Universitat de Girona

DATA-DRIVEN APPROACHES FOR EVENT DETECTION, FAULT LOCATION, RESILIENCE ASSESSMENT, AND ENHANCEMENTS IN POWER SYSTEMS

Laiz Souto

Per citar o enllaçar aquest document:
Para citar o enlazar este documento:
Use this url to cite or link to this publication:
http://hdl.handle.net/10803/671402

http://creativecommons.org/licenses/by-nc/4.0/deed.ca

Aquesta obra està subjecta a una llicència Creative Commons Reconeixement-NoComercial

Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial

This work is licensed under a Creative Commons Attribution-NonCommercial licence

Universitat de Girona

Doctoral Thesis

Data-Driven Approaches for Event Detection, Fault Location, Resilience Assessment, and Enhancements in Power Systems

Laiz Souto

2020

DOCTORAL THESIS

Data-Driven Approaches for Event Detection, Fault Location, Resilience Assessment, and Enhancements in Power Systems

Laiz Souto

2020

DOCTORAL PROGRAM IN TECHNOLOGY

Supervised by:
Dr. Joaquim Melendez

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Certificat de direcció de tesi

Universitat de Girona

El Prof. Dr. Joaquim Meléndez i Frigola, de la Universitat de Girona,

DECLARO:

Que el treball titulat *Data-driven approaches for event detection, fault location, resilience assessment, and enhancements in power systems*, que presenta Laiz Souto per a l'obtenció del títol de doctora, ha estat realitzat sota la meva direcció i que compleix els requisits per al dipòsit.
I, perquè així consti i tingui els efectes oportuns, signo aquest document.

Signatura:

Joaquim Meléndez Frigola
Girona, 12/11/2020

Certificat de direcció de tesi

Universitat de Girona

El Prof. Dr. Joaquim Meléndez i Frigola, de la Universitat de Girona,

DECLARO:

Que el treball titulat *Data-driven approaches for event detection, fault location, resilience assessment, and enhancements in power systems*, que presenta Laiz Souto per a l'obtenció del títol de doctora, ha estat realitzat sota la meva direcció i que compleix els requisits per poder optar a Menció Internacional.
I, perquè així consti i tingui els efectes oportuns, signo aquest document.

Signatura

To my parents, my brother,
and my extended family overseas.

Acknowledgements

This thesis bears my name alone, but is made possible by a number of people whom I would like to thank. Those people are from the academic world and funding sources, besides my family and friends.

From the academic world, I would like to acknowledge everybody who assisted me during my doctoral program at the Universitat de Girona (UdG) and during my visit at the University of Texas at Austin (UT-Austin). First and foremost, I would like to thank Prof. Joaquim Melendez for the opportunity to join his laboratory as a Ph.D. student at UdG and Prof. Surya Santoso for the opportunity to join his laboratory as an exchange visitor at UT-Austin. Thank you for the great opportunities to develop and grow as a researcher and for all the windows open. I would also like to thank both of them for their guidance, dedication, kindness, and time that I truly appreciated.

I would also like to give special thanks to Dr. Sergio Herraiz for his support with the event detection algorithms; to the grid resilience project team members (Prof. Erhan Kutanoglu, Prof. John Hasenbein, Prof. Zong-Liang Yang, Dr. Carey W. King, Brent Austgen, Wen-Ying Wu, Joshua Yip, Vinicius Cunha, Ashutosh Shukla, Sabiha Tabassum, and Kyoung Kim) for the great teamwork and all the interesting discussions; to Dr. James Broda for his support with Bayes methods; to the OpenLV project team for allowing me to use their energy data; to Prof. Joan Colomer, Prof. Beatriz Lopez, Roberto Petite, Begona Fernandez, Cristina Rubirola, Dr. Ferran Torrent, Pablo Gay, and Sara Murla for the support with general matters at the UdG; to my colleagues with the Control Engineering and Intelligent Systems group not previously acknowledged for all the good times (and food and beer); and to my colleagues with the Electrical and Computer Engineering department of the UT-Austin for their assistance during my stay (and coffee and snacks).

In addition, I would like to thank all reviewers for their valuable feedback. I would also like to thank the IEEE Power & Energy society, IEEE Women in Engineering, IEEE Young Professionals, CIGRE (especially my colleagues from Working Groups C3.21 and D2.52), and ESIG for all interesting materials and discussions. Your contributions not only improved substantially the quality of my thesis and its related

publications, but also helped me to find meaning in doing this research and in what is next. Thanks for being passionate and enthusiastic about scientific progress.

Regarding funding and collaborations, I hereby acknowledge that this research was supported by the University of Girona scholarship program (IFUdG2019 edition) and mobility scholarship (MOB2020 edition) and by the European Union's Horizon 2020 research and innovation framework, call LCE-01-2016-2017, under the auspices of the project "Renewable penetration levered by Efficient Low Voltage Distribution grids", grant agreement number 773715. It also benefited from discussion and collaboration with the project "Defending the Electricity Infrastructure against Extreme Weather Events, Now and in the Future", funded by The University of Texas at Austin Energy Institute. I would also like to acknowledge the organising committee of the CompSust DC 2019 and IEEE T&D Latin America 2020 for awarding me a travel grant and the CELI's emPOWER20 conference for the sponsored ticket. Thank you all for providing support and resources for me to disseminate my research to a broader audience.

On the top of this, I would like to thank my family and all my friends for their unconditional support throughout these years. Especially my parents, my brother, and my extended family, whom I missed a lot – but words are not enough to convey my gratitude, so I will keep it short. Thank you for pushing me forward, for always showing interest, and for always being there for me, despite the distance and different time zones. You are my motivation to be my very best and my fuel for the journey. Eventually, we reached this milestone together.

"Rise to the challenge."
(One of my favourite English idioms)

Related Publications

Journal articles

- Souto, L., Melendez, J. & Herraiz, S. (2020). Monitoring of Low Voltage Grids with Multilayer Principal Component Analysis. In International Journal of Electrical Power and Energy Systems, vol. 125, Feb. 2021, 106471.

- Souto, L. & Melendez, J. (2020). Statistical Detectability and Isolability of Events in PMU Data – with Extensions to Other Sources of Power System Data. Submitted to: IEEE Transactions on Power Systems.

- Souto, L. & Broda, J. (2020). Network-Informed Bayesian Approach to Minimize the Risk of Faults in Power Distribution Networks. Submitted to: International Journal of Electrical Power and Energy Systems.

- Souto, L., Yip, J., Cunha, V., Austgen, B., Wu, W.Y., Kutanoglu, E., Hasenbein, J., Yang, Z.L., King, C., & Santoso, S. (2020). Power System Resilience to Floods: Modelling, Impact Assessment, and Long-Term Mitigation Strategies. Submitted to: International Journal of Electrical Power and Energy Systems.

Conference proceedings

- Souto, L. & Santoso, S. (2020). Overhead versus Underground: Designing Power Lines for Resilient, Cost-Effective Power Distribution Networks under Windstorms. In Resilience Week 2020, Salt Lake City, UT, 2020, pp. 1-5.

- Souto, L. & Santoso, S. (2020). Evaluation of Power System Resilience Enhancements in Low-Income Neighborhoods. In 2020 IEEE PES Transmission & Distribution Conference and Exhibition Latin America (T&D-LA), Montevideo, 2020, pp. 1-5.

- Souto, L. (2020). Fault Behavior of Power Distribution Networks with Distributed Generation and Uncertainties. In 2020 IEEE PES Transmission & Distribution Conference and Exhibition Latin America (T&D-LA), Montevideo, 2020, pp. 1-5.

- Souto, L., Herraiz, S. & Melendez, J. (2020). Performance Comparison of Quantitative Methods for PMU Data Event Detection with Noisy Data. In 2020 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), The Hague, 2020, pp. 1-5

- Souto, L., Melendez, J. & Herraiz, S. (2020). Fault Location in Power Distribution Networks using Arbitrary Similarity Criteria in the Principal Component Subspace. In 2020 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 2020, pp. 1-5.

- Souto, L., Melendez, J. & Herraiz, S. (2020). Comparison of Principal Component Analysis Techniques for PMU Data Event Detection. In 2020 IEEE Power and Energy Society General Meeting, Montreal, QC, Canada, 2020, pp. 1-5.

- Souto, L., Melendez, J. & Herraiz, S. (2020). Fault Location in Low Voltage Smart Grids Based on Similarity Criteria in the Principal Component Subspace. In 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 2020, pp. 1-5.

- Souto, L., Torrent, F., Herraiz, S. & Melendez, J. (2019). Detection of Voltage Fluctuations in Low Voltage Power Distribution Networks with Principal Component Analysis. 10th Protection, Automation and Control (PAC) World Conference, Glasgow, UK, 2019, pp. 1-8.

Workshops and presentations

- Souto, L. (2020). Overhead versus Underground: Designing Power Lines for Resilient, Cost-Effective Power Distribution Networks under Windstorms. In Resilience Week 2020, Salt Lake City, UT (lightning talk).

- Souto, L. (2020). Comparison of Principal Component Analysis Techniques for PMU Data Event Detection. In 2020 IEEE Power and Energy Society General Meeting, Montreal, QC, Canada (student poster contest).

- Souto, L. (2019). Decision Support System Framework for Self-Healing and Reconfiguration in Power Distribution Networks with DER. In CompSust DC 2019, Pittsburgh, PA (lightning talk).

- Souto, L. (2019). Detection, Location, and Diagnosis of Power System Faults and Disturbances with Multivariate Statistical Analysis. In CompSust DC 2019, Pittsburgh, PA (poster presentation).
- Souto, L. (2019). Fault detection with Principal Component Analysis in Power Distribution Networks with Distributed Generation. In III Conference of Doctoral Research of the University of Girona 2019, Girona (oral presentation).
- Souto, L. (2019). Fault Detection and Location in Low Voltage Power Distribution Networks. In OpenLV meeting and the Future of Local Energy, Exeter, UK (oral presentation).

Acronyms

AC Alternating current
DER Distributed Energy Resource
DG Distributed Generation
EENS Expected Energy Not Supplied
EU European Union
FACTS Flexible AC Transmission System
FFT Fast Fourier Transform
HAND Height Above Nearest Drainage
HIFLD Homeland Infrastructure Foundation-Level Data
HILP High-Impact, Low-Probability
HV High Voltage
IED Intelligent Electronic Device
IEEE Institute of Electrical and Electronics Engineers
LV Low Voltage
MV Medium Voltage
PCA Principal Component Analysis
PMU Phasor Measurement Unit
PV Photovoltaic solar generation
p.u. per unit
RES Renewable Energy Source
ROCOF Rate of Change of Frequency
RPAD Relative Phase Angle Difference
SAIDI System Average Interruption Duration Index
SM Smart Meter
SNR Signal-to-Noise Ratio
SPE Square Prediction Error statistics
SVM Support Vector Machine
T^2 Hotelling's T^2 statistics
WAMS Wide Area Monitoring System
ZIP Impedance, Current, Power

List of Figures

List of Tables

Contents

Abstract

This thesis presents the study and development of distinct data-driven techniques to support event detection, fault location, and resilience assessment towards enhancements in power systems. It is divided in three main parts as follows. The first part investigates improvements in power system monitoring and event detection methods with focus on dimensionality reduction techniques in wide-area monitoring systems. The second part focuses on contributions to fault location tasks in power distribution networks, relying on information about the network topology and its electrical parameters for short-circuit simulations over a range of scenarios. The third part assesses enhancements in power system resilience to high-impact, low-probability events associated with extreme weather conditions and human-made attacks, relying on information about the system topology combined with simulations of representative scenarios for impact assessment and mitigation. Overall, the proposed data-driven algorithms contribute to event detection, fault location, and resilience assessment, relying on electrical measurements recorded by intelligent electronic devices, historical data of past events, and representative scenarios, together with information about the network topology, electrical parameters, and operating status. The validation of the algorithms, implemented in MATLAB, is based on computer simulations using network models implemented in OpenDSS and Simulink.

Keywords— Bayes methods, distributed generation, electric variables measurement, event detection, fault location, power distribution, power system analysis computing, power system faults, power system simulation, power transmission, probability, resilience, resiliency, smart grids, statistics, substations, uncertainty

Resum

Aquesta tesi presenta l'estudi i el desenvolupament de diferents tècniques basades en dades per recolzar les tasques de detecció d'esdeveniments, localització de falles i resiliència cap a millores en sistemes d'energia elèctrica. Els continguts es divideixen en tres parts principals descrites a continuació. La primera part investiga millores en el monitoratge de sistemes d'energia elèctrica i mètodes de detecció d'esdeveniments amb enfocament en tècniques de reducció de dimensionalitat en *wide-area monitoring systems*. La segona part se centra en contribucions a tasques de localització de falles en xarxes elèctriques de distribució, basant-se en informació sobre la topologia de la xarxa i els seus paràmetres elèctriques per a simulacions de curtcircuit en una varietat d'escenaris. La tercera part avalua millores en la resiliència de sistemes d'energia elèctrica davant esdeveniments d'alt impacte i baixa probabilitat associats amb condicions climàtiques extremes i atacs provocats per humans, basant-se en informació sobre la topologia del sistema combinada amb simulacions de escenaris representatius per a l'avaluació i mitigació de l'impacte. En general, els algoritmes proposats basats en dades contribueixen a la detecció d'esdeveniments, la localització de falles, i l'augment de la resiliència de sistemes d'energia elèctrica, basant-se en mesuraments elèctrics registrades per dispositius electrònics intel.ligents, dades històriques d'esdeveniments passats i escenaris representatius, en conjunt amb informaciò sobre la topologia de la xarxa, paràmetres elèctrics i l'estat operatiu. La validació dels algoritmes, implementats en MATLAB, es basa en simulacions computacionals utilitzant models de xarxa implementats en OpenDSS i Simulink.

Keywords— Mètodes de Bayes, generació distribuïda, mesurament de variables elèctriques, detecció d'esdeveniments, localització de falles, distribució d'energia elèctrica, anàlisi computacional de sistemes energia elèctrica, falles de sistemes energia elèctrica, simulació de sistemes d'energia elèctrica, transmissió d'energia elèctrica, probabilitat, resiliència, xarxes intel.ligents, estadística, subestacions, incertesa

Resumen

Esta tesis presenta el estudio y el desarrollo de distintas técnicas basadas en datos para respaldar las tareas de detección de eventos, localización de fallos y resiliencia hacia mejoras en sistemas de energía eléctrica. Los contenidos se dividen en tres partes principales descritas a continuación. La primera parte investiga mejoras en el monitoreo de sistemas de energía eléctrica y métodos de detección de eventos con enfoque en técnicas de reducción de dimensionalidad en *wide-area monitoring systems*. La segunda parte se centra en contribuciones a tareas de localización de fallos en redes eléctricas de distribución, basándose en información acerca de la topología de la red y sus parámetros eléctricos para simulaciones de cortocircuito en una variedad de escenarios. La tercera parte evalúa mejoras en la resiliencia de sistemas de energía eléctrica ante eventos de alto impacto y baja probabilidad asociados con condiciones climáticas extremas y ataques provocados por humanos, basándose en información sobre la topología del sistema combinada con simulaciones de escenarios representativos para la evaluación y mitigación del impacto. En general, los algoritmos propuestos basados en datos contribuyen a la detección de eventos, la localización de fallos, y el aumento de la resiliencia de sistemas de energía eléctrica, basándose en mediciones eléctricas registradas por dispositivos electrónicos inteligentes, datos históricos de eventos pasados y escenarios representativos, en conjunto con información acerca de la topología de la red, parámetros eléctricos y estado operativo. La validación de los algoritmos, implementados en MATLAB, se basa en simulaciones computacionales utilizando modelos de red implementados en OpenDSS y Simulink.

