primary target  $T_{\tilde{n}}$  with a relaxation to avoid numerical issues which can lead to non-convergence. The relaxation is considered with a tolerance parameter  $\zeta$  on the primary target  $T_{\tilde{n}}$  (e.g., of  $\zeta = 0.001$ ). This optimisation minimises the mean squared error from the generated sample  $X^*$  to the targets of the selected secondary variables  $\Omega^{Q^*} \subset \Omega^Q$ .

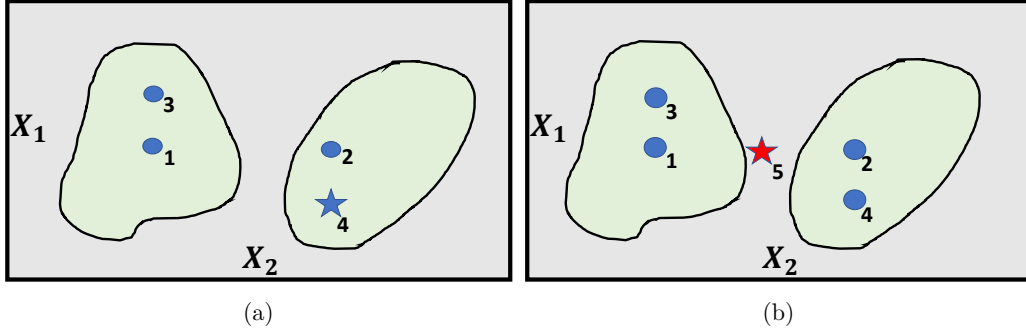


Figure 3.1: (a) GAPSPLIT generates new samples by attempting to split maximal gaps. (b) GAPSPLIT converges to an infeasible sample when the primary target (★) is located in the infeasible region. Max gap is found in  $X_2$

□ Input space    ● Feasible space    ● Infeasible region

$w_n = \frac{1}{r_n^2}$  is a normalisation parameter that re-weights all variables equally.  $g(X) \leq 0$  are the constraints that define the feasible space including power system constraints such as power flow equations, nodal balance equations, generator active and reactive limits, and bus voltage and phase angle limits. This optimisation aims to consider the physical constraints of the power system and to split the gaps between the previously generated samples (that is why the algorithm is called GAPSPLIT). The optimisation in (3.7) returns a new sample, the optimised OC  $X^*$ . This sample  $X^*$  is added to the set of samples  $\Omega^G \leftarrow X^*$ , and the next iteration is started.

The algorithm terminates when a user-specified criterion is met, for example, when a specified number of samples  $\mathcal{S}^G$  have been generated. Then, a new sample is only generated if  $|\Omega^G| \leq \mathcal{S}^G$ , otherwise the sampling algorithm stops.

### 3.2.2 Issues with GAPSPLIT

Two issues arise when using the above split-based approach for sampling power system OCs. The first issue is the low coverage of the physical feasible space and the second issue is the computational inefficiency when generating a large number of samples.

The first issue of low coverage is the result of GAPSPLIT converging to infeasible regions. GAPSPLIT may converge to such infeasible regions when the search space (feasible space) is

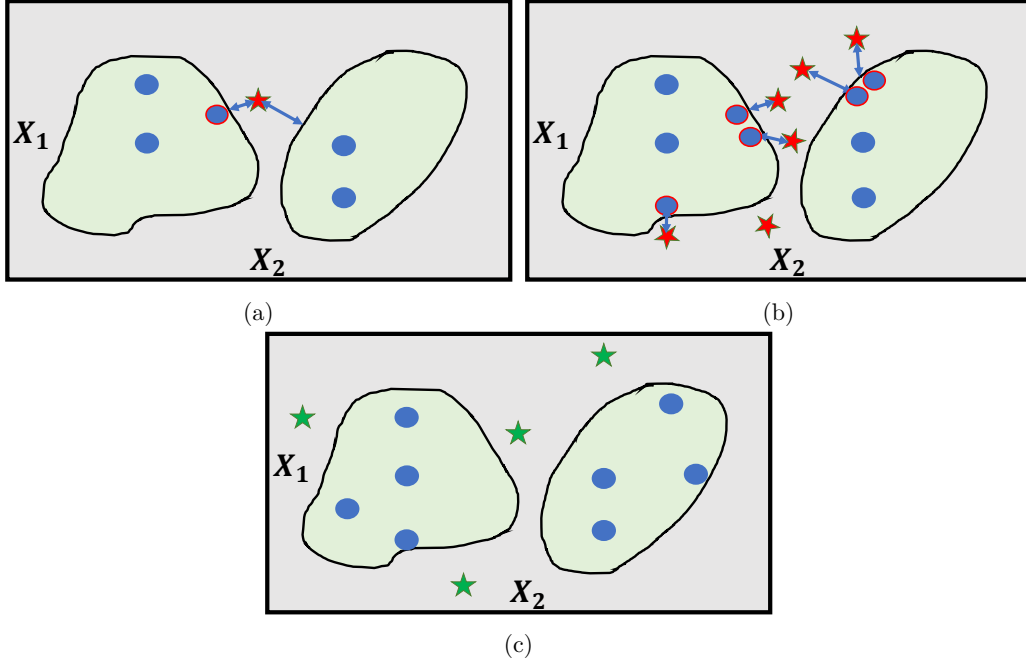


Figure 3.2: (a) GAPSPPLIT\* uses only secondary targets (★) to minimise the shortest distance to the feasible region. (b) GAPSPPLIT\* with only secondary targets (★) generates samples around the boundary of the feasible space (●) when  $T_n$  is located in infeasible regions. (c) GAPSPPLIT\* re-directs sampling to other regions of the physical space by memorising infeasible targets (★)

non-convex and disconnected as in power systems. The feasible space in power systems is the set of OCs that satisfy all operational equality and inequality constraints in  $g(X) \leq 0$  [126]. When the GAPSPPLIT algorithm locates the primary target  $T_{\tilde{n}}$  in an infeasible region, the algorithm results in an infeasible optimisation (3.7) and returns an infeasible sample  $X^*$  as the constraint  $g(X) \leq 0$  when  $\lim_{X_{\tilde{n}} \rightarrow T_{\tilde{n}}} g(X) > 0$  is not met. This issue is illustrated in Fig. 3.1. In Fig. 3.1a, the primary target  $T_{\tilde{n}}$  of the 4th candidate sample is in the feasible space, and GAPSPPLIT successfully generates a corresponding sample. However, when trying to generate the following 5th candidate sample in Fig. 3.1b, the primary target  $T_{\tilde{n}}$  is in the infeasible region where  $\lim_{X_{\tilde{n}} \rightarrow T_{\tilde{n}}} g(X) > 0$ . Hence, GAPSPPLIT is unable to generate the 5th candidate sample as  $\lim_{X_{\tilde{n}} \rightarrow T_{\tilde{n}}} g(X) > 0$ . As a result, the maximal gap  $\Delta X_{\tilde{n}}^{(max)}$  from Eq. (3.4) does not change in subsequent iterations as all the gaps  $\Delta X_n^{(u)}$  between previously generated samples remain unchanged, and consequently, GAPSPPLIT converges to that infeasible sample (5th candidate sample in Fig. 3.1b). The second issue is the computational bottleneck of sequential sampling approaches in high-dimensional settings. This

Table 3.2: Computational analysis of the **GAPSPLIT** algorithm where  $|\Omega^N|$  is a constant and  $|\Omega^G|$  corresponds to the number of iterations. The sorting step is the computational bottleneck of the algorithm.

Steps			Computation time
Sort $\Omega_n^{Go} \forall n$	in Eq.(3.1)		$\mathcal{O}( \Omega^N  \Omega^G  \log  \Omega^G )$
Compute $\Delta X_n^{(u)}$	in Eq.(3.2)		$\mathcal{O}( \Omega^N  \Omega^G )$
Compute $\Delta X_n^{(max)}$	in Eq.(3.3)		$\mathcal{O}( \Omega^N  \Omega^G )$
Compute $X_{\tilde{n}}^{(max)}$	in Eq.(3.4)		$\mathcal{O}( \Omega^N )$

bottleneck is particularly critical in power systems that have a large number of variables  $|\Omega^N|$  and require a large number of samples  $|\Omega^G|$ . The computational bottleneck of some sequential sampling approaches, such as **GAPSPLIT**, is that they often need to analyse a large number of previously generated samples  $\Omega^G$  in each iteration. Table 3.2 analyses the computational requirements in each iteration for the **GAPSPLIT** algorithm in Big- $\mathcal{O}()$  notation to demonstrate this issue. In each iteration, the sorting of samples  $\Omega^G$ , computing of gaps  $\Delta X_n^{(u)}$  and maximal gaps  $\Delta X_n^{(max)}$  steps have complexities of  $\mathcal{O}(|\Omega^N||\Omega^G| \log |\Omega^G|)$ ,  $\mathcal{O}(|\Omega^N||\Omega^G|)$  and  $\mathcal{O}(|\Omega^N||\Omega^G|)$  respectively. The key bottleneck is the sorting step which grows  $\mathcal{O}(|\Omega^N||\Omega^G| \log |\Omega^G|)$  as the size of  $|\Omega^G| \rightarrow a$ , where  $a \gg 1$  is a large number.

### 3.2.3 Proposed split-based approach: **GAPSPLIT\***

The proposed **GAPSPLIT\*** approach improves the **GAPSPLIT** approach with two modifications to address each of the above issues as follows.

#### Exclusive sampling of secondary variables and introducing the set of infeasible samples

The first proposed modification of **GAPSPLIT\*** approach is twofold (i) sampling exclusively with secondary variables  $\Omega^{Q*}$  and (ii) considering infeasible samples in the subsequent progressions of **GAPSPLIT\*** algorithm. This modification addresses the first issue of low coverage when the feasible space is non-convex and disconnected as in power systems.

The modification (i) of sampling exclusively with secondary variables  $\Omega^{Q^*}$  is to discard the hard constraint on the primary variable  $\tilde{n}$  in optimisation (3.7). Therefore, the optimisation simplifies to

$$\begin{aligned} & \underset{X^*}{\text{minimise}} && \sum_{n \in \Omega^{Q^*}} w_n (X_n^* - T_n)^2 \\ & \text{subject to} && g(X^*) \leq 0, \end{aligned} \tag{3.8}$$

where the objective is to minimise the mean squared error of  $X^*$  to secondary targets  $T_n$ ,  $\forall p \in \Omega^{Q^*}$  of the secondary variables. The effect is illustrated in Fig. 3.2a where a target is located in the infeasible region where  $\lim_{X_n \rightarrow T_n} g(X) > 0$ . When comparing Fig. 3.1b (GAPSPLIT) with Fig. 3.2a (GAPSPLIT\*), GAPSPLIT would converge to an infeasible sample when the primary target  $T_{\tilde{n}}$  is located in the infeasible region. However, GAPSPLIT\* addresses this issue by removing the hard constraint on  $\tilde{n}$  and minimising the distance to the optimised feasible OC, marked with a red circle in Fig. 3.2a. This minimisation of distances generates samples around the boundary of the feasible space and close to each other as illustrated in Fig. 3.2b. This accumulation does not support effectively covering the full feasible space.

The modification (ii) addresses the issue of accumulating infeasible samples presented in Fig. 3.2b. This modification (ii) re-directs the sampling to other regions of the feasible space by considering previously encountered infeasible samples. The algorithm of GAPSPLIT\* with modification (ii) is similar to that described in Sec. 3.2.1 with the crucial difference being the iteration step when the solution to the optimisation (3.7) is infeasible. Here, GAPSPLIT\* stores (memorises) the targets that led to the infeasible solutions, and subsequently uses them to avoid sampling at these infeasible targets again. Fig. 3.2c presents a visual illustration of this approach. The set  $\Omega^{G'}$  is the set of infeasible samples and  $\Omega^{G''} = \Omega^{G'} \cup \Omega^G$  is the set of all feasible  $\Omega^G$  and infeasible  $\Omega^{G'}$  samples. Fig. 3.3 shows the algorithmic flowchart of this key difference in GAPSPLIT\* modification (ii). If the optimisation is infeasible  $\lim_{X_{\tilde{n}} \rightarrow T_{\tilde{n}}} g(X) > 0$ , then GAPSPLIT\* assigns the value of this infeasible primary target  $T_{\tilde{n}}$  to the primary variable, and the minimal values  $X_n^{LB}$  to all other secondary variables

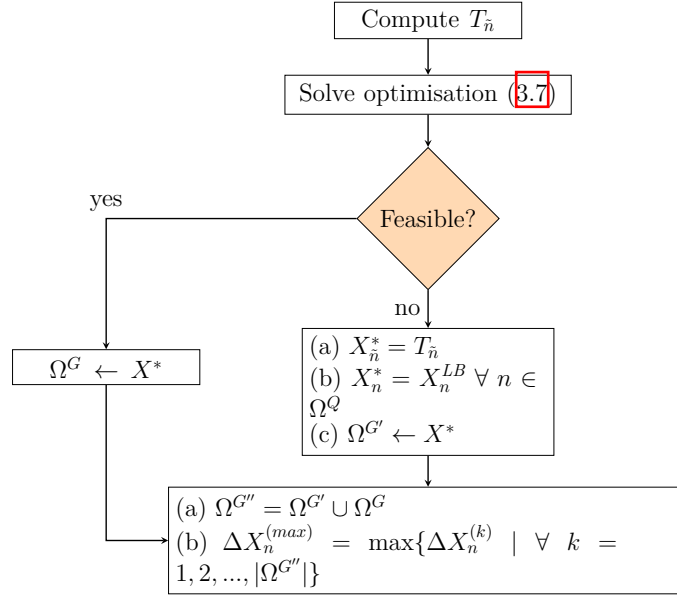


Figure 3.3: The iteration step of the proposed modification (ii) of GAPSPLIT\* that introduces the set of infeasible samples  $\Omega^{G'}$

$\Omega^Q$  of the infeasible sample

$$\begin{aligned} X_{\tilde{n}}^* &= T_{\tilde{n}} \\ X_n^* &= X_n^{LB} \quad \forall n \in \Omega^Q. \end{aligned} \tag{3.9}$$

Using the lower bound  $X_n^{LB}$  is arbitrary. As the aim is to preserve the gaps in the other variables, using the upper bound  $X_n^{UB}$  would have the same effect. Subsequently, GAPSPLIT\* adds this infeasible sample  $X^*$  to the set of infeasible samples  $\Omega^{G'} \leftarrow X^*$ . This step implicitly stores the information that the primary target  $T_{\tilde{n}}$  is in the infeasible region and allows GAPSPLIT\* to disregard the corresponding gap  $\Delta X_{\tilde{n}}^{(max)}$  between  $X_{\tilde{n}}^{(d)}$  and  $X_{\tilde{n}}^{(d-1)}$  in subsequent iterations, and therefore avoids converging to that infeasible sample. Finally, GAPSPLIT\* computes the next maximal gaps

$$\Delta X_n^{(max)} = \max\{\Delta X_n^{(k)} \mid \forall k = 1, 2, \dots, |\Omega^{G''}|\} \tag{3.10}$$

by using feasible and infeasible samples  $\Omega^{G''}$ , and continues with Eq. (3.4) and subsequent steps in Sec. 3.2.1.

### Efficient sorting

The second proposed modification of **GAPSPLIT\*** approach is the efficient sorting of gaps to identify the largest gap in each iteration. This modification addresses the second issue of scalability of the **GAPSPLIT** approach in high-dimensional settings as in the power system. The algorithm starts with an initialisation step, followed by iterations that terminate when a stopping criterion is satisfied.

Initially, **GAPSPLIT\*** assigns a set  $\Omega^F$  to maintain an ordered set of all gaps across all variables which has the cardinality  $|\Omega^F| = |\Omega^N| \times |\Omega^G|$ . This ordered set contains all gaps  $\Delta X_n^{(u)}$  from Eq. (3.2) for all variables  $n$ . The set is

$$\Omega^F = \{\Delta X^{(k)} \mid \forall k = 1, 2, \dots, |\Omega^N| \times |\Omega^G|, \Delta X^{(k)} \geq \Delta X^{(k-1)}\}, \quad (3.11)$$

where  $\Delta X^{(k)}$  is the  $k$ th largest gap across all variables and all samples. The notation of the gap  $\Delta X^{(k)}$  drops the index for the sample  $u$  and for the variable  $n$  for simplicity reasons. The sample index  $u$  and the variable index  $n$  can be retrieved with the two mappings  $\mathcal{U}(k)$  and  $\mathcal{P}(k)$ , respectively. The overall largest gap is the last element of the ordered set  $\Delta X^{(|\Omega^F|)}$  which avoids using the max operators in Eqs. (3.3)–(3.4).

In each iteration, **GAPSPLIT\*** locates the primary target  $T_{\tilde{n}}$  at the centre of this overall largest gap  $\Delta X^{(|\Omega^F|)}$ . **GAPSPLIT\*** obtains  $d = \mathcal{U}(|\Omega^F|)$ ,  $\tilde{n} = \mathcal{P}(|\Omega^F|)$ , and the primary target

$$T_{\tilde{n}} = \frac{X_{\tilde{n}}^{(d+1)} - X_{\tilde{n}}^{(d)}}{2} + X_{\tilde{n}}^{(d)}, \quad (3.12)$$

where the samples  $X_{\tilde{n}}^{(d+1)}$  and  $X_{\tilde{n}}^{(d)}$  form the gap  $\Delta X^{(|\Omega^F|)}$ . Subsequently, **GAPSPLIT\*** selects the secondary variables, for instance with Eq. (3.6), and then solves optimisation (3.7) to obtain the optimised OC, the new sample  $X^*$ . Subsequently, **GAPSPLIT\*** copies this

generated sample  $X^*$  in total  $|\Omega^N|$  times and inserts one copy each into the sets

$$\Omega_n^{Go} \leftarrow X^* \quad \forall n \in \Omega^N \quad (3.13)$$

using the bisection method, which can only be used as the sets  $\Omega_n^{Go}$  are ordered. The position of the insertions in the corresponding sets is the map  $\mathcal{I}(n)$  such that  $X_n^{(\mathcal{I}(n)-1)} \leq X_n^* \leq X_n^{(\mathcal{I}(n)+1)}$ . Note that this bisection insertion step is the key advancement as it replaces the sorting step required in each iteration of **GAPSPLIT**. The reader may recall that the sorting step is the key bottleneck of **GAPSPLIT** as per analysis in Table 3.2. However, the bisection method requires only a computational time of  $\mathcal{O}(|\Omega^N| \log |\Omega^G|)$  in the worst case. Hence, this efficient bisection step with  $\mathcal{O}(|\Omega^N| \log |\Omega^G|)$  replaces the inefficient sorting step with  $\mathcal{O}(|\Omega^N| |\Omega^G| \log |\Omega^G|)$ . Following the insertion, **GAPSPLIT\*** generates  $2|\Omega^N|$  new gaps at

$$\begin{aligned} \Delta X_n^{(a)} &= X_n^{(\mathcal{I}(n)+1)} - X_n^* \\ \Delta X_n^{(b)} &= X_n^* - X_n^{(\mathcal{I}(n)-1)}. \end{aligned} \quad (3.14)$$

Subsequently, **GAPSPLIT\*** inserts these  $2|\Omega^N|$  new gaps in  $\Omega^F \leftarrow \Delta X_n^{(a)}, \Omega^F \leftarrow \Delta X_n^{(b)} \forall p \in \Omega^N$  by using the bisection method, as well. **GAPSPLIT\*** limits the cardinality of the set  $|\Omega^F| \leq \rho$  to avoid memory issues when the size of  $|\Omega^G| \rightarrow a$ , where  $a \gg 1$  is a large number. In response to this threshold, if  $|\Omega^F| > \rho$ , then **GAPSPLIT\*** drops the smallest  $2|\Omega^N|$  gaps in each iteration

$$\Omega^F \setminus \Delta X^{(k)} \mid \forall k = 1, 2, \dots, 2|\Omega^N|, \quad (3.15)$$

such that  $|\Omega^F| \leq \rho$  is satisfied at all times. **GAPSPLIT\*** terminates when sufficient samples are created  $|\Omega^G| \geq S$ .

### 3.3 Measuring performance of samplers

*Generic sampling* focuses on covering the feasible space with the generated samples  $\Omega^G$ . A performance metric of such samplers should quantify the coverage of feasible space, which also is a metric for the quality of samples in  $\Omega^G$ . Such a metric for coverage can also serve as a criterion to stop sampling when the feasible space is sufficiently sampled. However, this criterion is not implemented for the case studies in this thesis.

The  $\mathcal{COV}$  metric measures the performance of the GAPSPLIT sampler in [118]

$$\mathcal{COV} = 1 - \frac{1}{|\Omega^N|} \sum_{n=1}^{|\Omega^N|} \frac{\Delta X_n^{(max)}}{r_n} \quad (3.16)$$

representing the average relative maximal gap  $\Delta X_n^{(max)}$  in  $|\Omega^N|$  dimensions. To illustrate this metric, the  $\mathcal{COV}$  metric has a minimal value  $\mathcal{COV} = 0$  when all samples  $\Omega^G$  are stacked on top of each other, and a maximal value  $\mathcal{COV} = 1$  when an infinite number of samples  $\Omega^G$  are uniformly distributed. For example,  $\mathcal{COV} = 0.75$  indicates that the relative maximum gap is 25 % on average over all variables  $\Omega^N$ .

The drawback of analysing sample distributions using the  $\mathcal{COV}$  metric is that the analysis focuses on the marginal (univariate) and not the multivariate distribution. Hence, using  $\mathcal{COV}$  as a performance metric to assess samplers in high-dimensional settings may result in a poor characterisation of multivariate sample distributions, which is important when using an optimisation procedure to generate the samples (as we will demonstrate in the case study). The example in Fig. 3.4 illustrates this drawback of using  $\mathcal{COV}$  to measure coverage. The samples in the two figures, Fig. 3.4a and Fig. 3.4b, are clearly differently distributed but have the same marginal distributions in both dimensions. However, as the  $\mathcal{COV}$  metric only assesses the marginal univariate distribution (in this example, projecting the points to one dimension), it calculates the same  $\mathcal{COV}$  values for the two figures, thereby ignoring the difference in the two bivariate distributions. Hence, the  $\mathcal{COV}$  metric is an unsuitable measure of the coverage of samples. Generally in the literature, point-to-point coverage measures fail to quantify how well samples are distributed relative to one another

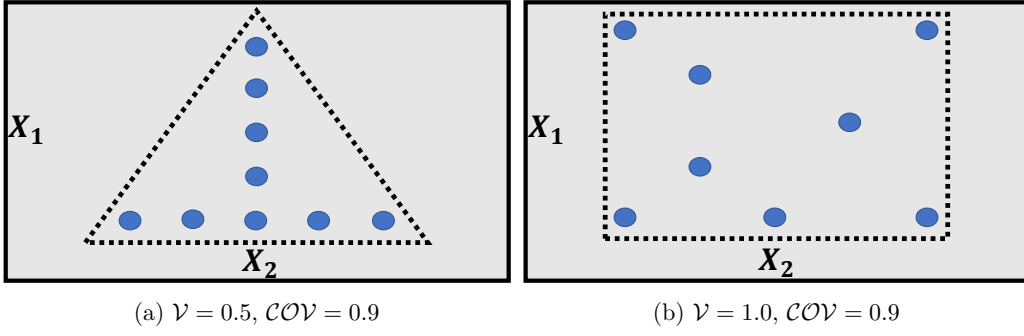


Figure 3.4: The samples (blue circles) in (a) and (b) have the same  $\mathcal{COV}$  but different  $\mathcal{V}$  values. The proposed convex hull metric (dotted black line) is suitable for measuring multivariate coverage.

in high dimensional settings and do not account for the distribution of samples in a region [123][122].

Conversely, assessing the volume occupied by the samples seems to be a suitable approach to measure the performance of multivariate sample distributions. The proposed  $\mathcal{V}$  metric based on computing an approximation of the convex hull volume can overcome the drawback of the  $\mathcal{COV}$  metric. The convex hull of a set of samples  $\Omega^G$

$$CH = \left\{ \xi^1 X^{(1)} + \dots + \xi^r X^{(r)} \left| \sum_{u=1}^r \xi^u = 1, X^{(u)} \in \Omega^G, \xi^u \geq 0 \right. \right\} \quad (3.17)$$

is the smallest convex set that contains all other samples defined in some  $|\Omega^N|$ -dimensional space, where  $X^{(1)}, X^{(2)}, \dots, X^{(r)}$  are independent samples in some Euclidean space  $\mathcal{R}^{|\Omega^N|}$ , and  $\xi^u$  are real numbers. The samples  $X^{(1)}, X^{(2)}, \dots, X^{(r)}$  are the vertices of the convex hull as they enclose all other samples. The index  $r$  represents the number of samples that form the vertices of the convex hull, where  $r \leq |\Omega^G|$ . In this work uses the Qhull algorithm [127] to compute the  $\mathcal{V}$  metric that measures the convex hull volume occupied by the generated samples  $\Omega^G$ .

Subsequently, the volume of the convex hull with vertices  $X^{(1)}, X^{(2)}, \dots, X^{(r)}$  is

$$\mathcal{V} = \left| \frac{1}{r!} \det \begin{pmatrix} X^{(2)} - X^{(1)} & X^{(3)} - X^{(1)} & \dots & X^{(r)} - X^{(1)} \end{pmatrix} \right| \quad (3.18)$$

which further resolves to

$$\left| \frac{1}{r!} \det \begin{pmatrix} X^{(1)} & X^{(2)} & \dots & X^{(r)} \\ 1 & 1 & \dots & 1 \end{pmatrix} \right| \quad (3.19)$$

This formula is the simplified version of the convex hull volume. However, computing the  $\mathcal{V}$  in higher dimensions ( $|\Omega^N| > 6$ ) is intractable, as it is a P-hard problem. This work circumvents this issue by randomly selecting a subset of the input variables  $\hat{\Omega} \subset \Omega^N; |\hat{\Omega}| \leq 6$  to compute the volume. This random selection allows us to approximate the  $\mathcal{V}$  in higher dimensions, as we will show in the case study section. This limitation of scalability to higher dimensions mean that certain regions of the feasible space that are "relevant" may be ignored when computing the volume. The notation  $\mathcal{V}_{|\Omega^N|=|\hat{\Omega}|}$  denotes the  $|\hat{\Omega}|$ -dimensional convex hull volume for a set of samples. The reader may refer to the text in [128] and [129] for further information on convex hulls and their associated volumes.

Finally, we consider other state-of-art coverage metrics in [121]. Specifically, the coefficient of variation between all samples  $X^u, X^{u_j} \in \mathbb{R}^{|\Omega^N|}$  is  $\lambda = (|\Omega^G| \frac{\sum_{u=1}^{|\Omega^G|} \gamma_u^2}{(\sum_{u=1}^{|\Omega^G|} \gamma_u)^2})^{1/2}$ , where  $\gamma = \min_{u \neq u_j} |X^u - X^{u_j}|$ . The smaller the value of  $\lambda$ , the more uniform the distribution of samples and  $\lambda = 0$  signifies a perfect uniform mesh. We also consider the point norm distribution,  $h = \max_{i=1,2,\dots,|\Omega^G|} h_u$ , where  $h_u = \max_{y \in V_u} |X^u - y|$ , where  $h_u$  is the maximum distance between a sample-point  $X^u$  and the points that enclose the cell of its Voronoi tessellation  $V_u$ . Here also, the smaller the value of  $h$ , the more uniform is the distribution. The scalability of  $\lambda$  as the number of samples  $|\Omega^G| \rightarrow a$  increase, where  $a \gg 1$  is a large number is challenging as it requires  $\mathcal{O}(|\Omega^G|^2)$  computations. The scalability of the  $h$  coverage measure in high dimensional settings is similar to that of the proposed  $\mathcal{V}$  metric, which can ignore "relevant" regions of the feasible space.  $\lambda$  and  $h$  are included in the case study comparison of this chapter for the sake of completeness.

### 3.4 Case study

In this section, firstly, this work investigates the suitability of the proposed **GAPSPLIT\*** approach to generate representative power systems OCs in comparison to random sampling ( $\mathcal{RS}$ ) and minimised generation cost minimum generation cost ( $\mathcal{MGC}$ ) approaches. Secondly, this work investigates the performance of the proposed **GAPSPLIT\*** approach to address the low coverage issue of **GAPSPLIT** when generating OCs for power systems. Thirdly, this work shows the suitability of the proposed  $\mathcal{V}$  metric to measure coverage of samples generated by the proposed **GAPSPLIT\*** approach. Fourthly and finally, this work discusses the scalability of the proposed  $\mathcal{V}$  metric to higher dimensions and the computational time of the proposed **GAPSPLIT\*** approach on the IEEE 118-bus system.

#### 3.4.1 Test System and assumption

The case studies consider the IEEE 6-bus [130] and IEEE 68-bus [131] test systems. Subsequently, using the IEEE 118-bus system [132], a case study presents a scalability study that considers a DC approximation of the power flow. The networks studied in this chapter do not have any renewable generation connected. To generate the load profiles, the active loads were sampled from a multivariate Gaussian distribution (via Monte Carlo sampling) and assume the correlation between loads to follow Pearson's correlation with a correlation coefficient of 0.75. The distribution was then converted to a marginal Kumaraswamy(1.6,2.8) distribution using inverse transformation. The reactive loads at the buses scale linearly with active loads by a factor of 0.15 ( $\frac{KVAR}{KW} \approx 0.15$ ). To create the generator profiles,  $X_n^*$   $\forall n \in \Omega^N$ , different sampling approaches, including the proposed split-based sampling,  $\mathcal{MGC}$ , and  $\mathcal{RS}$  approaches attempt to solve an optimisation problem that balances generator output with randomly generated loads. The proposed split-based sampling follows the sampling procedure described in Sect. 3.2.3, whereas the  $\mathcal{MGC}$  approach solves the optimal power flow of the AC-model. Finally, the  $\mathcal{RS}$  approach involves sampling generator profiles using the LHS procedure and accepting either the LHS generated profile

$X_n^*$ , or a perturbation  $X_n^* + \delta_{X_n}$ , where  $\delta_{X_n} \forall n \in \Omega^N$  are slack variables in the optimisation. The AC models of the networks are used to ensure feasible OCs representing the steady-state operation of the system under AC assumptions.  $|\Omega^G| = 1,000$  initial OCs were selected for each sampling approach. This number of OCs is arbitrary and was chosen to study the dynamic security profiles of the initial OCs, where other stopping criteria such as coverage of the feasible space will in practice determine the size of  $|\Omega^G|$ . Then, for each pre-fault OC, their corresponding post-fault security labels were simulated with time-domain dynamic simulations. For the simulations, the initial conditions included the pre-fault variables for active and reactive power generations, and active and reactive power loads. The dynamic simulation considered a three-phase fault on line 31 – 38 for the IEEE-68 bus system with a clearance time of 0.5 s. Subsequently, the simulations were analysed and the post-fault transient security label was computed. The label of an OC was either secure  $\mathcal{Y}_j^m = 0, \forall j \in \Omega^J$  when all phase angle differences between any two generators were less than  $180^\circ$  within the 10 s simulation time after the fault, otherwise, the OC was insecure  $\mathcal{Y}_j^m = 1, \forall j \in \Omega^J$ . To see the generation of database in the context of the final use case, ML models were trained using the generated data. There, the pre-fault OCs and post-fault security labels were used as training databases for quantifying the performance of the trained ML models on testing data. Different ML models, including feed-forward Artificial Neural Network (ANN), Support Vector Machine (SVM), boosting algorithms (Xgboost and Adaboost), and Decision Trees (DTs), were trained as example ML models. The model hyperparameters were not tuned as only the relative performance of the models on different datasets is of concern. The ANNs had three hidden layers with 60, 30, and 10 neurons, respectively, and were trained with a stochastic gradient descent optimiser using the package PyTorch 1.10.0 [133]. The SVM training used a linear kernel, and the boosting algorithms had 50 estimators using the package *scikit-learn* 0.18.1 [134]. DTs were trained with the CART algorithm [135] from the package *scikit-learn* 0.18.1 [134] in Python 3.5.2. The default training settings were selected except using gini impurity instead of entropy to measure the quality of the splits. The data-set was split into training/testing sets in ratio of 75 % / 25 %. 5-fold cross-validation was applied to address under-/overfitting. Subse-

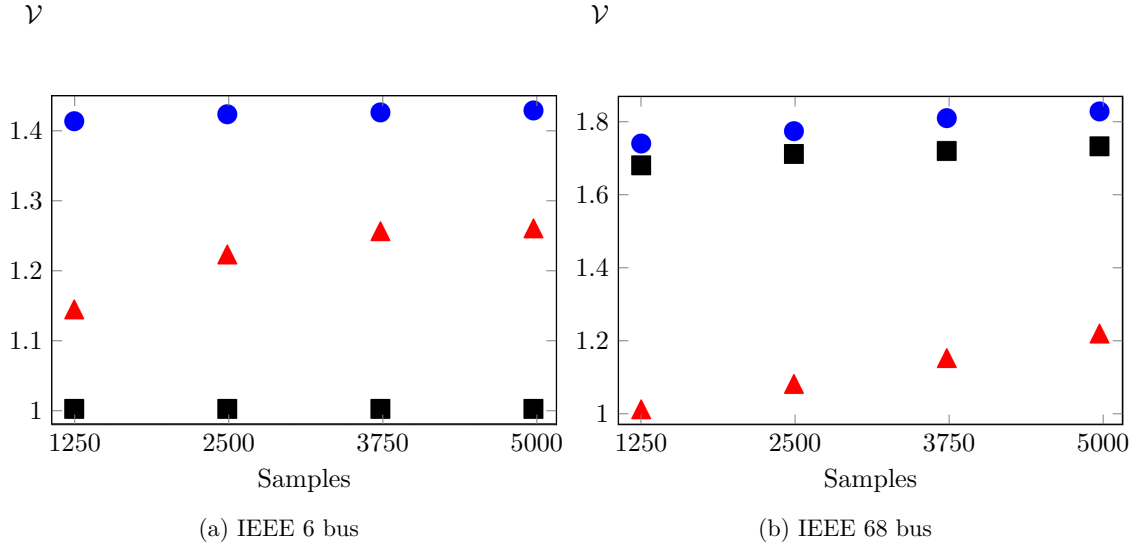


Figure 3.5: (a) The  $\nu$  of samples generated using the proposed modification (i) (■), the proposed modification (ii) (●), and  $\mathcal{RS}$  (▲).

quently, the Platt method was used to calibrate the score-output  $S$  of the classifier [136]. The ML models were evaluated with metrics as the testing accuracy =  $\frac{Tp+Tn}{Tp+Tn+Fp+Fn}$ , precision =  $\frac{Tp}{Tp+Fp}$ , specificity =  $\frac{Tn}{Tn+Fp}$ , and  $F1$ -score =  $\frac{2*precision*specificity}{precision+specificity}$ , where  $Tp$  and  $Tn$  are correctly classified positive and negative OCs, and  $Fp$  and  $Fn$  are incorrectly classified negative and positive OCs. Additionally, the fraction of insecure OCs  $\Pi_{\Omega^G}^1 = \frac{|\Omega_1^G|}{|\Omega^G|}$  was computed,  $|\Omega_1^G|$  is the number of OCs in  $\Omega^G$  with label 1.

The non-linear optimisation problems were implemented using the package Pyomo 5.6.8 [137] in Python 3.7.4 and solved using IPOPT 3.13.2 [138]. All studies except the scalability section were carried out on a Dell XPS 139360 running an Intel(R) Core(TM) i5-8250U processor with 8 GB installed RAM. The scalability study was carried out on a Windows Server 2008 R2 Enterprise running an Intel(R) processor with 96 GB installed RAM. The dynamic simulations are implemented in Julia 1.6.4 with the packages *PowerSystems.jl* [139], *PowerSimulationsDynamics.jl* [140]. The simulations were solved with the IDA package from Sundials solvers [141]. All dynamic simulations were performed on a standard machine with six cores and 16GB RAM.

### 3.4.2 Effective sampling with GAPSPLIT\*

In this study, this work contrasts the performance of the candidate approaches (the proposed GAPSPLIT\*,  $\mathcal{MGC}$ , and  $\mathcal{RS}$ ) in generating representative power systems OCs, which results appear in Fig. 3.5. Concretely, the figure depicts the 3D- $\mathcal{V}$  covered by 5000 OCs generated with the candidate approaches in the IEEE 6- and 68-bus systems. The  $\mathcal{V}$  values in the figure are normalised with the minimum value for each test system such that  $\hat{\mathcal{V}} = \frac{\mathcal{V}}{\min(\mathcal{V})}$ . In the proposed GAPSPLIT\* approach, this work considers  $|\Omega^{Q^*}| = 0.3|\Omega^N|$  secondary variables.

As evidenced by Fig. 3.5, the proposed GAPSPLIT\* cover a higher 3D- $\mathcal{V}$  than the  $\mathcal{RS}$  approach, as much as 40 % and 55 % more in the IEEE 6- and 68-bus systems, respectively. In contrast, as evidenced by Table 3.3, the proposed GAPSPLIT\* approach cover a significantly higher 3D- $\mathcal{V}$  than the  $\mathcal{MGC}$  approach, in the order of  $15\times$  and more than  $10^{11}\times$  magnitude, respectively, in the IEEE 6- and 68-bus systems. Admittedly, the poor performance of the  $\mathcal{MGC}$  approach is a reflection of its objective function in solving the optimisation problem. Thus, the samples generated by the  $\mathcal{MGC}$  approach will only cover a small volume even as the approach generates more OCs, as a result of choosing the same cheap generator combinations to minimise cost.

For a more exhaustive evaluation, this work investigates the performance of the proposed GAPSPLIT\* and  $\mathcal{RS}$  approaches considering 10'000 random variable selections  $\{\hat{\Omega} \subset \Omega^N\}$ , where  $|\hat{\Omega}| = 3$ . The results are summarised in Table 3.3, which shows the  $\mathcal{V}_{|\Omega^N|=3}$  of the candidate approaches. Overall, the proposed GAPSPLIT\* approach cover 10 % and 37.5 % more volume than the  $\mathcal{RS}$  approach in the IEEE 6- and 68-bus systems, respectively.

These results imply that the proposed GAPSPLIT\* is suitable for generating a wide range of OCs, which is necessary to enrich the database, especially as the integration of intermittent renewable energy sources becomes the norm.

Table 3.3:  $\mathcal{V}$  of 5000 samples computed for 10'000 random variable subsets  $\{\hat{\Omega} \subset \Omega^N\}$ , where  $|\hat{\Omega}| = 3$  for different sampling approaches

Approach	$\mathcal{V}$	
	IEEE 6 bus	IEEE 68 bus
GAPSPLIT*	$(0.77 \pm 0.60)$	$(407 \pm 474)$
$\mathcal{MGC}$	$(0.05 \pm 0.24)$	$< 10^{-11} (< 10^{-11})$
$\mathcal{RS}$	$(0.70 \pm 0.55)$	$(296 \pm 250)$

### 3.4.3 Addressing GAPSPLIT issues

In this study, this work investigates the performance of the first proposed modification of the GAPSPLIT\* approach (Sect. 3.2.3) to address the low coverage issue of GAPSPLIT (Algorithm (i)). This work contrasts the proposed modification (ii) that introduces the set of infeasible samples  $\Omega^{G'}$ , the proposed modification (i) that utilises only secondary variables  $\Omega^{Q*}$ , and regular GAPSPLIT.

To preface this comparison, regular GAPSPLIT can generate on average three and six unique OCs in the IEEE 6-bus and IEEE 68-bus systems, respectively, before converging to an infeasible region. Subsequently, the maximal gap  $\Delta X_{\tilde{n}}^{(max)}$  (Eq. (3.4)) remains the same, and the algorithm is unable to generate any more feasible OCs. The comparison with the proposed modification (ii) is demonstrated by the results in Fig. 3.6. Concretely, the figure shows the share of infeasible OCs  $\beta = \frac{|\Omega^{G'}|}{|\Omega^{G''}|}$  in the IEEE 6- and 68-bus systems as the candidate sampling approaches generate many OCs  $|\Omega^G| \rightarrow a$ , where  $a \gg 1$  is a large number. As evidenced by Fig. 3.6, the proposed modification (ii) has a higher value of  $\beta$  in earlier iterations for both systems that decrease as more OCs are generated. The value of  $\beta$  decreases from 19.5 % when  $|\Omega^{G''}| = 210$  to 4.9 % when  $|\Omega^{G''}| = 15593$  in the IEEE 6-bus system, and from 29.3 % when  $|\Omega^{G''}| = 3622$  to 21.3 % when  $|\Omega^{G''}| = 9366$  in the IEEE 68-bus system. This downward trend of  $\beta$  indicates an improved performance of the proposed modification (ii) as the algorithm generates more OCs. The proposed modification (ii) works for both small and relatively large systems as  $\beta$  decreases when  $|\Omega^S|$  grows in both systems. In contrast, the share of infeasible samples  $\beta$  increases in both systems for regular GAPSPLIT. On the other hand, from this perspective of  $\beta$ , the proposed modification (i) has

the best performance as it generates on average only one infeasible OC in both systems. Modification (ii) also avoids converging to infeasible regions. Admittedly, the modification (i) trivially avoids converging to infeasible regions as its optimisation discards the hard constraint on the primary target. Additionally, its objective function aims to minimise the distance to the candidate targets. However, in terms of generating OCs in a non-convex and disconnected feasible space, as is the case in power systems, these results indicate that the first proposed modification of GAPSPLIT\* improves on the low coverage issue of GAPSPLIT.

Furthermore, this work notes that modification (ii) is preferred to modification (i) in small systems by the results in Fig. 3.7b and Fig. 3.7a. The figures show a scatter-plot of OCs generated by the modifications (ii) and (i), respectively. As evidenced by the figures in Fig. 3.7, the OCs generated with modification (ii) cover the entire feasible space and not only the boundaries, and is thus the preferred approach. However, this preference of modification (ii) over modification (i) is not entirely visible in larger systems. In larger systems (e.g., IEEE 68-bus), there is a higher share of infeasible OCs  $\beta$  (e.g.,  $\beta = 5.1\%$  and  $23.5\%$ , respectively, in the IEEE 6- and 68-bus systems). This higher value of  $\beta$  in the IEEE 68-bus system denotes an increase in the number of iterations required by the algorithm before termination, and invariably, an increase in the computation time of modification (ii). On that note of computation time, modification (i) is suitable for large networks. However, the  $\mathcal{V}$  comparison between the two approaches in Fig. 3.5b indicates that coverage of OCs generated using the proposed modification (ii) is marginally better than modification (i). In the rest of the manuscript, unless otherwise stated, this work consider the proposed modification (ii) as GAPSPLIT\*.

It is also worth highlighting that GAPSPLIT\* sampling with both primary and secondary variables is preferred over GAPSPLIT\* sampling with only primary variables, as demonstrated by the distribution of samples in Figs. 3.8b 3.8a.

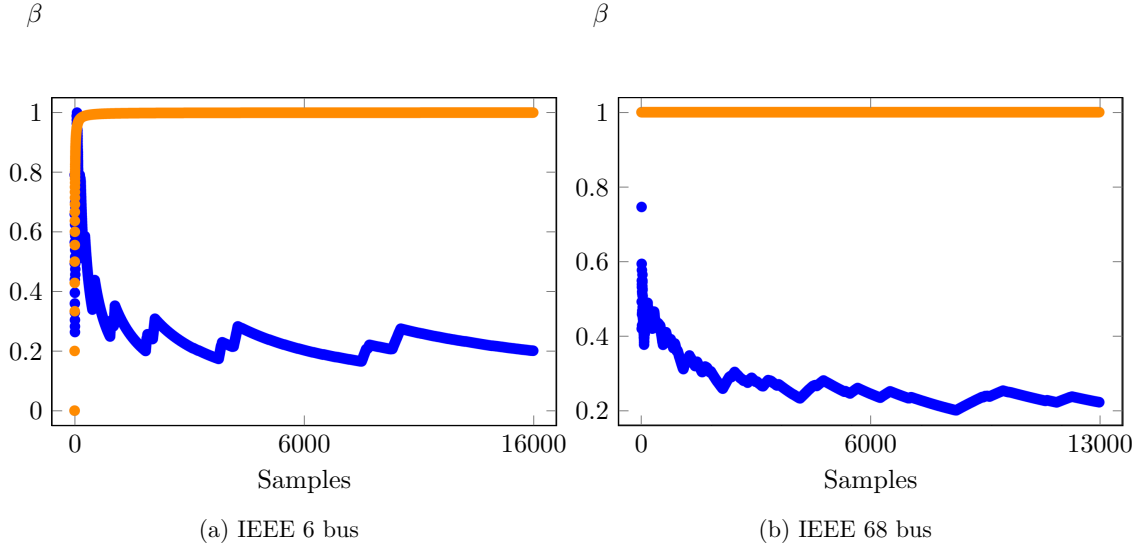


Figure 3.6: The value of  $\beta = \frac{|\Omega^{G'}|}{|\Omega^{G''}|}$  for the proposed modification (ii) (●) reduces while regular GAPSPLIT's (●) increases as more samples are generated in both (a) small and (b) larger systems.

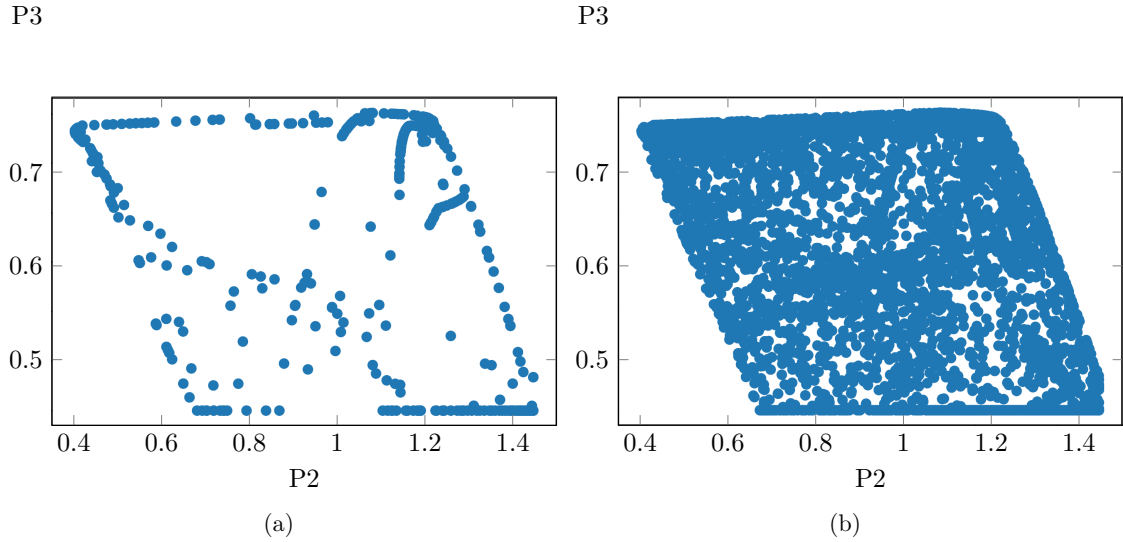


Figure 3.7: Modification (ii) better distributes samples across the feasible space than modification (i) in small systems, respectively in (b) and (a).

### 3.4.4 Measuring performance of samplers

This case study contrasts the proposed  $\mathcal{V}$  metric with the  $\mathcal{COV}$  metric to measure coverage of the feasible space by generated OCs. As an illustrative example on the IEEE 68-bus system, this work uses the two candidate metrics to compute coverage of different multivariate distributions in Fig. 3.8a and Fig. 3.8b. Concretely, the figures depict a scatter-plot of OCs generated using GAPSPLIT\* sampling with only primary variables and GAPSPLIT\* sampling with both primary and secondary variables, respectively.

As evidenced by Fig. 3.8, the  $\mathcal{COV}$  value is the same in both Fig. 3.8a and Fig. 3.8b, while the  $\mathcal{V}$  value is approximately 100 % higher in Fig. 3.8b than in Fig. 3.8a. This result shows that the  $\mathcal{COV}$  does not distinguish between different multivariate distributions.

For a more exhaustive evaluation, this work considers four different sets of 5000 OCs that are generated by varying the number of secondary variables  $|\Omega^{Q*}| = \{0, 1, 2, 3\}$  in the GAPSPLIT\* algorithm. Subsequently, this work computes the  $\mathcal{COV}$  and  $\mathcal{V}$  of the different sets considering 10'000 random variable selections  $\{\hat{\Omega} \subset \Omega^N\}$ , where  $|\hat{\Omega}| = 3$ . The results are in Fig. 3.9, which depicts a scatter-plot of the proposed  $\mathcal{V}$  metric against the  $\mathcal{COV}$  metric for the same sets of OCs. Overall, the proposed  $\mathcal{V}$  can distinguish the coverage of 'good' from 'bad' sample distributions while  $\mathcal{COV}$  can not. Concretely, the proposed  $\mathcal{V}$  has a wider range of values (0.25, 2.50) and higher standard deviation of 0.74 than the  $\mathcal{COV}$  metric with values ranging between (0.9985, 0.9995) and standard deviation of  $< 0.001$ . Additionally, other coverage metrics like  $\lambda$  and  $h$  range between (0.65, 1.02) and (9.89, 9.99), respectively, with standard deviations of 0.18 and 0.03 for the same dataset. There, just as  $\mathcal{COV}$ ,  $h$  can not distinguish 'good' from 'bad' sample distributions, while  $\lambda$  can distinguish. However, as pointed out in Sec. 3.3, the metric  $\lambda$  is not suitable as it does not computationally scale well to large number of samples. The  $\mathcal{V}$  metric considers the multivariate distribution of OCs to measure coverage, and that makes it a better metric to quantify the spread of OCs across multidimensional space. Table 3.4 summarises the comparison between  $\mathcal{COV}$ ,  $\lambda$ ,  $h$  and the proposed  $\mathcal{V}$  metrics for the different sets of sample distributions, showing that the proposed  $\mathcal{V}$  is more suitable to measure coverage by differentiating the distinct sets of OCs.

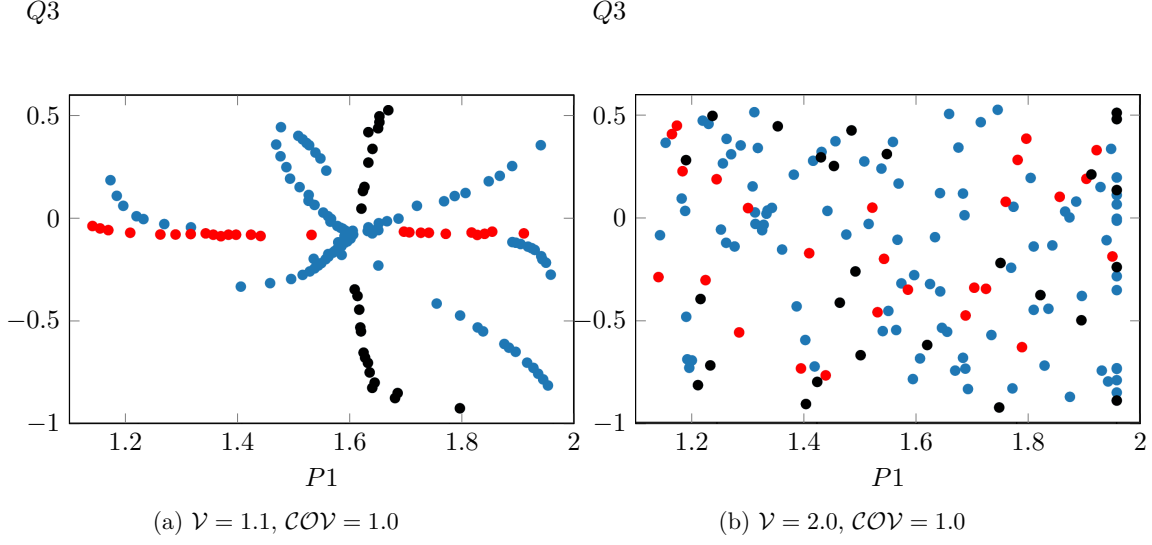


Figure 3.8: (a) GAPSPLIT\* with only primary targets (shown for  $P1$ ( $\bullet$ ) and  $Q3$ ( $\bullet$ )) generates samples that are distributed along the axis of that variable and do not cover the entire feasible space. (b) Proposed use of GAPSPLIT\* uses primary and secondary targets to cover the entire feasible space

Table 3.4: The  $\mathcal{V}$ ,  $\mathcal{COV}$ ,  $h$  and  $\lambda$  values of four different sets of 5000 samples computed for 10'000 random variable selections  $\{\hat{\Omega} \subset \Omega^N\}$ , where  $|\hat{\Omega}| = 3$

$ \Omega^{Q*} $	$\mathcal{V}$	$\mathcal{COV}$	$h$	$\lambda$
0	$(0.49 \pm 0.36)$	$0.99(< 0.01)$	$(9.84 \pm 0.16)$	$(0.75 \pm 0.15)$
1	$(0.99 \pm 0.78)$	$0.99(< 0.01)$	$(9.75 \pm 0.24)$	$(0.60 \pm 0.06)$
2	$(1.02 \pm 0.72)$	$0.99(< 0.01)$	$(9.90 \pm 0.07)$	$(0.46 \pm 0.04)$
3	$(1.09 \pm 0.86)$	$0.99(< 0.01)$	$(9.83 \pm 0.18)$	$(0.49 \pm 0.04)$

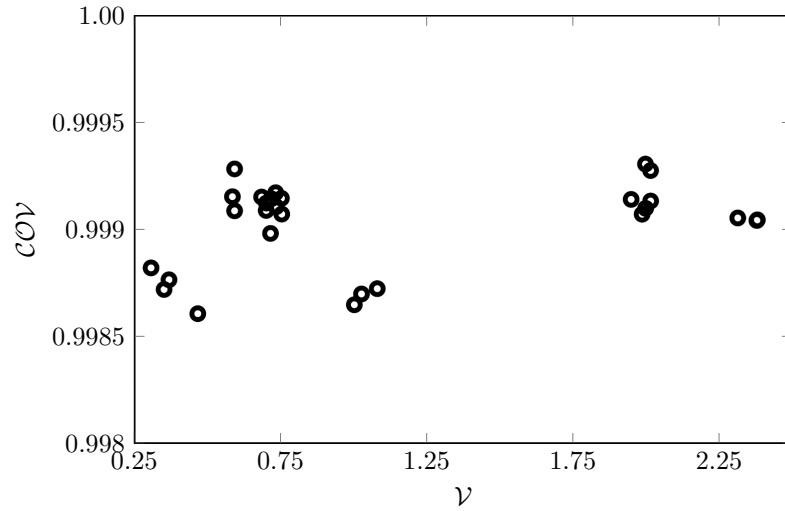


Figure 3.9: Samples that have similar  $\mathcal{COV}$  values are differentiated with the proposed  $\mathcal{V}$

Table 3.5: Computation time to generate 100'000 samples in IEEE 118-bus generator space

Approach	Average time	Total time
<b>GAPSPLIT*</b>	$(0.57 \pm 0.20) \text{ s}$	16 h
<b>GAPSPLIT</b>	$(0.78 \pm 0.24) \text{ s}$	21 h
<i>MGC</i>	$(0.60 \pm 0.20) \text{ s}$	17 h
<i>RS</i>	$(0.64 \pm 0.28) \text{ s}$	17 h

### 3.4.5 Computational Performance & Scalability

This case study tests the computational performance and scalability of the proposed split-based sampling approach and the proposed coverage metric to larger systems. The performance was tested for the number of OCs generated and the size of the power system (number of dimensions of variables). In this study, **GAPSPLIT** was modified to store infeasible OCs for comparison (to prevent early convergence to an infeasible OC), while the proposed **GAPSPLIT\*** approach is modified as described in Sect. 3.2.3. To study the scalability of the  $\mathcal{V}$  metric, on the IEEE 118-bus system, 100 random subsets of variables with dimension sizes  $|\Omega^N| = \{2 - 7\}$  and sample size  $|\Omega^G| = 5000$  are drawn and the  $\mathcal{V}$  is computed for each subset of variables. The choice of dimension sizes  $|\Omega^N| = \{2 - 7\}$  was influenced by the upper-limit of computing the  $\mathcal{V}$  metric in higher dimension  $|\Omega^N| > 6$  and computing the  $\mathcal{V}$  metric of 100 samples was tractable for  $|\Omega^N| > 3$ . Fig. 3.10a shows that the random selection of variables does not influence the  $\mathcal{V}$  and the mean and median values of the normalised  $\mathcal{V}$  are suitable to approximate the  $\mathcal{V}$  for dimension sizes  $|\Omega^N| = \{2 - 7\}$ . Fig. 3.10b shows the relationship between the  $\mathcal{V}$  of random subsets for dimensions 3 and 7. The correlation shows that the average value of  $\mathcal{V}_{|\Omega^N|=3}$  is sufficient to approximate  $\mathcal{V}_{|\Omega^N|=7}$  for the IEEE-118 bus test system, and therefore can extrapolate that computing a reduced  $\mathcal{V}$  is a good estimator for  $\mathcal{V}$  in higher dimensions.

Table 3.5 shows the computational times to generate 100'000 OCs with different approaches on the IEEE 118-bus system. **GAPSPLIT** takes 21 hours, in contrast to 17 hours by the *MGC* and *RS* approaches. Albeit, the OCs generated using **GAPSPLIT** cover a 30 % larger volume than OCs generated using *RS*. This increase in total time for **GAPSPLIT** is as a result of increased time to sort  $\Omega^G$  and find the maximal gap  $\Delta X_{\tilde{n}}^{(max)}$  as more OCs are generated.

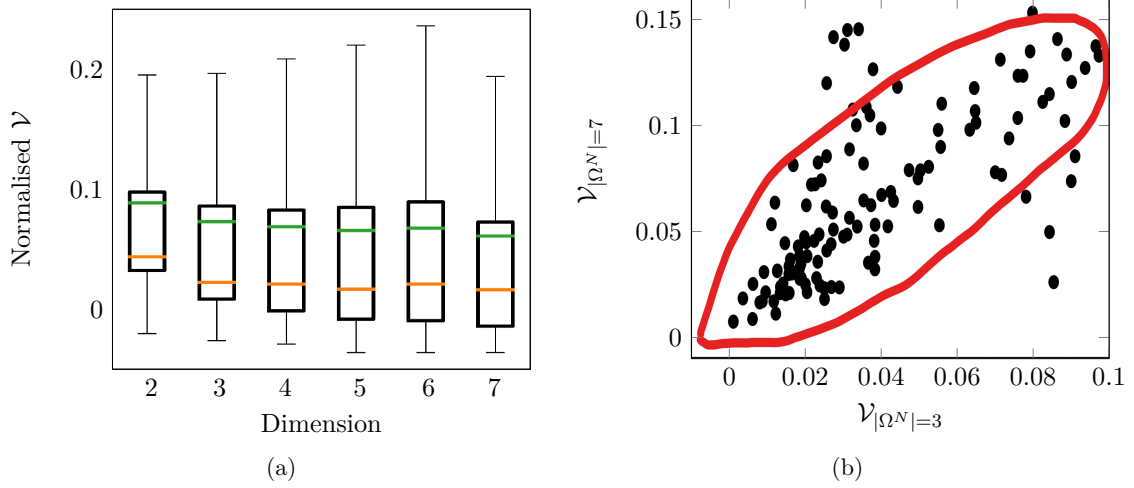


Figure 3.10: (a) The normalised  $\mathcal{V}$  mean (—) and median (—) values of random subset selections are similar across different dimensions (b) There is some correlation between the  $\mathcal{V}$  of a random subset and the  $\mathcal{V}$  of the full dimension

Table 3.6: ANN trained for dynamic security on 1000 OCs from IEEE 68-bus system

Approach	$\Pi^1$	$F1$ -score	Accuracy	Precision	Specificity
<b>GAPSPLIT*</b>	93.3 %	99.5 %	98.4 %	99.9 %	99.9 %
<b><math>\mathcal{RS}</math></b>	84.4 %	93.7 %	91.2 %	93.5 %	90.0 %

The moving average (with a sliding window of 1000) of the time it takes to sample an OC for **GAPSPLIT** shows a linear increase over time, with a slope angle  $\angle \frac{\Delta t}{\Delta s}$  of  $45^\circ$ . The proposed second modification of the **GAPSPLIT\*** approach from Sect. 3.2.3 mitigates this increase in time by regulating the size of the set of gaps and efficiently sorting newly generated OCs.

### 3.4.6 Dynamic Security and Machine learning

This case study tests the generated data when applied to the intended use case of ML-based DSA on the IEEE 68-bus system. The dynamic security labels of OCs from **GAPSPLIT\*** and  **$\mathcal{RS}$**  were simulated, and different ML models including SVM, Adaboost, Xgboost, DT, and ANN were trained.  $\Pi^1$  is the share of infeasible OCs in the dataset. The results in Table 3.6 show that the generated data from **GAPSPLIT\*** results in better performances when training an ANN across the metrics of test accuracy,  $F1$ -score, precision, and specificity by 7.2%, 5.8%, 6.4% and 9.9%, respectively. **GAPSPLIT\*** generated more insecure OCs

Table 3.7: DT trained for dynamic security on 1000 OCs from IEEE 68-bus system

Approach	$\Pi^1$	F1-score	Accuracy	Precision	Specificity
<b>GAPSPLIT*</b>	93.3 %	99.6 %	99.2 %	99.6 %	99.6 %
<b><math>\mathcal{RS}</math></b>	84.4 %	99.2 %	98.8 %	99.0 %	99.5 %

Table 3.8: F1-score for 5 different ML models trained on 1000 OCs from the IEEE 68-bus system. Each type of model is trained 100 times.

Approach	SVM	Adaboost	Xgboost	DT	ANN
<b>GAPSPLIT*</b>	99.0 %	99.5 %	99.2 %	99.5 %	98.8 %
<b><math>\mathcal{RS}</math></b>	96.0 %	98.0 %	98.6 %	98.2 %	91.9 %

which can enhance the prediction accuracy of predicting insecure OCs. Maximising the accuracy for insecure OCs and reducing false negatives is important as these type of errors can lead to power blackouts which are significantly worse than false positives. For DTs the values remained similar in **GAPSPLIT\*** and  **$\mathcal{RS}$**  as shown in Tab 3.7. For a more exhaustive comparison, each of SVM, Adaboost, Xgboost, DT, and ANN models were trained 100 times on data from **GAPSPLIT\*** and  **$\mathcal{RS}$** . The results in Table 3.8 show that the generated data results in marginally better performance across the two approaches for SVM, Adaboost, Xgboost, and DT models. A 6.9 % improvement is recorded for the ANN.

### 3.5 Conclusion

A systematic approach to creating representative databases is pivotal to the adoption of ML methods for real-time (dynamic-) security assessment. This work proposes a novel split-based sampling approach **GAPSPLIT\*** to generate representative samples that systematically explore the feasible space of power systems. The key feature of the split-based sampling is the ability to consider model-based constraints  $g(X) \leq 0$  when generating a sample (OC) in the optimisation. When using this sampling approach for power systems, the physical constraints can be considered for the steady-state in the form of the AC network power flow constraints, as used when optimising the generator dispatches in an ACOPF model. The proposed split-based sampling aims for diverse data by jumping from

one part of the solution space to another underrepresented part to cover a larger space (distribution) with fewer OCs. In the IEEE 68-bus system, samples generated using the proposed split-based sampling cover 37.5% more volume than with  $\mathcal{RS}$ . The proposed  $\mathcal{V}$  is better suited than distance-based metrics to quantify the performance of a *generic sampler* and differentiate good from bad sample distributions. The proposed split-based approach takes 0.57 s on average to generate samples for the IEEE 118-bus system. Future work will involve exploiting historical data in the sampling procedure to generate new OCs that improve the information gain of the classifier. There, the proposed algorithm as a sequential process shall consider another variable that creates balanced datasets. The vision is to use this proposed algorithm as a baseline then consider “active learning” that can use discriminative information on the class distribution in the sequential sampling process.

## Chapter 4

# Trade-offs in Generating Balanced Datasets

*“There are no solutions, there are only trade-offs.”*

— Thomas Sowell

This chapter presents a novel, unified approach for generating high-quality datasets for training machine-learned (ML) models for real-time security assessment in power systems. The proposed approach balances the trade-off between historically relevant operating conditions (OCs) and rare but feasible OCs. Unlike conventional methods that rely on historical records or generic sampling, the proposed approach results in datasets that generalise well beyond similar distributions. The proposed approach is validated through experiments on the IEEE 118-bus system, where an ML model trained on data generated using the proposed approach achieved 97 % accuracy in predicting the security label of rare OCs, outperforming baseline approaches by 41 % and 20 %. This work is crucial for deploying reliable machine-learned models for real-time security assessment in power systems undergoing decarbonisation and integrating renewable energy sources.

## 4.1 Introduction

In Chapter 3, this thesis argued for a renewed interest in data generation approaches [21, 55, 56, 57, 59] to produce representative datasets, especially as low *quality* data leads to training inaccurate models. Chapter 2 emphasized the complexity of the data generation challenge primarily underscored by the three contrasts of *historical relevance*, *coverage* and *discriminative relevance* [42, 1] that define *quality* datasets.

An example of these contrasts is between *historical relevance* and *coverage*. Approaches maximising *coverage* (e.g. [55, 56, 57, 59]) aim to uniformly span all the possible feasible OCs (region A of Fig. 4.1) and do not consider the dependency structures between variables thereby sacrificing *historical relevance* (region B of Fig. 4.1). While approaches that consider dependency structures of variables via *historical relevance* (e.g. copula modelling to capture complex non-gaussian marginal distributions and non-linear multivariate dependencies [42, 21], autoencoders and conditional variational autoencoders [44]) do not consider other feasible but rare OCs (region C of Fig. 4.1) thereby sacrificing *coverage*. Another contrast is between *historical relevance* and *discriminative relevance*. Approaches that focus on *historical relevance* (region B of Fig. 4.1) aim to mimic typical power systems operations and retain variable dependency structures but do not generalise to OCs not in historical records. While approaches that focus on *discriminative relevance* (line D of Fig. 4.1) aim to target security decision boundaries, so-called high information content regions but do not consider dependency structures between variables that represent typical power system OCs. A further contrast is between *coverage* and *discriminative relevance*.

Approaches that focus on maximising *coverage* may miss out on regions that span the security boundary. While approaches that focus on *discriminative relevance* assume a stationary security decision boundary and do not generalise to "rare" OCs (region C of Fig. 4.1).

The existing gap of all the aforementioned state-of-the-art approaches for generating synthetic datasets for power system security assessment is the lack of a unified approach

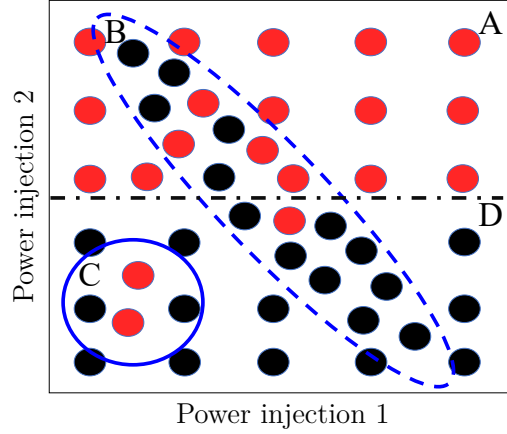


Figure 4.1: Quality datasets balance historical relevance (—), coverage (□), and discriminative relevance (---) so that ML-models can be trained for accurately predicting secure (●) and insecure (●) OCs.

to efficiently combine all three contrasts of *quality* datasets. The Wasserstein distance [142] addresses this gap and leverages advances in optimal transport research [143] to deal with the comparison of distributions. This ability to compare distributions allows trading-off *historical relevance* and *coverage* of generated OCs thereby preserving relevant dependency structures while generating rare OCs. The Wasserstein distance [142] can be thought of in 1-dimension as the earth mover’s distance and calculates how much work it takes to transport the mass of one distribution to another. Besides its intuitiveness, the Wasserstein distance can make meaningful comparisons between distributions with non-overlapping support such as the popular Kullback-Leibler divergence [144], and by extension the Jensen-Shannon divergence. Other metrics include the total variation and Cramér distances [145], which are generally considered as computationally efficient measures of distance, particularly for low-dimensional data. However, the Wasserstein distance is preferred for distributions with complex structure and high dimensionality [146]. While computing the Wasserstein distance [142] in high dimensions is non-trivial, the Sliced-Wasserstein distance [147] exploits the closed-form of projected one-dimensional distances and has acceptable statistical and asymptotic properties.