Keywords— Métodos de Bayes, generación distribuida, medición de variables eléctricas, detección de eventos, localización de fallos, distribución de energía eléctrica, análisis computacional de sistemas energía eléctrica, fallos de sistemas energía eléctrica, simulación de sistemas de energía eléctrica, transmisión de energía eléctrica, probabilidad, resiliencia, redes inteligentes, estadística, subestaciones, incertidumbre

Chapter 1

Introduction

This thesis is concerned with event detection, fault location, and power system resilience. It presents the study and development of data-driven techniques to support event detection, fault location, and resilience assessment for enhancements in the planning and operation of electric power systems. The scope is comprehensive and intends to provide good solutions for important concerns rather than perfect solutions for every single minor concern about event detection, fault location, and power system resilience. This chapter introduces the problem motivation, highlights the main contributions, and presents the outline of the thesis.

1.1 Electric power systems in transition, new problems, and promising uses of data to address them

Electric power systems are changing to address environmental concerns and comply with new regulatory frameworks in the energy sector. Environmental issues are paving the way for the energy transition and leading to the phasing-out of fossil fuels and an increasing penetration level of renewable energy sources for energy production, along with storage systems, energy systems integration, and energy efficiency initiatives. In turn, new regulatory frameworks are restructuring power systems while decentralizing operation, which brings new actors to the energy sector, such as local energy markets, energy communities, aggregators, and energy prosumers (that is, energy consumers and producers at the same time); and enables an increasing penetration level of distributed energy resources while reducing network congestion, particularly at distribution level. In this ever-changing scenario, power system planning and operation practices must be revisited accordingly to ensure a satisfactory level of quality of supply and efficient allocation of infrastructure and resources while respecting constraints related to feasibility studies and operational security.

To make matters worse, the increasing frequency and intensity of external high-impact, low-probability events, such as extreme weather events, natural disasters, and human-made attacks, are posing electricity infrastructures at a higher risk of collapse. This may lead to large power system outages with potentially catastrophic impacts to electric power infrastructures, economy, and society. To prevent this from happening and guarantee security of supply, power system planning and operation must also be resilient to extreme events of different nature.

In this scenario, enhanced observability capabilities are necessary to improve quality of supply, reduce interruption times, and optimize assets utilization in power grid infrastructures, along with improvements in power system resilience, now and in the future. Likewise, conventional power protection, automation, and control systems have to be redesigned to properly identify abnormal operating conditions of different nature (for example, caused by the intermittent nature of renewable energy sources and sudden, random, unpredictable changes in energy consumption patterns). Additionally, existing electricity infrastructures have to be upgraded to be resilient to extreme events of different nature. This is crucial for the grid to withstand these occurrences, while maintaining continuity and quality of supply and minimizing performance degradation, monetary losses, and damages to infrastructures.

Current research efforts have focused on digitalization of electric power systems with respect to monitoring, supervision, protection, and control systems, which may provide important pieces of information about the system at different voltage levels, geographical locations, and time intervals. Nonetheless, even if plenty of data are available at different locations, information related to minor events (for example, caused by small imbalances between supply and demand, grid reconfiguration, and equipment trips) is not readily available. Such events must be identified and handled properly, as they might otherwise affect quality of supply and lead to major failures. In this scenario, tools and methods enabling an online, automated, accurate detection of such occurrences, with adjustments to specific events of interest, are required to build more reliable, intelligent grids.

In this context, the application of computational intelligence to addressing event detection, fault location, and resilience assessment in power systems is appealing, as it enables to take advantage of a high processing capacity to incorporate knowledge of previous events and their characteristics. In particular, machine learning techniques may be helpful to capture relevant features and trends that support these activities, but their accuracy is strongly conditioned to the data, models, and assumptions in use. Hence, the application of machine learning techniques in these tasks has to be made with caution, otherwise it may lead to wrong results and misleading conclusions. Typical strengths and weaknesses of machine learning techniques are listed as follows in Table 1.1.

Thanks to the digitalization of electric power infrastructures, wide-area monitoring systems may also improve accuracy of event detection, fault location, and resilience assessment, relying on multiple measurements from different locations together with advanced communication and data processing capabilities. At present, however, most grids do not have a wide-area monitoring infrastructure because of the high implementation cost of

Table 1.1: Definitions – Strengths and weaknesses of machine learning techniques in power systems

Strengths	Weaknesses
Effective analysis of patterns in large amounts of data	"Garbage in, garbage out" when using bad or overly limited data
Optimization of complex systems and processes	Biases inherited in data, design, and uses
Can be integrated with physical models and other methods	Data models and patterns assumed to be persistent
	Finds correlation, not causality
	Fails to represent physical models as accurately as traditional methods

multiple measurement devices. As a result, existing applications are limited to specific locations or based on sparse estimation, which may produce inaccurate results over a range of scenarios.

Regardless of the digital capabilities of the grid, other pieces of information may be used to improve accuracy of event detection and fault location tasks and enhance power system resilience. Sources of data may include power system measurements, simulations of representative realistic scenarios, past events, and new evidences. In particular, the incorporation of expert knowledge and historical data of previous occurrences may be helpful to identify common locations and causes of power system events and to reduce the search area and select candidate locations of power system faults. In this context, an adjustable, one-size-fits-all method for event detection is desired to handle distinct events of interest in the grid. In addition, a probabilistic approach to solving the fault location problem is appealing, as it allows for prior beliefs to be updated when new evidence is available, which may improve accuracy of fault location with expandability to handle different network settings and operating conditions. Moreover, information from past events may be useful to enhance power system resilience to extreme conditions, including natural disasters, extreme weather events, and human-made attacks of different nature. This information may be integrated with power system data and models for impact assessment and mitigation.

1.2 Objectives

The main goal of this thesis is to develop novel data-driven techniques to support event detection, fault location, resilience assessment, and enhancements in power systems. To this extent, this thesis presents different approaches that are further compared in terms of advantages and disadvantages and demonstrated in case studies. Ultimately, the usage

of data to support power system engineering tasks is expected to prevent major outages and ensure quality and continuity of supply at acceptable levels.

Specific objectives of this thesis are listed as follows.

1. Propose a data-driven methodology for effective monitoring of smart grids.
2. Propose a one-size-fits-all data-driven approach for detection and diagnosis of distinct events of interest in power systems.
3. Propose a data-driven methodology for accurate fault location in smart grids, based on trends of data.
4. Propose a data-driven methodology to address uncertainties in fault location, based on power system simulations with accurate network models.
5. Propose a probabilistic method for fault location, relying on historical data of past events and information about the network topology and operating status.
6. Propose data-driven strategies to enhance long-term power system resilience to extreme events with usage of historical data from past events and realistic scenarios.
7. Demonstrate, test, and validate the proposed methods in case studies.

1.3 Outline

The rest of this thesis is structured as follows. The main theoretical contents are presented in Chapters 2 to 4, together with case studies, results, and discussions. Chapter 2 is devoted to event detection in power systems, whereas Chapter 3 is concerned about fault location, and Chapter 4 is concerned about power system resilience to extreme events. Finally, Chapter 5 presents the conclusions of the thesis and future perspectives, followed by the bibliography.

Chapter 2

Event Detection in Power Systems

This chapter describes state-of-the-art event detection techniques and presents the data-driven methods for event detection in power systems proposed in the articles derived from this thesis, with applications to modelling and monitoring tasks. The case studies feature the implementation of the proposed methods and comparisons between different methods.

2.1 Definitions

In this chapter, an event of interest may refer to any deviations from the normal operating conditions at different locations and voltage levels. Event detection is defined as the identification of an abnormal operating condition in the data, whereas event isolation (or diagnosis) is the identification of the variables involved in an abnormal operating condition detected (in terms of correlation and not causation). In this context, event detection and isolation methods are expected to contribute to the visualization and identification of patterns, outliers, and abnormal behaviours in huge amounts of data at relevant scales. For illustration, Fig. 2.1 (adapted from [82]) displays typical time ranges associated with distinct power system phenomena of interest.

Data-driven methods for event detection are particularly relevant to identify minor events (for example, caused by small imbalances between supply and demand and equipment tripping) that would be missed otherwise when screening for information in large amounts of data. Such events must be handled properly, as they may affect quality of supply and lead to major failures. Conversely, the contributions of data-driven methods to the detection of major events (for example, associated with specific protection functions) are limited. In this case, the protection devices act immediately to interrupt the event and keep the affected parts of the grid de-energized until the occurrence is cleared. For more information, refer to [5] for a complete list of standard protection devices and function numbers.

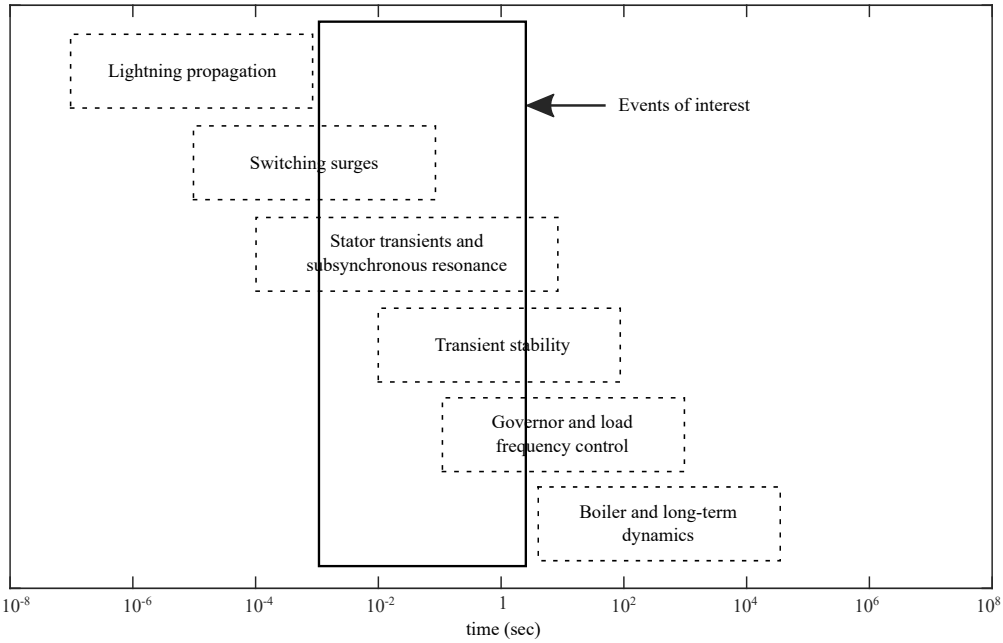

Figure 2.1: Definitions – Time ranges of distinct power system phenomena

2.2 State-of-the art methods

Event detection and diagnosis techniques used in power systems can be classified as model-based, signal-based, knowledge-based, and hybrid approaches [38, 39], described as follows.

- Model-based techniques require a model of the system or process for evaluation of the consistency between the measured outputs and the model predicted outputs. They can be categorized as deterministic (e.g., solvable with linear matrix inequalities) and stochastic (e.g., Kalman filters).

- Signal-based techniques rely on measured signals for feature extraction in either or both time domain and frequency domain. Time-domain features include continuous monitoring of average, standard deviation, phases, slope, peak, and root mean square magnitudes (e.g., min-max difference, differential function), whereas frequency-domain features include spectrum analysis (e.g., discrete Fourier transform, Yule-Walker spectral method) [10].

- Knowledge-based methods rely on large amounts of historical data for extraction of the underlying knowledge of a given process and can be categorized as quantitative or qualitative. Qualitative methods (e.g., expert systems, qualitative trend analysis) are based on a set of rules, whereas quantitative methods rely on pattern recognition and can be statistical (e.g., principal component analysis, independent component analysis, partial least squares, support vector machines) or non-statistical (e.g., neural networks, fuzzy logic).

- Hybrid approaches consist of a combination of the strategies previously described.

The usage of signal-based and statistical quantitative knowledge-based methods is advised when the monitored process is non-stationary and requests adaptation to changing operating conditions. Model-based techniques require an accurate input-output description which is not always available, whereas qualitative and non-statistical quantitative knowledge-based methods are process specific and computationally more expensive than statistical quantitative methods to be re-trained. Given the dynamic nature of electric power systems, this thesis investigates signal-based and statistical quantitative methods for event detection. Signal-based and statistical quantitative methods commonly used for event detection are described as follows.

2.2.1 Fast Fourier transform

The FFT algorithm computes the discrete Fourier transform of a signal to quantify its frequency content. Essentially, it transforms a sequence $\{x_n\} := x_0, x_1, \cdots, x_{N-1}$ of complex numbers into another sequence of complex numbers $\{X_k\} := X_0, X_1, \cdots, X_{N-1}$ defined by (2.1):

$$X_k = \sum_{n=0}^{N-1} x_n . \exp -\frac{i.2.\pi}{N} k.n = \sum_{n=0}^{N-1} x_n . \left[\cos\left(\frac{2\pi}{N} k.n\right) - i. \sin\left(\frac{2\pi}{N} k.n\right)\right]. \tag{2.1}$$

For an in-depth explanation, refer to [23].

2.2.2 Yule-Walker Spectral method

The Yule-Walker Spectral method computes the power spectral density of a signal using the auto-regressive Yule-Walker method. An auto-regressive model of order $n \in \mathbb{N}$ is given by (2.2), where $\varphi_i \in \mathbb{R}$ are the auto-regression parameters and $\epsilon_k \in \mathbb{R}_+$ is the error:

$$X_k = \sum_{i=1}^{n} \varphi_i X_{k-i} + \epsilon_k. \tag{2.2}$$

The Yule-Walker equations are given by (2.3), where $\gamma_m \in \mathbb{R}$ is the auto-covariance of X_k and $\sigma_\epsilon \in \mathbb{R}_+$ is the standard deviation of the input noise, with $m = 0, \cdots, p \in \mathbb{N}$:

$$\begin{cases} \gamma_m = \sum_{i=1}^{p} \varphi_i \gamma_{m-i}, i > 0 \\ \gamma_m = \sum_{i=1}^{p} \varphi_i \gamma_{m-i} + \sigma_\epsilon^2, i = 0 \end{cases} . \tag{2.3}$$

For a thorough explanation, see [106].

2.2.3 Min-Max difference

This method computes the difference between maximum and minimum values of the signal within the time window. Thereby, for a sequence of $N \in \mathbb{N}$ samples $\{x_n\} := x_1, \cdots, x_N$, define $\overline{x} = \max x_i \in \{x_n\}$, $\underline{x} = \min x_i \in \{x_n\}$, $\forall i \in \{1, \cdots, N\}$, and compute the min-max difference with $\overline{x} - \underline{x}$. For more details, see [10].

2.2.4 Difference and approximate derivative

This method computes the maximum difference between consecutive samples of the signal within the time window. The calculation is analogous to the ROCOF; for a sequence of $N \in \mathbb{N}$ samples $\{x_n\} := x_1, \cdots, x_N$, the maximum difference is given by $\max(x_i - x_{i-1})$, $\forall i \in \{2, \cdots, N\}$. For more details, refer to [10].

2.2.5 Support Vector Machines

This method finds the best hyperplane that separates all data points representing normal operation from those representing abnormal operation and detects anomalies with use of this classification hyperplane. Thereby, given a training data set of $N \in \mathbb{N}$ points $(x_1, y_1), \cdots, (x_N, y_N)$, where $y_i = \{-1, 1\}$, $\forall i = 1, \cdots, N$, indicates the class which the m-dimensional point $\vec{x}_i \in \mathbb{R}^m$ belongs to, the goal is to find the maximum-margin hyperplane that divides the points for which $y_i = -1$ from those for which $y_i = 1$. With a standardized data set, these hyperplanes can be described by (2.4):

$$\begin{cases} \vec{w}\vec{x} - b = 1 \ \ \text{(anything on or above this boundary belongs to label 1)} \\ \vec{w}\vec{x} - b = -1 \ \ \text{(anything on or below this boundary belongs to label -1)} \end{cases} . \tag{2.4}$$

These constraints can be written as (2.5):

$$\begin{cases} \vec{w}\vec{x} - b \geq 1 \ \ \text{(if y=1)} \\ \vec{w}\vec{x} - b \leq -1 \ \ \text{(if y=-1)} \end{cases} \rightarrow y_i(\vec{w}\vec{x}_i - b) \geq 1. \tag{2.5}$$

The distance between these two hyperplanes is $\frac{2}{\|\vec{w}\|}$. Thereby, to maximize this distance, it is necessary to minimize $\vec{w} \in \mathbb{R}^m$ subject to (2.5).