### 4.1.1 Contributions

The contribution of this chapter is the combination of the three properties of *historical relevance*, *coverage*, and *discriminative relevance* that were individually considered to generate *quality* datasets in previous research. For the first time, this chapter proposes a novel unified approach that considers all three properties that define *quality* datasets to generate information-rich and historically relevant datasets while considering rare OCs to train reliable models for power system security. The proposed approach leverages advances in optimal transport research and introduces the Wasserstein distance as a metric. The proposed metric allows to efficiently combine historically relevant OCs modelled with copulas and rare OCs modelled using state-of-the-art split-based generic sampling. Additionally, the proposed approach uses entropy to redirect sampling to other regions of the feasible space.

## 4.2 Single criterion-based sampling approaches

The three existing single criterion-based sampling approaches have strengths and limitations for security assessments. Security assessment  $X \rightarrow \mathcal{Y}$  takes as an input the power system OCs  $X$  and outputs security labels  $\mathcal{Y}$  for a set of probable contingencies  $\Omega^J$ . The security labels  $\mathcal{Y}_j^m \in \{0, 1\} \forall j \in \Omega^J$  represent secure and insecure OCs, respectively, where  $m$  is the index of OC  $X^m$ . Typically, the input variables are the static pre-fault set-point of all generators and loads. These input variables define the OC  $X^m$  and are bounded by the power system's physical limits, such as generator limits, line limits, and complex network constraints.

The first two types of sampling approaches, *historical* and *generic sampling*, focus on generating representative pre-fault OCs  $X$ . *Importance sampling* approach focuses on generating  $X$  aiming at the inverse  $\mathcal{Y} \rightarrow X$  by targeting information-rich regions  $\hat{\alpha} \subseteq \alpha$ , where  $\alpha$  denotes the feasible space containing all possible OCs  $X$ . Each approach subsequently focuses on a specific property and has an associated metric to measure sampling quality.

### 4.2.1 State-of-the-art historical sampling

*Historical sampling* focuses on the *historical relevance* property of sampling quality. *Historical sampling* approaches aim to generate similar OCs to observed historical data by learning the underlying distribution and the dependency structures of historical data. The set of historical observations is  $\Omega^H$  with the data  $X_n^m \forall n \in \Omega^N, m \in \Omega^H$ , where  $\Omega^N$  is the set of power systems variables (e.g. loads and injections). For simplicity, let us denote  $X_n$  as a vector of all observations for the  $n$ -th variable,  $X^m$  as a data vector for the  $m$ -th observation and  $p = |\Omega^N|$  as the cardinality of set  $\Omega^N$ . From a statistical perspective, the historical data in  $\Omega^H$  are assumed to be drawn from an unknown true distribution, represented by the continuous random variable  $\mathbf{X}$ . A *historical sampling* approach starts by approximating this true distribution's random variable  $\mathbf{X}$  by fitting a statistical model  $\hat{\mathbf{X}}$  to the observed data  $\Omega^H$ . Then, the approach applies Monte Carlo (MC) sampling. MC sampling is widely used to randomly sample probability distributions to generate new data. A challenge of *historical sampling* is the separation of marginal distributions from a multivariate distribution with non-linear dependencies, as power system OCs are typically load and generator injection profiles with non-linear dependencies such as renewables. Previous works [148, 42, 149] consider copula-based sampling models (CSM) in power systems, as copulas can separate the dependency structure of marginal distributions from a multivariate distribution. Another challenge is that OCs can be observed in disjoint clusters that follow distinct statistical characteristics due to unique power system modes, e.g., considering different seasons or times of the day. Previous work in [21] partitions the observed data  $\Omega^H$  into distinct clusters of similar characteristics profiles. This work approximates  $\mathbf{X}$  using a combination of copulas and clustering.

A  $d$ -dimensional copula,  $C : [0, 1]^p \rightarrow [0, 1]$  is a cumulative distribution function (CDF) with uniform marginals that provides a suitable way to separate the marginal distributions of  $X_n$  from their dependency structure. The multivariate CDF,  $F_X$ , with marginal

distributions  $F_1, \dots, F_p$  is then

$$F_X(X_1, \dots, X_p) = C\left(F_1(X_1), \dots, F_p(X_p)\right) \quad (4.1)$$

which represents Sklar's theorem [150]. The copula,  $C$  is unique if all the marginal distributions  $F_1, \dots, F_p$  are continuous. Without loss of generality to other variations of copulas, a single  $d$ -dimensional Multivariate Gaussian (MG) copula can model the dependency structure parameterised by the correlation matrix  $Q$  as

$$C_Q(\mathbf{u}) := \Phi_Q\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p)\right) \quad (4.2)$$

where  $\Phi(\cdot)$  is the standard univariate normal CDF,  $\Phi_Q(\cdot)$  is the joint CDF of a MG variable with mean  $\mu = 0$ , covariance matrix  $\Sigma = Q$ , with uniform marginals  $u_p = [0, 1]$ , and  $\mathbf{u} = (u_1, \dots, u_p)$ . The correlation matrix can be transformed using the Cholesky decomposition  $Q = A^T A$ , where  $A$  is the lower triangular matrix of  $Q$ .

The sampling from a MG CSM follows Algorithm 2, where the set of OCs generated with *historical sampling*  $\Omega^O = \{\}$  is initially empty. The algorithm then partitions the observed data  $\Omega^H$  into one of  $\mathcal{L} \in \mathbb{N}$  pre-defined clusters. Then, for each independent cluster  $l \leq \mathcal{L}$ , the algorithm generates a random MG variable  $\mathbf{Z} \sim \text{MG}_p(\mathbf{0}, \mathbf{I}_p)$ , where  $\mathbf{I}_p$  is the  $d$ -dimensional identity matrix. The algorithm then determines  $\beta = A^T \mathbf{Z}$ , and computes  $\mathbf{U} = (\Phi(\beta_1), \dots, \Phi(\beta_p))$ , whose distribution represents the MG copula from Eq. (4.2) s.t.  $\text{Prob}(U_1 \leq u_1, \dots, U_p \leq u_p) = \Phi_Q(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_p))$ . Using the copula property of invariance under monotonic transformations, a resulting random OC  $m$  with data vector  $\hat{X}^m = (\hat{X}_1^m, \dots, \hat{X}_p^m)$  is obtained via the standard inverse transform method along each dimension such that  $\hat{X}_n^m = F_n^{-1}(U_n)$ . The CSM stops after  $\mathcal{S}^O \in \mathbb{N}$  OCs are generated and added to  $\Omega^O \leftarrow m$ .  $|\Omega^O|$  is the cardinality of the set.

An advantage of *historical sampling* is that the set of generated OCs  $\Omega^O$  retains the dependency structure of observed historical records  $\Omega^H$ . A limitation of *historical sampling* is that  $\hat{\mathbf{X}}$  can only generate OCs  $\Omega^O$  that are statistically similar to OCs in historical records

**Algorithm 2** Copula-based Historical Sampling

---

**Require:**  $\Omega^H, \Omega^O = \{\}, \mathcal{S}^O, l = 1, \mathcal{L} \in \mathbb{N}$

- 1: Segment  $\Omega^H$  into  $\mathcal{L}$  disjoint clusters
- 2: Compute Spearman correlation matrix  $Q$
- 3: Compute the Cholesky decomposition  $Q = A^T A$
- 4: **for**  $l \leq \mathcal{L}$  **do**
- 5:     **while**  $|\Omega^O| \leq (\lceil \frac{\mathcal{S}^O}{\mathcal{L}} \rceil \times l)$  **do**
- 6:         Compute  $\mathbf{Z} = \text{MG}_p(\mathbf{0}, \mathbf{I}_p)$
- 7:         Determine  $\boldsymbol{\beta} = A^T \mathbf{Z}$
- 8:         Compute  $\mathbf{U} = (\Phi(\beta_1), \dots, \Phi(\beta_p))$
- 9:         Compute  $\hat{X}_n^m = F_n^{-1}(U_n) \forall n \in N$
- 10:          $\Omega^O \leftarrow m$
- 11:     **end while**
- 12:      $l = l + 1$
- 13: **end for**

---

$\Omega^H$ .

### 4.2.2 State-of-the-art generic sampling approach

*Generic sampling* focuses on maximising *coverage* of a broad spectrum of varying and physically feasible OCs. *Generic sampling* approaches in power systems typically involve stratified sampling, such as the Latin Hypercube Sampling [55, 56, 57] and, recently, the sequential split-based sampling [59].

The generic sampling model (GSM) aims to generate a set of OCs  $\Omega^G$  that are uniformly distributed across the entire feasible space while ensuring that the physical constraints that represent power systems equality  $g(X^{\tilde{m}}) = 0$  and inequality  $h(X^{\tilde{m}}) \leq 0$  constraints (e.g., nodal balance, line flow, voltage, phase angle, and generator limits) are met for each generated OC  $\tilde{m} \in \Omega^G$  [59]. This algorithm is the modification (ii) in Section 3.2.3 of Chapter 3.

An advantage of *generic sampling* is that the set of generated feasible OCs  $\Omega^G$  covers a much larger volume  $\mathcal{V}$  of the feasible space as compared with the set of OCs  $\Omega^O$  generated by the CSM in Sec. 4.2.1,  $\mathcal{V}_{\Omega^G} \gg \mathcal{V}_{\Omega^O}$ .  $\mathcal{V}_{\Omega}$  is the volume covered by the OCs in  $\Omega$ . This increase in volume results from the exploration of new OCs that are not presented in historical records.

A limitation of *generic sampling* is that the set of generated OCs  $\Omega^G$  is missing relevant information like dependency structures between variables. As the correlation information from the correlation matrix  $Q$  is not considered, many generated OCs may be irrelevant, either as probable OCs or in enhancing the discriminative information for the mapping  $X \rightarrow \mathcal{Y}$ .

### 4.2.3 State-of-the-art importance sampling approach

*Importance sampling* approaches focus on the *discriminative relevance* property of sampling quality and aim to maximise the information content for the security assessment  $X \rightarrow \mathcal{Y}$ , as the goal of ML-based security analysis is the correct prediction of security labels for OCs. These approaches assume the existence of an information-rich region (area around line D of Fig. 4.1)  $\hat{\alpha} \subseteq \alpha$  as a subset of the feasible space  $\alpha$  that can be explicitly specified [55, 51] or obtained from initial OCs such that the entropy  $\mathcal{E}$  of the set of OCs (e.g.  $\Omega^O$ ) is maximised [49, 37]

$$\mathcal{E} = \sum_{i=1}^b -\Pi_{\Omega^O}^i \log_2 \Pi_{\Omega^O}^i \quad (4.3)$$

$b$  is the number of disparate labels, usually  $b = 2$  for secure and insecure labels.  $\Pi_{\Omega^O}^i = \frac{|\Omega_i^O|}{|\Omega^O|}$  is the share of OCs that have label  $i$ , where  $|\Omega_i^O|$  is the number of OCs in  $\Omega^O$  with label  $i$ , i.e.  $\Omega^O = \bigcup_{i=1}^b \Omega_i^O$ . The maximisation of entropy allows *importance sampling* approaches to generate datasets according to the probability distribution of the security boundary area.

An advantage of *importance sampling* approaches is the generation of balanced datasets by sampling on both sides of the security decision boundary (see line D of Fig. 4.1) as the result of interpolating between secure and insecure OCs [49]. Otherwise, approaches fit a multivariate distribution of OCs within the secure feasible space to generate new OCs [56]. This crucial advantage ensures that the resulting database does not suffer from a class imbalance that can affect ML models' performance. A limitation of *importance sampling*

is the exclusion of large regions of the feasible space to focus on a specific region of interest  $\hat{\alpha} \subseteq \alpha$  in high-dimensional feasible spaces  $\mathbb{R}^d, d \gg 1$ . Additionally, *Importance sampling* does not consider the dependency structures of power system variables and can miss on relevant information like different operating modes and seasonality.

Other *importance sampling* approaches included directed walk methods and game theoretic approaches, that aim to sample around the decision boundary, for instance in small-signal stability analysis [55, 52]. These methods involve approaching the boundary and collecting samples in its vicinity. However, in this thesis, such directed walk approaches were not considered due to their reliance on assuming a static decision boundary. Instead, the focus was on utilising entropy information to infer the decision boundary directly from the available data. The choice of maximizing entropy as the approach in this work was motivated by its ability to capture the overall uncertainty and information content of the data. By maximizing entropy, the aim is to find a decision boundary that maximizes the information content, allowing for a more comprehensive exploration of the system's behavior. This approach provides a more general framework for decision boundary inference, as it does not rely on specific assumptions about the underlying dynamics or require explicit sampling near the boundary.

### 4.3 Proposed unified sampling approach

The proposed approach has two phases: a knowledge discovery phase of generating feasible and diverse pre-fault OCs  $X$  and a dataset enrichment phase to generate pre-fault OCs relevant for security assessment  $X \rightarrow \mathcal{Y}$ . In the knowledge discovery phase (phase A in Fig. 4.2), the proposed approach trades off copula-based historically relevant OCs and uniformly distributed OCs. This trade-off combines the two properties of maximising *coverage* while retaining *historical relevance*. In the dataset enrichment phase, the proposed approach identifies the most information-rich region of the feasible space to initialise new pre-fault OCs using entropy.

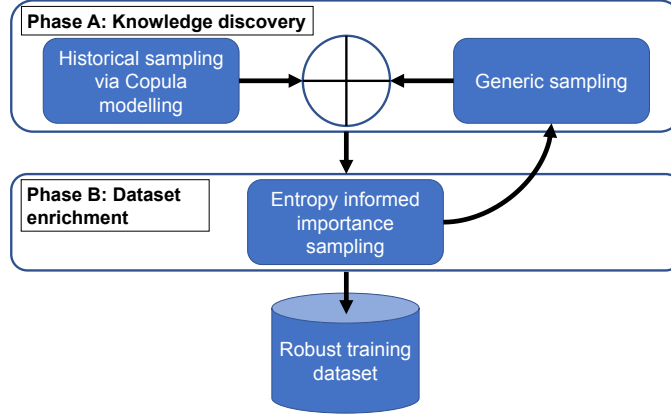


Figure 4.2: Proposed unified sampling considering all three properties of *quality* datasets. Phase A combines generic and copula-based historical sampling to generate historically relevant and diverse pre-fault OCs. Phase B directs generating labels to entropy-rich regions.

### 4.3.1 Knowledge discovery: trading off historical and rare OCs

The trade-off in the knowledge discovery phase between *historical relevance* and *coverage* of OCs minimises the maximum distance between the distribution of the generated OCs and a target probability distribution. Here, the target distribution is the historical distribution which retains the variable dependency structures, for instance, between generation and loads. The proposed Wasserstein distance helps to find a good trade-off by computing the distance between two probability measures.

**Definition:** We define  $\mathcal{P}_p(\mathbf{X}) \forall p \geq 1$  as the set of probability distributions  $\mathcal{P}_p(\mathbf{X}) = \{\eta \in \mathcal{P}(\mathbf{X}) : \int_{\mathbf{X}} \|\mathbf{X}\|^p d\eta(\mathbf{X}) < +\infty\}$ . The  $p$ -th order Wasserstein distance  $\mathcal{W}_p \forall \eta, \nu \in \mathcal{P}_p(\mathbf{X})$  is

$$\mathcal{W}_p(\eta, \nu) \equiv \inf_{\gamma \in \Gamma(\eta, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{X} - \hat{\mathbf{X}}\|^p d\gamma(\mathbf{X}, \hat{\mathbf{X}}) \quad (4.4)$$

where  $\Gamma(\eta, \nu)$  is the set of all joint probability distributions on  $\gamma$  defined on  $\mathbb{R}^d \times \mathbb{R}^d$  with

respective marginal distributions  $\eta$  and  $\nu$ . The Wasserstein distance has a closed form of

$$\mathcal{W}_p = \left( \int_0^1 |F_\eta^{-1} - F_\nu^{-1}|^p d\gamma(\eta, \nu) \right)^{1/p} \quad (4.5)$$

for one-dimensional measures, where  $F_\eta$  and  $F_\nu$  represent the respective cumulative dis-

---

**Algorithm 3** Proposed Unified Sampling: phase A

---

**Require:**  $\Omega^A = \{\}$ ,  $\Omega^O = \{\}$ ,  $\tau$ ,  $\Omega^H$ ,  $\hat{\mathbf{X}}$

- 1: Execute Alg. (2) to generate set of OCs  $\Omega^O$  using  $\hat{\mathbf{X}}$
  - 2:  $\Omega^A \leftarrow \tilde{m} : g(X^{\tilde{m}}) = 0, h(X^{\tilde{m}}) \leq 0 \forall \tilde{m} \in \Omega^O$
  - 3: **while**  $\mathcal{W}_2(\Omega^O, \Omega^A) \leq \tau$  **do**
  - 4:     Execute generic sampling modification (ii) of Section 3.2.3 to generate set of OCs  $\Omega^G$
  - 5:      $\Omega^A \leftarrow \Omega^G$
  - 6: **end while**
- 

tributions of  $\eta$  and  $\nu$ . Two sets of OCs are assumed,  $\underline{\Omega}^h$  and  $\overline{\Omega}^h$  with data matrices  $\underline{X}$  and  $\overline{X}$ , where the data  $\underline{X}^m$  corresponds to  $m \in \underline{\Omega}^h$  (and equivalently for  $\overline{\Omega}^h$ ). The Wasserstein distance between the corresponding data  $\underline{X}$  and  $\overline{X}$  is

$$\mathcal{W}_2(\underline{\Omega}^h, \overline{\Omega}^h) = \min_{\gamma \in \Gamma(\underline{X}, \overline{X})} \sum_{\underline{m}=1}^{|\underline{\Omega}^h|} \sum_{\overline{m}=1}^{|\overline{\Omega}^h|} \gamma_{\underline{m}, \overline{m}} D(\underline{X}^{\underline{m}}, \overline{X}^{\overline{m}}), \quad (4.6)$$

where  $\gamma$  is a joint distribution over the two matrices,  $\Gamma(\underline{X}, \overline{X})$  is the set of all joint distributions, and  $D(\underline{X}^{\underline{m}}, \overline{X}^{\overline{m}})$  is the Euclidean distance between the  $\underline{m}$ -th row of  $\underline{X}$  and the  $\overline{m}$ -th row of  $\overline{X}$ . Here, the second-order Wasserstein distance is the sum of the distances between each pair of OCs multiplied by the amount of probability mass that must be moved and effectively measures the "distance" between the two matrices of OCs in terms of how much "work" must be done to transform one matrix of OCs into the other. An example in Fig. 4.3 visualises the Wasserstein distance between two probability distributions,  $P(\overline{X})$  and  $P(\underline{X})$ . The larger the Wasserstein distance between any two probability distributions, the more dissimilar they are. The proposed Wasserstein distance thus allows comparing the probability distributions of synthetically generated OCs  $\Omega^A$  with historical data  $\Omega^H$ . This comparison ensures that the distribution of synthetic OCs does not deviate beyond a user-defined threshold distance  $\tau \in \mathbb{N}$  from the historical data distribution, ensuring the

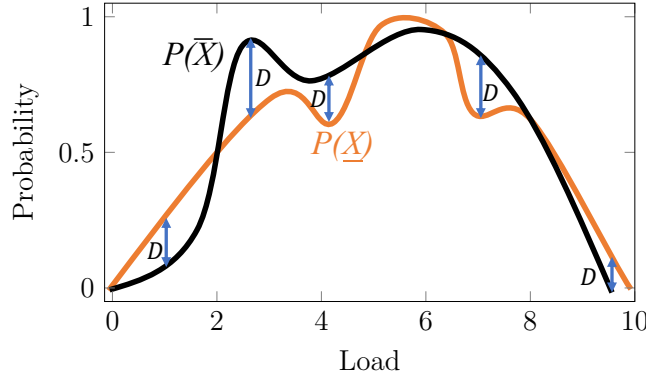


Figure 4.3: Wasserstein distance between two probability distributions  $P(\bar{X})$  and  $P(\underline{X})$  is analogous to the amount of ‘work’ to transform one distribution into another. ‘Work’ is the distance moved multiplied by the probability ‘mass’ at that distance.

dependency between power loads and generators is preserved. This proposal thus allows synthetic OCs to retain the two properties of *historical relevance* and *coverage*.

Algorithm (3) presents the proposed trade-off between *coverage* and *historical relevance* using the proposed Wasserstein distance. The algorithm starts with an empty set of generated OCs  $\Omega^A = \{\}$ . Subsequently, a copula-based model  $\hat{\mathbf{X}}$  generates historically relevant OCs  $\Omega^O$  based on available historical data  $\Omega^H$ . Only the OCs  $\tilde{m} \in \Omega^O$  that satisfy the power system constraints  $g(X^{\tilde{m}}) = 0, h(X^{\tilde{m}}) \leq 0$  (e.g., nodal balance, line flow, voltage, phase angle, and generator limits) are considered and serve as a baseline target distribution. The OCs that satisfy the power system constraints populate the set  $\Omega^A \leftarrow \Omega^O$ . The algorithm sorts the set  $\Omega^A$  in ascending order for each variable. Then, for a pre-defined Wasserstein distance threshold  $\tau \in \mathbb{N}$ , inserts new “rare” OCs using the CSM in Algorithm (2) such that  $\mathcal{W}_2(\Omega^O, \Omega^A) \leq \tau$ , which substitutes the stopping criterion that considers  $\mathcal{S}^G$  in generic sampling modification (ii) of Section 3.2.3.

### 4.3.2 Dataset enrichment and labelling: importance sampling in the context of feasibility

The dataset enrichment and labelling phase of the proposed approach incorporates an entropy-informed re-sampling of pre-fault OCs shown as phase B in Fig. 4.2. This phase

**Algorithm 4** Proposed Unified Sampling: phase B**Require:**  $\Omega^A, \Omega^B = \{\}, \mathcal{S}^B, K$ 

- 1: Segment  $\Omega^A$  into  $T$  disjoint clusters
- 2: Identify  $\Omega_{\hat{t}}^A : \mathcal{E}_{\hat{t}} = \max(\mathcal{E}_1, \dots, \mathcal{E}_t)$
- 3: Compute  $[\underline{\rho}_n, \bar{\rho}_n], \forall n \in \Omega^N$  for cluster  $\Omega_{\hat{t}}^A$  with  $\mathcal{E}_{\hat{t}}$
- 4:  $\Omega^B \leftarrow \Omega_{\hat{t}}^A$
- 5: **while**  $|\Omega^B| \leq \mathcal{S}^B$  **do**
- 6:     Execute generic sampling modification (ii) of Section 3.2.3 :  $\underline{\rho}_n \leq X_n^{\hat{m}} \leq \bar{\rho}_n, \forall n \in \Omega^N$
- 7:      $\Omega^B \leftarrow \Omega^G$
- 8: **end while**

aims to improve the knowledge discovery phase by focusing on information-rich regions of the feasible space.

Algorithm 4 presents phase B of the proposed unified sampling approach. After the initial generation of  $|\Omega^A|$  pre-fault OCs  $\Omega^A$  using Algorithm 3, dynamic simulations  $X \rightarrow \mathcal{Y}$  are performed to obtain the security labels

$$\mathcal{Y}_j^m = \{0, 1\}, \forall m \in \Omega^A, \forall j \in \Omega^J \quad (4.7)$$

for the set of probable contingencies  $\Omega^J$ . Subsequently, for each contingency  $j \in \Omega^J$ , the algorithm segments the OCs in  $\Omega^A$  using K-means clustering into  $T \in \mathbb{N}$  clusters. An example in Fig. 4.4 illustrates this phase of the proposed approach on three clusters  $T = 3$ .

The set of OCs

$$\Omega^A = \bigcup_{\forall t \in \Omega^C} \Omega_t^A \quad (4.8)$$

is segregated into  $T$  distinct clusters where  $\Omega_t^A$  is the set of OCs in  $\Omega^A$  belonging to cluster  $t$  and  $\Omega^C$  is the set of all clusters  $|\Omega^C| = T$ . In Fig. 4.4, the circles show the entailing OCs  $\Omega_t^A$  belonging to each cluster  $t = 1, 2, 3$ . Each of these clusters  $t$  has an entropy  $\mathcal{E}_t$  computed as in Eq. (4.3) substituting  $\Omega^O = \Omega_t^A$ . The cluster  $\hat{t}$  is the cluster that has the highest entropy

$$\mathcal{E}_{\hat{t}} = \max(\mathcal{E}_t | \forall t \in \Omega^C) \quad (4.9)$$

with OCs  $\Omega_{\hat{t}}^A$ . Subsequently, to improve the dataset generated from the knowledge discovery phase  $\Omega^A$ , the algorithm focuses on data  $X_n^m, m \in \Omega_{\hat{t}}^A$  and compute the bounds of the

cluster with the maximum entropy  $\Omega_t^A \subset \Omega^A$  to form a hypercube.

$$\underline{\rho}_n = \min(X_n^m | \forall m \in \Omega_t^A) \forall n \in \Omega^N \quad (4.10)$$

$$\bar{\rho}_n = \max(X_n^m | \forall m \in \Omega_t^A) \forall n \in \Omega^N \quad (4.11)$$

Using the bounds  $[\underline{\rho}_n, \bar{\rho}_n], \forall n \in \Omega^N$  as additional inequality constraints

$$\underline{\rho}_n \leq X_n^{\tilde{m}} \leq \bar{\rho}_n, \forall n \in \Omega^N, \quad (4.12)$$

in optimisation (3.7), the algorithm generates new pre-fault OCs using the generic sampling modification (ii) of Section 3.2.3. The sampling stops after generating a user-defined  $\mathcal{S}^B$  OCs. The final training dataset is then augmented by the pre-fault OCs  $\Omega^B$  and their respective labels  $\mathcal{Y}_j^m = \{0, 1\}, m \in \Omega^B, j \in \Omega^J$  obtained using dynamic simulations for each contingency. The sensitivity of the number of clusters  $|\Omega^C|$  to the target region  $\Omega_t^A$  should be considered in the analysis. It is important to find a balance between the granularity of the target region  $\Omega_t^A$  and the computational cost associated with a higher number of clusters  $|\Omega^C|$ . Increasing the number of clusters allows for more precise bounds of the target region, but it comes with the potential risk of missing out on other relevant regions and incurring additional computational costs. On the other hand, using a lower number of clusters expands the search space but may also include irrelevant regions. Therefore, the choice of the number of clusters should be carefully evaluated to ensure a reasonable trade-off between accuracy and computational efficiency.

## 4.4 Case Study

This section examines the effectiveness of the proposed unified sampling approach in generating historically relevant, representative, and balanced datasets for training ML models for security assessment. The first study investigates the trade-off between *historical relevance* and *coverage*. The second and third studies focus on the *historical relevance* and

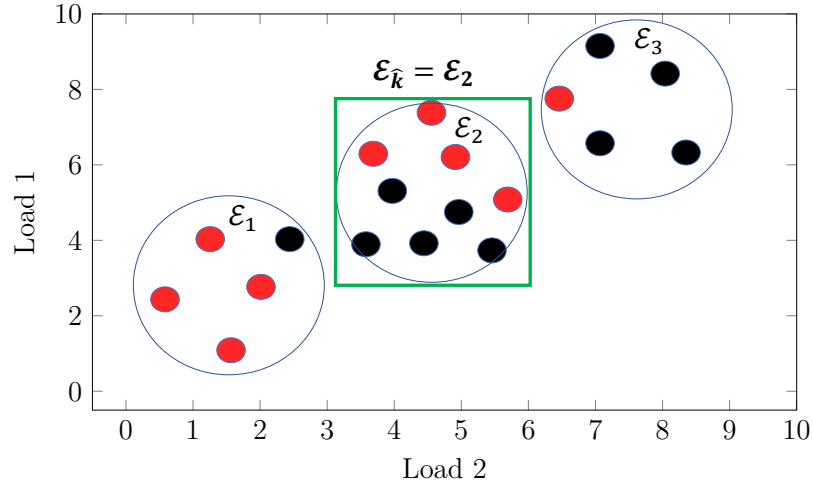


Figure 4.4: Proposed phase B computes the entropy of clusters with OCs that are secure (●) and insecure (●). The cluster with the largest entropy initialises the bounds for resampling using the generic sampling modification (ii) of Section 3.2.3.

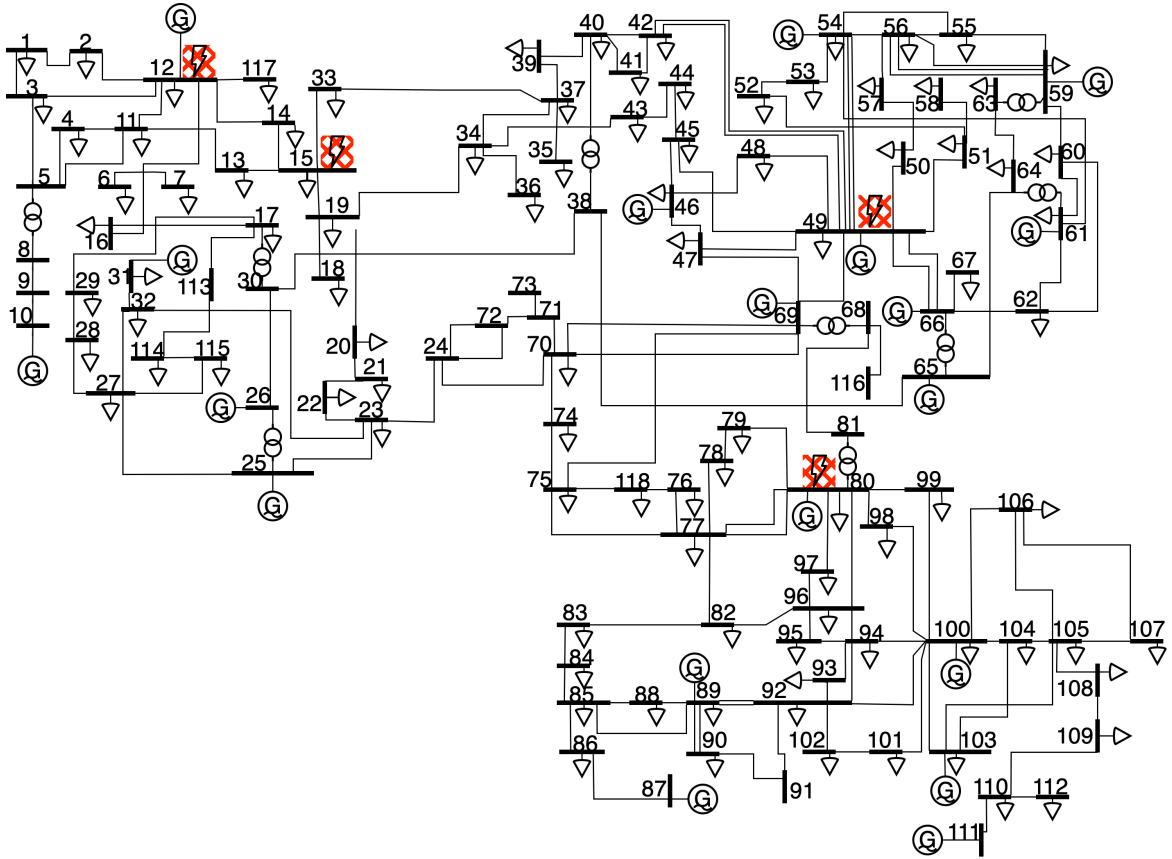


Figure 4.5: Single line diagram of the IEEE 118 bus test system showing fault locations at buses 12, 15, 49 and 80.

*coverage* metrics, respectively. The fourth and fifth studies analyse the results of security assessments using ML models trained on different databases. The final study presents the results of balancing the label distribution.

#### 4.4.1 Test system and assumptions

The case studies use the IEEE 118-bus system [132], where the observed historical data contain 14,250 measurements at 5 min intervals for a period in 2012 provided by the French Transmission SO, RTE. The original dataset spanned over 7,000 load points and 200 wind turbines and was projected relative to the upper limit in the snap-shot onto the IEEE 118-bus test system [151]. The modified IEEE 118-bus system only considers historical wind outputs without converter interfaced models for the dynamic studies. For generating pre-fault OCs, a DC approximation (of the OPF) is sufficient to demonstrate the challenge and proposed solution. Therefore, the studies use only active power loads to model and validate the compared approaches. If required, reactive power loads can be similarly modelled. The studies considered  $\mathcal{L} = 10$  disjoint clusters to capture different operation modes. Subsequently, 5,000 OCs were generated using all the approaches in contention. The baselines are the historical model (HM), CSM, and GSM, against the proposed unified sampling model (USM). The HM is data from historical records, the CSM, GSM and the proposed USM (A) and USM (B) approaches are as described in Algorithms (2), generic sampling modification (ii) of Section 3.2.3, (3), (4) respectively. The proposed USM (A) and USM (B) were studied separately, denoting phases A and B.  $T = 10$  clusters were used when studying USM (B). The observed historical data were randomly split into training, and testing sets in the ratio of 80:20, and the training set was used to build the CSM.

For the transient studies, a three-phase fault is simulated at bus 12, 15, 49 and 80 for all OCs at time 0.5 s. The fault is cleared by opening the line between buses 12 and 14 after 0.2 s. The transient stability was analysed for 10 s. The post-fault OC was considered as secure ( $\mathcal{Y}_j^m = 0$ ) if the difference between any two generator phase angles is less than  $180^\circ$ , otherwise, the OC is considered insecure ( $\mathcal{Y}_j^m = 1$ ). The results shown in this study are for

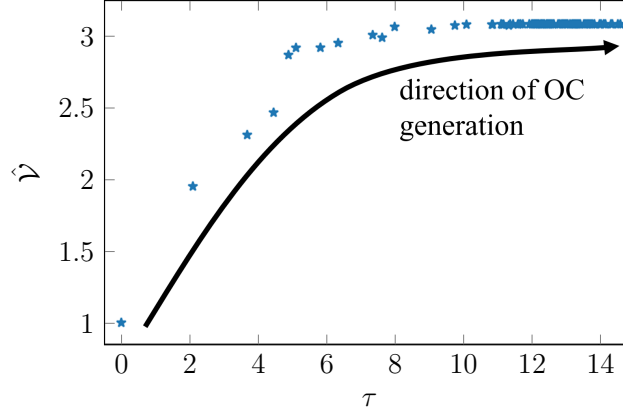


Figure 4.6: The relationship between relative volume  $\hat{\mathcal{V}}$  and Wasserstein distance threshold  $\tau$ .  $\hat{\mathcal{V}}$  increases with the number of generated OCs and plateaus around  $\tau \geq 10$ .

transient stability and can easily be extended to consider other security indices. Albeit, the complexities of considering other dynamic phenomena are not trivial. Investigating various aspects of system dynamics introduces additional complexities in terms of modeling, data requirements, and computational resources. However, this is not the main focus of the thesis.

For the security assessment, decision trees (DTs) were trained using the CART algorithm [135]. The training settings were set to their default values, except for using gini impurity to measure the quality of splits instead of entropy and limiting the maximum depth of the trees to 5. The choice of tree depth and gini impurity hyperparameters could be optimised, however, as all models are trained using the same settings, this does not affect the relative performance of the models on different datasets. The data was split into a training set and a testing set in a 75:25 ratio, with the feature variable  $X$  and the labels  $\mathcal{Y}$  serving as the inputs for training the classifier. To address underfitting or overfitting, 10-fold cross-validation was applied and one DT was trained for each contingency  $\forall j \in \Omega^J$ . The study considers the F1-score  $= \frac{2Tp}{2Tp + Fp + Fn}$  to measure the test accuracy of the DTs, where  $Tp$ ,  $Fp$ ,  $Fn$  are the true positives, false positives and false negatives, respectively. The standard nonparametric two-sample tests from the literature are used to measure *historical relevance*, the Kolmogorov-Smirnov (K-S) test and the multivariate energy test. The K-S test investigates whether the generated data from  $\hat{\mathbf{X}}$  can reconstruct the marginal distributions of the true distribution,

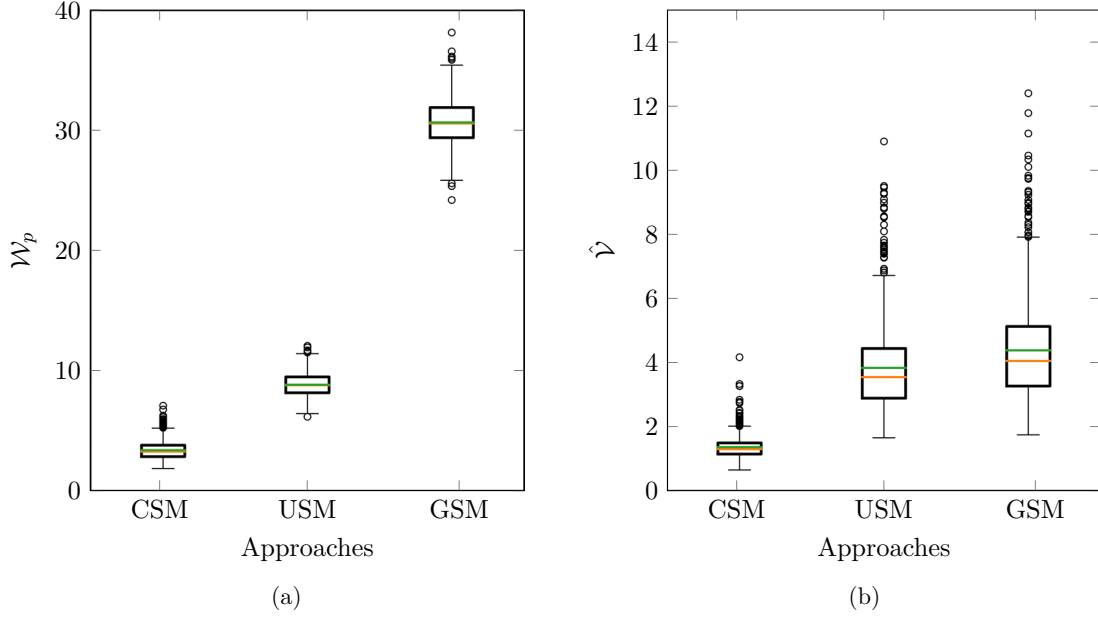


Figure 4.7: Boxplots showing mean (—) and median (—) values of 1,000 randomly selected OCs from different approaches corresponding to (a) the proposed  $\mathcal{W}_p$  and (b) the normalised  $\mathcal{V}$ .

while the energy tests show how much variable dependency is maintained in the generated dataset. Finally, The convex hull volume  $\mathcal{V}$  occupied by the generated dataset is used to compare the *coverage* of OCs generated by the three models.

All optimisation problems were implemented using the package Pyomo 5.6.8 [137] in Python 3.7.4, and the DC approximation was solved using Gurobi 9.5.0 [152] while using IPOPT 3.13.2 [153] for the AC models of the networks. The DTs were trained with the scikit-learn package version 0.18.1 [134]. The ODEs were solved to simulate the transients using *odeint* in *scipy*. All studies were conducted on a standard Windows HP desktop running an Intel(R) processor with 64 GB of RAM.

#### 4.4.2 Trading-off historical relevance and coverage

This section studies the trade-off between *historical relevance* and *coverage* properties of the proposed USM. This trade-off is achieved via the proposed Wasserstein distance  $\mathcal{W}_p$ . By adjusting the threshold  $\tau$  of the acceptable Wasserstein distance  $\mathcal{W}_p$ , the proposed USM can generate new OCs that explore the feasible space.

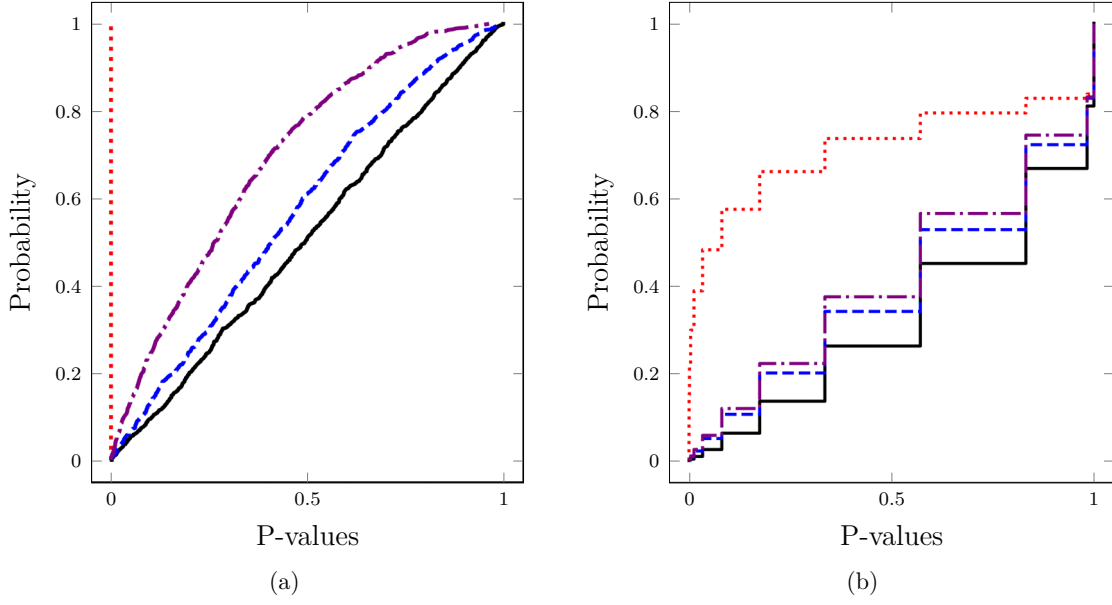


Figure 4.8: CDF of p-values corresponding to historical data (—), CSM (---), GSM (···), and the proposed USM (A) (— ·) for (a) energy tests and (b) K-S tests.

The result in Fig. 4.6 shows the coverage metric,  $\hat{\mathcal{V}}$ , and the threshold  $\tau$  on the Wasserstein metric, which measures similarity to historical data as new samples are generated.  $\tau$  plateaus around  $\tau \approx 10$ . As a result, subsequent case studies consider the USM approach until  $\tau \leq 10$ . Fig. 4.7a shows many experiment variations, specifically, the distribution of 1,000  $\mathcal{W}_p$  tests between 250 randomly selected OCs from the observed historical dataset  $\Omega^H$  and the baseline approaches, the proposed USM, CSM, and GSM. The closer the averages are to 0, the closer the distributions are to the historical distribution. The proposed USM is between CSM and GSM, balancing these two approaches by varying the threshold  $\tau$  of the acceptable Wasserstein distance  $\mathcal{W}_p$  to the target distribution. This trade-off allows the proposed USM to generate historically relevant datasets that sufficiently cover the feasible space simultaneously.

### 4.4.3 Historical relevance

This case study investigates the *historical relevance* of the proposed USM (A) and the baseline approaches (HM, CSM, and GSM) using standard statistical two-sample tests. The study randomly generates 1,000 sets of 0.5% of the generated data and compares

them to a set of randomly generated 0.5 % of observed historical test data. The distribution of the p-values from the 1,000 energy tests and 128,000 K-S tests for each approach are compared and presented in Figures 4.8a & 4.8b, respectively. For both tests, the null hypothesis assumes data from any two disparate approaches in comparison come from the same model and follow a similar distribution. Therefore, the p-values should be uniformly distributed. As a baseline for comparison, the solid black line in Figures 4.8a & 4.8b represents the CDF of the p-values that compare data randomly drawn from the observed historical training dataset and the test data. The larger the maximum difference between the CDFs of p-values, the more dissimilar the two datasets are.

The results show that the CSM approach outperforms GSM, as its p-values in the K-S and energy tests come closest to the solid black line of HM. This result is consistent with previous research as CSM can effectively model and capture marginal distributions and dependency structure of variables in higher dimensions. GSM has all 0 p-values for all tests and does not capture the marginal distribution as well, as evidenced by the K-S tests in Fig. 4.8b. Notably, in Figures 4.8a & 4.8b, the proposed USM (A) preserves most of the marginal distributions while maintaining a significant variable dependency information, even though the CSM outperforms in that regard.

#### 4.4.4 Coverage

This case study assesses the *coverage* of generated datasets from the proposed USM (A), CSM and GSM. The results shown in Fig. 4.7b present the volume  $\mathcal{V}$  covered by the generated OCs of the different approaches for 1,000 different realisations of variable selection,  $\hat{\Omega} \subset \Omega^N$ , with a cardinality of  $|\hat{\Omega}| = 3$ . The  $\mathcal{V}$  values in the figure are normalised by the minimum volume value computed using the historical data,  $\Omega^H$ , for the selected  $\hat{\Omega}$  variables, such that  $\hat{\mathcal{V}} = \frac{\mathcal{V}}{\min(\mathcal{V}_{\Omega^H})}$ . Specifically, the figure shows that the datasets generated by the proposed USM (A) and GSM have similar volume coverage, which on average is significantly larger by as much as  $4\times$  more volume than that of CSM-generated datasets. Notably, the proposed USM approach presents the best trade-off among the tested ap-

Table 4.1: Results for contingency representing a three-phase fault at bus 12

Training data	Testing data (F1-score)			$\Pi^1$
	historical	generic	rare	
HM	$0.89 \pm 0.02$	$0.60 \pm 0.21$	$0.55 \pm 0.30$	0.14
CSM	$0.84 \pm 0.04$	$0.56 \pm 0.20$	$0.41 \pm 0.23$	0.13
GSM	$0.76 \pm 0.05$	$0.99 \pm 0.01$	$0.71 \pm 0.09$	0.15
USM (A)	$0.88 \pm 0.03$	$0.96 \pm 0.01$	$0.98 \pm 0.01$	0.17

Table 4.2: Results for contingency representing a three-phase fault at bus 15

Training data	Testing data (F1-score)			$\Pi^1$
	historical	generic	rare	
HM	$0.88 \pm 0.02$	$0.49 \pm 0.20$	$0.44 \pm 0.15$	0.13
CSM	$0.82 \pm 0.02$	$0.71 \pm 0.16$	$0.66 \pm 0.17$	0.14
GSM	$0.75 \pm 0.07$	$0.97 \pm 0.02$	$0.69 \pm 0.11$	0.15
USM (A)	$0.86 \pm 0.02$	$0.96 \pm 0.02$	$0.98 \pm 0.01$	0.18

proaches as it covers nearly the same volume as the GSM (90 %), while also providing the additional benefit of high historical relevance, as studied in Sect. 4.4.3.

#### 4.4.5 Security assessment for out-of-distribution OCs

This study compares the performance of ML models trained on datasets from the baseline approaches and tests the models on three types of OCs: historical, generic (for data uniformly covering the feasible space) and rare (for data deviating from typical historical distribution).

The results in Tables 4.1-4.4 show the average and standard deviation of F1-scores for 100 DT models, where the training and testing data come from the different baseline approaches. The "rare" OCs in the testing set were randomly selected from data generated using the proposed USM (A). The models trained on the proposed USM (A) data achieved an average F1-score of at least 96 % on these "rare" OCs, outperforming models trained on data from HM and CSM by 54 % and 57 %, respectively, in contingencies 15 and 12. In contrast, models trained on data from HM and CSM had the worst performance on these "rare" testing OCs as this training data does not consider such rare cases and is biased

towards specific data from HM and CSM (e.g. does not generalise well to data from USM (A) that are considered as "rare"). GSM aims to cover the feasible space uniformly, and the DTs trained on this data performs well in some contingencies (e.g., 49 and 80) but still does underperform when using data from the proposed USM (A) for the DTs. These results suggest that the proposed USM (A) can support the development of models that generalise better to OCs from other distributions. For further comparisons, Table 4.1 presents an example for contingency 12, where the DT models trained on datasets generated by the proposed USM (A) achieved a high accuracy (F1-score of at least 86 %), outperforming models trained on data from CSM and GSM in predicting uniformly distributed and historical OCs, respectively, by as much as 40 % and 12 %. A similar analysis can be made for the other three contingencies in Tables 4.2-4.4.

The proposed USM (A) datasets resulted in DT models that are (nearly) as accurate as DTs models with training and testing data from the same distribution. For example, the USM (A) based DT models were within 2 % accuracy of the HM and CSM-based DTs in contingencies 12 and 15. Also, the F1-scores tested on historical HM data on contingencies 49 and 80 showed USM (A)-based DTs are as high as HM-based DTs (Table 4.4). In comparison, models trained on GSM datasets have an accuracy within 13 % when tested on the same historical data. Additionally, models trained on USM (A) datasets were found to have a maximum deviation of 4 % accuracy from the best model performance in all contingencies when tested on generic sampling data, where GSM had the best performance. In contrast, models from HM or CSM can have an accuracy deviation of up to 48 % in all contingencies when tested on the same data. Importantly, these results show that the proposed USM (A) performs with high accuracy across all tested datasets, showing a high level of generalisability to data from other distributions within the test settings.

#### 4.4.6 Security assessment for similar distribution OCs

This study examines the performance of ML models trained and tested on data from similar distributions.

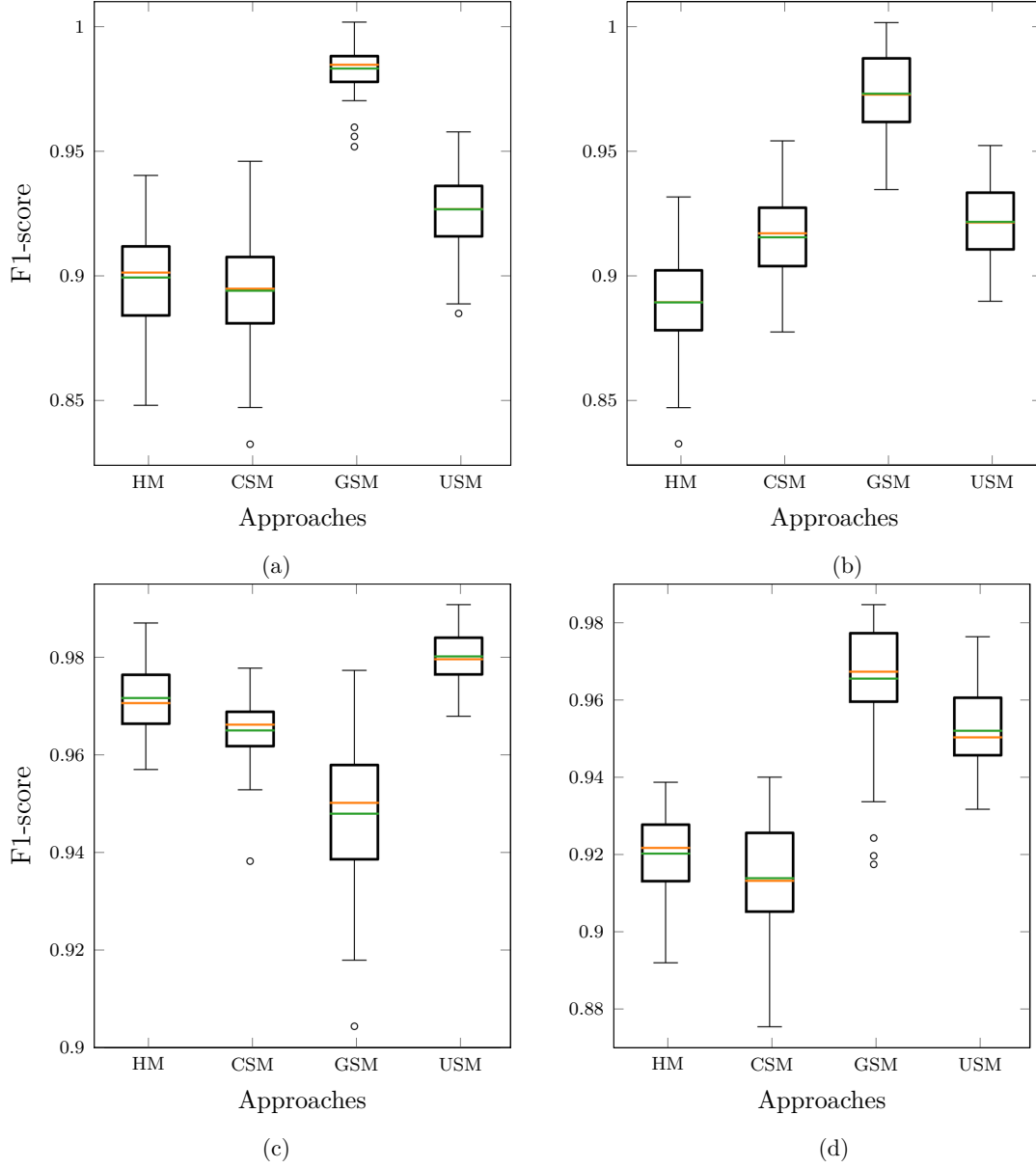


Figure 4.9: Boxplots showing mean (—) and median (—) values for 100 models trained with different datasets to predict the post-fault status of a three-phase fault near buses (a) 12 (b) 15 (c) 49 (d) 80. Both training and testing data follow a similar distribution.

Table 4.3: Results for contingency representing a three-phase fault at bus 49

Training data	Testing data (F1-score)			$\Pi^1$
	historical	generic	rare	
HM	$0.97 \pm 0.01$	$0.65 \pm 0.16$	$0.60 \pm 0.14$	0.47
CSM	$0.97 \pm 0.01$	$0.64 \pm 0.15$	$0.66 \pm 0.17$	0.46
GSM	$0.93 \pm 0.03$	$0.95 \pm 0.01$	$0.93 \pm 0.05$	0.30
USM (A)	$0.98 \pm 0.01$	$0.91 \pm 0.01$	$0.99 \pm 0.01$	0.48

Table 4.4: Results for contingency representing a three-phase fault at bus 80

Training data	Testing data (F1-score)			$\Pi^1$
	historical	generic	rare	
HM	$0.92 \pm 0.01$	$0.53 \pm 0.21$	$0.46 \pm 0.18$	0.27
CSM	$0.91 \pm 0.01$	$0.58 \pm 0.35$	$0.56 \pm 0.25$	0.27
GSM	$0.83 \pm 0.06$	$0.97 \pm 0.01$	$0.80 \pm 0.09$	0.18
USM (A)	$0.92 \pm 0.01$	$0.94 \pm 0.01$	$0.96 \pm 0.02$	0.29

The results in Fig. 4.9 show the distribution of F1-scores for 100 DT models trained using different datasets to predict the post-fault status of the system following four separate three-phase faults. The results show that all databases can provide input to train accurate models for testing data from the same distribution, with an F1-score  $\geq 88\%$ . The results indicate that for contingencies with a more balanced share of labels, such as the fault on bus 49 where the label distribution  $\Pi_{\Omega^A}^1 = 0.48$ , the proposed USM (A) outperforms the other approaches (HM, CSM and GSM). However, for other contingencies, the model trained on GSM datasets has better performance, as the OCs are more uniformly distributed. Notably, the label distribution for GSM datasets for contingency 49 is  $\approx 34\%$  lower than the other approaches (assuming an ideal distribution  $\Pi^1 = 0.5$ ) which explains the relatively poor performance compared to the other models, albeit with an F1-score  $\approx 95\%$ .

#### 4.4.7 Balancing the distribution

This section studies the performance of USM (B) as the share of labels  $\Pi^1$  could impact the performance of models and is thus an important factor to consider when generating training OCs (as also Sec. 4.4.5 showed). Therefore, this study focuses on contingencies where the share of labels  $\Pi^1 \ll 0.5$ , namely contingencies 12 and 14. USM (B) is limited

Table 4.5: Performance of USM (A) and USM (B) approaches for contingencies representing a three-phase fault at buses 12 and 15.

Approach	Contingency 12		Contingency 15	
	F1-score	$\Pi^1$	F1-score	$\Pi^1$
USM (A)	$0.92 \pm 0.02$	0.17	$0.91 \pm 0.02$	0.18
USM (B)	$0.94 \pm 0.01$	0.27	$0.95 \pm 0.01$	0.26

to generate 20 % of  $|\Omega^A|$ , e.g.,  $\mathcal{S}^B = 0.2|\Omega^A|$ .

The results in Table 4.5 show the performance of 100 DT models trained with data from the two approaches, USM (A) and USM (B). The results show that in contingency 12, the share of labels improved from  $\Pi_{\Omega^A}^1 = 0.17$  to  $\Pi_{\Omega^B}^1 = 0.27$  to and the corresponding F1-score from 0.92 to 0.94. Similarly in contingency 15, the share of labels improved from  $\Pi_{\Omega^A}^1 = 0.18$  to  $\Pi_{\Omega^B}^1 = 0.26$ . Also, this improved label share improved the F1-score from 0.91 to 0.95.

## 4.5 Conclusion

A crucial challenge is generating high-quality datasets for training machine-learned models for real-time security assessment in power systems. Conventional approaches have failed to generate datasets that generalise well beyond similar distributions, resulting in models that are not always accurate. To overcome this challenge, a novel unified approach is proposed for generating datasets that balance *historical relevance*, *coverage*, and *discriminative relevance*. The proposed approach balances historically relevant operating conditions (OCs) with rare but feasible OCs, leading to datasets representing a diverse set of possible OCs. The dynamic simulations required in the phase B of the proposed USM approach increases the computational implications for generating OCs. Since the data generation phase is performed offline, this increase in computational time can be tolerated. However, implementing "smart" strategies to enhance the efficiency of the sampling procedure, including using active learning to select OCs whose dynamic response would yield the most information can help alleviate the increase in computational time. Experimental results on

the IEEE 118-bus system demonstrate the effectiveness of the proposed approach. The model trained on data generated using the proposed approach achieved an F1-score of 91% for different contingencies and 96% accuracy in predicting the security label of rare OCs, outperforming baseline approaches. Future work can consider the difference in cost distribution of wrong predictions to improve the sampling approach, where false alarms must be avoided. This work is an important step towards developing new tools that enable the adoption of machine learning for sensitive tasks such as security assessment in power systems.

# Chapter 5

## Model Selection

*“When you compare to choose from many options, it’s important to be clear about what you want and what you’re willing to give up to get it.”*

— Unknown

Power systems transport an increasing amount of electricity, and in the future, involve more distributed renewables and dynamic interactions of the equipment. The system response to disturbances must be secure and predictable to avoid power blackouts. The system response can be simulated in the time domain. However, this dynamic security assessment (DSA) is not computationally tractable in real-time. Particularly promising is to train decision trees (DTs) from machine learning (ML) as interpretable classifiers to predict whether the system-wide responses to disturbances are secure. In most research, selecting the best DT model focuses on predictive accuracy. However, it is insufficient to focus solely on predictive accuracy. Missed and false alarms have drastically different costs, and as security assessment is a critical task, interpretability is crucial for operators. This work considers the multiple objectives of interpretability, varying costs, and accuracies for DT model selection. This work proposes a rigorous workflow to select the best classifier. In addition, this work presents two graphical approaches for visual inspection to illustrate the selection sensitivity to probability and impacts of disturbances. This work proposes cost

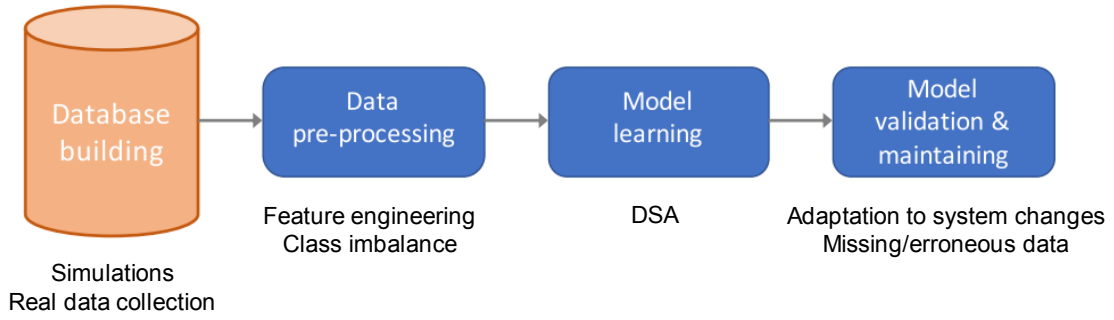


Figure 5.1: Overall workflow for ML-based security assessment [1].

curves to inspect the selection, combining all three objectives for the first time. Case studies on the IEEE 68 bus system and the French system show that the proposed approach allows for better DT-selections, with an 80 % increase in interpretability (in terms of reducing the number of nodes in tree-based algorithms), 5 % reduction in expected operating cost in tree-based algorithms, while making almost zero accuracy compromises. The proposed approach scales well with larger systems and can be used for models beyond DTs. Hence, this work provides insights into criteria for model selection in a promising application for methods from artificial intelligence (AI).

### 5.0.1 Selection of the machine learning model

Recall that the ML approach (Fig. 5.1) to security assessment is to predict the outcome of the stability analysis [22]. This prediction can replace the analysis itself, and the key benefit is that the prediction is instantaneously available. As this benefit is promising, many variations of the ML approaches were proposed [1]. The most common approach is to use a binary classifier as a model, which subsequently predicts whether an operating condition is stable or unstable (in situations where the entire DSA outcome is used, then secure/insecure).

The selection and evaluation of the DT model involve finding the model with the highest performance out of a large set of trained models. Typically, the performance is measured by testing how the model performs on data that is not part of the training. For instance, a

testing set is used to compute the testing error (the ratio of inaccurate predictions). Other performance measures like the  $F_1$  score [66] allow for a harmonic balance of the precision and recall for different errors [106] as used in [29], or the G-mean score that computes the geometric mean as used in [28]. Also, graphical approaches can be used to select models, such as the precision-recall (PR) curve [107] or receiver operating characteristic (ROC) curve [108] as applied to DSA in [31, 109]. However, selecting a model based on a single criterion may be sub-optimal. In DSA, the following three need to be considered:

Firstly, predicting errors for the different classes can have various impacts. A missed alarm is much more severe than a false alarm. Missing an alarm can result in power blackouts and load shedding that have high expected costs, however, a false alarm may require only preventive and corrective control measures (e.g., generation re-dispatch) to be taken that are significantly cheaper. Considering the different impacts of errors is important when training the model specifically if the training database is imbalanced in the classes [1]. This difference in the impact of errors renders several performance criteria unsuitable, such as the test error. The ROC curve or  $F_1$  score may be more suitable. However, the expected outage cost to the end-customer should be considered as the performance metric in security assessments [154] and none of these scores allows for directly measuring the performance in terms of expected costs. Computing the outage cost is difficult as it depends on the disturbance and the load condition [155, 154]. However, estimates of the costs are considered in cost-sensitive learning by adjusting the decision threshold when predicting with the model [156, 157]. Cost-sensitive learning was applied to DSA in combination with ensemble DTs [158], with deep learning [159] and with SMOTE as the imbalance challenge addressed is similar [28].

Secondly, it is crucial to consider the interpretability (and complexity) of the selected model. Models that are high in their complexity are traditionally not interpretable for operators [1] and this renders them unsuitable for the application to DSA. Albeit, recent advances including in SHAP values have shown interpretability methods for complex ML models are promising [110]. Operators responsible for the critical task of security assessments may prefer interpretable DT models in their decision-support tools such that

manual inspection remains possible and errors can be identified [38, 42]. Therefore, the interpretability of the models needs consideration when selecting models for DSA.

Thirdly, it is also crucial to consider frequent changes in the system. Frequent changes in system parameters may require changes in the selection of models. For instance, the weather changes frequently, and with that, the likelihood of contingencies [111, 112]. If the probability is high, an operator may select more conservative DT models than at times with low likelihood. In practice, 1000s of models may be used in real-time [42] and a fast, adaptable selection process is needed.

### 5.0.2 Contributions

The contribution of this work is to propose a rigorous workflow that considers all three aforementioned specific needs (accuracy, interpretability, and cost sensitivity) to select the best classifier for the application of security assessment, and in addition two approaches for graphical inspection to demonstrate the selection sensitivity to key parameters such as probability and impact of disturbances. The cost-curves [69], which are based on ideas from cost sensitive-learning [156] are introduced. This workflow allows for the first time, selecting the best machine-learned DT model for DSA in response to frequent changes in expected costs or likelihood of contingencies, as well as their uncertainty. The proposed workflow is fast, simple, and effective in selecting the best models.

The proposed workflow is studied on the IEEE 68-bus system and further extended to the French transmission system. The challenge of selecting models/classifiers based on predictive accuracy is presented, resulting in sub-optimal selections. Subsequently, the benefits of the proposed multi-objective selection procedure are demonstrated. Finally, the cost-curve approach is showcased.

The remainder of the chapter is structured as follows. Sec. 5.1 presents the three different objectives when selecting a model for DSA. Sec. 5.2 presents the proposed workflow and the two graphical approaches that allow considering all three objectives together. Finally,

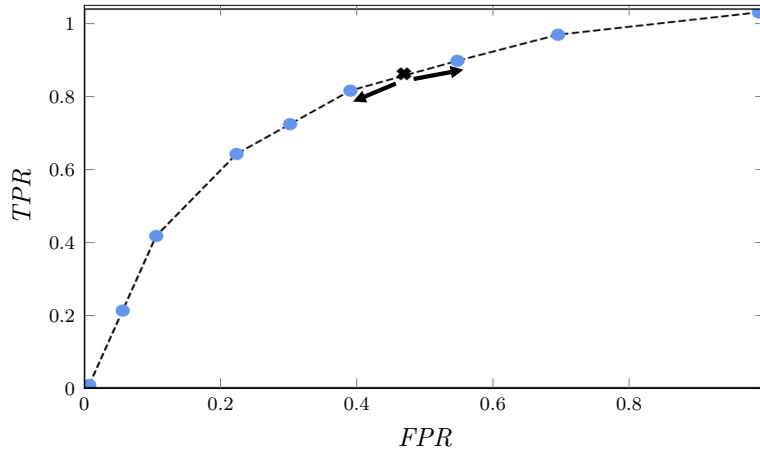


Figure 5.2: The ROC curve for evaluating a classifier. Different combinations of  $TPR$  and  $FPR$  are obtained by varying the decision threshold  $Z$  (shown with arrows). The perfect classifier would be in the top left corner having  $TPR = 1$  and  $FPR = 0$ .

Sec. 5.3 presents the case study and Sec. 5.4 concludes this work.

## 5.1 Objectives in selecting DT classifiers for DSA

In supervised ML, the procedure of learning and selecting a classifier starts with defining the models and hyperparameters to study. After these are defined, the range of hyperparameters is typically explored with a cross-validation search procedure to address under- and overfitting of the data. In the cross-validation search, typically many different combinations of values for the hyperparameters are explored, and for each, a classifier is trained using the gini index or entropy in DTs. The result of this exploration of combinations of hyperparameter values is a set of classifiers (one per combination). Subsequently, one classifier of this set must be selected. For this selection, various metrics can be used. Typically, the validation accuracy (an approximation of the testing accuracy) is chosen. As pointed out earlier, choosing the performance of the testing accuracy may not represent the needs for DSA to balance the different costs of inaccuracies, interpretability, and robustness.

### 5.1.1 Minimising the effects of inaccuracies

One of the objectives of selecting classifiers is to minimise the effect of inaccurate predictions. Two types of inaccurate predictions exist: missed alarms (false positives) and false alarms (false negatives), and these have different effects/costs. The symbols  $FP$  and  $FN$  correspond to the absolute numbers of inaccurately predicted operating conditions, of the types of false positives and false negatives, respectively. The specific effects (or impacts, costs) of these inaccurate predictions are  $C_{FN}$  for false-negative predictions and  $C_{FP}$  for false-positive predictions, where  $C_{FP} \gg C_{FN}$  as missed alarms have higher impacts than false alarms. For missed alarms, these involve the expected outage cost to the end-consumer, and for false alarms, these are the expected costs for unnecessary preventive/corrective control actions. Accurate predictions have zero costs, hence  $C_{TP} = C_{TN} = 0$  for true negatives ( $TN$ ) and true positives ( $TP$ ), respectively. The cost ratio is

$$\gamma = \frac{C_{FP}}{C_{FP} + C_{FN}}. \quad (5.1)$$

Typically, the objective in binary classification is to minimise the test error  $\frac{FP+FN}{FP+FN+TP+TN}$ , however, firstly, the test error can not directly be modelled; hence in the training, an approximation for the test error is typically used (e.g., the training error, entropy). Secondly, by minimising this objective, this imbalance in the cost/impact of the two different types of inaccuracies is not considered. Another way to evaluate and train a classifier in binary classification is to maximise the  $F_1$  score.