If the data do not allow for a separating hyperplane, Kernel functions and soft margins can be used to separate as many data points as possible. For a thorough explanation, refer to [90].

2.2.6 Principal Component Analysis

This method builds a statistical model of the data with dimensionality reduction and detects anomalies by projecting the subsequent data onto the projection subspace. The methodology is introduced as follows, considering linear approximations. If the relations describing the system are non-linear, Kernel functions can be used to separate the projection subspace from the residual subspace [83]. For an in-depth explanation, see [71].

Building the PCA model

Let $\mathbf{X} \in \mathbb{R}^{n \times m}$ be the $n \times m$ observation matrix displayed in (2.6) with $n \in \mathbb{N}$ observations – referred to the number of samples of phasor quantities – and $m \in \mathbb{N}$ variables – referred to PMU locations – supposed to be centred (zero mean) and scaled (unit variance):

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix}. \tag{2.6}$$

Then, compute the covariance matrix of $\mathbf{X}$ and apply eigenvalue decomposition to obtain two $m \times m$ matrices $\mathbf{V} \in \mathbb{R}^{m \times m}$ (whose columns are the eigenvectors and contain the principal components) and $\mathbf{\Lambda} \in \mathbb{R}^{m \times m}$ (diagonal matrix whose elements express the variability in the direction of each principal component or column of $\mathbf{V}$) with (2.7):

$$\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}} = \frac{1}{n-1}\mathbf{X}^{\mathbf{T}}\mathbf{X}. \tag{2.7}$$

Dimensionality reduction in the number of variables can be performed by retaining $r \in \mathbb{N}$ principal components of $\mathbf{V}$ ($r < m$) with the largest eigenvalues. Then, the $m \times m$ matrix $\mathbf{V}$ becomes an $m \times r$ matrix $\mathbf{P} \in \mathbb{R}^{m \times r}$ which defines a projection space of lower dimension representing the r most significant principal components.

The usage of $\mathbf{P}$ instead of $\mathbf{V}$ to transform $\mathbf{X}$ into the principal components representation space results in a projection onto a space of lower dimension in which some information contained in the original data is lost. As a matter of a fact, since $\mathbf{V}$ is a unitary matrix, the inverse operation is carried out with the transpose, i.e., $\mathbf{V}\mathbf{V}^{\mathbf{T}} = \mathbf{I}$, but $\mathbf{P}$ is not unitary, therefore $\mathbf{P}\mathbf{P}^{\mathbf{T}} \neq \mathbf{I}$. The scores and the transformation of scores into the original data with $\mathbf{P}$ can be calculated with (2.8):

$$\begin{cases} \mathbf{t} = \mathbf{x}\mathbf{P} \quad \text{and} \quad \hat{\mathbf{x}} = \mathbf{t}\mathbf{P}^{\mathbf{T}} \\ \mathbf{T} = \mathbf{X}\mathbf{P} \quad \text{and} \quad \hat{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathbf{T}} \end{cases}. \tag{2.8}$$

where $\mathbf{t} \in \mathbb{R}^r$ and $\hat{\mathbf{x}} \in \mathbb{R}^m$ denote the score and projection of a single observation $\mathbf{x} \in \mathbb{R}^m$ (top) and $\mathbf{T} \in \mathbb{R}^{n \times r}$ and $\hat{\mathbf{X}} \in \mathbb{R}^{n \times m}$ denote the score matrix and projection

matrix of the whole dataset $\mathbf{X}$ (bottom). The difference between $\mathbf{X}$ and $\hat{\mathbf{X}}$ is the residual matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times m}$ which resumes the information contained in the $m-r$ components from the residual space for each observation and can be calculated with the residual matrix $\tilde{\mathbf{P}} \in \mathbb{R}^{m \times m}$. Thereby, the complete PCA model can be described as in (2.9):

$$\tilde{\mathbf{X}} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X}\left(\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathbf{T}}\right) = \mathbf{X}\tilde{\mathbf{P}}. \tag{2.9}$$

Detection in the projection subspace and residual subspace

Event detection in the projection subspace is evaluated with the T^2 statistics. It computes a weighted distance of the projected data to the centre of the model using the scores t_i calculated for r variables and λ_i as a weight, as in (2.10), where $T_x^2 \in \mathbb{R}_+$ denotes the T^2 statistics:

$$T_x^2 = \sum_{i=1}^{r} \frac{t_i^2}{\lambda_i}. \tag{2.10}$$

For a single observation $\mathbf{x}$ whose score vector is $\mathbf{t}$, $\mathbf{T_x^2}$ is given by (2.11):

$$T_x^2 = \mathbf{t}{\mathbf{\Lambda_r}}^{-1}\mathbf{t}^{\mathbf{T}}, \tag{2.11}$$

with

$${\mathbf{\Lambda_r}}^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_r} \end{bmatrix}. \tag{2.12}$$

The statistical threshold $T_{thresh}^2 \in \mathbb{R}_+$ is calculated analytically with (2.13):

$$T_{thresh}^2 = \frac{r\left(n^2-1\right)}{n\left(n-r\right)} F_\alpha\left(r, n-r\right), \tag{2.13}$$

where $\alpha \in [0,1]$ is the confidence level and $F_\alpha\left(r, n-r\right) \in \mathbb{R}_+$ is the critical point of the Fischer-Snedecor distribution for r and $n-r$ degrees of freedom. Any result that surpasses T_{thresh}^2 is tagged as faulty for the T^2 statistics.

In turn, event detection in the residual subspace is evaluated with the SPE statistics. It evaluates the variation out of the projection space defined by the r principal components through the error component $\tilde{\mathbf{x}}$, as in (2.14), where $Q_x \in \mathbb{R}_+$ denotes the SPE statistics of an observation $\mathbf{x}$:

$$Q_x = \left(\mathbf{x} - \hat{\mathbf{x}}\right)\left(\mathbf{x} - \hat{\mathbf{x}}\right)^T = \|\tilde{\mathbf{x}}\|^2. \tag{2.14}$$

The statistical threshold $Q_{thresh} \in \mathbb{R}_+$ is calculated analytically with (2.15):

$$Q_{thresh} = \theta_1 \left[1 + \frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + \frac{h_0 \theta_2 \left(h_0 - 1\right)}{\theta_1^2} \right]^{\frac{1}{h_0}}, \tag{2.15}$$

with

$$\theta_k = \sum_{i=r+1}^{m} \lambda_i^k, k = \{1, 2, 3\} \quad \text{and} \quad h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}, \tag{2.16}$$

where $c_\alpha \in \mathbb{R}_+$ is the normal deviation for a confidence level $\alpha \in [0, 1]$. Any result that surpasses Q_{thresh} is tagged as faulty for the SPE statistics.

Selection of principal components to the projection subspace

The following subsections describe five different criteria often applied in the literature to select an appropriate value of r for a given statistical model.

Kaiser criterion In this method, r is selected such that all principal components whose eigenvalues are below the average variance are dropped from the matrix $\mathbf{P}$, in agreement with (2.17). In other words, this criterion consists in retaining all r principal components whose variance is larger than one, as $\mathbf{X}$ is a scaled matrix. This ensures that each principal component selected contains at least as many information as a single original variable in terms of variance:

$$r := \{\max \ j \in \{1, \cdots, m\} \mid \lambda_j \geq \frac{1}{m} \sum_{i=1}^{m} \lambda_i\}. \tag{2.17}$$

Automatic scree plot In this method, the eigenvalues λ_i are plotted decreasingly as a function of their element number i in the matrix $\mathbf{\Lambda}$. The chosen value of r then corresponds to the eigenvalue whose distance to the origin of the coordinate system is the shortest, in agreement with (2.18). The idea is to search for an elbow in the plot, which always displays a downward curve, from which the eigenvalues are approximately equal:

$$r := \{i \in \{1, \cdots, m\} \mid \min_i \sqrt{\lambda_i^2 + i^2}\}. \tag{2.18}$$

Explained variance In this method, a minimum percentage of the total variance $Var\,(\%) \in [0, 100]$ is previously defined and r is taken as the smallest integer satisfying (2.19):

$$r := \frac{1}{m} \sum_{i=1}^{r} \lambda_i \times 100 \geq Var\,(\%)\,. \tag{2.19}$$

Variance reconstruction error In this method, further explained in [74], the optimal value of r is determined by the minimum VRE, in agreement with (2.20), considering a faulty observation $\mathbf{x_f} \in \mathbb{R}^m$ represented by an m-dimensional unitary vector $\xi_\mathbf{i} \in \mathbb{R}^m$ multiplied by an event magnitude $f \in \mathbb{R}_+$ and the correlation matrix of reconstruction error $\mathbf{R} \in \mathbb{R}^{m \times m}$. This procedure results in the best reconstruction of the variables, as the VRE decreases monotonically in the residual subspace and increases in the projection subspace with the number of principal components, and the selection of r can be adjusted to detect specific events of interest defined by $\mathbf{x_f}$:

$$r := \{j \in \{1, \cdots, m\} \mid \min_j {\xi_\mathbf{i}}^T \mathbf{R} \xi_\mathbf{i}\}. \tag{2.20}$$

Statistical detectability In this criterion, r is chosen such that the smallest detectable events can be detected statistically in the projection subspace and residual subspace, according to Section 2.2.6, considering that $T_x^2 \geq T_{thresh}^2$ must hold so that an event can be detected with T^2 statistics and $Q_x \geq Q_{thresh}$ must hold so that an event can be detected with SPE statistics. For a faulty observation $\mathbf{x_f}$, r is selected such that (2.21) holds with a single PCA model, otherwise r is selected such that (2.22) holds with two PCA models built separately for T^2 and SPE statistics. This criterion enables adjustments to detect specific events of interest defined by $\mathbf{x_f}$ with both T^2 and SPE statistics:

$$r := \{\min\ i \in \{1, \cdots, m\} \mid T_{x_f}^2 \geq T_{thresh}^2\ and\ Q_{x_f} \geq Q_{thresh}\}, \tag{2.21}$$

or

$$r := \begin{cases} \{\min\ i \in \{1, \cdots, m\} \mid T_{x_f}^2 \geq T_{thresh}^2\} & \text{with } \mathrm{T}^2 \text{ statistics} \\ \{\min\ i \in \{1, \cdots, m\} \mid Q_{x_f} \geq Q_{thresh}\} & \text{with SPE statistics} \end{cases}. \tag{2.22}$$

This criterion considers the worst-case scenario to build one-size-fits-all PCA models for the overall grid. This choice shall be able to detect the smallest theoretical values of $T_{x_f}^2$ and Q_{x_f} computed with each individual variable possibly involved in the event for 2.21 or 2.22 and consequently presents the highest theoretical statistical detectability without requiring different PCA models for each set of variables.

Isolation in the projection subspace and residual subspace

If an event is detected with either or both T^2 and SPE statistics, analysis proceeds with isolation in the projection subspace and/or residual subspace.

Isolation in the projection subspace is performed through the contribution analysis of T^2 statistics, which consists of computing the influence of each variable x_j in the calculated value of T_x^2 that exceeds T_{thresh}^2 to identify those involved in the occurrence. Considering each score t_i of $\mathbf{t}$ in (2.11) as the contribution of the original variables weighted by the corresponding components of the i^{th} principal component, the total contribution of x_j to T_x^2 can be obtained as the sum of individual contributions of x_j to t_i written as (2.23):

$$contr_{T_x^2}(x_j) = \sum_{i=1}^{r} contr(t_i, x_j) = \sum_{i=1}^{r} \frac{t_i x_j P_{j,i}}{\lambda_i}. \tag{2.23}$$

In turn, isolation in the residual subspace is performed through the contribution analysis of SPE statistics, which consists of computing the influence of each variable x_j of $\mathbf{x}$ in the calculated values of Q_x that exceeds Q_{thresh} to identify those involved in the occurrence. From (2.14), the contribution of x_j to Q_x is given by (2.24):

$$contr_{Q_x}(x_j) = (x_j - \hat{x}_j)^2 = \tilde{x}_j^2. \tag{2.24}$$

Finally, it is possible to isolate an event in x_j if (2.23) surpasses the individual threshold $T_{j,thresh}^2$ and/or (2.24) surpasses the individual threshold $Q_{j,thresh}$. Such conditions are given in (2.25), where $T_{j,thresh}^2$ and $Q_{j,thresh}$ are determined by objective criteria (e.g., average values of (2.23) and (2.24) over the time window). The variables x_j presenting the largest relative contributions are identified as the most affected by the event:

$$contr_{T_x^2}(x_j) > T_{j,thresh}^2 \quad \text{and/or} \quad contr_{Q_x}(x_j) > Q_{j,thresh}. \tag{2.25}$$

2.2.7 Literature review

Different event detection methods have been proposed in the power system literature. Among those based in signal processing techniques, [10] suggested a combination of the Fast Fourier Transform, Yule-Walker, Matrix-pencil, and minimum-maximum difference methods to retrieve information related to minor events in PMU data related to frequency and RPAD. It conducted a joint analysis in MATLAB with use of PMU data from the Texas Synchrophasor Network and considered that real events should be identified by three or more methods simultaneously. For each selected method, peak magnitudes greater than three times the standard deviation were labelled as possible events, based on examples of real occurrences in the Texas Synchrophasor Network. However, the detection results are conservative and produce undesirable missed detections.

Power distribution networks typically rely on power quality analysers for monitoring purposes, as presented in [25, 19, 91], among others. Although the measurements are reliable, the burden is heavy when the number of monitored variables is too high, assuming a centralised monitoring and control architecture. In LV networks, monitoring infrastructures based on Internet of Things have also been used for power monitoring and

energy conservation, as in [49]. However, their implementation may be impractical in some settings due to high cost of installation and shabby communication infrastructures.

Other knowledge-based approaches designed to handle generic events in PMU data were presented in [87]-[85], among others. Notably, an event discovery and labelling framework based on matrix profile was introduced in [87]. In [61], synchrophasor measurements were used along with a forecasting-aided state estimator for feature extraction and identification plus correction of invalid PMU data to avoid wrong event detection. An unsupervised ensemble learning approach based on anomaly scores was developed in [109] for fast, scalable bad data and event detection in PMU data. In [53], a rule-based data-driven method was proposed for wide-area fault detection in power systems using synchrophasor data, relying on rules created with PMU data. In [85], a centralized algorithm was proposed to identify irregular output power from renewable sources, using PMU data from different zones to obtain sequences of short-time local outlier probabilities. However, their implementation may be impractical in some settings due to the curse of dimensionality and limited processing capabilities.

Dimensionality reduction techniques, such as PCA, are advantageous to process vast amounts of data and screen for events of interest in WAMS. These methods are appealing to handle large amounts of data collected by many devices installed at different locations, assuming centralized control, and also at a specific location, assuming decentralized control and many monitored variables [110]. Moreover, they may be used to screen for generic events whose nature is unknown [18] and to deal with data quality issues, such as missing data and spikes caused by hardware or communication issues [40]. Furthermore, as they rely on the magnitudes of the monitored variables, they can be used in combination with frequency analysis techniques, when frequency data and/or high sampling metering devices are available, or as an alternative to methods relying on the frequency spectrum of the signals otherwise.

The usage of statistical tests is acceptable to identify outliers in power system data [18, 10]. When it comes to PCA models, the combined usage of T^2 and SPE statistics for event detection is complementary and may increase the detection capability. The former measures the Mahalanobis distance of the projected data to the centre of the model and detects deviations from the average values, whereas the latter measures the square distance of the observation to the projection subspace and detects deviations that break the correlation structure of the data. Relying on those statistics, [18] identified the variables involved in generic power system disturbances and their magnitudes, [54] focused on islanding detection, [43] distinguished islanding events from system-wide disturbances in power systems with high penetration of distributed generation, and [75] classified islanding, loss-of-load, and loss-of generation events with a moving window approach, as in [15], to allow for continuous monitoring with improved situational awareness.