The  $F_1$  score equally weights precision and recall, however, it does not consider the different costs of the two classes [66], similarly as the training error and entropy. One approach to consider both inaccuracies is to use the ROC curve. The ROC curve is a graphical approach that shows the true positive rate  $TPR = \frac{TP}{TP+FN}$  and the false positive rate  $FPR = \frac{FP}{TN+FP}$  [108]. Various combinations of  $TPR$  and  $FPR$  are computed by varying the threshold  $Z$  that the classifier uses for prediction. This threshold  $Z$  is used to obtain the predicted class. Initially, the classifier outputs a score  $S \in [0, 1]$ . Subsequently, this score is compared

against the threshold  $Z$  (the default value is  $Z = 0.5$ ) to determine the predicted class. If the score  $S \geq Z$ , then the prediction is the positive class  $Y = 1$ , otherwise negative  $Y = 0$ . Hence, for a testing set, the combinations of  $(TPR, FPR)$  values are computed for varying  $Z$ s. Then, these points build the ROC curve.

The ROC curve is used to evaluate classifiers with cost-sensitivity as shown in Fig. 5.2. Each point corresponds to a single classifier where the decision threshold was varied in  $Z = [0, 1]$ . It is possible to estimate the optimal  $\tilde{Z}^*$  from an effect/cost minimising viewpoint. The objective of minimising costs/impacts of inaccurate predictions is

$$\Sigma = FN * \Pi_+ * C_{FN} + FP * \Pi_- * C_{FP}, \quad (5.2)$$

where

$$\Pi_- = \frac{N_-}{N_- + N_+} \quad \text{and} \quad \Pi_+ = \frac{N_+}{N_- + N_+} \quad (5.3)$$

are the two class distributions of positive  $N_+$  and negative  $N_-$  points in the testing set [157]. Typically  $N_+$  and  $N_-$  are not exactly known, however,  $\Pi_+$  and  $\Pi_-$  can be assumed to be similar to the distributions in the training set or inferred from a validation process/set. Note that in DSA, often  $\Pi_+ \gg \Pi_-$  is an additional (class) imbalance to  $C_{FP} \gg C_{FN}$ . For the class imbalance problem, the methods developed in this thesis, (Chapters 3 & 4) could improve the class distribution. For the cost imbalance, although  $C_{FN}$  and  $C_{FP}$  are unknown, estimates could be assumed  $\tilde{C}_{FN}$  and  $\tilde{C}_{FP}$  and the expected costs are

$$\tilde{\Sigma} = FN * \Pi_+ * \tilde{C}_{FN} + FP * \Pi_- * \tilde{C}_{FP}. \quad (5.4)$$

Then, the estimated cost ratio  $\tilde{\gamma}$  is computed by using Eq. (5.1) and the estimates  $\tilde{C}_{FN}$  and  $\tilde{C}_{FP}$ . Subsequently, the expected costs from Eq. (5.4)

$$N\tilde{\Sigma} = \frac{(1 - TP) * \Pi_+ * (\frac{1-\tilde{\gamma}}{\tilde{\gamma}}) + FP * (1 - \Pi_+)}{\Pi_+ * (\frac{1-\tilde{\gamma}}{\tilde{\gamma}}) + (1 - \Pi_+)}. \quad (5.5)$$

This (normalised) expected cost is minimised when selecting the decision threshold at the

estimated cost-optimal point as

$$\tilde{Z}^* = \frac{\Pi_- * \tilde{C}_{FP}}{\Pi_- * \tilde{C}_{FP} + \Pi_+ * \tilde{C}_{FN}}. \quad (5.6)$$

### 5.1.2 Maximising the interpretability of DTs

The second objective of a classifier for DSA is to maximise interpretability. If the learning approach is interpretable, then human experts, here the system operator, can build trust in using these approaches. The classifier is interpretable when the learned model can be understood and offers insights into the process of how a prediction is being made. This requires models and data that are non-complex. For instance, the model complexity can be described by the type of parametrisation, the number of hyperparameters, and the number of features. In DTs, the type of parametrisation is a linear splitting scheme. As this linear splitting scheme is not complex, DTs are known for their interpretability. In DTs, the number of hyperparameters can be measured as the number of nodes/splits as each node involves a single hyperparameter. Hence, the number of nodes is a measure for the model complexity (and inversely interpretability). For trading-off the model complexity with accuracy, regularisation is typically used in training for the purpose of avoiding overfitting. Hence, in a similar way accuracy and interpretability can be traded-off as in [38].

### 5.1.3 Maximising the robustness of classifiers

The third objective is to maximise the robustness of the classification decisions under uncertainties in the input parameters, as costs/impacts and probabilities of contingencies. The cost/impact of contingencies on the system  $\tilde{C}_{FN}$ ,  $\tilde{C}_{FP}$  and  $\tilde{\gamma}$  may be wrongly estimated and these uncertainties result in sub-optimal prediction decisions. Some classifiers may be more prone to these uncertainties than others, hence, it is proposed to consider this uncertainty to select classifiers that are less prone.

The objective is to minimise Eq. (5.2). However, actually, Eq. (5.4) is minimised as the cost ratio  $\gamma$  is unknown and the estimate  $\tilde{\gamma}$  is used. The impact of the uncertainty in these estimates can be studied in a sensitivity analysis involving comparing the expected normalised costs  $N\tilde{\Sigma}$  from Eq. (5.5) with the normalised actual costs

$$N\Sigma = \frac{(1 - TP) * \Pi_+ * (\frac{1-\gamma}{\gamma}) + FP * (1 - \Pi_+)}{\Pi_+ * (\frac{1-\gamma}{\gamma}) + (1 - \Pi_+)} \quad (5.7)$$

that is similarly derived as Eq. (5.5). These two costs  $N\tilde{\Sigma}$  and  $N\Sigma$  are compared for various errors  $\Delta\gamma$  in the cost-ratios, where  $\tilde{\gamma} = \gamma \pm \Delta\gamma$ . This sensitivity analysis involves computing the false negatives and false positives  $(FN, FP) = f(Z)$  for a given test set, where  $f$  is the prediction function of the classifier and the decision threshold  $Z$  is varied. Then, the normalised actual costs  $N\Sigma$  are computed from Eq. (5.7) using  $\gamma$  and the normalised expected costs  $N\tilde{\Sigma}$  from Eq. (5.5). Subsequently, the differences in these costs are compared, and the various estimation errors  $\Delta\gamma$  are studied to understand the sensitivity of this cost difference. These studies can be repeated for multiple classifiers and various decision thresholds to find the best combination of classifiers and threshold  $Z$  most insensitive (robust) against uncertainties in the costs.

## 5.2 Multi-criteria selection of classifiers

This section presents the proposed rigorous workflow to select the best DT classifier, and subsequently, introduces two approaches for graphical inspection to illustrate and show the selection sensitivity to estimated parameters. The first graphical approach modifies the ROC curve and allows considering the first two objectives: to minimise inaccurate predictions and maximise the interpretability as shown in Fig. 5.10. The second approach modifies the cost-curves, as shown in Fig. 5.5. The proposed modification to the second approach allows trading-off on all three objectives.

The proposed workflow has two parts: (i) offline training and (ii) online selection. The first part is the proposed offline workflow to train and prepare classifiers for application in the

online selection workflow. The offline workflow of using the proposed cost-curve approach is illustrated in Fig. 5.3. Initially, many classifiers are trained by varying some (hyper-)parameters, such as the DT depth, and each classifier training follows cross-validation resulting in  $\mathcal{N}$  candidate DTs,  $\Omega = \{\mathcal{C}_i^{(p)}, \forall i = 1, 2, \dots, \mathcal{N}\}$ , and  $p \in \mathfrak{R}$  is the number of DT nodes. Subsequently, the task is to select the best DT classifier from this set  $\Omega$  according to the introduced criteria of Sec. 5.1.

The second part is the proposed online selection workflow to consider realtime information when selecting a classifier. The probabilities of contingencies  $\Pi_+, \Pi_-$ , the expected cost of contingencies  $\tilde{C}_{FP}, \tilde{C}_{FN}$ , and the probability cost function's (PCF) range of interest  $[PCF^L, PCF^U] \in [0, 1]$  are updated with real time data. The range  $[PCF^L, PCF^U]$  represents various expected combinations of contingency probabilities and expected contingency costs. Then, for each DT  $\mathcal{C}_i^{(p)}$ , the normalised expected costs  $N\tilde{\Sigma}$  are computed with Eq. (5.5) within the range of  $[PCF^L, PCF^U]$  at  $K$  user-defined, equidistant steps. The average normalised expected cost is then computed as  $\overline{N\tilde{\Sigma}} = \frac{\sum_{i=1}^K N\tilde{\Sigma}}{K}$ , where  $\overline{N\tilde{\Sigma}}$  measures on average the expected impact (e.g. cost of loss of load) of wrongly classifying the security status of an operating condition. Subsequently, the average normalised actual cost  $\overline{N\Sigma}$  ( $N\Sigma$  computed with Eq. (5.7)) is calculated assuming  $\tilde{\gamma} = \gamma \pm \Delta\gamma$  within the same  $[PCF^L, PCF^U]$  range and  $K$  equidistant steps, where  $\overline{N\Sigma}$  measures on average the actual impact of wrongly classifying the security status of an operating condition. This step is done to compare the sensitivity of a classifier  $\mathcal{C}_i^{(p)}$  to estimation errors  $\Delta\gamma$  of the cost ratio  $\tilde{\gamma}$  that is assumed. The number of steps  $K$  is selected by the user, and the more steps are selected, the better the resolution of both actual  $\overline{N\Sigma}$  and expected  $\overline{N\tilde{\Sigma}}$  costs. Finally, the proposed, optimal DT  $\mathcal{C}_i^{(p)}$  is the one with the minimum average relative costs across the range of  $[PCF^L, PCF^U]$  as  $\mathcal{G}(\Sigma) = \frac{\overline{N\tilde{\Sigma}} - \overline{N\Sigma}}{\overline{N\Sigma}}$ , and DT nodes  $p \leq a$ , where  $a$  is a user-defined criterion that constraints the number of DT nodes, and keeping  $a$  small aims for high interpretability. This is the optimal DT as the contingency probabilities and contingency expected costs are uncertain.

The final cost-curve provides a graphical illustration of the sensitivity of selecting the best classifier to minimise costs of missed and false alarms considering inaccurate estimations

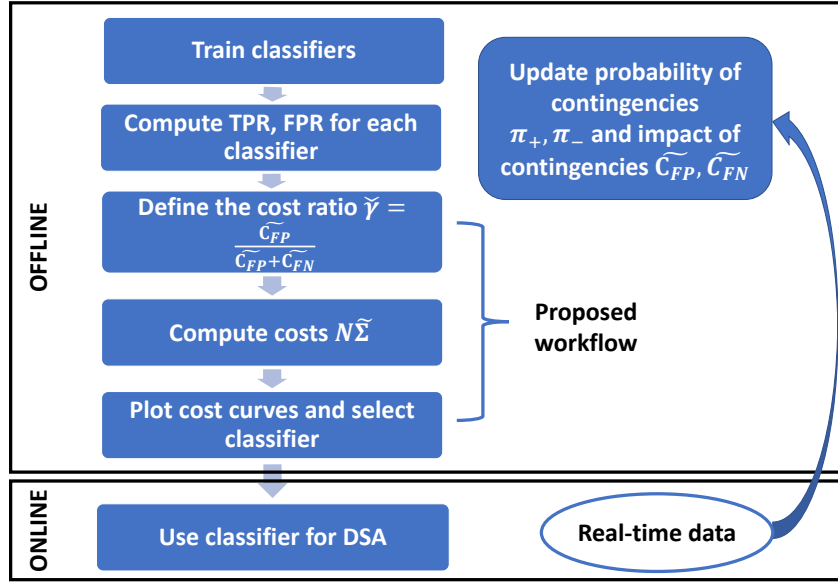


Figure 5.3: Proposed cost-curve workflow to select the best DT classifier offline. The selection can be updated with real-time data.

of the costs  $\tilde{C}_{FP}, \tilde{C}_{FN}$  and probabilities of contingencies  $\Pi_+, \Pi_-$ . The DT  $\mathcal{C}_i^{(p)}$  whose cost-curve has the least variation in the range  $[PCF^L, PCF^U]$  is the most suitable DT and would be the ideal DT from the perspective of cost uncertainty. The two proposed graphical approaches are introduced as follows.

### 5.2.1 Visual inspection using the ROC curve

A color bar is used for the interpretability on the ROC curve, and the interpretability of a DT model (here, in terms of the number of nodes) correlates inversely with the DT model complexity. The ROC curve allows studying the performance of a binary classifier when adjusting the decision threshold  $Z$  as illustrated in Fig. 5.2. The advantage of the ROC curve is the visual ability to compare classifiers across ranges of various  $Z$ s instead of a single point comparison that does not allow for variability, such as in selecting based on computing the cost-optimal  $\tilde{Z}^*$  using Eq. (5.6). Classifier comparison using the ROC curve starts by drawing each classifier's ROC curve. Subsequently, the cost-optimal decision thresholds  $\tilde{Z}^*$  are marked. An example is presented in Fig. 5.2.

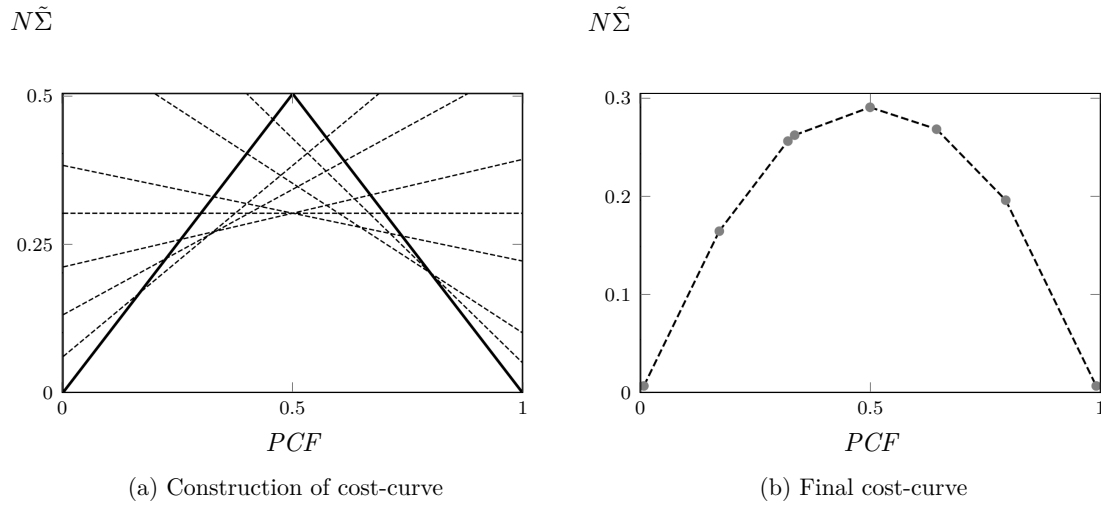


Figure 5.4: Construction of cost-curves: In (a), each dashed linear curve corresponds to one blue point in the ROC curve Fig. 5.2. The combined minimal expected cost  $N\tilde{\Sigma}$  of (a) is where the selected  $Z$  equalled the cost-optimal decision threshold  $\tilde{Z}^* = PCF$  and is shown in (b).

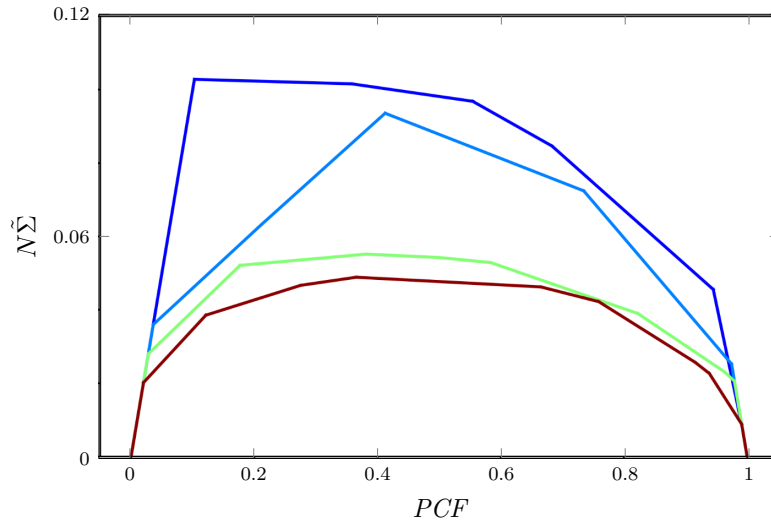


Figure 5.5: Proposed cost-curve based selection considering multiple criteria. The lime classifier has the best trade-off in terms of interpretability and expected cost in the relevant interval  $0.8 \leq PCF \leq 1$ .

In this example, the DT classifier corresponding to the blue curve is the best in terms of interpretability (fewer nodes), and the brown is the best in terms of accuracy (closest to the top left corner). However, when considering these two objectives in the ROC curve together, the best trade-off is the classifier corresponding to the cyan curve. Although the interpretability is slightly worse than the blue classifier, the  $TPR$  is significantly better (almost 0.05 higher). Also, the classifiers represented by the brown and green curves only offer marginal improvements in terms of  $TPR$ . However, they are worse in terms of interpretability. Therefore, the cyan curve is the best classifier in this example. This suggested selection procedure can be quickly and visually performed by an operator to trade-off the cost-optimality, accuracy, and interpretability using a single graphical approach.

### 5.2.2 Proposed cost-curve approach for graphical inspection

The proposed approach modifies cost-curves to include information on the sensitivity of inaccurate estimations of the costs/impacts. The cost-curve shows the normalised expected cost  $N\tilde{\Sigma}$  from Eq. (5.5) with varying probability cost function  $PCF$ . This is the main difference to the ROC curve where the expected costs of inaccurate predictions are not directly presented such as in the cost-curve. The proposed modification of the cost-curve allows selecting the best classifier by considering all three aforementioned objectives.

The construction of the cost-curve starts with the ROC curve of the classifier. The ROC curve is constructed from a set of  $(TPR, FPR)$  points corresponding to applying different decision thresholds  $Z$  to the score-output  $S$  of the classifier as illustrated in Fig. 5.2. Subsequently, the cost-curve aims to investigate the normalised expected costs for each of these points dependent on changes in  $\tilde{C}_{FN}$ ,  $\tilde{C}_{FP}$ ,  $\Pi_-$  and  $\Pi_+$ . Changes in these are considered by using a single parameter, the probability cost function, defined as:

$$PCF = \frac{\Pi_- * \tilde{C}_{FP}}{\Pi_- * \tilde{C}_{FP} + \Pi_+ * \tilde{C}_{FN}}. \quad (5.8)$$

Subsequently, the normalised expected cost from Eq. (5.5) are

$$N\tilde{\Sigma} = (FP - FN) * PCF + FN, \quad (5.9)$$

where  $FP$  and  $FN$  can be directly computed from the ROC values  $TPR$  and  $FPR$ . Then, Eq. (5.9) is the linear equation connecting the points  $(PCF, N\tilde{\Sigma}) = (0, FN)$  and  $(PCF, N\tilde{\Sigma}) = (1, FP)$ . The constructions of the lines corresponding to the blue points in Fig. 5.2 are presented in Fig. 5.4a. The lower envelope of the cost-curve is the minimum costs that can be obtained. This lower envelope represents selecting the cost-optimal  $\tilde{Z}^*$  where the  $Z$  applied to the classifier equalled  $Z = \tilde{Z}^* = PCF$ . This final minimal cost-curve is presented in Fig. 5.4b. The cost-curve shows the classifier performance across all cost-class distributions whereas the ROC curve allows presenting only a single cost-optimal point. Subsequently, the proposed approach is to use cost-curves for selecting the best classifier along with the relevant range of cost-class distributions as shown in Fig. 5.5 to ensure the best selection according to the discussed multiple objectives.

The sensitivity of errors in estimating the cost ratio  $\tilde{\gamma}$ , on the selection based on cost-curves can be studied by computing the corresponding actual cost-curve for  $N\Sigma$  based on the actual cost  $\gamma$  assuming an error  $\Delta\gamma$  as pointed out in Sec. 5.1.3. Firstly, recall that the estimated cost ratio  $\tilde{\gamma}$  influences the choice of  $\tilde{Z}^*$  and that the cost-curve is defined by  $Z = \tilde{Z}^*$ . If the choice of estimated cost ratio  $\tilde{\gamma}$  differs from the actual cost ratio  $\gamma$ , then the normalised actual cost can be computed based on the actual value of  $\gamma$  and can be presented similarly as the estimated cost-curve. The PCF values do not differ for the normalised expected and actual costs as the threshold  $\tilde{Z}^*$  is based on the expected cost ratio of  $\tilde{\gamma}$ . This modification of the cost-curve allows for considering the cost/impact of estimation errors in  $\tilde{\gamma}$  and represents another advantage over the ROC curve. Subsequently, the proposed cost-curve approach can consider interpretability using color schemes in the same way as in the ROC curve approach. Consequently, this approach considers all 3 objectives as criteria for visual inspection.

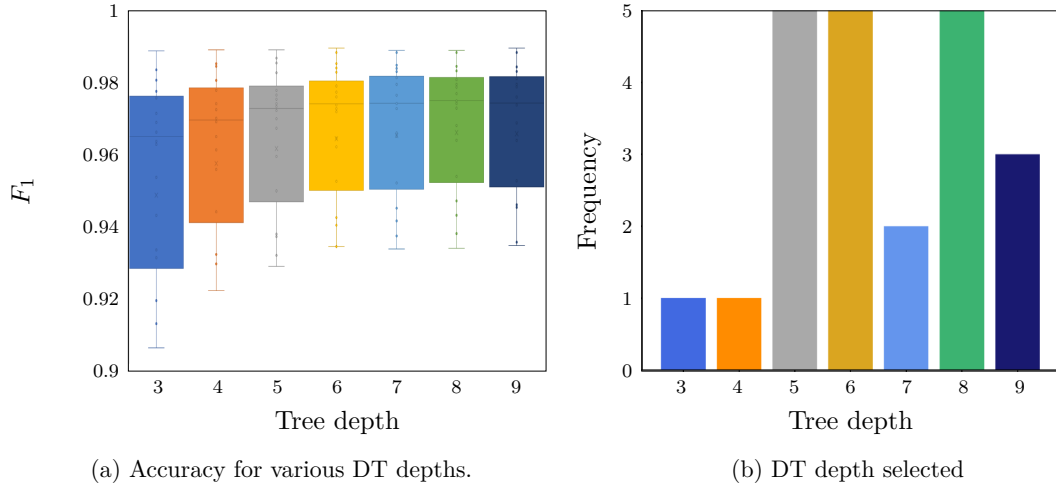


Figure 5.6: The predicting performance of classifiers with 7 different DT depths. The box plots in (a) show min, max, interquartiles, and median for  $22 \times 7 = 154$  trees. Using (a), the 22 classifiers with the highest  $F_1$  score are selected and the selection-frequencies are in (b).

## 5.3 Case study

In this case study, firstly, the challenge of sub-optimal selections when using a single objective in selecting classifiers is studied. Secondly, the effectiveness of the proposed workflow that considers multiple objectives concurrently to select the best DT classifier is studied, with visualisation using the two proposed graphical approaches. Thirdly, the proposed workflow is studied when considering the uncertainty of the cost estimations. Finally, the computational times of the proposed workflow are studied and the limitations are discussed.

### 5.3.1 Test system and assumption

The case study is mainly carried out on the IEEE 68-bus system and the scalability is demonstrated in a study of the French transmission system.

The first data set for this case study was generated using the network data from the IEEE 68-bus system [131].  $N_D = 12000$  operating conditions were sampled as follows. The active loads were sampled from a multivariate Gaussian distribution (via Monte Carlo sampling), and the correlation was assumed to follow Pearson's correlation with a correlation coefficient of 0.75. The distribution was converted to a marginal Kumaraswamy(1.6,2.8)

distribution using the method of inverse transformation. The AC-model of the network was used to compute the active and reactive powers of generators to ensure feasible operating conditions. This sampling process of the IEEE 68-bus system results in 12000 samples, where each describes the operating condition of the power system in a steady-state. Then, all phase angles and voltages, reactive and active power flows, and the reactive and active power injections were used to construct the feature vector  $X \in \mathcal{R}^{N_D \times N_F}$ , where  $N_F = 438$  was the number of features. For each of these 12000 operating conditions, the transient stability response to faults was simulated using MATLAB Simulink.  $N_C = 22$  different three-phase line outages were simulated as event-type events on the pre-fault steady-state condition and the faults are cleared after 0.1 s. The simulation time was 10 s on a standard desktop computer. If at any point in time, the difference of any phase angles of the generators was larger than  $180^\circ$ , then the operating condition for that particular contingency was considered unstable, and the corresponding element of the label matrix  $Y \in \{0, 1\}^{N_D \times N_C}$  was set at 0 for unstable, otherwise stable 1. In total  $N_D \times N_C = 12000 \times 22 = 264000$  simulations were performed as each of the 22 contingencies need to be independently simulated.

The second data set of the French transmission system was used. The French transmission system had 1955 transmission lines, 798 transformers, 1886 buses, 411 generators and 127 shunt elements. This data-set consisted of  $N_D = 16722$  operating conditions in a feature vector  $X \in \mathcal{R}^{N_D \times N_F}$ , where  $N_F = 35873$  is the number of features. To generate a single data point required 56 s time on a computer cluster [42] and 1980 different contingencies were analysed where each required a single time-domain simulation. Subsequently, the time-domain simulations were assessed and 9 reliability metrics were computed, including overload, loss of synchronisation, over/under-voltages, small-signal stability, transient stability, et cetera. More details can be found in [42, 96]. This second data set was used to demonstrate the scaling of the proposed approach.

The subsequent processes of the training workflows (feature selection, DT training, et cetera.) were carried out on a Dell XPS 13 9360 running an Intel(R) Core(TM) i5-8250U processor with 8 GB installed RAM. DTs were trained with the CART algorithm [135]

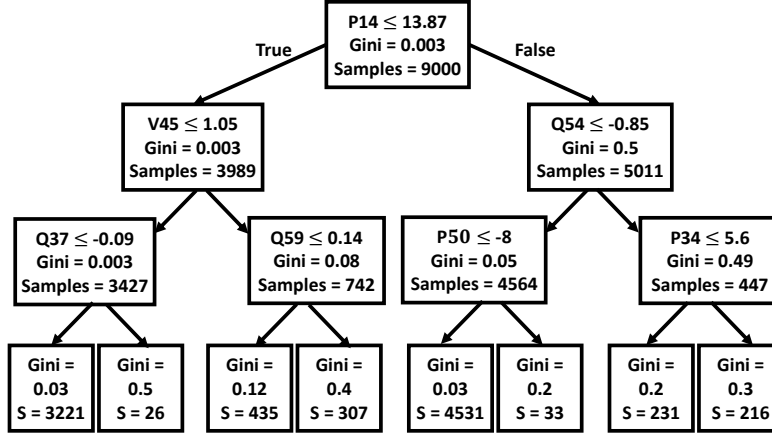


Figure 5.7: Tree structure of a DT with DT depth of 3. P and Q are the active and reactive power-flow between buses, and V is the voltage at the bus.

from the package *scikit-learn* 0.18.1 [134] in Python 3.5.2. The default training settings were selected except using gini impurity instead of entropy to measure the quality of the splits. The data-set was split into training/testing sets in ratio of 75 % / 25 %. The feature variable  $X$  was used and the labels  $Y$  were used as the input for the training of the classifier, however, for each contingency, a single DT was trained (in total 22 DTs). 5-fold cross-validation was applied to address under-/overfitting. Subsequently, the Platt method was used to calibrate the score-output  $S$  of the classifier [136].

### 5.3.2 Selections based on single criteria

The first study illustrates the effect on the model interpretability when predictive accuracy is used for selecting the model. Typically, the  $F_1$  score or the test accuracy is used for selecting the model. In this study, firstly, DTs were trained for the depths  $\{1 - 20\}$  for each of the 22 contingencies. Subsequently, 5-fold cross-validation was used to select the best DT depth based on the highest  $F_1$  score. Fig. 5.6a presents the  $F_1$  accuracy values for all different tree depths involved in this study showing that an increase in tree depth on average results in higher  $F_1$  scores, however, over-fitting occurred for larger depths than 9 and no tree was selected with a depth larger than 9. Fig. 5.6b shows the exact breakdown of the final selected classifiers and most of the selected 22 DT depths were

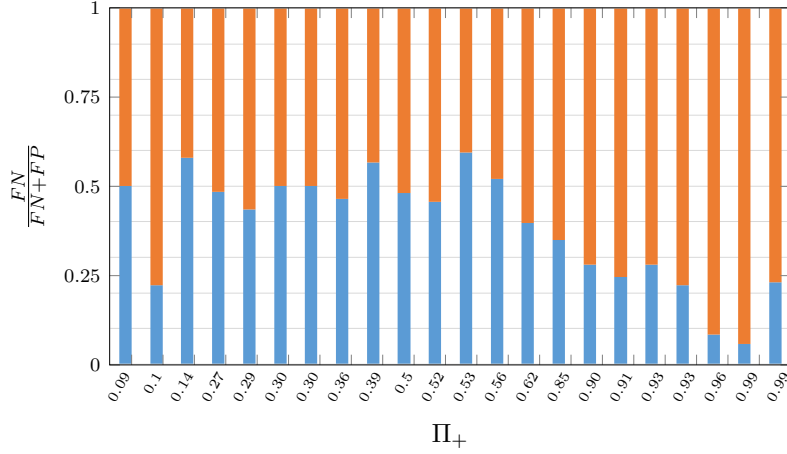


Figure 5.8: Ratio between missed (orange) and false (blue) alarms for the 22 DTs without considering the impact of different costs.  $\Pi_+$  is the class ratio of positive conditions.

around larger depths of [5, 9]. However, the DT structures of larger trees are not easily readable (interpretable). For example, a DT with DT depth = 3 (Fig. 5.7) has 15 nodes. Conversely, a DT with DT depth = 9 has 100 nodes. When focusing only on the predicting accuracy, the user may select the DT with depth = 9 as the  $F_1$  score of 0.975 outperforms the tree with depth = 3 having an  $F_1$  score of 0.96. However, when focusing only on the interpretability, the operator may select the tree with DT depth = 3 with a lower  $F_1$  score of 0.96. These two criteria are contrasting and require a suitable trade-off.

The second study investigates the impact when considering neither the difference in expected costs nor the class imbalances. Typically classifiers are trained to minimise the test inaccuracy, and this is insensitive to the differences in costs and the imbalance and is therefore prone to more missed alarms than false alarms. The 22 DTs from the first study were used, and their relationship between missed and false alarms was investigated in Fig. 5.8. In this training procedure, neither the impact of costs nor class imbalances were considered. It shows: when the imbalance is large, for instance when many more positive than negative operating conditions are in the database (toward high  $\Pi_+$ ), the share of missed alarms increases significantly. This is not in favor of operators, as typically the expected cost for missed alarms is large than that of false alarms  $C_{FP} \gg C_{FN}$ , hence the operators aim to avoid missed alarms. Here, it is assumed that  $C_{FP} : C_{FN} = 2 : 1$ . However,

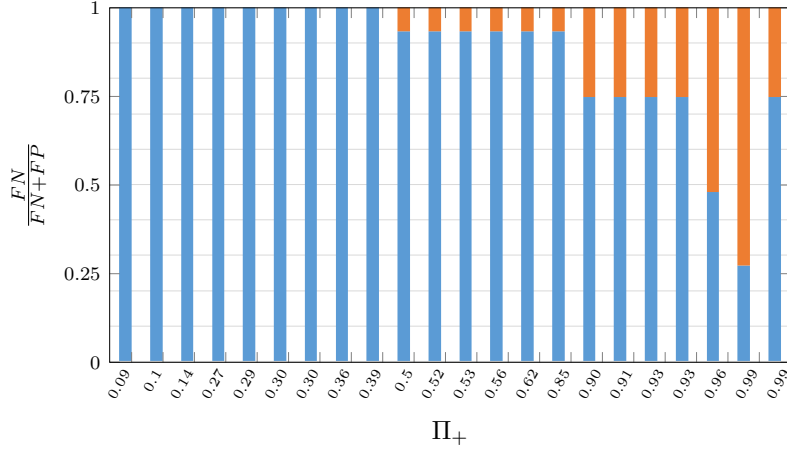


Figure 5.9: Ratio between missed(■) and false(■) alarms for the 22 DTs by considering the impact of different costs.

when the 22 DTs and a shifted decision threshold  $\tilde{Z}^*$  from Eq. (5.6) was used as described in Sec. 5.1.1, then the ratio of missed alarms decreased significantly as demonstrated in Fig. 5.9. Ideally, when considering the costs and imbalances, the ratio of missed and false alarms would have been constant for all different contingencies with different imbalances. However, Fig. 5.9 shows an increase of missed alarms toward high class imbalances of high  $\Pi_+$ . The reason is that these trained classifiers for high  $\Pi_+$  have more knowledge available on positive than on negative operating conditions and are therefore more accurate on positive conditions. Cost-sensitive learning aims to address that imbalance, however, can never fully address it. This highlights another trade-off that needs to be made between minimising test inaccuracy and the impact of different costs. These two studies showed that considering a single criterion is insufficient when selecting a classifier.