Although several PCA-based methods have been proposed to detect and isolate power system disturbances, none of them allows for pertinent adjustments in the principal components that ensure detection of minor events. In [100, 24, 27], the selection of principal components is based on a predefined variance threshold, whereas [27] also relies on a

clustering process to search for datasets with similar operating conditions and screen for events. In this context, [54] applies two statistical tests within a PCA-based approach for islanding detection, further extended in [75] to detect and classify islanding, loss-of-load, and loss-of-generation events in real time, using only one principal component to build the PCA model over a sliding window and adapt to the most recent operating conditions as soon as they become available. However, the information contained in the first principal component represents the steady-state part of the signal and does not ensure detection of minor events, as the information related to disturbances is often in other principal components.

To this extent, the statistical detectability of events was previously studied in [30] to minimize the variance reconstruction error and perform anomaly detection in chemical processes with use of SPE statistics, and generalized in [62] to determine unified sufficient conditions for PCA-based event detectability and isolability. In this context, an integrated approach allowing for adjustments to specific events of interest in any network topologies, monitoring infrastructures, and operating conditions, is desired.

With regards to monitoring tasks, the application of digital technology for real-time observability of LV networks is still impractical due to their inner complexity – radial topology, heterogeneous lines, high spatial density of customers, and unbalanced phases. Implementation would require a massive effort of measuring, collecting and processing the input data. This is particularly challenging when events of very different durations are considered, as a multiple time-scale resolution is necessary to properly identify distinct power system phenomena. Therefore, a strategy capable of detecting and diagnosing generic power system disturbances and abnormal behavioural patterns in a standard, coherent, coordinate way at multiple time scales is necessary as a first step to ensure reliable operation of LV smart grids. In this context, dimensionality reduction techniques may also simplify monitoring tasks, as shown in the next sections.

2.2.8 Contributions

The following articles derived from this thesis explore contributions to the state-of-the-art methods for event detection in power systems with applications to modelling and monitoring tasks.

- Souto, L., Melendez, J. & Herraiz, S. (2020). Monitoring of Low Voltage Grids with Multilayer Principal Component Analysis. In International Journal of Electrical Power and Energy Systems, vol. 125, Feb. 2021, 106471.

- Souto, L. & Melendez, J. (2020). Statistical Detectability and Isolability of Events in PMU Data – with Extensions to Other Sources of Power System Data. Submitted to: IEEE Transactions on Power Systems.

- Souto, L., Herraiz, S. & Melendez, J. (2020). Performance Comparison of Quantitative Methods for PMU Data Event Detection with Noisy Data. In 2020 IEEE

PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), The Hague, 2020, pp. 1-5

- Souto, L., Melendez, J. & Herraiz, S. (2020). Comparison of Principal Component Analysis Techniques for PMU Data Event Detection. In 2020 IEEE Power and Energy Society General Meeting, Montreal, QC, Canada, 2020, pp. 1-5.

- Souto, L., Torrent, F., Herraiz, S. & Melendez, J. (2019). Detection of Voltage Fluctuations in Low Voltage Power Distribution Networks with Principal Component Analysis. 10th Protection, Automation and Control (PAC) World Conference, Glasgow, UK, 2019, pp. 1-8.

In particular, contributions to the well-established PCA-based methods are investigated due to the dimensionality reduction in the number of monitored variables. Section 2.3 presents an example of event detection with PCA. Next, Section 2.4 further investigates the statistical detectability and isolability criterion and presents a methodology adjustable to distinct power system events of interest, whereas Section 2.5 presents comparisons made between different methods for event detection in terms of correct results. Afterwards, Section 2.6 presents an automated monitoring strategy in which the PCA models are built on multiple sliding windows and sized (in terms of length and sampling time) according to the type of phenomena to detect at each sliding window. The methodology is validated in case studies and further discussed.

2.3 Example of event detection with PCA: detecting voltage fluctuations in LV distribution networks with PCA

In this example, the PCA-based event detection method described in Section 2.2.6 is applied to detect voltage fluctuations in an LV distribution network. Thus, the methodology evaluates statistically the PMU data recorded at the substation nodes of a real-based power distribution network simulated in MATLAB. First, it computes a statistical model that represents the normal operating conditions with a reduced number of variables (projection space). Then, anomalies can be detected by projecting the subsequent observations onto the modelled space and verifying consistency of observations with respect to the normal operation model.

It is noteworthy that the reference model representing normal operating conditions did not have to be adjusted to handle changes in energy production and consumption profiles in this example, as the simulations were run in steady state. In real-world scenarios, however, the reference model is expected to be adjusted with some periodicity to handle variations in energy production and consumption profiles without detecting false negatives or missing true positives.

The power distribution network, illustrated in Fig. 2.2, represents a rural LV network in Catalonia, Spain. It consists of a primary distribution feeder with branches connecting the substation node to the customers (i.e., local energy producers and/or consumers). The distribution substation has 138.64 kW of contracted power from a 250 kVA transformer Dyn11 (400 V secondary), with industrial and residential energy consumption profiles, and DG from solar PV panels.

In this particular LV network, although other sources of power system data could have been used, such as power quality monitors, the usage of PMUs is chosen within the RESOLVD project. For more information, refer to [3]. In this context, the choice for PCA is motivated by the benefits of dimensionality reduction in a setting with many monitored variables and limited data storage and communication capabilities.

Two different PMU settings are evaluated: (1) a single PMU installed at the substation node and (2) PMUs installed at the substation and DG nodes (illustrated in Fig. 2.2). The statistical model representing normal operation is built with phase voltage magnitudes (in other words, $m = 3$ in the first setting and $m = 9$ in the second setting) sampled every 100 ms over a 1-day data frame window (with $n = 24 \times 60 \times 60 \times 10 = 864000$ samples per day). In turn, abnormal operation is represented by voltage sags and swells simulated under normal operation during a day to explore the sensibility at different time instants. The voltage magnitudes recorded at the substation and DG nodes are illustrated in Fig. 2.3, in which each phase (A, B, C) is plotted with a different colour (blue, red, yellow). The value of $r = 1$ is chosen with a single PMU installed, whereas $r = 2$ is chosen with multiple PMUs installed, as those principal component express over 99% of the total variability (given by (2.19)) under normal operation in both PMU settings, with $\alpha = 0.95$.

The calculated values of the T^2 and SPE statistics for phase voltage magnitudes when a single PMU is installed are illustrated in Fig. 2.4, whereas the calculated values of the T^2 and SPE statistics for phase voltage magnitudes with PMUs installed at the substation and DG buses are illustrated in Fig. 2.5, together with their statistical thresholds. In both cases, it can be noticed that some values, associated with the voltage sags and swells illustrated in Fig. 2.3, present a different statistical pattern that surpasses the thresholds and is clearly distinguishable from normal operation.

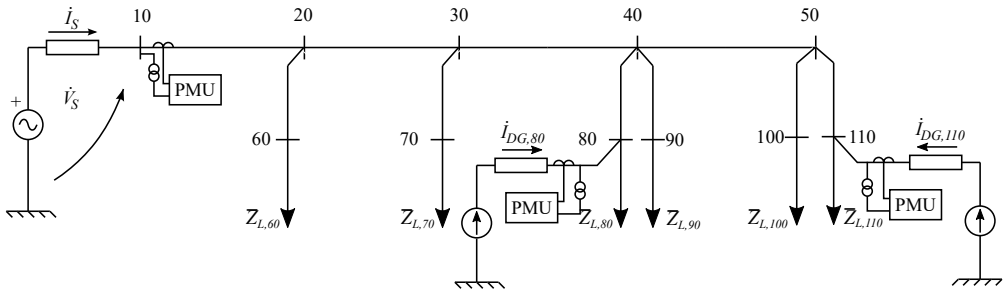

Figure 2.2: Case study – Network topology with PMUs installed at the substation and DG nodes

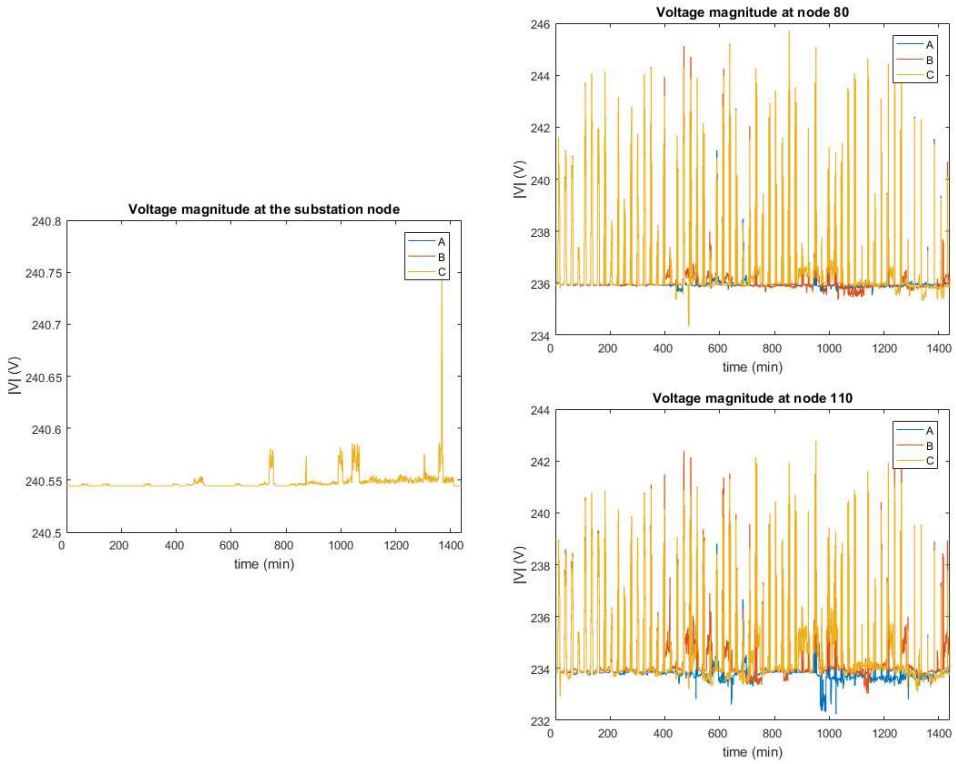

Figure 2.3: Case study – Voltage magnitudes at the substation and DG nodes over a day

2.3.1 Discussion

The results of Section 2.3 show that the PCA-based method is capable of detecting voltage sags and swells with both T^2 and SPE statistics and usage of distinct phasor quantities as input data. On the one hand, the PCA model built with a single PMU performs better than the PCA model built with three PMUs with respect to the number of correct event detections with the SPE statistics and the time intervals of event detection with the T^2 statistics; on the other hand, it performs worse with respect to the number of correct event detections with the T^2 statistics.

Although the SPE statistics is expected to perform better than the T^2 statistics due to the large variability of the abnormal voltage data in the residual subspace, it only occurs in the setting with a single PMU when it comes to the number of different event detections. The good performance of the T^2 statistics is due to the total variability of the abnormal voltage data expressed in the projection space. Thereby, the combined usage of the T^2 and SPE statistics is recommended to detect events in PMU data.

Moreover, the choice of r shows that the model built with data collected by a single PMU is well represented by the eigenvector associated with the largest eigenvalue, as it represents a three-phase, symmetrical network. In turn, the model built with data collected

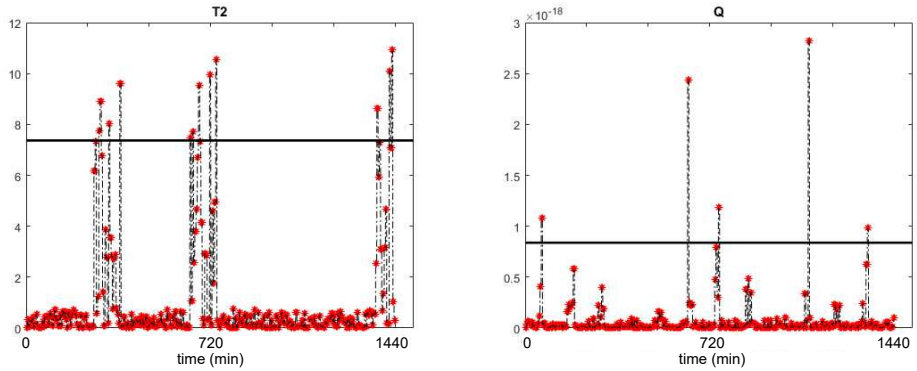

Figure 2.4: Results – T^2 and SPE values (red stars) and statistical threshold (solid black lines) with a single PMU installed

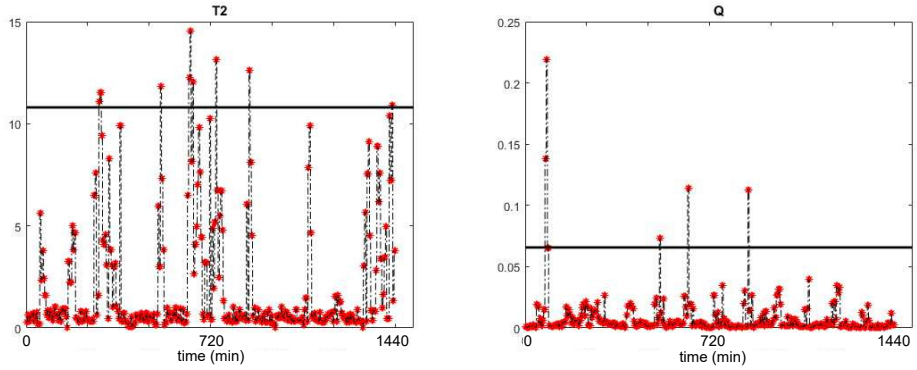

Figure 2.5: Results – T^2 and SPE values (red stars) and statistical threshold (solid black lines) with multiple PMUs installed

by three PMUs is well represented by the eigenvectors associated with the two largest eigenvalues, as they represent a three-phase, symmetrical network with two different supply sources. As a result, events can be easily detected with significant dimensionality reduction, as the number of variables is reduced from 3 to 1 with a single PMU in use and from 9 to 2 with three PMUs in use. As the simulations were run in steady state, the reference model did not have to be adjusted to handle changes in energy production and consumption profiles.

On the whole, the PCA model built with a single PMU is computationally more efficient than the PCA model built with three PMUs, as it requires half the number of principal components to express the same variability of data. The good performance of the PCA model built with a single PMU can also be attributed to the information captured by the eigenvector associated with the largest eigenvalue, represented by one major data trend. Although this representation provides a clear distinction between the projection subspace and the residual subspace, the usage of a single PMU provides a less accurate representation of the distribution network, as it does not include information about the

DG nodes. As this efficiency with respect to the number of principal components is not relevant for monitoring, the setting providing a more accurate representation of the network is recommended.

It is noteworthy that a real-time implementation of the algorithm with $n = 60 \times 10 = 600$ samples per second ($n = 24 \times 60 \times 60 \times 10 = 864000$ samples per day) is conditioned to the execution times at the substation site and the data acquisition and communication infrastructure available in the grid. At present time, the algorithm to be implemented in the rural LV network under consideration will rely on a previous data processing step to overcome limitations concerning slow execution times and data acquisition and communication infrastructures.

2.4 Statistical detectability and isolability

This subsection presents a methodology for detection and isolation of events in power system data based on the statistical detectability (that is, the extent to which something is detectable) and isolability (that is, the extent to which something is isolable) with PCA. The results of this analysis are used to define a new method to select an appropriate value of r to be retained, in agreement with the minimum detectable event with T^2 and SPE statistics in specific directions of interest. The next sections describe the assessment of statistical detectability in the projection subspace and residual subspace in specific directions of interest, then used to define an appropriate value of r, and the assessment of statistical isolability in the events detected. Furthermore, power system information retrieval is considered to proceed with inspection of the location and root causes of the occurrences.

2.4.1 Assessment of statistical detectability in the projection subspace

To evaluate statistical detectability in the projection subspace, consider a generic faulty observation $\mathbf{x_f}$ defined as a variation with respect to a normal observation $\mathbf{x_o}$ in the direction ξ (representing a unit vector) with magnitude f, as in (2.26):

$$\mathbf{x_f} = \mathbf{x_o} + \mathbf{\Delta x_f} = \mathbf{x_o} + f\xi. \tag{2.26}$$

As $\mathbf{x_o}$ represents normal operating conditions, calculation of (2.11) with $\mathbf{x_o}$ falls in the projection subspace. For simplicity, assuming that x_o is an average observation that falls in the centre of the projection subspace, it holds that $T^2_{x_o} \approx 0$ and $\mathbf{x_o}$ can be neglected from the calculations.

Substituting (2.26) in the score matrix $\mathbf{T}$ and in (2.11), taking the magnitude of the resulting quadratic equation, and considering that $T^2_x \geq T^2_{thresh}$ must hold for event

detection with T^2 statistics, then the minimum detectable event with T^2, with magnitude f_{min}, satisfies (2.27):

$$\|f_{min}\xi\mathbf{P}\mathbf{\Lambda_r}^{\frac{-1}{2}}\|^2 \geq T^2_{thresh}. \tag{2.27}$$

To proceed with the assessment using T^2 statistics, consider $\hat{\xi}$ given by (2.28) as an r-dimensional vector:

$$\hat{\xi} = \xi\mathbf{P}\mathbf{\Lambda_r}^{\frac{-1}{2}}. \tag{2.28}$$

For a given ξ and T^2 statistics, it holds that

1. it is possible to obtain the magnitude from $f = \frac{\sqrt{T_x^2}}{\|\hat{\xi}\|}$;
2. an event is detectable if $\|\hat{\xi}\| \neq 0$ (that is, when ξ is not orthogonal to the projection hyperplane) and $f_{min} \geq \frac{\sqrt{T^2_{thresh}}}{\|\hat{\xi}\|}$ (that is, when event magnitude is larger than the threshold);
3. as real events are upper-bounded by a finite maximum magnitude in the original data, a detectable event can only assume values in the interval $\frac{\sqrt{T^2_{thresh}}}{\|\hat{\xi}\|} \leq f \leq f_{max}$ (usually limited by range of measurements or physical constraints of the system);
4. f_{min} is inversely proportional to the module of (2.28) and directly proportional to the variability (i.e., eigenvalues of Λ_r) in the direction ξ;
5. the larger the eigenvalue in the direction ξ, the larger f_{min} and the better the detection capability of the model.

2.4.2 Assessment of statistical detectability in the residual subspace

To evaluate statistical detectability in the residual subspace, consider $\mathbf{x_f}$ as in (2.26). As $\mathbf{x_f}$ is a deviation from normal operating conditions represented by $\mathbf{x_o}$ in the direction ξ and $\mathbf{x_o}$ falls within the thresholds (2.13) and (2.15), then $\tilde{\mathbf{x_o}} \approx 0$.

Likewise, substituting (2.26) in the residual matrix $\tilde{\mathbf{X}}$ and in (2.14), taking the magnitude of the resulting quadratic equation, considering that $Q_x \geq Q_{thresh}$ must hold for event detection with SPE statistics, and that in the worst-case scenario it requires a magnitude twice as big to move outside the threshold in the opposite direction, then the minimum detectable event with SPE statistics, with magnitude f_{min}, satisfies (2.29):

$$\|f_{min}\xi\tilde{\mathbf{P}}\|^2 \geq 2Q_{thresh}. \tag{2.29}$$

To proceed with the assessment using SPE statistics, consider $\tilde{\xi}$ given by (2.30) as an m-dimensional vector:

$$\tilde{\xi} = \xi\tilde{\mathbf{P}} = \xi\left(1 - \mathbf{P}\mathbf{P}^{\mathbf{T}}\right). \tag{2.30}$$

For a given direction ξ and SPE statistics, it holds that

1. it is possible to obtain the magnitude from $f = \frac{\sqrt{Q_x}}{\|\tilde{\xi}\|}$;
2. it is detectable if $\|\tilde{\xi}\| \neq 0$ and $f_{min} \geq \frac{\sqrt{2Q_{thresh}}}{\|\tilde{\xi}\|}$ in the worst-case scenario;
3. as real events are upper-bounded by a finite maximum magnitude in the original data, a detectable event can only assume values in the interval $\frac{\sqrt{2Q_{thresh}}}{\|\tilde{\xi}\|} \leq f \leq f_{max}$;
4. f_{min} is inversely proportional to the module of (2.30);
5. the detection capability of the model does not depend on the magnitude of the eigenvalues.

2.4.3 Statistical detectability in specific directions of interest

In power systems, specific directions $\xi_x \in \mathbb{R}^m$ of interest include individual phases and/or all phases of a three-phase system, with one or multiple devices involved within a WAMS infrastructure. Thereby, the statistical detectability shall be assessed in these specific directions associated with distinct events of interest.

It is noteworthy that the statistical detectability criteria (2.27) and (2.29) depend on the centring and scaling of the statistical model. Consequently, the magnitude f_{min} in each direction of the original data is perceived in different ways. Considering the m-diagonal matrix of standard deviations $\mathbf{\Sigma} \in \mathbb{R}^{m\times m}$ and the m-dimensional vector $\bar{\mathbf{x}}_{\mathbf{x}} \in \mathbb{R}^m$ containing the mean values of the orignal variables used to build the PCA model, the relation between $\mathbf{x_{f,x}}$ and $\mathbf{x_f}$ is given by (2.31):

$$\mathbf{x_f} = \left(\mathbf{x_{f,x}} - \bar{\mathbf{x}}_{\mathbf{x}}\right)\mathbf{\Sigma}^{-1}. \tag{2.31}$$

Replacing (2.26) in (2.31), it holds that $\xi_x = \frac{\xi\mathbf{\Sigma}}{\|\xi\mathbf{\Sigma}\|}$ and $f_x = f\|\xi\mathbf{\Sigma}\|$ in the original data. Therefore, if ξ_x is known in the original data, its statistical detectability can be evaluated in terms of ξ_x. In this case, the goal is to find the minimum f_x that can be detected in ξ_x.

2.4.4 Selection of principal components

It is possible to build PCA models that are sensitive enough to ensure statistical detectability of specific events of a given magnitude f_{min} and direction ξ without previous knowledge of the nature of the events. For this to be possible, the relation between (2.27), (2.29) and the thresholds defined in (2.13), (2.15) has to be evaluated as a function of r so that the most appropriate r can be chosen. The challenge is to select adequate values of r such that (2.27) and (2.29) are satisfied, considering that the wrong detections (i.e., false positives or normal observations labelled as faulty) and missed detections (i.e., false negatives or faulty observations not detected) depend on α and f_{min} for all directions of interest. To this extent, satisfying (2.27) and (2.29) with maximum dimensionality reduction is desired to simplify monitoring tasks.

To ensure detection of the smallest events of interest with both T^2 and SPE statistics, for each ξ, the choice of r is made either with (2.32) or with (2.33); if there does not exist an r satisfying (2.27) and (2.29) at the same time, that is, ensuring simultaneous event detection with both T^2 and SPE statistics, then (2.32) is applied such that r is defined in separate for the projection and residual subspaces, otherwise, (2.33) is applied:

$$r := \begin{cases} \{\min\ i \in \{1, \cdots, m\} |\ (2.27) \text{ holds with } \mathrm{T}^2\} \\ \{\min\ i \in \{1, \cdots, m\} |\ (2.29) \text{ holds with SPE}\} \end{cases}, \tag{2.32}$$

or

$$r := \{\min\ i \in \{1, \cdots, m\} |\ (2.27) \text{ and } (2.29) \text{ hold}\}. \tag{2.33}$$

Additionally, it is possible to make a global choice of r for the overall network according to (2.32) or (2.33) such that the resulting PCA models enable detection of the smallest f_{min} computed with (2.27) and (2.29) for all directions of interest. For a given r, $\mathbf{P}$ and $\mathbf{\Lambda}_r$ are fixed and any arbitrary direction ξ can be expressed as a linear combination of those representing faults in a single variable of the original space (i.e., m-dimensional vectors $[1, 0, \cdots, 0], [0, 1, \cdots, 0], \cdots, [0, 0, \cdots, 1]$). Thereby, detection of events in any direction is guaranteed by designing the models to be sensitive to f_{min} in the poorest primary direction, in which the smallest values of the left side of (2.27) and (2.29) are calculated. As the resulting matrices in (2.27) and (2.29) are positive definite, it holds that f_{min} can always be found for some primary direction ξ representing a single variable of the original space.

2.4.5 Assessment of statistical isolability

It holds that a detectable event can be always isolated, as the statistical detectability is a result of individual contributions $x_j,\ j = \{1, \cdots, m\}$ of $\mathbf{x}$ that together produce violation of constraints (i.e., $T_x^2 > T_{thresh}^2$ with T^2 statistics and $Q_x > Q_{thresh}$ with SPE statistics). However, an isolable event cannot be always detected, as (2.25) may hold with $T_x^2 \leq T_{thresh}^2$

and with $Q_x \leq Q_{thresh}$. Thus, event detection must always occur with either or both T^2 and SPE statistics so that event isolation can be performed coherently.

Considering $\hat{\xi}_j$ as the j^{th} element of $\hat{\xi}$, an event is isolable with T^2 statistics if (2.34) holds:

$$\|f_{min}\hat{\xi}_j\|^2 \geq T^2_{j,thresh}. \tag{2.34}$$

Likewise, considering $\tilde{\xi}_j$ as the j^{th} element of $\tilde{\xi}$, an event is isolable with SPE statistics if (2.35) holds:

$$\|f_{min}\tilde{\xi}_j\|^2 \geq 2Q_{j,thresh}. \tag{2.35}$$

2.4.6 Power system information retrieval

The isolation procedure previously described returns the variables involved in the detected occurrence. Despite this, additional information from the grid is still required to further identify the root causes of an occurrence and prevent it from happening again. In this section, this limitation is overcome with information retrieval from the grid, whenever available, as soon as an event is detected and isolated with the procedure previously presented.

Additional pieces of information from the grid may include details about components, operating status, exact occurrence times, and different electrical quantities. They should be retrieved with the system operator, utility companies, and equipment owners for all the isolated variables, whenever possible, so that their individual contributions can be understood. Ultimately, this procedure facilitates the identification of the root causes of an event, directing the search to the isolated variables.

2.4.7 Flow diagram of the proposed method

Fig. 2.6 displays the flowchart of the proposed strategy over a generic sliding window $k \in \mathbb{N}$. Note that the method can be applied with or without overlapping time windows. The algorithm starts with the computation of the PCA model, followed by the evaluation of the statistical detectability in all possible directions and the selection of principal components which are able to detect the smallest detectable event in the directions of interest. Next, event detection is verified in the projection subspace using T^2 statistics and in the residual subspace using SPE statistics. Afterwards, the detected events are isolated with contribution analysis and additional information is retrieved from the grid to proceed with the search for their location and root causes. Finally, computation continues to the next time window.

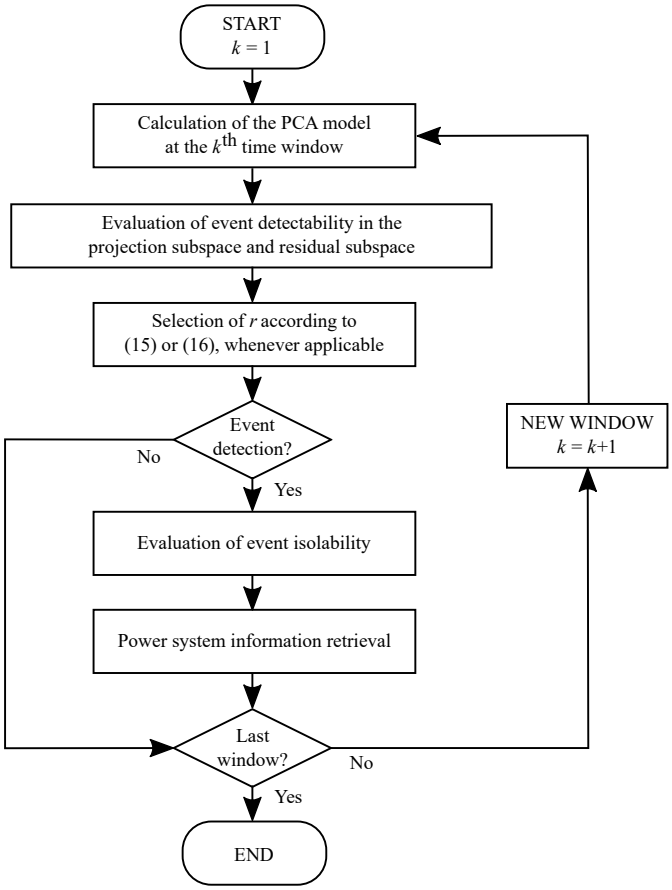

Figure 2.6: Statistical detectability – Flow diagram of the method

2.4.8 Case studies

This section provides examples of the statistical detectability assessment in specific directions of interest and for the selection of principal components. Next, the applicability of the proposed method is verified in a 1-hour data file from the Independent Texas Synchrophasor Network. The analysis presents an overview of the Texas Synchrophasor Network with emphasis on the PMUs used to build the PCA models (for more information, refer to [42], delimits the events of interest, describes the analysis coded and run in MATLAB, and presents the corresponding statistical detectability and isolability results.

Example of statistical detectability in specific directions of interest

This section illustrates the calculation of the statistical detectability of a generic event in specific directions of interest. Consider the voltage phasor magnitudes of Fig. 2.7 gathered by a single IED at LV distribution level, with $\bar{\mathbf{v}}_{\mathbf{x}} = [400.3, 400.3, 400.0]^T$ V, $\mathbf{\Sigma} = diag\,(2.0784, 3.0343, 3.2137)$, $r = 1$, and $\alpha = 0.95$ defined *a priori*, and the resulting PCA

model given by $\mathbf{P} = [-0.6098, -0.5573, -0.5635]^T$, $\mathbf{\Lambda}_1 = diag\,(2.3859)$, $T^2_{thresh} = 5.0899$, $Q_{thresh} = 2.6247$ from Section 2.2.6. For an event in the direction $\xi_x = [0, 0, 1]^T$, (2.28) results in $\hat{\xi} = [-0.3796]$ and the minimum detectable events in (2.27) are $f_{min} = 6.1848$ and $f_{min,x} = 19.8762$ V; and (2.30) results in $\tilde{\xi} = [-0.3510, -0.3084, 0.6780]^T$ and the minimum detectable events in (2.29) are $f_{min} = 1.9594$ and $f_{min,x} = 6.2971$ V.