### 5.3.3 Multi-criteria selection with modified ROC approach

In this study, the proposed selection workflow is investigated using the modified ROC approach. This workflow allows for a visual inspection of the accuracy performance and interpretability at the same time when selecting the classifier. Firstly, a variety of classifiers were trained with tree depths of  $\{2, 3, 4, 5\}$ . To select the best classifier, the ROC approach was applied as follows: the  $TPR$  and  $FPR$  values for each tree were obtained by varying

the decision threshold  $\tilde{Z}^*$  within  $[0, 1]$ . The  $TPR$  and  $FPR$  values are obtained from the test set and each combination  $TPR$  and  $FPR$  values represent the classifier being used with a different decision threshold. Subsequently, their values were plotted for each of the 4 trees in Fig. 5.10. The color spectrum shows the different levels of interpretability (or DT depths). The blue tree is most interpretable while the brown tree is the least interpretable. Subsequently, the cost-optimal points for each tree were marked with the  $X$  symbol for an assumed cost ratio  $\tilde{C}_{FP} : \tilde{C}_{FN} = 2 : 1$ . These points represent the cost-optimal use of the given tree with the specified cost ratio. The ROC approach can be used to select the best among these 4 cost-optimal DTs. For instance, the lime and cyan curves (and points) are noteworthy. Both classifiers have similar values for  $TPR$  and  $FPR$ . However, the classifier represented in the cyan color offers higher interpretability and is the final classifier. It would be difficult to see the similarity of these two curves by analysing the predicting performances based on data and hence the proposed selection is more robust as the entire curve can be assessed while keeping the focus on the comparison of the cost-optimal points. The proposed approach increased interpretability by 82% (on average 30 DT nodes in the proposed approach and 54 nodes when focusing only on accuracy) while the accuracy slightly decreased (0.960 in the proposed approach and 0.964 otherwise).

### 5.3.4 Multi-criteria selection with cost-curves

In this study, the proposed selection workflow with graphical inspection using the cost-curve approach is in focus. The cost-curve approach allows for visual inspection of the accuracy performance, the interpretability, and quantifies the expected costs of misclassifications. The same DT classifiers with tree depths of  $\{2, 3, 4, 5\}$  were used as in the previous study. The construction of the cost-curves followed the steps described in Sec. 5.2.2. The four resulting cost-curves are presented in Fig. 5.5. Showing these curves allows comparing classifiers across different intervals of costs instead of a single point-wise comparison as in the ROC curve approach. This is useful as the class-cost distribution  $PCF$  can change frequently ( $PCF$  is a function of the likelihood of contingency and the outage costs calculated

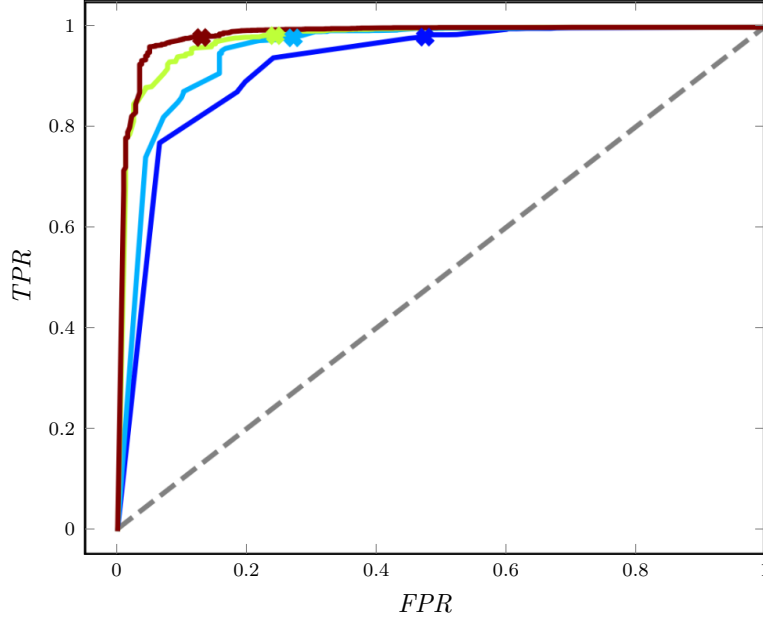



Figure 5.10: Selection based on modified ROC curves for multiple criteria. The DT classifier in cyan has the best trade-off in terms of cost-optimality (x on the curve) and interpretability (the number of DT-nodes from 0 to 50, where 0 is on the left side of the color scheme ).

using Eq. [5.8](#), where it is assumed that  $\tilde{C}_{FP} : \tilde{C}_{FN} = 2 : 1$ ). In this example, the relevant region is  $0.8 \leq PCF \leq 1$  as this is where high class imbalances are (toward large  $\Pi_+$ ). The brown classifier has the lowest expected costs  $\Sigma$  and the blue classifier has the highest interpretability. The expected cost of the brown and lime classifiers are similar within the relevant region, however, the lime classifier has a steeper curve for  $PCF$  values lower than around 0.9. Also here, an operator may select the lime classifier as it represents a good trade-off between interpretability and normalised expected cost in the relevant interval. This proposed novel cost-curve approach can be used by an operator to study cost-class distribution intervals, and goes beyond the point-wise comparison that would be possible with ROC-curves or other data-based comparisons. These are additional insights to the model-selection.

### 5.3.5 Reduction of loss of load costs under uncertainties

In this study, the proposed selection workflow is investigated when considering uncertainties in estimates of the cost parameter  $\tilde{\gamma}$ . An error of 20% was assumed, where  $\Delta\gamma = \pm 0.2\gamma$

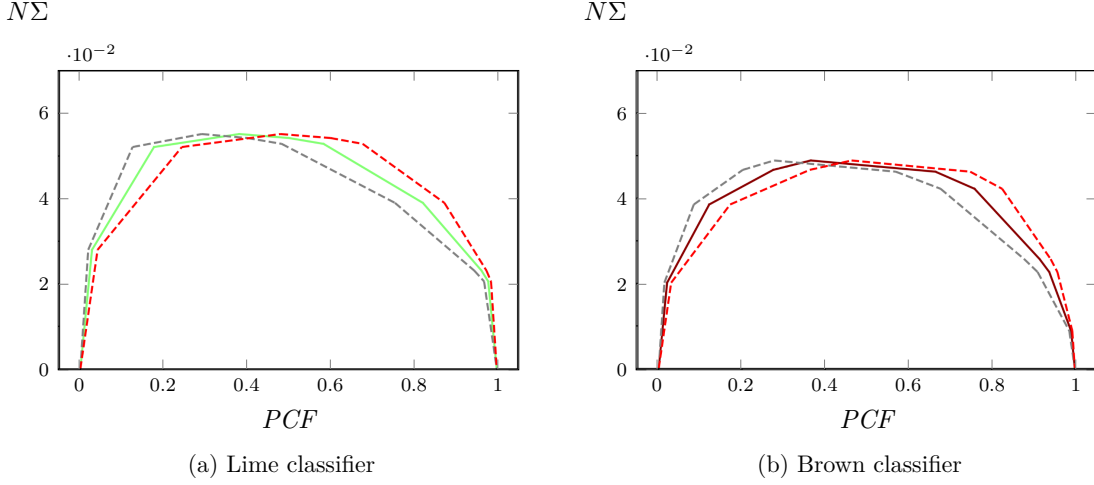


Figure 5.11: Sensitivity study on errors in estimations of parameter  $\tilde{\gamma}$ . The actual costs  $N\Sigma$  under the error of +20 % (---) and -20 % (---) are presented against the estimated costs  $N\tilde{\Sigma}$  that is shown as the baseline. In (a) and (b) are the lime and cyan DTs from Fig. 5.5.

for the two classifiers from the previous study (lime and brown of Fig. 5.5). The actual normalised cost for loss of load  $N\Sigma$  using  $\gamma$  for the various  $PCF$  values based on  $\tilde{\gamma}$  were computed for three cases as described in Sec. 5.2.2, where the baseline is  $\Delta\gamma = 0$  representing no error and the two error cases  $\Delta\gamma = \pm 0.2\gamma$ . Subsequently, the three cases are presented in Fig. 5.11. In the baseline case, the normalised expected cost  $N\tilde{\Sigma}$  for loss of loads equals the normalised actual costs  $N\Sigma$  of loss of loads. In cases with errors, the actual costs  $N\Sigma$  deviate from the expected costs  $N\tilde{\Sigma}$  in both classifiers. In the relevant region,  $0.8 \leq PCF \leq 1$ , the brown classifier shows a higher variability in the impact of parameter estimations. This is also demonstrated by analysing the average costs  $\overline{N\Sigma}$ ,  $\overline{N\tilde{\Sigma}}$ , and standard deviations  $\sigma_{N\Sigma}$ ,  $\sigma_{N\tilde{\Sigma}}$ . Table 5.1 shows the brown classifier has a higher standard deviation than the lime classifier. In addition, the relative change in operating costs for loss of load  $\mathcal{G}(\Sigma)$  is lower in the lime classifier 13 % versus 18 % for the brown classifier (in the case  $\Delta\gamma = +0.2\gamma$ ). Thus the lime classifier is more robust against uncertainties, estimation errors and is, therefore, the final selected classifier under this viewpoint.

$\gamma$	$-0.2\gamma$	$0$	$+0.2\gamma$
<b>Lime classifier (depth 4)</b>	2.7(.7)	3.1(.7)	3.5(.7)
<b>Brown classifier (depth 5)</b>	2.4(.7)	2.8(.8)	3.3(.9)

Table 5.1: Actual normalised costs  $N\Sigma$  (to the basis  $10^{-2}$ ) considering errors of 20 % in  $\tilde{\gamma}$ . The presented values are averages and standard deviations in the interval  $0.8 \leq PCF \leq 1$  of Fig. 5.11.

### 5.3.6 Computation time and scalability

The scalability of the proposed selection workflow is analysed both for larger systems and for comparing many DT classifiers.

Table 5.2 shows the computational times for the offline workflows and testing the classifiers online on both the IEEE 68-bus and French systems. The table shows the average time for ten contingencies selected at random that have a balanced class distribution ( $\Pi_- \geq 35\%$ ).

#### Scalability of proposed selection workflow to larger systems

The proposed selection workflow scales well with the size of the system. Table 5.2 shows the average time for plotting cost-curves is similar for both the IEEE 68 bus and French systems, 0.06 s and 0.07 s respectively. The data generation and feature selection are, however, dependent on the size of the network.

#### Comparing many classifiers with cost curves

Table 5.2 shows the data generation is the most computationally intensive step in the ML workflow with more than 99.9 % of the time. The DT training time, on average 0.4 s for the IEEE 68 bus system, and 2.4 s for the French system, is negligible in comparison. Thus, many DTs can be trained and afterward compared. However, as the proposed selection workflow allows for the comparison of a small number of classifiers at a time (in the order of 5), a pre-selection step can be added. For instance, a metric-based selection approach can be used (example F1-score), to reduce the number of candidate classifiers.

State	Process	IEEE 68 bus	French System (1886 bus)
Offline	Data generation	40 min	260 h
	Feature selection	$(4.0 \pm 0.8) \text{ s}$	$(8.7 \pm 3.8) \text{ min}$
	DT training	$(0.40 \pm 0.01) \text{ s}$	$(2.4 \pm 3.5) \text{ s}$
	DT testing	$(0.01 \pm 0.01) \text{ s}$	$(0.03 \pm 0.02) \text{ s}$
	Cost curve plotting	$(0.06 \pm 0.01) \text{ s}$	$(0.07 \pm 0.03) \text{ s}$
	Selection by operator	5 s	5 s
Online	Prediction	$< 0.01 \text{ s}$	$< 0.01 \text{ s}$

Table 5.2: Computation time for the offline and online workflows showing the proposed approach to compute cost curves for classifier selection scales well to large systems.

### 5.3.7 Discussion

The proposed selection workflow selects security rules with 5 % lower relative operational costs under uncertainties in the potential impacts of contingencies (e.g., loss of load). Additionally, the proposed workflow increases model interpretability by up to 80 %. This proposed workflow is a pivotal step toward manual inspection of security rules and supports operators building up the trust for using these rules in the critical task of DSA. The proposed workflow is fast and adaptive. It takes less than 1 s to plot the cost-curves and around 5 s to select the best security rule. In addition, the workflow allows studying sensitivities on the input parameters when the basis for choosing decisions changes, and in response, adjusts quickly to these changes. Such an adaptive approach for model selection is needed as the future power system is ever-shifting. For instance, the probabilities of contingencies and the impact of faults can change within hours. In binary classification, it is important to consider the costs associated with misclassifications. However, defining the cost of an incorrect prediction can be challenging if the model does not predict certain outcomes, such as loss of load. In such cases, it becomes necessary to contextualize the costs by incorporating relevant metrics that reflect the potential consequences of misclassifications. Metrics such as the value of loss load or customer average interruption duration index can be utilised to concretise the estimated and actual costs of misclassifications. By incorporating these metrics, a more comprehensive understanding of the implications and impact of incorrect predictions can be provided. This approach improves the usability of the metrics and models by providing a clearer picture of the costs associated with misclas-

sifications, even in scenarios where the model does not directly predict specific outcomes.

A first limitation of the proposed workflow is to conceptualise an intuition for the parameter sensitivities as they do not directly have a physical meaning. The sensitivity serves as a relative comparison between different security rules as a tool to compare them and decide on the best security rule. A second limitation is that the proposed workflow does not support different types of models (e.g., neural networks and DTs), as in this work, interpretability is defined for a single type of model, DTs. There, recent advancements on using SHAP values [110] offer new insights into a generalised definition of interpretability which can be used to compare amongst different types of models beyond DTs.

## 5.4 Conclusion

This proposed work showcased a promising application for methods from the field of AI which is DSA for power systems. This work also provides insights into the importance of metrics and criteria to learn models from AI for DSA and beyond. Those insights are transferable from DTs exercised in this work to other AI models. This work focuses on selecting a DT model for power system security assessment. Typically, a single selection criterion, the predictive accuracy is used, resulting in sub-optimal data-driven security rules. As a result, security rules are often not interpretable and can result in many missed alarms. These missed alarms have very high risks and economic costs for system operations. In response, this work proposes a rigorous selection workflow to consider multiple objectives in the model selection: accuracy, interpretability, and cost-robustness. The workflow increases interpretability by more than 80 % while making minimal compromises in the predictive accuracy. Likewise, the proposed workflow reduces expected relative operating costs by around 5 % with little compromise in the predictive accuracy. Other single-objective-based selection approaches miss such trade-offs, and finding these trade-offs is the key advantage of the proposed workflow. Also, the proposed workflow is fast and adaptive to new situations of system operation. The proposed workflow computes cost-curves within less than 0.1 s, and operators can select the best security rules based on analysing the sensitivity to

new situations. This adaptation is crucial as it increases interpretability through visual inspection and offers a high degree of situational awareness to the operators. In the future, this work shall consider selections across different types of models and include a general definition of interpretability.

## Chapter 6

# Machine Learning in Real-time Operation

*“AI is a tool. The choice about how it gets deployed is ours.”*

— Oren Etzioni

The classical formulation of the transmission switching problem as a mixed-integer problem is intractable for large systems in real-time control settings. Several heuristics have been proposed in the past to speed up the computation time, which only limits the number of switchable lines. In this chapter, this thesis presents a real-time switching heuristic based on Neural Networks that provides almost instantaneous switching actions. The findings are demonstrated on a case study of the IEEE 118-bus test system, and the results show that the proposed heuristic is robust to out of distribution data. Additionally, the proposed heuristic has significant computational savings while all other performance metrics like accuracy are similar to state-of-the-art machine learning methods proposed for transmission switching.

## 6.1 Introduction

Transmission switching (TS) involves the opening or closing of circuits or substations in a transmission network and has been used as a control mechanism by power system operators (SO) to improve voltage profiles and manage congestion on the electric network [73][160]. Although power networks are planned with redundancies to handle multiple contingencies and deal with the uncertainty of future operating conditions, in real-time operation, SOs could use TS to efficiently operate the network infrastructure according to the loading conditions and generator costs. From a reliability perspective, circuits can be switched off to improve cost in normal operations and brought back during contingencies [73][74]. Otherwise, TS happens while considering possible contingencies. Further work by Hedman [75] showed that TS does not inherently deteriorate reliability in the event on contingencies. Importantly the short-term operations savings can be up to 25 % of dispatch costs on the IEEE 118-bus system [73].

Fisher’s flagship paper [73] formulated the TS problem as a mixed-integer problem (MIP) that considers a DC optimal power flow (DCOPF) with binary variables tracking the state of lines as on or off. The formulation adopts the big-M method that constrains the line flows on switched-off lines to zero. As the SO schedules new dispatches every 5 minutes for the power system [161], solving for this optimised network topology and verifying as AC feasible ought to be done within that time frame. However, the computational burden of solving even the DC formulation of the TS problem in real-time prevents the adoption of TS in the control room as it is an NP-hard problem [162][163]. To quantify the magnitude of the search space, the IEEE 118-bus test case, which is small compared to real-world power systems, has 186 lines and thus  $2^{186}$  switching possibilities. Exploring such a search space is intractable with modern-day computing power.

Heuristic approaches based on greedy local search using sensitivity-based algorithms [164][165], constraint identification techniques [76], and power transfer distribution factors (PTDFs) [166], have been proposed to improve the computation speed. The sensitivity-based heuristics involve reducing the large search space of the original MIP problem into a computationally

tractable problem by ranking candidate switching actions. This ranking allows the easy exploration of possible 'relevant' line switches with the computationally inexpensive solution of multiple DCOPF problems. In [164][165] the dual problem of the DCOPF informs how the authors rank the switchable lines based on a line's tendency to improve the baseline dispatch cost. The baseline cost considers a topology with all lines as closed. A greedy local search of such a sensitivity-based ranking involves solving all possible DCOPFs for every possible line switching action [167][161]. High performance computing (as in [168]) and priority listing (as in [164][169]) may be used to reduce the computation burden. The main drawback of these heuristics is their scalability to large networks, as they only provide a limitation on the number of switching options and then greedily search the reduced solution space. Granted, very few lines result in the largest cost reductions [73][170]. An alternative sensitivity-based heuristic involves active constraint identification proposed by Crozier [76] to determine the need for TS in a congestion setting. The approach ranks the sensitivity of constraints on the dispatch cost when there is a mismatch between the DCOPF with economic dispatch (ED) and iteratively eliminates the constraints while solving a series of DCOPF problems. A different approach to the sensitivity-based heuristics avoids solving multiple DCOPFs and instead computes the PTDFs as done in ref [171]. However, these heuristics only work for a limited number of line switches and introduce sub-optimality resulting in higher normal operating costs than necessary.

Machine learning (ML) is promising to outperform heuristics in power system applications or make completely new applications feasible for the first time. For example, ML can simplify power system reliability studies [1] so they can run in real-time while avoiding sub-optimality otherwise introduced by heuristics. There, when selecting a supervised ML model for real-time reliability studies [95], the topological configuration can result in discrete changes in the underlying data distributions that challenge the learned ML-models [172]. Hence, exploring these discrete topological changes is an alternative that is then trained to an ML-model through reinforcement rather than in supervision [173]. Such explorations can enhance the operator's experience and heuristics that would otherwise never use the explored actions.

To use supervised ML within the OTS framework is a new proposal to improve solution times of the OTS problem, as the canonical MIP formulation introduced by Fisher [73] is intractable in practical settings. A major advantage of the ML approach is the reduction in computation time required to select a candidate topology, which makes it a suitable approach to applying in real-time by the SO. The proposed ML workflow for TS is shown in Fig. 6.1. There, the ML model is trained offline on historical operating data, and near-real time, proposes feasible topological configurations for possible operating conditions. After the physical feasibility of the proposed topology is confirmed, the proposed topology is configured in real-time. Previously, the authors in [167] use ML to predict a list of high priority candidate line switches. There, ML models including a decision tree, a  $k$ -nearest neighbour (KNN), and a feed-forward multilayer perceptron artificial neural network (ANN) identify sets of suitable lines for TS. Moving the research further, Johnson [161] proposed a KNN heuristic to explicitly learn  $k$ -best candidate topologies for different load profiles. There, a classifier learns the relationship between operating conditions and their optimal topologies and proposes  $k$  nearest candidate topologies to evaluate the best topology for a new load profile. The results show that the KNN heuristic has negligible computational bottlenecks and can provide significant cost savings on dispatch costs. However, the KNN model faces scalability issues related to  $k$ -neighbours in higher dimensions [174]. Additionally, the robustness of the KNN heuristic to out of distribution data is in question.

Neural Networks have previously been used in power system classification problems. [175] reviewed fault detection methodologies in power systems, and ANNs are suitable non-parametric models for the fault detection classification problem. Due to the non-linear activation functions of ANNs, they can capture non-linear dependencies of complex dynamic systems as power systems [167]. The perceptron-layered network has been a popular ANN architecture for classification problems in power systems. Particularly, the 3-layered radial basis function neural network (RBFNN) architecture is robust to inputs not in the training data [176]. This robustness makes it a favourite for power systems classification problems such as fault detection. RBFNNs [177] have been shown to perform well in terms of accuracy, interpretability, training data and time over other feed-forward architectures

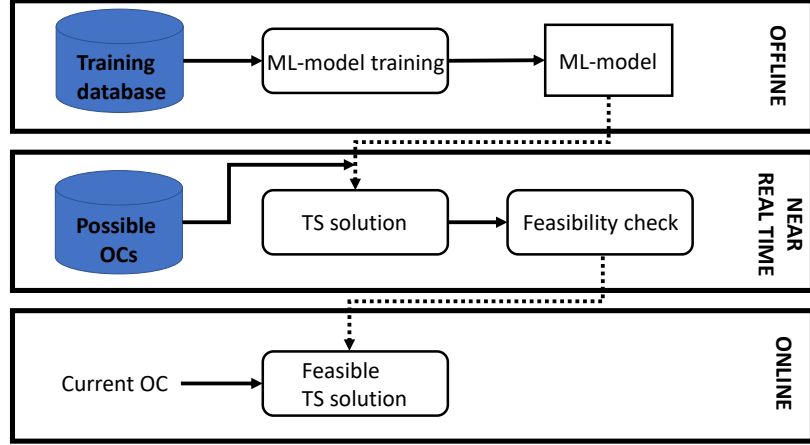


Figure 6.1: ML workflow for real-time topology change

like multilayer perceptron and probabilistic neural networks in other fields as well.

The contributions in this chapter are as follows: (i) we propose an RBFNN heuristic that is robust to noisy data to predict the TS solution in real-time settings considering  $\mathcal{J}$  maximum line switches. The proposed heuristic provides significant computational savings with similar accuracy performance as state-of-the-art ML heuristics for TS (ii) we propose a modified optimisation formulation of the TS problem to speed up the offline generation of training data.

The rest of the chapter is structured as follows: Section 6.2 introduces the proposed RBFNN heuristic and describes the RBFNN architecture. Firstly, the KNN and greedy search heuristics considered for comparison are briefly described. Section 6.3 presents a case study that compares the performance of the proposed RBFNN heuristic with KNN and greedy-search heuristics relative to a baseline best-known solution obtained using Gurobi-based heuristics. Section 6.4 concludes the chapter with an outlook for future work.

## 6.2 Heuristic approaches

This section introduces the proposed RBFNN heuristic and other state-of-the-art heuristics in the literature. The computational complexity of the heuristic approaches considered are

also presented.

ML-based heuristics assume the availability of solved TS instances  $\Omega^S$  to learn the mapping between input features (loading conditions) and output vectors (topology). Generating this set of solved TS instances  $\Omega^S$  often requires solving a relaxed MIP problem of DCOPF. This chapter adopts the MIP formulation in Fisher [73] slightly modified to penalise the objective function. The formulation considers a single period economic dispatch problem with binary variables to track switching decisions, where  $(z_l = 1)$  and  $(z_l = 0)$  indicate available and unavailable lines, respectively.

The proposed modified objective function in Eq. (6.1) explicitly considers the switching action alongside minimising the generation costs  $C_g P_g$ , thereby improving the solution time of the MIP problem. This explicit consideration of switching actions is the minimisation of the Euclidean distance  $w_l(z_l^* - \alpha_l)^2$  between the binary variables  $z_l$  and the base topology that considers all lines as available  $\alpha_l^{|\Omega|} \in \{1\}^{|\Omega|}$ , where  $w_l$  is a weighting parameter that determines the contribution of the switching action to the cost minimisation function, which can be tuned in the orders of magnitude of the cost function or scaled down to regulate the effect of the penalty contribution. The physical constraints that ensure the capacity of power systems are met are in Eqs. (6.2)-(6.5) and the big-M method in Eqs. (6.6)-(6.7) ensures that when a line is switched off ( $z_l = 0$ ), the line flow constraints on other lines that share same bus connections remain active.  $\theta_b, \theta_m \forall l$  represent nodal angles for a line  $l$  connecting buses  $b, m$  and  $b \neq m$ .  $\mathcal{J}$  in Eq. (6.8) sets an upper bound for the number of transmission lines that can be switched at all times.

All ML-based heuristics consider a set of TS solved instances  $\Omega^S$  generated using the formulation in Eqs. (6.1)-(6.9).

$$\begin{aligned} & \underset{P_g, z_l}{\text{minimise}} && \sum_g C_g P_g + \sum_l w_l (z_l - \alpha_l)^2 \\ & \text{subject to} && \end{aligned} \tag{6.1}$$

$$\theta_b^{min} \leq \theta_b \leq \theta_b^{max}, \forall b \tag{6.2}$$

**Algorithm 5** Proposed RBFNN heuristic for real-time TS

**Require:** set of solved TS instances  $\Omega^S = \{\{X_1, Y_1\}, \{X_2, Y_2\}, \dots, \{X_r, Y_r\}\}$

- 1: Train RBFNN using  $\Omega^S$
- 2: To solve a new instance  $X_{r+1}$ , use  $X_{r+1}$  as input into RBFNN to get  $Y_{r+1}$
- 3: Solve:

$$\begin{aligned} & \underset{Y_{r+1}^*}{\text{minimise}} \sum_{i=1}^{|\Omega|} (Y_{r+1}^* - Y_{r+1}^i)^2 \\ & \text{s.t. Eqs. (6.2)-(6.9) are satisfied} \\ & \text{where } Y_{r+1}^i \in \{0, 1\}, \end{aligned}$$

- 4:  $Y_{r+1} \leftarrow Y_{r+1}^*$
- 5: return  $Y_{r+1}$

$$P_g^{min} \leq P_g \leq P_g^{max}, \forall g \quad (6.3)$$

$$P_l^{min} z_l \leq P_l \leq P_l^{max} z_l, \forall l \quad (6.4)$$

$$-\sum_l P_l - \sum_g P_g - \sum_e P_e = 0, \forall b \quad (6.5)$$

$$B_l(\theta_b - \theta_m) - P_l + (1 - z_l)M \geq 0, \forall l \quad (6.6)$$

$$B_l(\theta_b - \theta_m) - P_l - (1 - z_l)M \leq 0, \forall l \quad (6.7)$$

$$\sum_l (1 - z_l) \leq \mathcal{J} \quad (6.8)$$

$$z_l \in \{0, 1\}, \alpha_l \in \{1\} \quad (6.9)$$

In the rest of this section, this work presents the proposed RBFNN heuristics and describe other heuristics considered for comparison as follows. These heuristics serve as baseline for state-of-art approaches to transmission switching and show how the proposed RBFNN differ in its approach.

### 6.2.1 Proposed RBFNN Heuristic

The proposed RBFNN heuristic is presented in *Algorithm 5*. The heuristic assumes the availability of many solved instances of the TS problem  $\Omega^S = \{\{X_1, Y_1\}, \{X_2, Y_2\}, \dots, \{X_r, Y_r\}\}$  that covers a range of possible operating conditions, where  $X_r$  represents the loading condition and  $f(X_r) = Y_r \in \{0, 1\}^{|\Omega|}$  are optimised topologies. Then, in real-time, the RBFNN

considers active load profiles  $X_{r+1}$  as inputs and outputs a topology  $Y_{r+1}$  that can minimise the dispatch cost considering TS. An example architecture of the RBFNN (presented in Fig. 6.2) is composed of 3 layers, the input layer, a non-linear hidden radial basis function (RBF) layer, and an output layer. The activation function of the non-linear hidden layer of the RBFNN architecture maps the input vector  $X$

$$\Phi(X) = e^{-\beta\|x-\mu\|^2} \quad (6.10)$$

to a Gaussian function such that the feature space spans a set of Gaussian neural nodes. The Gaussian function is well-suited for capturing non-linear relationships in the data, by assigning higher activation values to inputs that are closer to the center points and lower activation values to inputs that are farther away, which creates a smooth and continuous transformation of the input space, allowing the network to capture intricate patterns and non-linear relationships in the data. The node centers  $\mu$  are initialised with a K-means clustering algorithm, and the output of each node depends on  $\|x - \mu\|$ , such that similar inputs  $X$  in the Euclidean space generate similar outputs.  $\beta$  is a hyper-parameter of the hidden-layer to be tuned. The activation function of the output layer is a linear transformation of the weighted sum of all the outputs of the non-linear hidden layer  $w_{ij}$ . The training of the RBFNN with input vector  $X$ , hidden layer weights  $W$ , and output vector  $Y$  uses the Least Squares Linear Regression (LSLR) to obtain the weights

$$W = (X^T X)^{-1} X^T Y \quad (6.11)$$

The exponential activation function allows the RBFNN to provide confidence intervals of predictions which is an important feature as SOs prefer to be in the decision loop when using ML-based approaches [173].

Finally, in real-time, the RBFNN model accepts as input the current load profiles  $X_{r+1}$  and instantaneously outputs a TS solution  $Y_{r+1}$ . As it is possible that  $Y_{r+1}$  is not a feasible

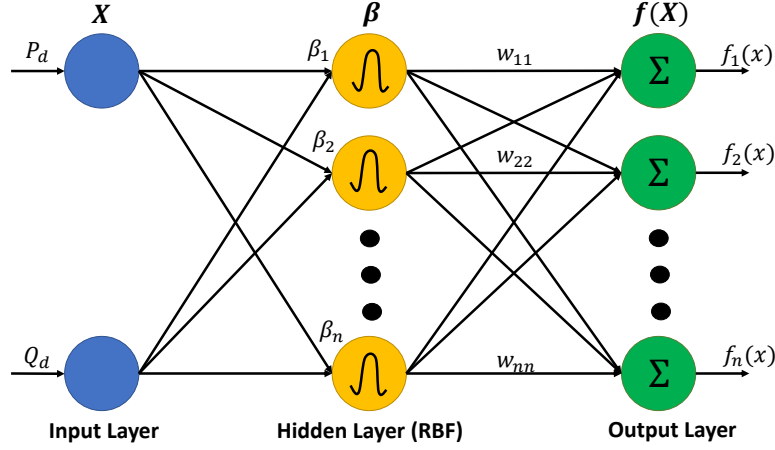


Figure 6.2: An example RBFNN architecture

solution, the algorithm solves the optimisation

$$\underset{Y_{r+1}^*}{\text{minimise}} \sum_{i=1}^{|\Omega|} (Y_{r+1}^* - Y_{r+1}^i)^2 \quad (6.12)$$

while satisfying the constraints in Eq. (6.2)-(6.7). Solving this optimisation returns the nearest feasible topology  $Y_{r+1}^*$  in the event that the TS solution obtained  $Y_{r+1}$  is not feasible.

### 6.2.2 KNN Heuristic

The KNN heuristic is based on Johnson [161], where a mapping exists between vectors of OCs  $X$  and the known DC optimal switching instances  $f(X)$  resulting in a set  $\Omega^S$  of TS solved instances. The KNN heuristic trains on the training data  $\Omega^S$ , and in real-time chooses  $k$  instances that are similar to a candidate OC  $X$  based on the Euclidean distance or  $l_\infty$ -norm. Subsequently,  $k$ -DCOPFs are solved and the TS instance with the lowest dispatch cost is returned as the selected switching action.

### 6.2.3 Sensitivity-based Greedy Search Heuristic

Another heuristic considered is the heuristic proposed in Fuller [164] to rank the lines according to the sensitivity parameter of the dual of the DCOPF (nodal price). Subsequently, the algorithm iterates over half the ranked lines in a sequential manner as done by Yang [167] termed the line enumeration algorithm. The algorithm starts with a baseline cost considering a topology with all lines in service. At the  $i$ -th iteration, the algorithm then solves  $|\Omega| - i$  DCOPF problems by opening single lines according to the line ranking and permanently switching off the line that results in the most cost improvement. Starting with  $i = 0$ , the algorithm continues until there is no cost improvement or the upper limit of switchable lines  $\sum_l (1 - z_l) = \mathcal{J}$  is satisfied.

### 6.2.4 Gurobic Heuristic

The Gurobi heuristic solves for the optimal TS problem described in Eqs. (6.1)-(6.9). This work also considers this formulation as a heuristic as there is a limit to the upper-bound of switched lines  $\sum_l (1 - z_l) \leq \mathcal{J}$ . This heuristic serves as the yardstick to compare the different heuristics.

### 6.2.5 Computational complexity

The computational complexity of the different heuristics differ depending on whether the comparison is offline or online. In the offline setting, the ML approaches generally require a significant amount of time that increases relative to the size of the test network, as computationally expensive MIPs need to be solved to curate the set  $\Omega^S$ . Training the RBFNN varies with the number of neurons in the architecture and the number of epochs, but the simple architecture of the RBFNN allows for a linear training time. The offline training time for the KNN heuristic is dependent on the  $k$ -neighbours. The offline computation for the sensitivity-based greedy local search heuristic is trivial as it requires only solving a DCOPF.

In the online setting where there is a huge constraint in time, an exhaustive search for the greedy search heuristic requires  $\mathcal{O}(|\Omega|^2)$  DCOPF solves and thus scales poorly in large systems. The KNN heuristic depends on the  $k$  DCOPFs required to be solved. The RBFNN heuristic provides almost instantaneous solutions that requires at most a single DCOPF solve to verify feasibility. Thus, it is expected that the proposed RBFNN heuristic will provide significant computational improvements in the online setting even with the feasibility verification step.

## 6.3 Case Study

In this section, this work showcases case studies to compare the proposed RBFNN heuristic with the KNN, sensitivity-based Greedy Search, and Gurobi-based heuristics. The first study investigates the performance of the heuristics to out of distribution data as a measure of robustness. In the second study, the real-time performance of the heuristics in respect to cost savings and computational time are compared. The third case study highlights the offline computational improvements of the modified MIP formulation. Finally, a discussion section concludes the case study section.