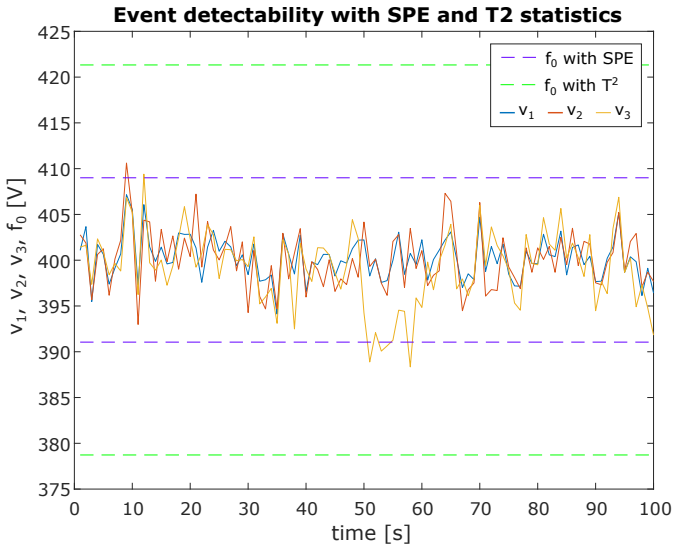

Figure 2.7: Results – Voltage phasor magnitudes (blue, red, and yellow) and minimum detectable magnitudes with SPE (dashed purple) and T^2 (dashed green) in the original data

Example of statistical detectability-based selection of principal components

This section illustrates the application of the statistical detectability for the selection of r principal components. Consider the voltage impulse sag illustrated in Fig. 2.8, with minimum detectable event magnitude $f_{min,x}$ set to 1% of the nominal voltage, maximum detectable event magnitude $f_{max,x}$ equal to 10% of the nominal voltage, and minimum detectable events calculated in the direction $\xi = [0, 0, 1, 0, 0, 0]^T$ of the dataset. The calculations of T^2 and SPE statistics with $f_{min,x}$ and the statistical thresholds calculated analytically with (2.13) and (2.15) as a function of r and α over this time window are illustrated graphically in Fig. 2.9 on the left, whereas the lower and upper limits of statistical detectability computed with $f_{min,x} = 1\%$ and $f_{max,x} = 10\%$, respectively, are displayed in logarithmic scale in Fig. 2.9 on the right. The solid lines represent the thresholds calculated with (2.13) and (2.15) with $\alpha = \{0.90, 0.95, 0.99\}$ and also the right side of (2.27) and (2.29), whereas the blue dotted lines represent the left side of (2.27) and (2.29) calculated with f (lower limit of detectability) and the black dotted lines represent the left side of (2.27) and (2.29) calculated with $f_{max,x}$ (upper limit of detectability).

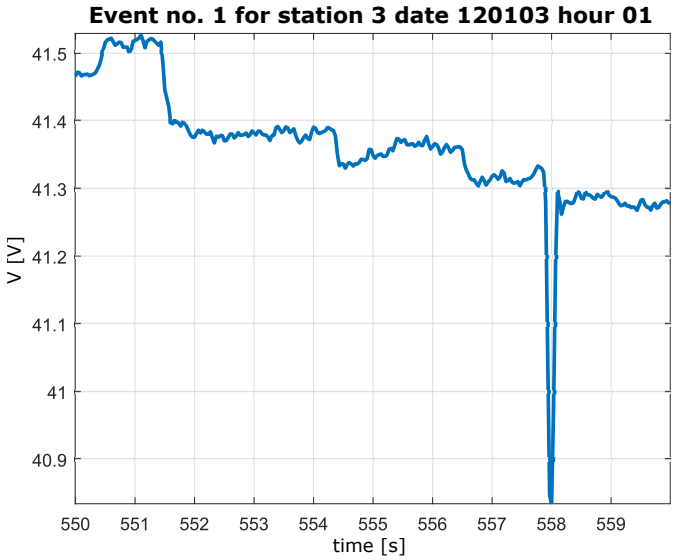

Figure 2.8: Case study – Example of voltage impulse

In this example, it can be seen in the graphs that detectability in the projection subspace occurs when $r \geq 2$ for $\alpha = \{0.90, 0.95\}$ and when $r \geq 4$ for $\alpha = 0.99$, whereas detectability in the residual subspace occurs when $r \leq 3$ for all tested values of α over this time window. As a consequence, $\alpha = 0.95$ and $r = 2$ are chosen to ensure event detection with usage of SPE statistics and T^2 statistics in a single PCA model.

Examples of statistical detectability and isolability in the Independent Texas Synchrophasor Network

In this section, the proposed event detection and isolation procedure is tested with a 1-hour data file from the Independent Texas Synchrophasor Network, publicly available in [10] (see website). The phasor quantities are sampled once every two wave cycles (that is, using the 30 Hz sampling frequency available in the monitoring infrastructure) for further assessments. This dataset contains 16 events of interest enumerated in Table 2.6, notably 6 single impulses and 4 multiple impulses that last no longer than a few milliseconds and 6 low-frequency oscillations that last no longer than a few hundred milliseconds.

The PMUs are placed in different zones of ERCOT, at distinct transmission and distribution voltage levels, and provide a good view of the power grid for its topology, illustrated in Fig. 2.10: McDonald in West Texas, Waco and Austin in Central Texas, Harris in East Texas, and UT3 and UTPan in Southern Texas. It is noteworthy that detection and isolation are restricted to the PMU observability and coverage area, which is known, and sampling rate (30 Hz, which enables analysis of events below 15 Hz). Additional information about the network topology and its electrical parameters are unknown and not necessary to perform the analysis.

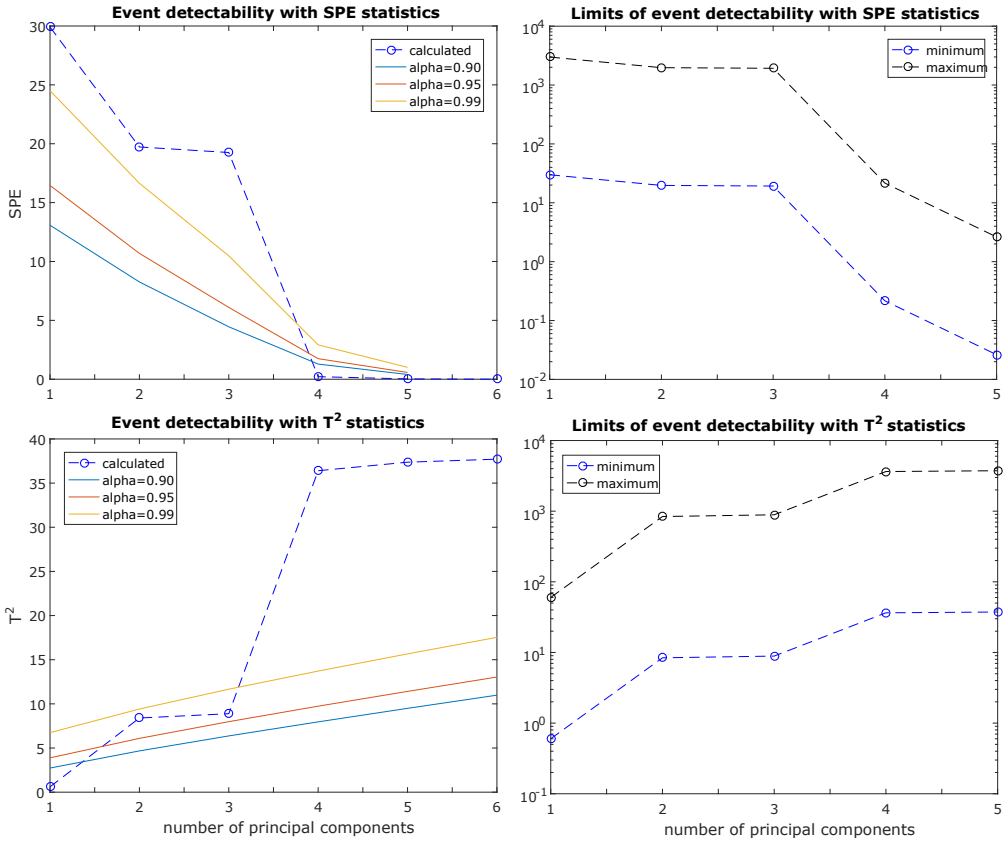

Figure 2.9: Results – Minimum detectable magnitude and analytical thresholds calculated with different α with SPE (top left) and T^2 (bottom left) and lower and upper limits of statistical detectability with SPE (top right) and T^2 (bottom right) vs. number of principal components

The PCA-based algorithm relies on voltage magnitude measurements collected over a time window to build a statistical model of the real-time network operating conditions. Only voltage magnitudes are used because PCA computes the strength of linear correlation between the original measurements and the statistical indicators, provided that they are nearly related through a few linear approximations under normal operation over a time window of short duration. The detectable events in the original data are characterized by a sudden variation larger than 1% of the nominal voltage at a given location. Therefore, the minimum event magnitude in the original data is set to 1% of the nominal voltage and the directions of interest are defined by the variables representing each PMU substation individually, regardless of the correlation structure of the data.

The event categories shown in Table 2.6 are numbered in the order they appear in the data file. The statistical analysis is run over a 5-second sliding window to detect distinct types of events and intrinsic characteristics with $\alpha = 0.95$, insofar as there are no missed detections and no more false positives than true positives over this length of time with

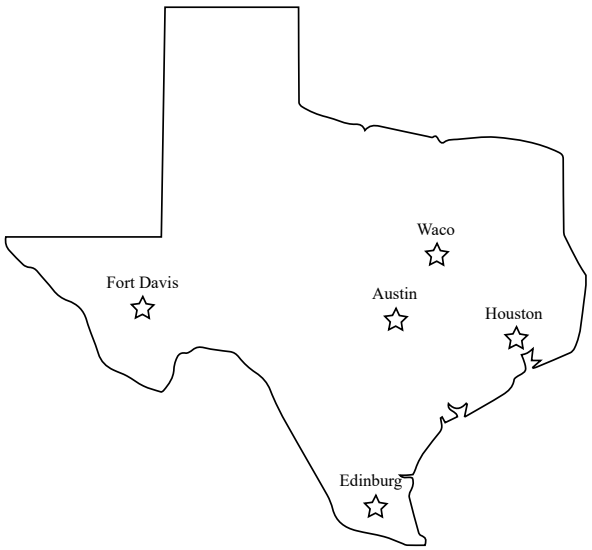

Figure 2.10: Case study – Map of the Independent Texas Synchrophasor Network

this α.

Table 2.1 shows the number of principal components r, the detection results obtained with T^2 and SPE statistics, and the criterion applied to determine r in the directions referred to the PMU substation responsible for the occurrence, whereas Table 2.2 shows the isolation results, that is, the substations that observe the events according to the constraint violations and their relative contributions in percentage to the overall constraint violation.

The results of Table 2.1 show that all voltage impulses and low-frequency oscillations are detected with at least one of the T^2 or SPE statistics over the directions where the event occurred when evaluated with a 5-second window and $\alpha = 0.95$. On the whole, more detections occur with usage of SPE statistics, which is expected due to the break in the linear correlation structure under abnormal operation. However, it can be noticed that there is no connection between the event categories listed in Table 2.6 and the correct detections with use of T^2 and/or SPE statistics.

2.4.9 Discussion

The procedure presented in Section 2.4 is expected to guarantee that all events of interest are detectable in both the projection and residual subspaces with maximum dimensionality reduction, as long as the detectability criteria comply with the power system operating constraints. The results shown in Section 2.4.8, Table 2.1, show that all events in Table 2.6 are detected with at least one statistics when the statistical detectability in the directions of interest (i.e., representing the PMU substation where the event occurred) is taken into consideration, under the circumstances described in Section 2.4.8. The combined usage

Table 2.1: Results – PCA-based event detection with a 5 s window

Event	PMU substation	T^2 statistics r	T^2 statistics detected?	SPE statistics r	SPE statistics detected?	Selection criterion
1	UTPan	2	Yes	2	Yes	joint
2	UTPan	4	No	3	Yes	separate
3	McDonald	2	Yes	2	Yes	joint
4	UTPan	2	Yes	2	Yes	joint
5	UTPan	3	No	2	Yes	separate
6	McDonald	3	Yes	1	Yes	separate
7	McDonald	2	Yes	2	Yes	joint
8	Waco	1	No	1	Yes	joint
9	UTPan	3	Yes	1	Yes	separate
10	UTPan	3	Yes	3	Yes	joint
11	UTPan	2	Yes	2	Yes	joint
12	UTPan	3	Yes	3	Yes	joint
13	UTPan	4	No	3	Yes	separate
14	UTPan	3	No	3	Yes	joint
15	McDonald	2	Yes	2	Yes	joint
16	UTPan	2	Yes	2	Yes	joint

of T^2 and SPE is expected to lead to a higher number of correct event detections and consequently increase the detection capability of the method, as previously discussed.

Moreover, the evaluation of statistical detectability may influence the choice of r for each time window, event, and direction of interest. Consequently, the use of two different PCA models might be necessary to detect events with T^2 and SPE statistics, as some events may not be detectable simultaneously in the projection subspace and residual subspace with any value of $r = \{1, \cdots, m\}$. This can be noticed in Table 2.1, which shows that events 2, 5, 6, 9, and 13 of Table 2.6 cannot be detected simultaneously in the projection subspace and residual subspace with any value of $r = \{1, \cdots, m\}$. In fact, events 2, 5, and 13 cannot be detected with T^2 statistics, but events 6 and 9 can be detected in separate when the statistical detectability is considered to compute different values of r in the projection subspace and residual subspace. Therefore, the evaluation of statistical detectability separately in the projection subspace and residual subspace may improve event detection.

The calculated values of r can be also attributed to the desired level of statistical detectability, which is set to detect events whose magnitudes are larger than a predefined threshold. Analogously, the detection of events whose magnitudes are smaller (or larger) than a specific threshold would result in a smaller (or larger) statistical detectability and consequently request different values of r in the chosen selection criterion.

Furthermore, the results are expected to be isolable with either or both statistics over one or more locations, including those where the events are recorded, regardless of the detection results. To this extent, Table 2.2 evinces that an isolable event cannot be always detected. Those results are also expected to provide a clear understanding of the substations affected, provided information about the observability and the coverage area of each IED/PMU/SM, as the contribution analysis computes the influence of each variable in the calculated values of T^2 and SPE statistics. Additional information about the network topology and its electrical parameters can also be helpful to evaluate different occurrences, if available, as an event can spread across the grid in different ways (e.g., through lines and transformers).

2.5 Performance comparison of PMU data event detection methods

This section makes a comparison of different methods for PMU data event detection. The events evaluated consist of anomalies of different nature whose causes are unknown. It is noteworthy that these methods are targeted at minor events that may lead to performance degradation, as major failures produce major effects on the system and are easily identifiable in the data.

2.5.1 Performance comparison of quantitative methods for PMU data event detection with noisy data

This section presents a comparative analysis of different event detection techniques with respect to the accuracy of correct detections in the presence of noisy data. The methods are tested in a 1-hour PMU data file from the Independent Texas Synchrophasor Network available in [10] (see website), relying on voltage magnitudes gathered at PMU substations at distinct transmission and distribution voltage levels (see [42] for more details about the network). The operating conditions of the grid are modelled over a 10-second sliding window basis, which improves the situational awareness of the analysis and allows for adaptation to the most recent operating conditions.

The events that can be found in the dataset are displayed in Table 2.3. They consist of low-frequency voltage oscillations which last no more than a few hundred milliseconds. Only events below 15 Hz are observable, as the PMU sampling rate is 30 Hz.

All the event detection techniques detect anomalies with use of the $3-\sigma$ rule, in accordance to the analysis performed in [10]. The PCA algorithm relies on SPE statistics with the VRE criterion, whereas the SVM classification considers that normal operation is represented by the average values of the data over a sliding time window and that a deviation of 2% from the average represents abnormal operation, based on the magnitude of the events recorded in the dataset. The noise consists of an additive disturbance rep-

resented by a white Gaussian noise δ with $var\,(\delta) = 0.25$ V^2 added to the original PMU data. This choice results in an SNR of 40 dB at the substation with the lowest voltage level, slightly below the 45 dB approximation suggested in [22].

The results of correct event detections in the original data are presented in Table 2.4, whereas the results of correct event detections in the presence of an additive noise are presented in Table 2.5.

2.5.2 Performance comparison of PCA techniques for PMU data event detection

In this section, the procedures described in Section 2.2.6 to select an appropriate number of principal components and perform event detection are tested with PMU data from the Independent Texas Synchrophasor Network, with the events listed in Table 2.6. Overall, they are expected to be isolated events at a single location (which ensures detection of the smallest events in the worst-case scenario) and occur once or less every 10 seconds. Their magnitudes are greater than 1% of the nominal voltage, varying from a few Volts at LV level (substation at Fort Davis) to a few thousand Volts at HV level (substations at Edinburg and Waco). Duration of disturbances is also heterogeneous, lasting from a few milliseconds (impulses) to a few hundred milliseconds (transients).

The PCA models are built online over a sliding window, relying on voltage magnitudes. The five different methods to select the number of principal components described in Section 2.2.6 are considered to build the PCA model and further tested and compared in terms of performance of T^2 and SPE statistics over a 10-second and 1-minute sliding window, supposedly associated with different types of events and intrinsic characteristics. As a result, 20 different scenarios were produced per detected event (i.e., 2 statistical tests times 5 selection criteria of r times 2 window lengths).

Tables 2.7 and 2.8 report the event detection results (yes/no) and the corresponding r obtained for each selection method and time window. A confidence level $\alpha = 0.95$ is chosen to calculate T^2 and SPE for all selection methods because it does not result in missed detections in this dataset with the statistical detectability criterion over a 10-second sliding window. The locations listed in Table 2.6 are considered to compute the statistical detectability in specific directions. In addition, the explained variance criterion is computed with an explained variance of 75% because of the average residual subspace defined by the other criteria.

2.5.3 Performance comparison between the statistical detectability and state-of-the-art PCA-based methods

This section compares the statistical detectability criterion described in Section 2.4 with the state-of-the-art PCA-based event detection methods proposed in the literature. Thus, the results reported in Table 2.1 are taken into consideration as follows.

- The usage of the first principal component, as in [75], would enable detection of events 6, 8, and 9 with the SPE statistics only.
- The usage of the predefined variance threshold of 90% proposed in [24] would result in $r = 4$ and allow for detection of events 1, 3, 4, 6, 7, 9, 10, 11, 12, 15, and 16 with the T^2 statistics only. Note that the detection capability increases with the T^2 statistics and decreases with the SPE statistics as r increases.
- The threshold criterion used in [27] would result in $r = 1$ or $r = 2$ and allow for detection of events 1, 3, 4, 5, 6, 7, 8, 9, 11, 15, and 16.

2.5.4 Discussion

Overall, the results of Section 2.5.1 indicate that the statistical knowledge-based event detection techniques present the best performance individually and are more robust to disturbances than the signal-based event detection techniques described in the state of the art. Based on this comparative analysis, the usage of knowledge-based techniques is recommended for accurate PMU data event detection.

Considering the correct detections reported in Table 2.4, the best performance is obtained with PCA, which detects all 20 events, followed by SVM, which detects 17 out of 20 events, whereas the results detected with FFT, Yule-Walker spectral analysis, min-max difference, and approximate derivative methods are complementary. In particular, the min-max difference detects 11 occurrences, whereas both the FFT and the approximate derivative detect 8 events and the Yule-Walker spectral analysis detects 4 events. In turn, considering the correct detections reported in Table 2.5, the best performance is obtained with SVM, which detects 15 out of 20 events, followed by PCA, which detects 12 out of 20 events. Additionally, it can be noticed in Table 2.5 that 18 out of 20 detections are achieved by PCA and SVM combined. Conversely, the Yule-Walker spectral analysis detects 3 events, whereas the min-max difference detects 2 events and both the FFT and the approximate derivative detect 1 event each.

It can be noticed that all the 16 voltage impulse events of Table 2.4 are detected with both PCA and SVM without noise. In this case, the best performance in the presence of noise is obtained with SVM, which detects 15 out of 16 events, followed by PCA, which detects 10 out of 16 events. Conversely, some step changes are missed by the PCA algorithm in the presence of noise and detected by both the Yule-Walker spectral method and the min-max difference algorithm.

In general, the additive noise is expected to affect most significantly those events detected at LV distribution level, for its magnitude. This is verified in Tables 2.4 and 2.5, as the results obtained for the sliding windows containing events at HV transmission level remain the same for all methods. It occurs because of the same SNR applied at both LV and HV levels. Moreover, large disturbances at HV transmission level are expected to be detected more easily than minor events at LV distribution level, regardless of noise.

In turn, the results shown in Section 2.5.2 suggest that the best method to select the

number of principal components is the statistical detectability criterion, considering the correct detections with both T^2 and SPE statistics for all lengths of time evaluated. This suggests that the combined use of T^2 and SPE statistics with the statistical detectability criterion may lead to a higher number of correct event detections, reduce the number of missed detections, and consequently increase the detection capability of the PCA model for different types of events.

In general, the SPE is expected to detect variations associated with changes in the correlation structure of the data and presents small values, whereas the T^2 is expected to detect deviations from the average normal operating conditions of the data and presents larger values. As a consequence, the SPE is more sensitive than the T^2 and tends to be a better indicator of abnormalities, since changes in the correlation structure of faulty observations are expected to be observed in the residual matrix. This explains why the SPE is expected to present the best performance when the VRE criterion is applied to select r, as it is aimed at minimizing the variance reconstruction error associated with the residual subspace to ensure detection of specific events of interest. However, this criterion may not be effective to detect events associated with deviations from the standard operating conditions.

It is noteworthy that all events of interest are detected with at least one of the T^2 or SPE statistics when a 10-second window is used to build the PCA model and that the highest number of correct event detections occurs when a 10-second window is applied to build the PCA model. It happens because the PCA models built over this length of time are more sensitive to the dynamics of the system and consequently are more suitable to detect events lasting no more than a few hundred milliseconds. Therefore, it is expected that a higher number of missed detections will occur with longer window sizes, which are associated with different phenomena and/or a static representation of the operating conditions of the grid. This explains why the results obtained with a 1-minute window are slightly worse than those obtained with a 10-second window in most cases. Notably, event number 8 is always missed with a 1-minute window, regardless of the method chosen to define an appropriate number of principal components.

Furthermore, a theoretical comparison between the five approaches previously described to select an adequate r indicates that the Kaiser criterion, the automatic scree plot, and the explained variance are defined quantitatively and arbitrarily, based on the variance of the data and thresholds that do not represent any characteristics of the events of interest. In turn, the statistical detectability and the VRE are defined qualitatively, according to specific event magnitudes and directions of interest, and can be adjusted to detect events defined by specific magnitudes f and direction vectors ξ. This explains why the results obtained with the statistical detectability criterion with both T^2 and SPE statistics are the most accurate, which implies that the detectability in the projection subspace and the residual subspace is the best indicator to select an adequate number of principal components and detect distinct events of interest.

The results shown in Section 2.5.3 also indicate that the statistical detectability criterion described in Section 2.4.8 produces a higher number of correct event detections

than the PCA-based methods proposed in the literature. This is due to the adjustments introduced by the methodology, which enable to detect events with distinct magnitudes and directions of interest; and again, to the combined usage of T^2 and SPE statistics, which enables to detect events with different statistical behaviours.

2.6 Monitoring of smart grids with multilayer PCA

This section presents a multilayer PCA strategy applied for modelling and monitoring LV grids and detecting disturbances over different time scales. The PCA models are built over a sliding time window to enable periodic updates reflecting changes in the generation and consumption patterns, considering the trade off between the expected duration of the events and the elapsed time between consecutive events of interest to select an adequate length of time. Event detection relies on the T^2 and SPE statistics and isolation relies on contribution analysis of those statistics.

The overall formulation considers distinct problems of interest, each of them requesting different electrical quantities and sampling rates and arranged hierarchically in accordance to its time scale. Input data may include voltage, current, impedance, power, or energy measurements acquired by any monitoring infrastructures or calculated from combinations of them. For a set of variables representing the same electrical quantity, these linear relations are assumed to be enough to represent the steady-state operation of the LV grid and enable detection of abnormal operating conditions at different layers. The data preparation (or preprocessing) is described in Section 2.6.1, together with some considerations about PCA-based modelling (considering the methodology previously described in Section 2.2.6). Next, the multilayer PCA strategy is described in Section 2.6.2.

2.6.1 Data preparation

The goal of this step is to prepare the data in a matrix structure suitable for PCA. Thus, given $m_{raw} \in \mathbb{N}^*$ original variables (electrical quantities) gathered by one or multiple IEDs over time at the sampling rate $f_{raw} \in \mathbb{R}_+^*$ (in Hertz), define $\ell \in \mathbb{N}^*$ hierarchical levels (layers) in such a way that the events of interest can be correctly identified and characterized over different lengths of time (e.g., second, minute, hour, etc.). Each layer $k \in \mathbb{N}^*$, $k = 1, \cdots, \ell$ is characterized by a length of time $\tau_k \in \mathbb{R}_+^*$ (total time which the monitoring lasts), an observation period $\sigma_k \in \mathbb{R}_+^*$ (if applicable), an observation time duration $\theta_k \in \mathbb{R}_+^*$ (with $\theta_k \leq \sigma_k$), and contains $n_k \in \mathbb{N}^*$ observations defined by $p_k \in \mathbb{N}^*$ samples of $m_k \in \mathbb{N}^*$ variables gathered at the sampling rate f_k. Thereby, a single observation in the k^{th} layer is a $p_k \times m_k$-dimensional array gathered in discrete time domain, hereby denoted by $\mathbf{x_k}(i)$, $i = \{1, \cdots, n_k\}$, forming an $n_k \times (p_k \times m_k)$ observation matrix $\mathbf{X}_k$ (as in (2.6)) such that (2.36) and (2.37) hold. Note that a particular case occurs when $\theta_k = \sigma_k = f_k^{-1}$, as the observations $\mathbf{x_k}(i)$ last a single sample and $\mathbf{X}_k$ is thereby an $n_k \times m_k$ observation matrix. Additionally, consider that $n_{k,\sigma}, n_{k,\theta} \in \mathbb{N}^*$ observations are gathered over σ_k and θ_k respectively:

$$n_k = \frac{\tau_k}{\sigma_k}, \tag{2.36}$$

and

$$p_k = \theta_k f_k. \tag{2.37}$$

It is noteworthy that the definition of θ_k allows for investigation of repetitive patterns lasting θ_k over time (e.g., daily and weekly energy consumption profiles). Thus, τ_k, σ_k, θ_k, f_k, and m_k are defined as design parameters for $k = 1, \cdots, \ell$ depending on the data organization required. In addition, assume that the layers are concatenated hierarchically over time such that $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_\ell$, with $\tau_{k+1} = j \times \tau_k$, $k = \{1, \cdots, \ell - 1\}$, for some $j \in \mathbb{N}^*$ Also, consider that screening is required every τ_k intervals, as long as abnormal behaviours are detected at the layers of shortest duration.

The data gathered by the IEDs might require a previous preprocessing step to suit to the analysis. To this extent, three different techniques described as follows can be applied and combined as a previous step to the PCA modelling and monitoring, depending on the relation between f_{raw} and τ_k, σ_k, and θ_k at the k^{th} layer. For illustration, Fig. 2.11 draws a comparison between them over the same τ_k, with different values of $\sigma_k = f_k^{-1}$ and θ_k chosen arbitrarily for each procedure: (a) time windowing, (b) filtering and re-sampling, and (c) multiway re-arrangement.

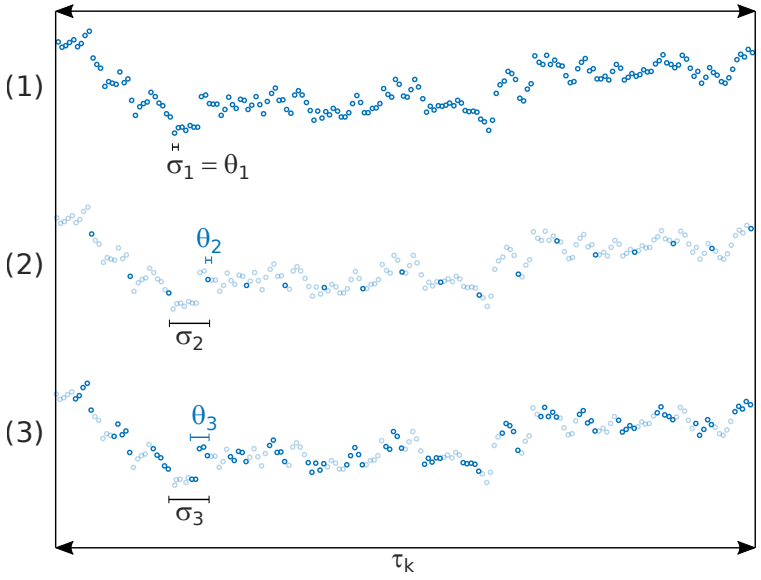

Figure 2.11: Multilayer – Comparison between distinct data organization procedures over τ_k

Time windowing This procedure, illustrated in Fig. 2.11 (a), consists of defining adequate time settings τ_k, σ_k, and θ_k for each layer $k = \{1, \cdots, \ell\}$. This choice is arbitrary and depends on the phenomena under evaluation, as shown in Section 2.6.3, but made automatically at the beginning of the analysis according to relevant time scales for humans. In this context, the following situations may occur: continuous observations ($\sigma_k = \theta_k$) and gapped observations ($\sigma_k > \theta_k$) over τ_k.

Filtering and re-sampling This procedure is recommended when re-sampling is required for the analysis every σ_k time instants to reduce the number of observations over τ_k, which is achieved by smoothing and re-sampling the original observations, with low-pass filtering required to avoid aliasing. It is exemplified in Fig. 2.11 (b), where the dark blue dots represent the re-sampled observations over the original discrete signal in light blue every $\sigma_k = f_k^{-1}$ time instants over τ_k.

Multiway re-arrangement This procedure is recommended when repetitive patterns lasting θ_k are expected to occur (e.g., daily, weekly) during specific periodic intervals defined by σ_k, as it allows to exploit possible correlations between time instants within θ_k. In this case, $p_k > 1$ samples acquired over θ_k (between $(i-1)\,\sigma_k$ and $\theta_k + (i-1)\,\sigma_k$) are concatenated to form a single observation $\mathbf{x_k}\,(i)$, $i = \{1, \cdots, n_k\}$. As a result, $\mathbf{x_k}\,(i)$ is a $p_k \times m_k$ matrix with as many rows as samples over θ_k and as many columns as variables, as illustrated in Fig. 2.11 (c).

PCA-based modelling and monitoring Having prepared the data in a matrix structure suitable for PCA, the algorithm described in Section 2.2.6 is deployed at layer level to build the statistical model of the data at each layer $k = 1, \cdots, \ell$. In this case, $\mathbf{X}$ is replaced with $\mathbf{X}_k$, an $n_k \times (p_k \times m_k)$ observation matrix (assumed to be centred and scaled in the PCA algorithm) of all selected IEDs at the k^{th} layer with n_k observations and $(p_k \times m_k)$ sampled variables referred to distinct electrical quantities gathered at f_k such that (2.36) and (2.37) hold over τ_k.

2.6.2 Sliding-window PCA

This section introduces the sliding window framework deployed at multiple layers. This approach increases the situational awareness of the analysis by allowing for simultaneous modelling and monitoring and implies that the PCA-based modelling and monitoring is run on the fly every τ_k time instants. Fig. 2.12 provides an overview of the PCA-based modelling and monitoring procedure for a generic layer k and shall be executed for $k = \{1, \cdots, \ell\}$. Assuming that the data are preprocessed and ready for evaluation, the algorithm starts with the creation of a time window of duration τ_k at the kth layer, once there are enough observations acquired over τ_k. Then, an $n_k \times (p_k \times m_k)$ observation matrix is arranged and the PCA model is built over τ_k. If an abnormal observation is

detected, the algorithm proceeds with diagnosis and continues to evaluate related occurrences up to the layer $k-1$ and the analysis proceeds with the creation of a new time window τ_k in the kth layer, as long as new data are received.