### 6.3.1 Test System and assumption

The case studies use the modified IEEE 118-bus test case in Blumsack [178]. The load profiles are generated via Monte Carlo sampling. The active loads were sampled from a multivariate Gaussian distribution and assume the correlation between loads to follow Pearson's correlation with a correlation coefficient of 0.75. The distribution was then converted to a marginal Kumaraswamy(1.6,2.8) distribution using inverse transformation. A  $\pm 10\%$  variation is considered in the load distribution for the training data-set. The optimisation in Eqs. (6.1)-(6.9) is then ran for each load profile with an upper-bound on switching actions  $\mathcal{J} \leq 5$ . The training data is then a feature space comprising 1000 active load profiles  $X_r \in \{P_e\}$  and labels that correspond to a binary sequence  $Y_r \in \{0,1\}^{|\Omega|}$ ,

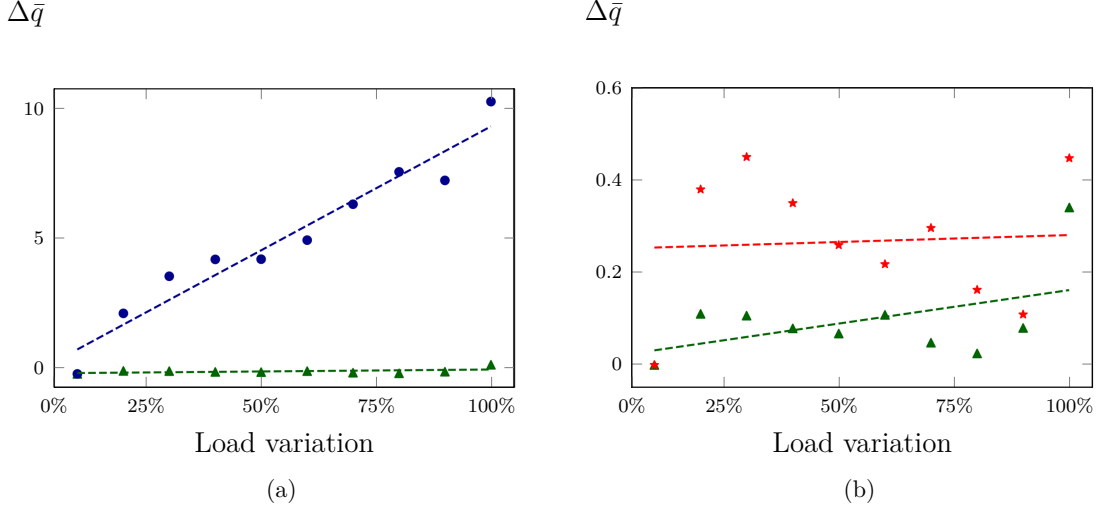


Figure 6.3: Change in average dispatch cost ( $\Delta \bar{q}$ ) with variation in load distribution for the proposed RBFNN ( $\blacktriangle$ ) and (a) KNN ( $\bullet$ ) (b) a feed-forward ANN ( $\star$ ) heuristics.

where  $Y_r^i = 0$  and  $Y_r^i = 1$  represents the  $i$ -th line as absent and present, respectively. In studying the real-time computational performance, all previously introduced heuristics are tested on 200 active load profiles with  $\pm 35\%$  variation in the load distribution. Unless otherwise stated, all studies consider the  $\pm 35\%$  variation in the testing data.

All optimisation problems were implemented using the package Pyomo 5.6.8 [137] in Python 3.7.4 and solved using Gurobi 9.5.0 [152]. The RBFNN is implemented using Keras package 2.8.0, with root mean squared propagation (RMSprop) as the optimiser and mean squared error as the loss function. The activation function between the input and hidden layer, and between the hidden and output layer is linear. A sigmoid activation function is included after the output layer such that all outputs lie in the range  $[0, 1]$ . The  $\beta$  parameters of the hidden layer were 0.001, and the training considered an epoch length of 200 and a batch size of 5.

### 6.3.2 Performance of heuristics on out of distribution data

This case study investigates the performance of the proposed RBFNN heuristic to out of distribution data. This study compares the proposed RBFNN heuristic against the KNN heuristic and a feed-forward perceptron ANN. The feed-forward ANN follows Algorithm (5)

similar to the proposed RBFNN with the difference only in the model architecture of 5 as opposed to 3 layers. All the heuristics were trained considering a  $\pm 10\%$  variation in the load distribution as described in Sec. 6.3.1. The baseline performance is considered using data not in the training set but following a similar  $\pm 10\%$  variation. Subsequently, the relative change in performance of the heuristics on test instances that differ from the  $\pm 10\%$  variation is compared.

The results are shown in Fig. 6.3a, which presents a scatter-plot that compares the change in average relative dispatch cost with different variations (10 %, 20 %, 30 %,  $\dots$ , 100 %) in load distribution for the proposed RBFNN, KNN and feed-forward ANN heuristics. The average relative change in dispatch costs considers a baseline where the training and testing data come from the same distribution of  $\pm 10\%$  variation. Subsequently the heuristics are tested for each variation in load distribution and compute the change ( $\Delta \bar{q} = \frac{\bar{q}-q}{q}$ ) from the baseline relative cost  $q$  for 200 test cases in  $\Omega^{S''}$ , where  $\bar{q} = \frac{\sum_{i=1}^{|\Omega^{S''}|} \frac{\bar{q}-q}{q}}{|\Omega^{S''}|}$  is the average relative cost for an heuristic.

The results in Fig. 6.3a show a minimum and maximum change in relative costs  $\Delta \bar{q}$  of 2.5 % and 34.6 %, respectively, corresponding to the  $\pm 20\%$  and  $\pm 100\%$  load variation for the proposed RBFNN heuristic. The linear regression line is horizontal, which denotes little to no change and thus robustness to out of distribution data. In contrast, the KNN heuristic has a minimum and maximum change  $\Delta \bar{q}$  of 224 % and 1000 %, respectively, corresponding to the  $\pm 20\%$  and  $\pm 100\%$  load variation. Additionally, the linear regression line makes an  $\approx 45^\circ$  with the horizontal axis, which suggests that the heuristic performance changes with different load variations. A further comparison between the proposed RBFNN heuristic with a feed-forward ANN heuristic for similar changes in load distribution in Fig. 6.3b shows that the proposed RBFNN has a similar regression line as a feed-forward ANN. There, the feed-forward ANN has a minimum and maximum change of 10.9 % and 45.2 %, respectively. While both the proposed RBFNN and the feed-forward ANN have an average relative cost  $\frac{\bar{q}-q}{q}$  of  $\leq 5\%$ , the average change  $\Delta \bar{q}$  over the different variations (10 %, 20 %,  $\dots$ , 100 %) for the proposed RBFNN is 9.7 % while that of the feed-forward ANN is 26.7 %. These

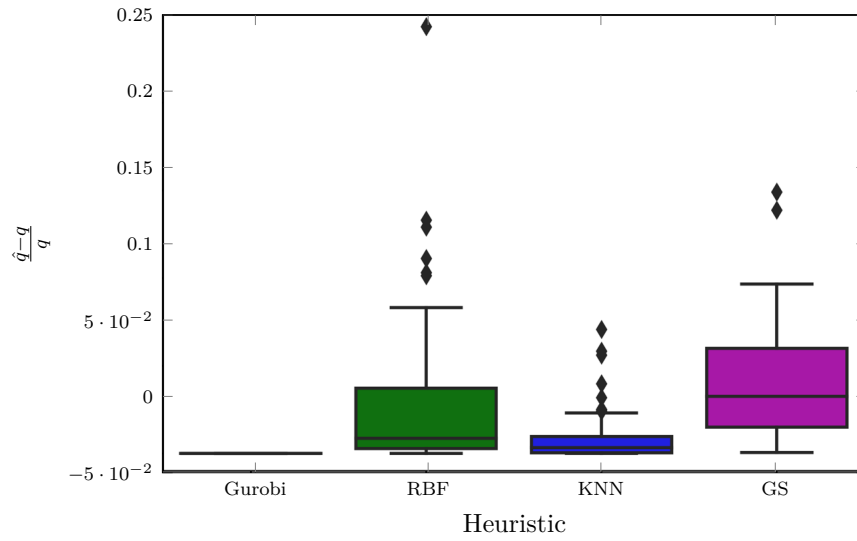


Figure 6.4: The relative dispatch cost to best known solution ( $\frac{\hat{q}-q}{q}$ ) for 200 test instances using different heuristics

Table 6.1: The relative cost and computation time and for different heuristics on 200 TS test instances

Heuristic	Relative cost ( $\frac{\hat{q}-q}{q}$ )	Time
Gurobi	$0.00 \pm 0.00$	$(31.31 \pm 61.38) \text{ s}$
RBFNN	$0.03 \pm 0.03$	$(0.57 \pm 0.21) \text{ s}$
KNN	$0.01 \pm 0.02$	$(6.25 \pm 0.25) \text{ s}$
Greedy Search	$0.04 \pm 0.03$	$(21.93 \pm 0.32) \text{ s}$

results support the robustness of the proposed RBFNN heuristic.

### 6.3.3 Real-time computational performance

This case study shows the computational results of the heuristics introduced in Sec. 6.2 in terms of dispatch cost savings and computational time. While the heuristics training was on data from the  $\pm 10\%$  variation, the testing data had a  $\pm 35\%$  variation. The box plots in Fig. 6.4 and Fig. 6.5, and the data in Table 6.1 summarise the results in terms of relative cost savings and computational time. The box plots indicate the median value as the middle line in the box, the first and third quartiles as the top and bottom lines in the box, respectively, and the outliers as individual points.

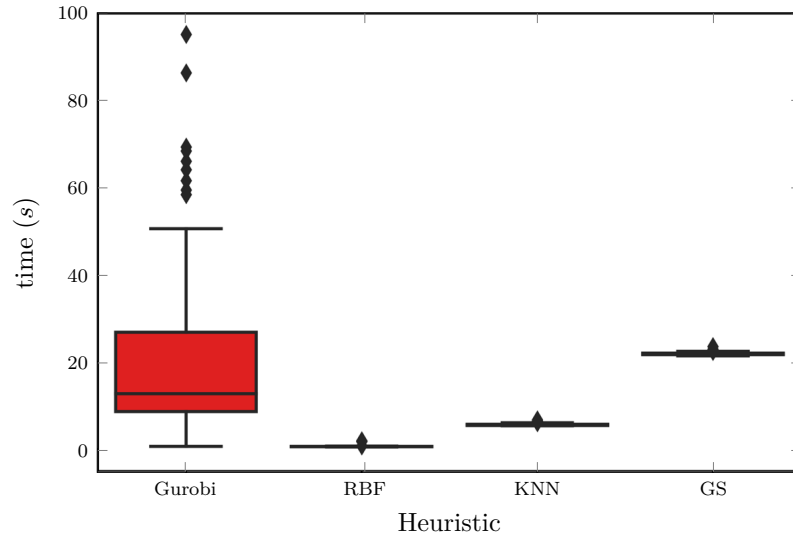


Figure 6.5: Computational time to obtain TS solutions for 200 test instances using different heuristics

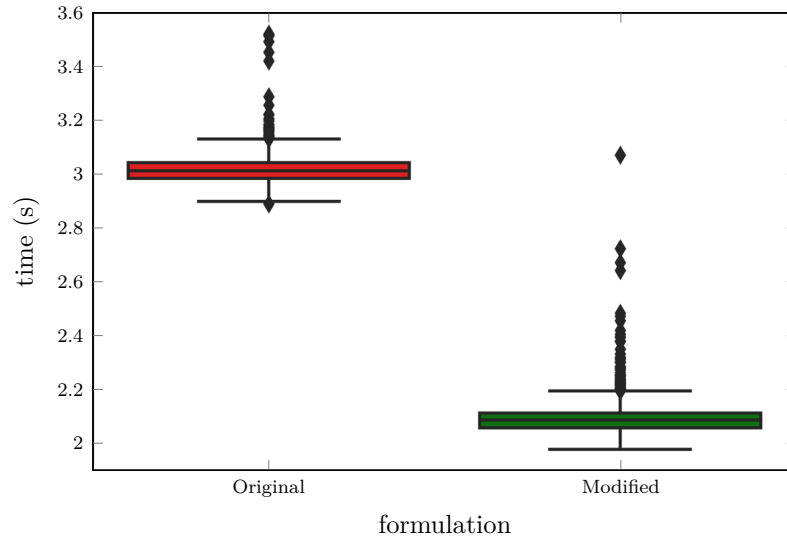


Figure 6.6: Solution time for different MIP formulations

### Cost performance

Fig. 6.4 compares the relative cost savings using the different heuristics for 200 test instances. The relative costs are calculated according to the best known dispatch cost (calculated using the Gurobi heuristic) as  $\frac{\hat{q}-q}{q}$ , where  $\hat{q}$  represents the dispatch cost obtained using an heuristic for TS, and  $q$  is the best known cost. Concretely, the figure shows that the proposed RBFNN heuristic with a median relative cost of 0.8 %, outperforms the

Table 6.2: Handling infeasible TS of the propsoed RBFNN heuristic for  $\mathcal{J} \leq 5$ 

Heuristic	Infeasible TS	Time	Relative cost ( $\frac{\hat{q}-q}{q}$ )
RBFNN	4 / 200	$(1.32 \pm 0.05)$ s	$0.08 \pm 0.10$
Gurobi	0 / 200	$(7.25 \pm 2.62)$ s	$0.00 \pm 0.00$

greedy search heuristics at 3.3 %, and is similar to the KNN heuristic at 0.3 % relative cost. This result is consistent with the data in Table 6.1 showing the mean and standard deviation of the relative costs for the different heuristics. The results in Table 6.1 show that the proposed RBFNN heuristic is on average 3 % close to the best known TS solution. This result suggests that the proposed RBFNN heuristic can propose good TS solutions.

### Computational time performance

The clear advantage of the proposed RBFNN heuristic is visible in the real-time computation time in Fig. 6.5 and Table 6.1. Concretely, Fig. 6.5 presents the distribution of the computation time it takes to solve the TS problem for 200 test instances. The proposed RBFNN heuristic has a median value of 1.2s, while the Gurobi, KNN and greedy search heuristics have median values of 13s, 6s, 22s, respectively. This result is supported by the data in Table 6.1 that shows the mean and standard deviation of the computation time for the different heuristics. There, the proposed KNN heuristic outputs a TS solution on average in 0.57s, while the KNN, Greedy Search and Gurobi heuristics average 6.25s, 21.93s, and 31.31s respectively. This massive computational saving makes the proposed RBFNN a suitable approach for real-time applications. These results also match the computational expectations described in Sec 6.2.5.

### 6.3.4 Guaranteeing feasible topology near real-time

Notably, the crucial comparison is between the proposed RBFNN and the KNN heuristics, as the scalability of the greedy search heuristic is poor due to  $\mathcal{O}(|\Omega|^2)$  DCOPF computations required. While the KNN offers a marginally better relative cost value, the value of  $k$

determines the number of DCOPF problems to solve and thus worsens the computational time. Sometimes, however, the initial predicted solution by the RBFNN is infeasible and in such cases the feasibility is ensured by solving Eq. (6.12) that returns the nearest feasible topology. In this case study, these infeasible solutions appeared in 4 out of 200 test instances (2%). For these 4 infeasible TS instances, the proposed heuristic took an average of 1.32 s to obtain the nearest feasible topology and the obtained result was within 8% of the best-known solution considering  $\mathcal{J} \leq 5$  line switches, as opposed to an average of 7.25 s required to solve the same MIP formulation with a similar relaxation. These results are summarised in Table 6.2 and importantly the proposed RBFNN solves for a feasible TS topology within 8% of the best-known solution while being more than  $5\times$  faster than solving Fisher's [73] MIP formulation.

### 6.3.5 Offline computational improvements

In this case study, the effect of penalising the objective function on the solution times of the MIP is investigated. A thousand TS solution instances of the MIP problem for a given load profile with  $\mathcal{J} \leq 3$  are considered. The original formulation does not consider the penalty term  $\sum_l w_l (z_l^* - \alpha_l)^2$  in Eq. 6.1, while the modified formulation is exactly as presented in Sec. 6.2. The results are presented as box-plots of solution times in seconds in Fig. 6.6. Concretely, the results show an average improvement of 0.90 s for the modified formulation, which is 43% faster than the original MIP formulation. This percentage improvement does not extrapolate a linear relationship as the maximum number of number of line switches  $\mathcal{J}$  increase. However, as the number of crucial line switches are usually few, this result is still relevant.

### 6.3.6 Key Findings

ML has shown tremendous promise for real-time TS. Previous literature [175] have shown the promise of ML models in fault detection classification problem, and the RBFNN model

has been shown to be robust to inputs not in the training data. This robustness makes it a favorite for power systems classification problems such as fault detection. In the case studies, the proposed RBFNN TS heuristic showed similar accuracy performance across different data distributions, with a maximum deviation of 34.6 % in contrast to the 1000 % deviation of the KNN model. RBFNN models have been shown to perform well in terms of accuracy, interpretability, training data and computation time. Similarly, RBFNN models show significant benefits of providing almost instantaneous TS options and thus are applicable in real-time settings by SOs. The main computational burden comes from generating the solved TS instances for training the RBFNN model which is done offline. However, this offline step is necessary for any ML workflow that require synthetic data to train a model. On a positive note, the simplicity of the RBFNN makes training the model relatively fast and its architecture allows for more interpretability of the model. The proposed RBFNN has an error margin of 3 % relative to the best known solution and provides the TS solution more than  $10\times$  faster than state of the art ML-based TS.

## 6.4 Conclusion

This work proposed a RBFNN heuristic for real-time TS. The proposed heuristic outperforms heuristics in the literature such as KNN models and sensitivity-based greedy search algorithms in terms of significant savings in computation time. The proposed RBFNN heuristic is fairly robust to noisy data (where the testing data has a different distribution than the training data) and performs as well as the state of art ML approach to TS. In situations where the proposed RBFNN outputs an infeasible topology, the proposed RBFNN solves for a feasible TS topology that is within 8 % of the best-known solution while being more than  $5\times$  faster than solving a standard MIP. The proposed approach is tested on the IEEE 118-bus test case and limited the maximum number of switchable lines to  $\mathcal{J} \leq 5$ . In larger systems, the maximum number of switchable lines may increase, albeit a small number of lines result in the largest cost reductions [73]. As the issue of contingencies is of utmost importance to the SO, further work will investigate learning a model to propose TS actions that are N-1 secure.

## Chapter 7

## Conclusion

*“There is no real ending. It’s just the place where you stop the story.”*

—Frank Herbert

*“What is our life? A play of passion,  
Our mirth the music of division.  
Our mothers’ wombs the tiring houses be  
Where we are dressed for this short comedy.  
Heaven the judicious sharp spectator is  
That sits and marks still who doth act amiss.  
Our graves that hide us from the searching sun  
Are like drawn curtains when the play is done.  
Thus march we playing to our latest rest,  
Only we die in earnest, that’s no jest.”*

—Sir Walter Raleigh

*“Stay hungry; stay foolish.”*

—Steve Jobs (more recently)

## 7.1 Summary of Thesis Achievements

The power system is undergoing a significant transformation driven by the need to decarbonise, decentralise and digitise the system. As power systems integrate more renewable energy sources to replace conventional thermal generation, the operational uncertainty increases further in an already complex system. Similarly, the direction of power flow evolves from unidirectional to bidirectional, as prosumers may actively wish to trade their electricity in this decentralised system. This bidirectional flow will require the development of new dynamic power markets as well as new services of flexibility services to handle increasing uncertainties. Historically, the system operator has relied on asset redundancy to conservatively operate equipment below rated capacity, thereby managing uncertainty. However, this approach is inefficient and may not be able to cope with the increased operational uncertainties introduced by the integration of renewable energy sources. Therefore, new approaches which are more effective and efficient are needed to manage these uncertainties, such as using advanced data analytics and machine learning algorithms.

The transition towards a decarbonised, decentralised and digitised power system necessitates the development of new operation and control techniques to manage the increased uncertainty inherent in integrating renewable energy sources. Instead of investing in new grid infrastructure, it is imperative to adopt innovative strategies for the existing grid. Specifically, the system operator can leverage the flexibility inherent in DERs at medium to low voltage levels. By harnessing the capabilities of these local assets, the system operator can substantially reduce operational costs and improve the security of the power system.

Although the digitalisation paradigm offers cost-effective solutions by avoiding asset redundancy and providing more data for system operators, it also brings along increased cybersecurity risks and possible inaccurate predictions. However, the digitalisation of the power system can support the adoption of data-driven methods in the control room, which can significantly enhance the situational awareness of the system operator and even propose appropriate control actions. The use of data-driven models is increasingly becoming

popular in power systems with recent advances and successes in other domains like biology (Deep Mind’s AlphaFold) and gaming (AlphaGo). The promising proposal of using these data-driven models to predict more uncertain dynamic phenomena that may be rare, the so-called ‘low probability, high impact’ events, can be the key to preventing power black-outs in electricity grids. In power systems, although the physics is known, for instance, the swing equations for the dynamics or Kirchhoff’s laws for the power flows, the system is a highly non-linear and complex interconnected system. There, data-driven approaches can tease out salient relationships and provide an advantage of speed. Additionally, baseline assumptions can be verified from the underlying physics. Further, there are opportunities for using data-driven methods that embed the known physics, which can be learned through regularising the loss function, for instance, as physics-informed learning [179]. Incorporating the established principles of network physics into ML approaches holds significant promise as it can enhance interpretability, robustness and generalisation.

However, applying data-driven approaches to the power system has some challenges, which this thesis attempted to address. These challenges include improving the quality of the training data, model selection and usage in real-time settings, and managing risks when using a machine-learned model. Building upon previous work on ML-based DSA and control [99, 4, 96], this thesis presented the challenges and corresponding contributions to address them. In this section, this thesis concludes by discussing future research directions.

### 7.1.1 Quality of training data

The first challenge addressed was on generating quality training data for ML-based DSA. This challenge is crucial for the adoption of ML-based approaches to DSA as the quality of the training data has a strong correlation with the quality of the trained models, and therefore, to improve the prediction performance of ML-based models for DSA, the training data must of high quality. The work presented in Chapters 3 & 4 addressed this challenge. In Chapter 3, this thesis presented a novel sampling method that sequentially generates physically feasible but ‘rare’ OCs that have not been previously observed in power systems

operations. The developed method aims to move beyond the limitations of historical observations to explore the full range of possible OCs. In Chapter 4, this thesis presented a novel approach based on Wasserstein distance and entropy to efficiently trade-off contrasting properties of quality datasets, trading off historically relevant with feasible but 'rare' OCs to create an enriched database capable of training a high-performing classifier. The proposed methods can generate datasets that can serve to benchmark the performance of ML models for security assessments.

Future research can focus on developing a sophisticated sampling procedure for different scenarios where the ML model to be used by the system operator then depends on the forecasted online conditions. A crucial feature of the proposed split-based sampling method is the sequential nature of synthetic sample generation. Hence, it can be adapted to consider multiple objectives by the system operator for creating different datasets according to the changing needs. For instance, the probabilities and impacts of contingencies can be considered as metrics to guide sampling from more 'relevant' regions of the feasible space, similar to the trade-off of different properties in Chapter 4. The proposed method can serve as a baseline for generating scenario-based datasets, as the sequential sampling process can consider the many requirements of the system operator by modifying the objective function. To generate datasets that generalise to different configurations, such as topology, Section 7.2.4 outlines possible research directions.

### 7.1.2 Model selection and usage for real-time application

The second challenge that this thesis addressed was the selection and usage of machine-learned models in the control room for real time operation. In this context, the system operator must balance multiple and often contrasting objectives in the training and selection process of the ML model. For instance, while high predictive accuracy is a prerequisite for these models, interpretability and sensitivity to changing risks and impacts are also crucial. In security assessments, interpretability is crucial for understanding the decision-making mechanism of a classifier, which is vital for building trust and adapting to

changing assumptions such as topological reconfiguration or data distribution. Additionally, verification of the feasibility of the solutions from these models [within an optimisation framework](#) can enable adoption by system operators. In Chapter [5](#), this thesis proposed a scalable workflow to trade-off between multiple objective criteria when selecting decision tree models for real time operation by system operators. Furthermore, in Chapter [6](#), this thesis presented a practical implementation of using a machine-learned model to optimise the cost of power system operations using topological control actions.

Future research on the selection of machine-learned models for real-time operation can further develop the concept of interpretability and establish a standardised approach that can be applied to various types of models, enabling a more rigorous and sophisticated selection process. An additional avenue for research is to incorporate the contingency probabilities and impacts into the data generation process and in model selection, creating a closed-loop framework of ML-based solutions [52] that better reflects the reality of power system operations.

Future work on ML-based real-time transmission switching can include extending the proposed method to consider probable contingencies in the solution, ensuring the optimised topology is secure against possible contingencies, such as  $N-1$  security. As data-driven models continue to be developed, it will be essential to conduct real-world demonstrations in collaboration with system operators to assess the applicability of these methods and identify new areas for research. These studies can provide insights into how system operators can efficiently utilise existing assets to operate the grid.

There are opportunities for developing machine-learned models that are not only accurate but also interpretable and sensitive to changing risks and impacts and the integration of these models into the decision-making process of the system operator in real time.

## 7.2 Future Research Directions

This section outlines a high-level outlook on potential future research directions for ML applications in power systems operations. While the section above summarises the incremental improvements from this thesis, the focus here is on three promising future directions. These directions include utilising knowledge of physics in the ML workflow, overlapping ML-based operational planning with asset deployment, and in the area of DSA, research on training models that generalise to different conditions such as faults and topology. These research directions are especially important given the role machine-learned can [model](#) play in securing low inertia (LI) systems. Below, this thesis briefly discusses the role of machine-learned models in securing low inertia systems and outlines the research directions.

### 7.2.1 Machine learning in low inertia systems

Most studies on ML-based security assessment assume power systems dominated by synchronous machines, characterized by high inertia (HI) [180]. However, the increasing penetration of renewable energy sources has led to a rise in converter-interfaced devices, resulting in lower inertia and faster dynamics [180]. This changing grid structure requires reevaluation of assumptions and conclusions in ML-based security assessment. The presence of converter-interfaced generators (CIGs) introduces faster dynamic time scales and changing variables, impacting stability classifications. Flexible loads, previously ignored transmission system dynamics, and the risk of large faults caused by High-voltage direct current (HVDC) lines further complicate the assessment [181, 180, 5, 3].

Integration of CIGs necessitates considering converter-driven stability, resonance stability, rotor angle, voltage, and frequency stability. CIGs associated with renewable sources offer fast primary frequency response critical for maintaining stability in low-inertia systems [180]. These changes must be studied within the ML-based security assessment framework. Two examples of these changes are the requirement for high-quality training data, including "rare" operating conditions (OCs), and the need for different models for various genera-

tor/converter insecurity patterns. Clustering dynamics and training unique classifiers for each cluster could address this challenge [101, 182].

A recent comprehensive analysis emphasized the feasibility and value of utilizing machine-learned models in low-inertia (LI) systems, indicating a stronger dependence on power generation changes for system security [101]. ML-based methods hold promise for efficiently securing LI systems, but standardized systems representing LI systems need continuous development and updates to ensure model validity. As power systems transition to LI systems, important research questions arise, and three future research directions are highlighted below.

### 7.2.2 Physics-informed learning

ML, particularly deep learning (DL), has shown great promise in recent years, both for operational planning and security assessment. However, DL is often marred with several challenges, including high requirements for training data, physically inconsistent solutions, and low generalisability and interpretability. Physics-informed neural networks (PINNs) can address these challenges by integrating physics-informed rules into DL methodology. PINN utilises scientific knowledge or physics laws to direct the optimisation, design of architecture, and implementation of deep neural networks [179]. The literature describes four types of PINN: (a) physics-informed loss functions, (b) physics-informed initialisation, (c) physics-informed architecture design, and (d) hybrid physics-DL models. In physics-informed loss functions, the idea is to attach physical interpretations to the variables present in the hidden or output layers of neural networks and subsequently add a term that represents these physical variables to the loss function through regularisation [183]. Physics-informed initialisation utilises transfer learning to enhance training efficiency and avoid getting stuck in a local minimum. First principles are utilised to provide good initial guesses, as good initial weights can influence the convergence rate [184]. Physics-informed architecture design optimises the neural network architecture and avoids redundant connections. Graph neural networks (GNNs) [185] are an example where the neurons, weights,

and outputs can physically represent real-world entities. Hybrid physics-DL models utilise a priori first principle knowledge to complement the function-approximating ability of DL models. The output of a physical model can be used as input to the neural network [186], thereby efficiently incorporating prior physical knowledge to choose, modify, and combine the most significant input variables of the neural network.

Research has shown promising results in using PINNs to improve state estimation, dynamic analysis, power flow calculation, anomaly detection, and optimal power flow [179]. For example, in state estimation, PINNs can overcome the limitations of the intractability and complexity of conventional methods for estimating large-scale three-phase imbalanced distribution networks. However, modelling unbalanced components and the coupling between phases in unbalanced distribution remains an open challenge. Further research can also consider using PINNs in future integrated energy systems, where understanding the dynamics of flow and different time characteristics of combined heat and power systems can unlock new insights into the cost-efficient utilisation of local flexibility. Moreover, PINNs can aid in the digitalisation paradigm and facilitate the integration of DERs, such as electric vehicles (EVs), with advanced metering and communication infrastructure.

The development of PINN is motivated by several potential advantages, such as improved accuracy, interpretability, physical consistency, generalisability, reduced search space of weights, improved training and convergence performance, and enhanced sampling efficiency, thus reducing the need for a large amount of training data. The use of PINN has shown promising results in current research to improve state estimation, dynamic analysis, power flow calculation, anomaly detection and optimal power flow [179]. For instance, in state estimation, PINNs can overcome the limitations of the intractability and complexity of conventional methods for estimating large-scale three-phase imbalanced distribution networks. However, modelling unbalanced components and the coupling between phases in unbalanced distribution remains an open challenge. Further research can also consider using PINNs in future integrated energy systems, where understanding the dynamics of flow and different time characteristics of combined heat and power systems can unlock new insights into the cost-efficient utilisation of local flexibility. Moreover, PINNs can also aid

in the digitalisation paradigm and facilitate the increasing integration of DERs, such as EVs, with advanced metering and communication infrastructure. As the world becomes more digitised, it is crucial to create networks that are both efficient and sustainable. This importance will only increase as more DERs are added in the future. Network operators must design innovative solutions to maintain a reliable power system. Consequently, PINNs offer valuable assistance for exploring these complex topics and will undoubtedly require further research.

### 7.2.3 ML-based network planning for secure operations

The decarbonisation of the transport and heating sector across the globe will see the massive adoption of technologies like EVs and hydrogen electrolyzers in power systems distribution networks. These technologies offer unique opportunities for their smart integration into electricity distribution networks and coordination with other storage devices to harness DERs for demand-side flexibility and autonomous local energy. In the digitalisation paradigm, smart controls are essential for efficiently utilising existing assets. As local assets aggregate to provide flexibility services, there are research opportunities at the intersection of coordinating the deployment of these assets for optimal operations. One research gap exists in developing probabilistic models and prediction tools to support city planners, network operators, and businesses deploy the required infrastructure. Ensuring power system security is crucial in system operation, and therefore, identifying critical nodes like transmission lines, distribution lines, and transformers as the grid evolves accommodating more renewable energy and DERs is vital. There may be higher peak load requirements imposed on the grid due to more DERs on the grid, such as EVs, leading to voltage violations in distribution networks. ML techniques can detect critical nodes based on PMU measurements in real-time, and probabilistic methods can support planning decisions for cost-effective and optimal grid reinforcement. For instance, the strategic deployment of charging infrastructure to reliably support the massive adoption of EVs can be coordinated alongside the operational needs of the system operator to efficiently run

the power system. Since EVs, alongside other DERs, can contribute to the secure operation of power systems [187], there is a research need for the development of data-driven distributed control of DERs for flexibility services and corrective control of local faults, as previous work indicates that the spatial distribution of inertia is vital [180]. Therefore there is a research gap in developing probabilistic models of usage patterns and prediction tools to support city planners, network operators, and businesses deploy the necessary infrastructure.

In the context of secure operations, corrective control is a promising approach to ensure system stability by efficiently utilising existing grid assets. EVs equipped with vehicle-to-grid (V2G) technologies can be utilised for corrective control, showing significant economic value [188]. New research directions include the use of corrective control in real-time operations to address transient instability [189]. For instance, reinforcement learning can be employed as a corrective control tool to identify optimal load-shedding strategies for different operating conditions. This approach can identify the crucial buses that are pivotal to system security and inform the deployment of infrastructure for local flexible assets such as EVs with V2G technologies and hydrogen electrolyzers. This research direction is particularly important in future low inertia systems, as local DERs can be used to provide ancillary services [190].

## 7.2.4 Robust training dataset

Another important possible future research direction considers the generalisability of the training data needed to develop robust models. This generalisability is not only related to the data distribution but also with regard to topology and different fault types. For instance, graph neural networks (GNNs), with their ability to learn from graph-structured data can model the electricity grid and attach physical meanings to the nodes and edge weights. GNNs are a type of PINNs, where the nodes represent buses and the edges are transmission lines. However, GNNs can be intractable when dealing with large graphs, containing millions of nodes and billions of edges. Recent research has proposed mini-batch

training of the GNNs [191] while attempting to capture the complex dependencies between the training samples that are inherently present in the graph structure. However, selecting which nodes to consider in the mini-batches is not trivial and differs from other applications such as natural language processing. There is a gap in developing novel discrete sampling strategies to select the "relevant" nodes.

Novel discrete sampling is also promising to estimate the flexibility boundary between the transmission and distribution, where the effects of discrete assets such as on-load tap changers and capacitor banks can create disjoint flexibility areas. However, current research assumes that these regions are convexly connected, leading to a research gap in search space exploration algorithms that take integer variables into account [192]. The significance of this research is expected to increase with the rise of DERs in the power system and the need for new tools in the control room that can facilitate the activation of these flexibility services.

# Bibliography

- [1] L. Duchesne, E. Karangelos, and L. Wehenkel, “Recent developments in machine learning for energy systems reliability management,” *Proceedings of the IEEE*, vol. 108, no. 9, pp. 1656–1676, 2020.
- [2] P. Kundur, “Power system stability,” *Power system stability and control*, pp. 8–1, 2007.
- [3] U. Markovic, O. Stanojev, P. Aristidou, E. Vrettos, D. S. Callaway, and G. Hug, “Understanding Small-Signal Stability of Low-Inertia Systems,” *IEEE Transactions on Power Systems*, pp. 1–1, 2021.
- [4] J. L. Cremer, “Probabilistic dynamic security assessment for power system control,” *PhD Thesis*, 2020.
- [5] N. Hatziargyriou, J. Milanovic, C. Rahmann, V. Ajjarapu, C. Canizares, I. Erlich, D. Hill, I. Hiskens, I. Kamwa, B. Pal, P. Pourbeik, J. Sanchez-Gasca, A. Stankovic, T. Van Cutsem, V. Vittal, and C. Vournas, “Definition and classification of power system stability – revisited & extended,” *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3271–3281, 2021.
- [6] P. Panciatici, G. Bareux, and L. Wehenkel, “Operating in the fog: Security management under uncertainty,” *IEEE Power and Energy Magazine*, vol. 10, no. 5, pp. 40–49, 2012.
- [7] G. Strbac, N. Hatziargyriou, J. P. Lopes, C. Moreira, A. Dimeas, and D. Papadaskalopoulos, “Microgrids: Enhancing the resilience of the european megagrid,” *IEEE Power and Energy Magazine*, vol. 13, no. 3, pp. 35–43, 2015.

- [8] F. Bellizio, W. Xu, D. Qiu, Y. Ye, D. Papadaskalopoulos, J. L. Cremer, F. Teng, and G. Strbac, "Transition to digitalized paradigms for security control and decentralized electricity market," *Proceedings of the IEEE*, 2022.
- [9] A. Rhodes, "Digitalisation of energy: An energy futures lab briefing paper," 2020.
- [10] A. Rashid and A. Chaturvedi, "Cloud computing characteristics and services: a brief review," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 2, pp. 421–426, 2019.
- [11] D. Fan, Y. Ren, Q. Feng, Y. Liu, Z. Wang, and J. Lin, "Restoration of smart grids: Current status, challenges, and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 143, p. 110909, 2021.
- [12] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213–225, 2019.
- [13] A. Churkin, J. Bialek, D. Pozo, E. Sauma, and N. Korgin, "Review of cooperative game theory applications in power system expansion planning," *Renewable and Sustainable Energy Reviews*, vol. 145, p. 111056, 2021.
- [14] S. Abapour, M. Nazari-Heris, B. Mohammadi-Ivatloo, and M. Tarafdar Hagh, "Game theory approaches for the solution of power system problems: A comprehensive review," *Archives of Computational Methods in Engineering*, vol. 27, pp. 81–103, 2020.
- [15] E. A. Soto, L. B. Bosman, E. Wollega, and W. D. Leon-Salas, "Peer-to-peer energy trading: A review of the literature," *Applied Energy*, vol. 283, p. 116268, 2021.
- [16] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, "Smart transmission grid: Vision and framework," *IEEE transactions on Smart Grid*, vol. 1, no. 2, pp. 168–177, 2010.
- [17] S. Brahma, R. Kavasseri, H. Cao, N. Chaudhuri, T. Alexopoulos, and Y. Cui, "Real-time identification of dynamic events in power systems using pmu data, and potential

- applications—models, promises, and challenges,” *IEEE transactions on Power Delivery*, vol. 32, no. 1, pp. 294–301, 2016.
- [18] M. L. Di Silvestre, P. Gallo, J. M. Guerrero, R. Musca, E. R. Sanseverino, G. Sciumè, J. C. Vásquez, and G. Zizzo, “Blockchain for power systems: Current trends and future applications,” *Renewable and Sustainable Energy Reviews*, vol. 119, p. 109585, 2020.
- [19] A. Hasankhani, S. M. Hakimi, M. Bisheh-Niasar, M. Shafie-khah, and H. Asadolahi, “Blockchain technology in the future smart grids: A comprehensive review and frameworks,” *International Journal of Electrical Power & Energy Systems*, vol. 129, p. 106811, 2021.
- [20] B. Donnot, I. Guyon, A. Marot, M. Schoenauer, and P. Panciatici, “Optimization of computational budget for power system risk assessment,” in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 2018, pp. 1–6.
- [21] I. Konstantelos, M. Sun, S. H. Tindemans, S. Issad, P. Panciatici, and G. Strbac, “Using vine copulas to generate representative system states for machine learning,” *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 225–235, 2018.
- [22] L. A. Wehenkel, *Automatic Learning Techniques in Power Systems*. Kluwer Academic Publishers, 1998.
- [23] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, “Financial time series forecasting with deep learning: A systematic literature review: 2005–2019,” *Applied soft computing*, vol. 90, p. 106181, 2020.
- [24] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz *et al.*, “Transformers: State-of-the-art natural language processing,” in *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 2020, pp. 38–45.
- [25] L. Wehenkel, T. Van Cutsem, and M. Ribbens-Pavella, “An artificial intelligence framework for online transient stability assessment of power systems,” *IEEE Transactions on Power Systems*, vol. 4, no. 2, pp. 789–800, 1989.

- [26] H. Ge, Q. Guo, H. Sun, and W. Zhao, "A model and data hybrid-driven short-term voltage stability real-time monitoring method," *International Journal of Electrical Power & Energy Systems*, vol. 114, p. 105373, 2020.
- [27] M. Amroune, "Machine learning techniques applied to on-line voltage stability assessment: a review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 273–287, 2021.
- [28] L. Zhu, C. Lu, Z. Y. Dong, and C. Hong, "Imbalance learning machine-based power system short-term voltage stability assessment," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2533–2543, 2017.
- [29] C. B. Saner, M. Kesici, Y. Yaslan, and V. I. Genc, "Improving the performance of transient stability prediction using resampling methods," in *IEEE International Conference on Electrical and Electronics Engineering*, 2019, pp. 146–150.
- [30] Z. Chen, X. Han, C. Fan, T. Zheng, and S. Mei, "A two-stage feature selection method for power system transient stability status prediction," *Energies*, 2019.
- [31] L. Moulin, A. A. Da Silva, M. El-Sharkawi, and R. J. Marks, "Support vector machines for transient stability analysis of large-scale power systems," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 818–825, 2004.
- [32] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1474–1483, 2010.
- [33] C. Liu, F. Tang, and C. Leth Bak, "An accurate online dynamic security assessment scheme based on random forest," *Energies*, vol. 11, no. 7, p. 1914, 2018.
- [34] H. Fan, Y. Chen, S. Huang, X. Zhang, H. Guan, and D. Shi, "Post-fault transient stability assessment based on k-nearest neighbor algorithm with mahalanobis distance," in *2018 International Conference on Power System Technology (POWERCON)*. IEEE, 2018, pp. 4417–4423.

- [35] Y. Zhou, Q. Guo, H. Sun, Z. Yu, J. Wu, and L. Hao, "A novel data-driven approach for transient stability prediction of power systems considering the operational variability," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 379–394, 2019.
- [36] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1935–1943, 2007.
- [37] V. Krishnan, J. D. McCalley, S. Henry, and S. Issad, "Efficient database generation for decision tree based power system security assessment," *IEEE Transactions on Power systems*, vol. 26, no. 4, pp. 2319–2327, 2011.
- [38] J. L. Cremer, I. Konstantelos, and G. Strbac, "From optimization-based machine learning to interpretable security rules for operation," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3826–3836, 2019.
- [39] C. Liu, K. Sun, Z. H. Rather, Z. Chen, C. L. Bak, P. Thøgersen, and P. Lund, "A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 717–730, 2014.
- [40] M. Sun, I. Konstantelos, and G. Strbac, "A deep learning-based feature extraction framework for system security assessment," *IEEE transactions on smart grid*, vol. 10, no. 5, pp. 5007–5020, 2018.
- [41] D. J. Hand, "Principles of data mining," *Drug Safety*, vol. 30, no. 7, pp. 621–622, 2007.
- [42] I. Konstantelos, G. Jamgotchian, S. H. Tindemans, P. Duchesne, S. Cole, C. Merckx, G. Strbac, and P. Panciatici, "Implementation of a massively parallel dynamic security assessment platform for large-scale grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1417–1426, 2017.
- [43] L. Zhu and D. J. Hill, "Data/model jointly driven high-quality case generation for power system dynamic stability assessment," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5055–5066, 2021.

- [44] G. Wang, J. Guo, S. Ma, X. Zhang, Q. Guo, S. Fan, and H. Xu, “Data-driven transient stability assessment with sparse pmu sampling and online self-check function,” *CSEE Journal of Power and Energy Systems*, 2022.
- [45] C. Ren and Y. Xu, “A fully data-driven method based on generative adversarial networks for power system dynamic security assessment with missing data,” *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5044–5052, 2019.
- [46] D. P. Kroese, T. Taimre, and Z. I. Botev, “Handbook of monte carlo methods,” *John Wiley & Sons*, vol. 706, 2013.
- [47] T.-e. Huang, Q. Guo, and H. Sun, “A distributed computing platform supporting power system security knowledge discovery based on online simulation,” *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1513–1524, 2016.
- [48] C. Hamon, M. Perninge, and L. Söder, “An importance sampling technique for probabilistic security assessment in power systems with large amounts of wind power,” *Electric power systems research*, vol. 131, pp. 11–18, 2016.
- [49] I. Genc, R. Diao, V. Vittal, S. Kolluri, and S. Mandal, “Decision tree-based preventive and corrective control applications for dynamic security enhancement in power systems,” *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1611–1619, 2010.
- [50] L. Zhu, D. J. Hill, and C. Lu, “Semi-supervised ensemble learning framework for accelerating power system transient stability knowledge base generation,” *IEEE Transactions on Power Systems*, vol. 37, no. 3, pp. 2441–2454, 2021.
- [51] R. Yan, G. Geng, and Q. Jiang, “Data-driven transient stability boundary generation for online security monitoring,” *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3042–3052, 2020.
- [52] J. Stiasny, S. Chevalier, R. Nellikkath, B. Sævarsson, and S. Chatzivasileiadis, “Closing the loop: A framework for trustworthy machine learning in power systems,” *arXiv preprint arXiv:2203.07505*, 2022.

- [53] S. Jafarzadeh and V. M. I. Genc, “Probabilistic dynamic security assessment of large power systems using machine learning algorithms,” *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 26, no. 3, pp. 1479–1490, 2018.
- [54] Q. Wu, T. J. Koo, and Y. Susuki, “Dynamic security analysis of power systems by a sampling-based algorithm,” *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 2, pp. 1–26, 2018.
- [55] F. Thams, A. Venzke, R. Eriksson, and S. Chatzivasileiadis, “Efficient database generation for data-driven security assessment of power systems,” *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 30–41, 2020.
- [56] A. Venzke, D. K. Molzahn, and S. Chatzivasileiadis, “Efficient creation of datasets for data-driven power system applications,” *Electric Power Systems Research*, vol. 190, p. 106614, 2021.
- [57] T. Joswig-Jones, K. Baker, and A. S. Zamzam, “Opf-learn: An open-source framework for creating representative ac optimal power flow datasets,” in *2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2022, pp. 1–5.
- [58] R. Mukherjee and A. De, “Real-time dynamic security analysis of power systems using strategic pmu measurements and decision tree classification,” *Electrical Engineering*, vol. 103, no. 2, pp. 813–824, 2021.
- [59] A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, “Split-based sequential sampling for real-time security assessment,” *International Journal of Electrical Power & Energy Systems*, vol. 146, p. 108790, 2023.
- [60] J. A. Swets, “Measuring the accuracy of diagnostic systems,” *Science*, vol. 240, no. 4857, pp. 1285–1293, 1988.
- [61] F. Provost and T. Fawcett, “Analysis and visualization of classifier performance with nonuniform class and cost distributions,” in *Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection & Risk Management*, 1997.

- [62] M. Kubat, S. Matwin *et al.*, “Addressing the curse of imbalanced training sets: one-sided selection,” in *Icml*, vol. 97. Nashville, USA, 1997, pp. 179–186.
- [63] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on knowledge and data engineering*, 2009.
- [64] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [65] G. Menardi and N. Torelli, “Training and assessing classification rules with imbalanced data,” *Data Mining and Knowledge Discovery*, 2014.
- [66] D. Hand and P. Christen, “A note on using the f-measure for evaluating record linkage algorithms,” *Statistics and Computing*, vol. 28, no. 3, pp. 539–547, 2018.
- [67] F. Provost and T. Fawcett, “Robust classification for imprecise environments,” *Machine learning*, vol. 42, no. 3, pp. 203–231, 2001.
- [68] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233–240.
- [69] C. Drummond and R. C. Holte, “Cost curves: An improved method for visualizing classifier performance,” *Machine learning*, vol. 65, no. 1, pp. 95–130, 2006.
- [70] G. Granelli, M. Montagna, F. Zanellini, P. Bresesti, R. Vailati, and M. Innorta, “Optimal network reconfiguration for congestion management by deterministic and genetic algorithms,” *Electric power systems research*, vol. 76, no. 6-7, pp. 549–556, 2006.
- [71] G. Schnyder and H. Glavitsch, “Integrated security control using an optimal power flow and switching concepts,” *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 782–790, 1988.
- [72] R. Bacher and H. Glavitsch, “Loss reduction by network switching,” *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 447–454, 1988.

- [73] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, "Optimal transmission switching," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1346–1355, 2008.
- [74] K. W. Hedman, R. P. O'Neill, E. B. Fisher, and S. S. Oren, "Optimal transmission switching—sensitivity analysis and extensions," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1469–1479, 2008.
- [75] —, "Optimal transmission switching with contingency analysis," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1577–1586, 2009.
- [76] C. Crozier, K. Baker, and B. Toomey, "Feasible region-based heuristics for optimal transmission switching," *Sustainable Energy, Grids and Networks*, p. 100628, 2022.
- [77] L. L. Grigsby, *Power system stability and control*. CRC press, 2007.
- [78] A. Fouad, F. Aboytes, V. Carvalho, S. Corey, K. Dhir, and R. Vierra, "Dynamic security assessment practices in north america," *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 1310–1321, 1988.
- [79] F. Capitanescu, J. M. Ramos, P. Panciatici, D. Kirschen, A. M. 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.
- [80] V. D. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, no. 9, pp. 1249–1274, 2000.
- [81] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziairgyriou, D. Hill, A. Stankovic, C. Taylor *et al.*, "Definition and classification of power system stability {IEEE/CIGRE} joint task force on stability terms and definitions," *IEEE transactions on Power Systems*, vol. 19, no. 3, pp. 1387–1401, 2004.
- [82] H.-D. Chiang, "Power system stability," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2001.

- [83] J. M. G. Alvarez, “Critical contingencies ranking for dynamic security assessment using neural networks,” in *2009 15th International Conference on Intelligent System Applications to Power Systems*. IEEE, 2009, pp. 1–6.
- [84] F. Capitanescu, “Critical review of recent advances and further developments needed in ac optimal power flow,” *Electric Power Systems Research*, vol. 136, pp. 57–68, 2016.
- [85] P. W. Sauer, K. L. Tomsovic, and V. Vittal, “Dynamic security assessment,” *Power system stability and control*, vol. 5, pp. 421–430, 2007.
- [86] L. Vanfretti and F. R. S. Sevilla, “A three-layer severity index for power system voltage stability assessment using time-series from dynamic simulations,” in *IEEE PES Innovative Smart Grid Technologies, Europe*, 2014, pp. 1–5.
- [87] Y. Pipelzadeh, R. Moreno, B. Chaudhuri, G. Strbac, and T. C. Green, “Corrective control with transient assistive measures: Value assessment for great britain transmission system,” *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1638–1650, 2016.
- [88] M. He, J. Zhang, and V. Vittal, “Robust online dynamic security assessment using adaptive ensemble decision-tree learning,” *IEEE Transactions on Power systems*, vol. 28, no. 4, pp. 4089–4098, 2013.
- [89] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, “Power system transient stability assessment based on big data and the core vector machine,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2561–2570, 2016.
- [90] J. James, D. J. Hill, A. Y. Lam, J. Gu, and V. O. Li, “Intelligent time-adaptive transient stability assessment system,” *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1049–1058, 2017.
- [91] Y. Zhang, Y. Xu, Z. Y. Dong, and R. Zhang, “A hierarchical self-adaptive data-analytics method for real-time power system short-term voltage stability assessment,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 74–84, 2018.

- [92] Y. Xu, Y. Dai, Z. Y. Dong, R. Zhang, and K. Meng, "Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems," *Neural Computing and Applications*, vol. 22, no. 3, pp. 501–508, 2013.
- [93] Q. Wang, F. Li, Y. Tang, and Y. Xu, "Integrating model-driven and data-driven methods for power system frequency stability assessment and control," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4557–4568, 2019.
- [94] T. Zhang, M. Sun, J. L. Cremer, N. Zhang, G. Strbac, and C. Kang, "A confidence-aware machine learning framework for dynamic security assessment," *IEEE Transactions on Power Systems*, 2021.
- [95] A.-A. B. Bugaje, J. L. Cremer, M. Sun, and G. Strbac, "Selecting decision trees for power system security assessment," *Energy and AI*, vol. 6, p. 100110, 2021.
- [96] J. L. Cremer and G. Strbac, "A machine-learning based probabilistic perspective on dynamic security assessment," *International Journal of Electrical Power & Energy Systems*, vol. 128, p. 106571, 2021.
- [97] Y. Liu, J. Wang, and Z. Yue, "Improved multi-point estimation method based probabilistic transient stability assessment for power system with wind power," *International Journal of Electrical Power & Energy Systems*, vol. 142, p. 108283, 2022.
- [98] P. N. Papadopoulos and J. V. Milanović, "Probabilistic framework for transient stability assessment of power systems with high penetration of renewable generation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3078–3088, 2016.
- [99] F. Bellizio, J. L. Cremer, M. Sun, and G. Strbac, "A causality based feature selection approach for data-driven dynamic security assessment," *Electric Power Systems Research*, vol. 201, p. 107537, 2021.
- [100] Y. Liu, X.-J. Shi, and Y. Xu, "A hybrid data-driven method for fast approximation of practical dynamic security region boundary of power systems," *International Journal of Electrical Power & Energy Systems*, vol. 117, p. 105658, 2020.

- [101] F. Bellizio, A.-A. B. Bugaje, J. L. Cremer, and G. Strbac, "Verifying machine learning conclusions for securing low inertia systems," *Sustainable Energy, Grids and Networks*, vol. 30, p. 100656, 2022.
- [102] F. R. S. Sevilla, Y. Liu, E. Barocio, P. Korba, M. Andrade, F. Bellizio, J. Bos, B. Chaudhuri, H. Chavez, J. Cremer *et al.*, "State-of-the-art of data collection, analytics, and future needs of transmission utilities worldwide to account for the continuous growth of sensing data," *International journal of electrical power & energy systems*, vol. 137, p. 107772, 2022.
- [103] M. He, J. Zhang, and V. Vittal, "A data mining framework for online dynamic security assessment: Decision trees, boosting, and complexity analysis," in *IEEE PES Innovative Smart Grid Technologies*, 2012, pp. 1–8.
- [104] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, 2002.
- [105] J. L. Cremer, I. Konstantelos, G. Strbac, and S. H. Tindemans, "Sample-derived disjunctive rules for secure power system operation," in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018, pp. 1–6.
- [106] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," *International Journal of Machine Learning Technology* 2:1, pp. 37–63, 2011.
- [107] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233–240.
- [108] T. Fawcett, "An introduction to {ROC} analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [109] M. Sadeghi, M. A. Sadeghi, S. Nourizadeh, A. M. Ranjbar, and S. Azizi, "Power system security assessment using adaboost algorithm," in *Proceedings North American Power Symposium*, 2009.

- [110] R. I. Hamilton and P. N. Papadopoulos, “Using shap values and machine learning to understand trends in the transient stability limit,” *IEEE Transactions on Power Systems*, 2023.
- [111] F. Xiao, J. D. McCalley, Y. Ou, J. Adams, and S. Myers, “Contingency probability estimation using weather and geographical data for on-line security assessment,” in *2006 International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2006, pp. 1–7.
- [112] R. Z. Fanucchi, M. Bessani, M. H. M. Camillo, J. B. A. London, and C. D. Maciel, “Failure rate prediction under adverse weather conditions in an electric distribution system using negative binomial regression,” in *International Conference on Harmonics and Quality of Power*, 2016, pp. 478–483.
- [113] M. Cui, J. Wang, J. Tan, A. R. Florita, and Y. Zhang, “A novel event detection method using pmu data with high precision,” *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 454–466, 2019.
- [114] J. Zhao, M. Netto, Z. Huang, S. Yu, A. Gomez-Exposito, S. Wang, I. Kamwa, S. Akhlaghi, L. Mili, V. Terzija, A. P. S. Meliopoulos, B. C. Pal, A. K. Singh, A. Abur, T. Bi, and A. Rouhani, “Roles of dynamic state estimation in power system modeling, monitoring and operation,” *IEEE Transactions on Power Systems*, pp. 1–1, 2020.
- [115] D. E. Kaufman and R. L. Smith, “Direction choice for accelerated convergence in hit-and-run sampling,” *Operations Research*, vol. 46, no. 1, pp. 84–95, 1998.
- [116] H. S. Haraldsdóttir, B. Cousins, I. Thiele, R. M. Fleming, and S. Vempala, “Chrr: coordinate hit-and-run with rounding for uniform sampling of constraint-based models,” *Bioinformatics*, vol. 33, no. 11, pp. 1741–1743, 2017.
- [117] P. A. Saa and L. K. Nielsen, “ll-achrb: a scalable algorithm for sampling the feasible solution space of metabolic networks,” *Bioinformatics*, vol. 32, no. 15, pp. 2330–2337, 2016.

- [118] T. C. Keaty and P. A. Jensen, “Gapsplit: Efficient random sampling for non-convex constraint-based models,” *Bioinformatics*, vol. 36, no. 8, pp. 2623–2625, 2020.
- [119] M. L. Jenior, T. J. Moutinho Jr, B. V. Dougherty, and J. A. Papin, “Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments,” *PLOS Computational Biology*, vol. 16, no. 4, 2020.
- [120] I. V. Nadal and S. Chevalier, “Optimization-based exploration of the feasible power flow space for rapid data collection,” *arXiv preprint arXiv:2206.12214*, 2022.
- [121] M. Gunzburger and J. Burkardt, “Uniformity measures for point sample in hypercubes,” *Rapp. tech. Florida State University (cf. p. 73)*, 2004.
- [122] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, “On the surprising behavior of distance metrics in high dimensional space,” *International Conference on Database Theory*, pp. 420–434, 2001.
- [123] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.
- [124] A. Zimek, E. Schubert, and H.-P. Kriegel, “A survey on unsupervised outlier detection in high-dimensional numerical data,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 5, pp. 363–387, 2012.
- [125] H. He, Y. Bai, E. A. Garcia, and S. Li, “Adasyn: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 2008, pp. 1322–1328.
- [126] D. K. Molzahn and I. A. Hiskens, “A survey of relaxations and approximations of the power flow equations,” 2019.
- [127] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, “The quickhull algorithm for convex hulls,” *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, 1996.
- [128] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

- [129] P. Stein, "A note on the volume of a simplex," *The American Mathematical Monthly*, vol. 73, no. 3, pp. 299–301, 1966. [Online]. Available: <http://www.jstor.org/stable/2315353>
- [130] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*. John Wiley & Sons, 2013.
- [131] B. Pal and B. Chaudhuri, *Robust control in power systems*. Springer Science & Business Media, 2006.
- [132] (2013) Illinois institute of technology (IIT), IEEE 118-bus system data. [Online]. Available: <http://motor.ece.iit.edu/Data/>
- [133] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.
- [134] 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," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
- [135] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees," *International Group*, vol. 432, pp. 151–166, 1984.
- [136] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
- [137] W. E. Hart, C. D. Laird, J.-P. Watson, D. L. Woodruff, G. A. Hackebeil, B. L. Nicholson, J. D. Sirola *et al.*, "Pyomo-optimization modeling in python," vol. 67, 2017.
- [138] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.

- [139] J. D. Lara, C. Barrows, D. Thom, D. Krishnamurthy, and D. Callaway, “Powersys-tems. jl—a power system data management package for large scale modeling,” *SoftwareX*, vol. 15, p. 100747, 2021.
- [140] R. Henriquez-Auba, J. D. Lara, D. S. Callaway, and C. Barrows, “Transient simulations with a large penetration of converter-interfaced generation: Scientific computing challenges and opportunities,” *IEEE Electrification Magazine*, vol. 9, no. 2, pp. 72–82, 2021.
- [141] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward, “Sundials: Suite of nonlinear and differential/algebraic equation solvers,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 31, no. 3, pp. 363–396, 2005.
- [142] E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, “Approximate bayesian computation with the wasserstein distance,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 81, no. 2, pp. 235–269, 2019.
- [143] C. Villani, *Optimal transport: old and new*. Springer, 2009, vol. 338.
- [144] B. Jiang, “Approximate bayesian computation with kullback-leibler divergence as data discrepancy,” in *International conference on artificial intelligence and statistics*. PMLR, 2018, pp. 1711–1721.
- [145] M. G. Bellemare, I. Danihelka, W. Dabney, S. Mohamed, B. Lakshminarayanan, S. Hoyer, and R. Munos, “The cramer distance as a solution to biased wasserstein gradients,” *arXiv preprint arXiv:1705.10743*, 2017.
- [146] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in *International conference on machine learning*. PMLR, 2017, pp. 214–223.
- [147] K. Nadjahi, V. De Bortoli, A. Durmus, R. Badeau, and U. Şimşekli, “Approximate bayesian computation with the sliced-wasserstein distance,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5470–5474.

- [148] S. Hagspiel, A. Papaemmannouil, M. Schmid, and G. Andersson, “Copula-based modeling of stochastic wind power in europe and implications for the swiss power grid,” *Applied energy*, vol. 96, pp. 33–44, 2012.
- [149] N. Zhang, C. Kang, C. Singh, and Q. Xia, “Copula based dependent discrete convolution for power system uncertainty analysis,” *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 5204–5205, 2016.
- [150] M. Sklar, “Fonctions de repartition an dimensions et leurs marges,” *Publ. inst. statist. univ. Paris*, vol. 8, pp. 229–231, 1959.
- [151] M. Sun, J. Cremer, and G. Strbac, “A novel data-driven scenario generation framework for transmission expansion planning with high renewable energy penetration,” *Applied Energy*, vol. 228, pp. 546–555, 2018.
- [152] . Gurobi Optimization, “Gurobi optimizer reference manual,” 2018.
- [153] A. Wächter and L. T. Biegler, “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming,” *Mathematical Programming*, vol. 106, no. 1, pp. 25–57, 2006.
- [154] D. Kirschen and D. Jayaweera, “Comparison of risk-based and deterministic security assessments,” *IET Generation, Transmission & Distribution*, vol. 1, no. 4, pp. 527–533, 2007.
- [155] J. McCalley, S. Asgarpour, L. Bertling, R. Billinion, H. Chao, J. Chen, J. Endrenyi, R. Fletcher, A. Ford, C. Grigg *et al.*, “Probabilistic security assessment for power system operations,” in *IEEE PES General Meeting*, 2004, pp. 212–220.
- [156] C. Elkan, “The foundations of cost-sensitive learning,” in *International joint conference on artificial intelligence*, vol. 17, 2001, pp. 973–978.
- [157] N. Nikolaou, N. Edakunni, M. Kull, P. Flach, and G. Brown, “Cost-sensitive boosting algorithms: Do we really need them?” *Machine Learning*, vol. 104, no. 2-3, pp. 359–384, 2016.

- [158] F. Hang, S. Huang, Y. Chen, and S. Mei, "Power system transient stability assessment based on dimension reduction and cost-sensitive ensemble learning," in *IEEE Conference on Energy Internet and Energy System Integration*. IEEE, 2017, pp. 1–6.
- [159] Y. Zhou, W. Zhao, Q. Guo, H. Sun, and L. Hao, "Transient stability assessment of power systems using cost-sensitive deep learning approach," in *IEEE Conference on Energy Internet and Energy System Integration*, 2018, pp. 1–6.
- [160] M. Khanabadi and H. Ghasemi, "Transmission congestion management through optimal transmission switching," in *2011 IEEE power and energy society general meeting*, 2011, pp. 1–5.
- [161] E. S. Johnson, S. Ahmed, S. S. Dey, and J.-P. Watson, "A k-nearest neighbor heuristic for real-time dc optimal transmission switching," *arXiv preprint arXiv:2003.10565*, 2020.
- [162] K. Lehmann, A. Grastien, and P. Van Hentenryck, "The complexity of dc-switching problems," *arXiv preprint arXiv:1411.4369*, 2014.
- [163] B. Kocuk, H. Jeon, S. S. Dey, J. Linderoth, J. Luedtke, and X. A. Sun, "A cycle-based formulation and valid inequalities for dc power transmission problems with switching," *Operations Research*, vol. 64, no. 4, pp. 922–938, 2016.
- [164] J. D. Fuller, R. Ramasra, and A. Cha, "Fast heuristics for transmission-line switching," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1377–1386, 2012.
- [165] P. A. Ruiz, J. M. Foster, A. Rudkevich, and M. C. Caramanis, "Tractable transmission topology control using sensitivity analysis," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1550–1559, 2012.
- [166] P. A. Ruiz, E. Goldis, A. M. Rudkevich, M. C. Caramanis, C. R. Philbrick, and J. M. Foster, "Security-constrained transmission topology control milp formulation using sensitivity factors," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1597–1605, 2016.

- [167] Z. Yang and S. Oren, "Line selection and algorithm selection for transmission switching by machine learning methods," in *IEEE Milan PowerTech*, 2019, pp. 1–6.
- [168] A. Papavasiliou, S. S. Oren, Z. Yang, P. Balasubramanian, and K. Hedman, "An application of high performance computing to transmission switching," in *2013 IREP Symposium Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid*, 2013, pp. 1–6.
- [169] M. Soroush and J. D. Fuller, "Accuracies of optimal transmission switching heuristics based on dcopf and acopf," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 924–932, 2013.
- [170] C. Barrows, S. Blumsack, and R. Bent, "Computationally efficient optimal transmission switching: Solution space reduction," in *IEEE Power and Energy Society General Meeting*, 2012, pp. 1–8.
- [171] P. A. Ruiz, A. Rudkevich, M. C. Caramanis, E. Goldis, E. Ntakou, and C. R. Philbrick, "Reduced mip formulation for transmission topology control," in *2012 50th annual Allerton conference on communication, control, and computing (Allerton)*. IEEE, 2012, pp. 1073–1079.
- [172] F. Bellizio, J. L. Cremer, and G. Strbac, "Machine-learned security assessment for changing system topologies," *International Journal of Electrical Power & Energy Systems*, vol. 134, p. 107380, 2022.
- [173] D. C. Marcu and C. Grava, "The impact of activation functions on training and performance of a deep neural network," in *2021 16th International Conference on Engineering of Modern Electric Systems (EMES)*. IEEE, 2021, pp. 1–4.
- [174] N. Kouiroukidis and G. Evangelidis, "The effects of dimensionality curse in high dimensional knn search," in *2011 15th Panhellenic Conference on Informatics*. IEEE, 2011, pp. 41–45.
- [175] S. A. Aleem, N. Shahid, and I. H. Naqvi, "Methodologies in power systems fault detection and diagnosis," *Energy Systems*, vol. 6, no. 1, pp. 85–108, 2015.

- [176] A. M. Khalkho and D. K. Mohanta, “Rbfnn assisted transient instability-based risk assessment of cyber-physical power system,” *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107787, 2022.
- [177] H. Lassoued, R. Ketata, and S. Yacoub, “Ecg decision support system based on feedforward neural networks,” *International Journal on Smart Sensing and Intelligent Systems*, vol. 11, no. 1, pp. 1–13, 2018.
- [178] S. Blumsack, “Network topologies and transmission investment under electric-industry restructuring,” Ph.D. dissertation, Carnegie Mellon University, 2006.
- [179] B. Huang and J. Wang, “Applications of physics-informed neural networks in power systems-a review,” *IEEE Transactions on Power Systems*, 2022.
- [180] F. Milano, F. Dörfler, G. Hug, D. J. Hill, and G. Verbič, “Foundations and challenges of low-inertia systems,” in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–25.
- [181] J. Fang, H. Li, Y. Tang, and F. Blaabjerg, “On the inertia of future more-electronics power systems,” *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 7, no. 4, pp. 2130–2146, 2018.
- [182] P. N. Papadopoulos and J. V. Milanović, “Methodology for online identification of dynamic behavior of power systems with an increased amount of power electronics interface units,” *CSEE Journal of Power and Energy Systems*, vol. 5, no. 2, pp. 171–180, 2019.
- [183] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, “Physics-informed neural networks for power systems,” in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.
- [184] X. Jia, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, and V. Kumar, “Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles,” *ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–26, 2021.

- [185] W. Liao, B. Bak-Jensen, J. R. Pillai, Y. Wang, and Y. Wang, “A review of graph neural networks and their applications in power systems,” *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 345–360, 2021.
- [186] Y. Chen, S. M. Mazhari, C. Chung, S. O. Faried, and B. C. Pal, “Rotor angle stability prediction of power systems with high wind power penetration using a stability index vector,” *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4632–4643, 2020.
- [187] J. Romero-Ruiz, J. Pérez-Ruiz, S. Martin, J. Aguado, and S. De la Torre, “Probabilistic congestion management using evs in a smart grid with intermittent renewable generation,” *Electric Power Systems Research*, vol. 137, pp. 155–162, 2016.
- [188] F. Bellizio, J. L. Cremer, and G. Strbac, “Transient stable corrective control using neural lyapunov learning,” *IEEE Transactions on Power Systems*, 2022.
- [189] Y. Wei, A.-A. B. Bugaje, F. Bellizio, and G. Strbac, “Reinforcement learning based optimal load shedding for transient stabilization,” in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 2022, pp. 1–5.
- [190] A. Blatiak, F. Bellizio, L. Badesa, and G. Strbac, “Value of optimal trip and charging scheduling of commercial electric vehicle fleets with vehicle-to-grid in future low inertia systems,” *Sustainable Energy, Grids and Networks*, vol. 31, p. 100738, 2022.
- [191] D. Zheng, C. Ma, M. Wang, J. Zhou, Q. Su, X. Song, Q. Gan, Z. Zhang, and G. Karypis, “Distdgl: distributed graph neural network training for billion-scale graphs,” in *2020 IEEE/ACM 10th Workshop on Irregular Applications: Architectures and Algorithms (IA3)*. IEEE, 2020, pp. 36–44.
- [192] J. Silva, J. Sumaili, R. J. Bessa, L. Seca, M. Matos, and V. Miranda, “The challenges of estimating the impact of distributed energy resources flexibility on the tso/dso boundary node operating points,” *Computers & Operations Research*, vol. 96, pp. 294–304, 2018.