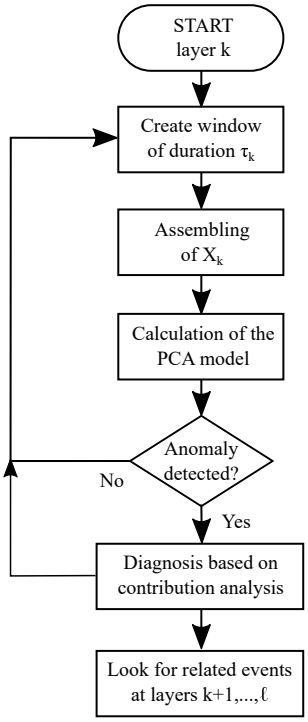

Figure 2.12: Multilayer – Flow diagram of the method in a generic layer k

2.6.3 Case study: the OpenLV project network

This section provides examples of the multilayer PCA strategy described in Section 2.6, tested in MATLAB using data from the OpenLV project network at Marshfield Village, South Gloucestershire, United Kingdom. It presents an overview of the monitoring applied to the OpenLV network in the next paragraphs, followed by an example of sliding-window PCA-based monitoring, the main analysis description and results. For more information about the OpenLV project and the chosen area, see [2].

A map of the approximate locations of the IEDs in this network (OpenLV substations 43, 44, 69, and 70) is displayed in Fig. 2.13. They are installed at LV distribution level (415 V) in different areas of a three-phase unbalanced network, where single-phase solar PV panels are also installed, and record average values of phase voltage and line current magnitudes and active and reactive power and energy every minute (i.e., $f_{raw} = 1$ min^{-1}). Therefore, $m_{raw} = 72$ variables in total (4 IEDs $\times$ 3 phases/IED $\times$ 6 quantities/phase). The measurements aggregate energy consumption from distinct households and buildings and energy production from solar PV panels whose rated power and connection points are

known.

Figure 2.13: Case study – Map of IED locations within the Marshfield network

Example of sliding-window monitoring in multilayer PCA

This section provides an example of sliding-window PCA within a multilayer implementation, considering the goal of detecting distinct hourly and daily variations in energy consumption patterns. In this example, input data consist of single-phase active power measurements gathered every minute (i.e., $f_{raw} = 1$ min^{-1}) by a single IED (i.e., $m_{raw} = 3$ in a three-phase system). For illustration, a single-phase active power profile is plotted in Fig. 2.14 over distinct lengths of time (solid yellow lines). In the next paragraphs, consider that the layers $k = \{1, 2\}$ refer to hourly and daily variations in power consumption patterns displayed in the bottom and top graphs of Fig. 2.14, respectively.

In this scenario, consider the introduction of two time windows $\tau_1 = 1$ day and $\tau_2 = 91$ days enabling a PCA-based monitoring every day and over a season, respectively. As a result, in a season, τ_1 would be run 91 times on a sliding-time basis and τ_2, just once. Without data organization, anomalies can only be detected at specific minutes, as $f_{raw} = 1$ min^{-1}. This choice results in $n_2 = 131,040$ observations (solid yellow line at the top graph of Fig. 2.14) and $n_1 = 24 \times 60 = 1,440$ observations (solid yellow line at the bottom graph of Fig. 2.14). To prevent this, observations can be filtered and re-sampled (solid blue line in the bottom graph of Fig. 2.14) and/or re-arranged multiway (pink boxes in the top and middle graphs of Fig. 2.14). In $k = 1$, re-sampling of average values over an hour with $f_1 = 1$ h^{-1} and $\sigma_1 = \theta_1 = 1$ h results in $p_1 = 1$ from (2.37) and $n_1 = \frac{24}{1} = 24$ observations from (2.36). In $k = 2$, multiway re-arrangement to evaluate a specific day of the week with $\sigma_2 = 7$ days, $\theta_2 = 1$ day, and $f_2 = 1$ h^{-1} results in $p_2 = 24 \times 1 = 24$ from (2.37) and $n_2 = \frac{91}{7} = 13$ from (2.36). As a result, $\mathbf{X}_1$ is a 24×3 matrix and $\mathbf{X}_2$ is a $13 \times (24 \times 3)$ matrix.

Additionally, in a multilayer implementation of PCA-based modelling and monitoring, τ_1 and τ_2 can be concatenated such that the anomalies detected over the shortest lengths of time (τ_1) can be related to those detected over the longest lengths of time (τ_2). This is performed as a bottom-up procedure, as illustrated in Fig. 2.12, from $k+1$ to ℓ.

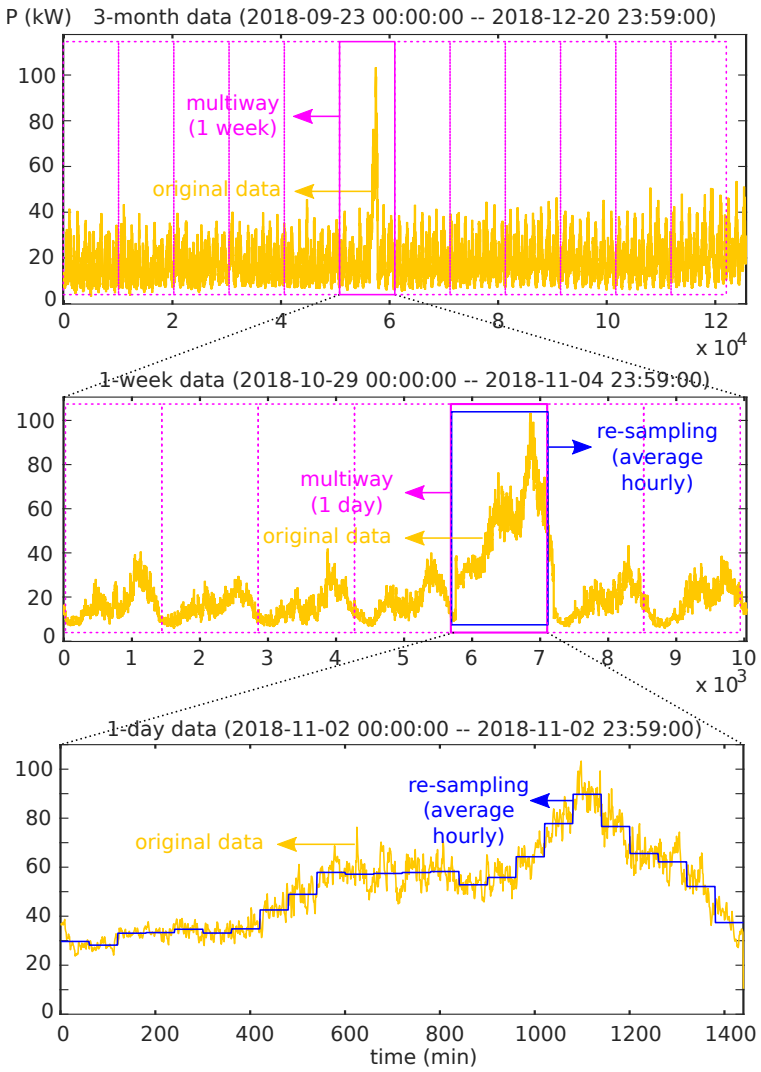

Figure 2.14: Case study – Example of data organization over power profiles: original data in yellow, multiway re-arrangement in pink, and re-sampling averages in blue

Alternatively, consider that the layers $k = \{1, 2, 3\}$ refer to hourly, weekly, and seasonal variations in power consumption patterns displayed in the bottom, middle, and top graphs of Fig. 2.14, respectively, with $m_{raw} = 1$ and $f_{raw} = 1$ min^{-1}. The goal of identifying daily profile changes over the season (with $\tau_3 = 91$ days for exact integer calculations) with $f_3 = f_{raw}$ and $\theta_3 = \sigma_3 = 1$ min results in $n_3 = 131,040$ observations (solid yellow

line at the top graph), which would not be adequate for the task. As an option, the observations could be re-arranged multiway to evaluate the data from specific weekdays (e.g pink box in the second graph) without re-sampling, which provides $f_2' = 1$ min^{-1}, $\theta_2 = 1$ day, $n_2' = 24 \times 60 = 1,440$, $m_2' = m_{raw} = 1$, $m_2 = 1 \times 1,440 = 1,440$, $\tau_2 = 91$ days, $\sigma_2 = 7$ days, $f_2 = \frac{1}{7}$ day^{-1}, and $n_2 = \frac{91}{7} = 13$, or averaged every hour over a day with re-sampling for a joint evaluation of the data gathered at specific hours of the day (e.g., solid blue line at the bottom graph) with $\tau_1 = 1$ day, which provides $\sigma_1 = 60$ min, $f_1 = \frac{1}{60}$ min^{-1}, $\theta_1 = 60$ min, and $n_1 = \frac{1,440}{60} = 24$. As a result, $\mathbf{X}_1$ is a $131,040 \times 1$ matrix, $\mathbf{X}_2$ is a $1,440 \times 13$ matrix, and $\mathbf{X}_3$ is a 24×1 matrix.

Analysis description and results

This section describes the analysis conducted to identify and characterize distinct abnormal behavioural patterns associated with electricity production and consumption over time. Next, it presents examples of detection and diagnosis of abnormal power consumption patterns using the multilayer PCA strategy presented in Section 2.6 with OpenLV data recorded on November 02, 2018.

The multilayer PCA strategy is aimed at identifying and characterizing distinct abnormal behaviours associated with electricity production and consumption over time, from fast load behavioural changes at minute resolution (which can be seen in the reactive power measurements, denoted by Q, changing from/to inductive to/from capacitive) to slow profile changes at day resolution (which can be seen in the variations in the active power measurements, denoted by P). The layers and problems of interest are defined accordingly in Table 2.10, together with their corresponding f_k, θ_k, σ_k, τ_k, and the input data. The analysis relies on OpenLV data recordings from July 17, 2018 to June 30, 2019. Since $f_{raw} = 1$ min^{-1}, only events that last a few minutes or longer are visible in the input data, which represent single phase quantities acquired separately at the three phases of each substation. In accordance to the SAIDI values in the EU [26], this analysis considers that abnormal phenomena should last no longer than 5 % of the total duration of the analysis.

It is noteworthy that not all measured quantities are required for the evaluation of a specific problem of interest. For instance, phase voltage and line current magnitudes are better indicators of power quality problems, whereas active and reactive power and energy are better indicators of energy behavioural patterns. To some extent, the power and energy measurements contain redundant information, as the energy quantities derive from their corresponding power quantities. Based on the measurement-based techniques summarized in [14], and considering that power is a better indicator of average behaviours over time than energy, by definition, the usage of active and reactive power is adopted for the problems of interest defined in Table 2.9 and further in Table 2.10. The active and reactive power measurements are further adjusted to the time scales shown in Fig. 2.1.

Layer 1 does not require *a priori* re-sampling or multiway re-arrangement due to its short duration; in contrast, layers 2 and 3 do, for their long duration. Thereby, in layers

2 and 3, the IED data are averaged (over an hour in layer 2 and over a day in layer 3) to represent a single measurement with a lower sampling rate, and further re-arranged multiway in layer 3 to evaluate repetitive patterns on a specific day of the week. This previous data organization procedure is necessary whenever the data inputs have to be adjusted to a specific problem of interest.

The PCA models are built on the fly, as soon as new data are ready for analysis over τ_k at the ℓ layers described in Table 2.10. Moreover, power system operation over a day is further divided in two periods, before noon and after noon, as a typical day presents two different load peaks, one in the morning and the other in the afternoon (see Fig. 2.15). This choice is made to catch these two daily load peaks in separate within the statistical models, such that the evaluation of one peak does not interfere with the other, whenever suitable. This procedure is adopted in layers 2 and 3, which contain half the number of observations of the whole day.

A confidence level $\alpha = 0.95$ is chosen for the whole analysis, as it results in a few observations surpassing (2.13) and/or (2.15). This means that the conclusion reached by the experiment will actually be wrong (that is, result in false positives or false negatives) in 5% of the tests. Further investigation is required to discard false positives and negatives (i.e., wrong event detections and missed event detections). For instance, in layer $k = 1$, there are $n_1 = 60$ observations, of which 3 are expected to be wrongly classified; in $k = 2$, there are $n_2 = 84$ observations, of which at least 4 are expected to be wrongly classified; and in $k = 3$, there are $n_3 = 26$ observations, of which at least 1 is expected to be wrongly classified.

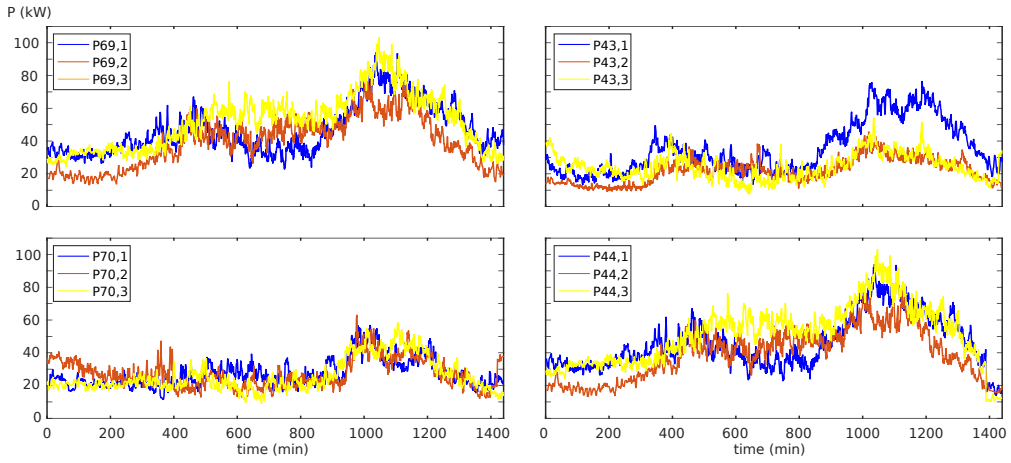

Figure 2.15: Case study – Single-phase daily active power profiles on November 02, 2018

For layers 1 to 3 of Table 2.10, anomaly detection results are displayed in Figs. 2.16 to 2.18. Contributions in terms of T^2 and SPE statistics for the specific events highlighted in Figs. 2.16 to 2.18 are illustrated graphically in Figs. 2.19 to 2.21 in the projection

subspace (top chart) and residual subspace (bottom chart), respectively. Although there are a few points outside the square area in Figs. 2.16 to 2.18, this section focuses on the occurrences highlighted in Figs. 2.16 to 2.18 to illustrate the methodology. All graphs display the statistical thresholds calculated individually for each substation variable (i.e., active power P or reactive power Q at phases 1, 2, and 3 of substations 43, 44, 69, and 70), represented by the solid black line, together with the contributions of each substation variable to the calculated values of T^2 and SPE, represented by the column charts. Although there are more abnormal observations in Figs. 2.16 to 2.18 than those highlighted, the analysis focuses on those points in particular.

In these examples, the individual contributions that violate the statistical thresholds of T^2 indicate that the reactive power variables $Q_{43,1}$, $Q_{43,3}$, $Q_{44,1}$, and $Q_{69,1}$ are the main involved in the statistically abnormal behaviour in Fig. 2.19 from 07:05 PM to 07:10 PM; and that the active power variables $P_{69,1}$, $P_{69,2}$, and $P_{69,3}$ are involved in the abnormal behaviour in Fig. 2.20 and that none of the IEDs is involved in Fig. 2.21. Further evaluation of the network topology suggests that constraint violations with T^2 statistics are probably due to the solar PV panels connected to the feeders of the substations involved in the occurrence followed by an increase in electricity consumption. In fact, the presence of inverter-based generation changes the behaviour of the network from inductive to capacitive over a few minutes (Fig. 2.19) while increasing the injection of active power in the network (Fig. 2.20), whereas an increase in electricity consumption, reflected in the active power demand, is related to changes in the standard operation of the grid over longer intervals. In turn, the individual contributions that violate the statistical thresholds of SPE indicate that the reactive power variables $Q_{43,1}$, $Q_{70,1}$, and $Q_{70,3}$ are the main involved in the abnormal behaviour in Fig. 2.19 from 07:05 PM to 07:10 PM; that the active power variable $P_{69,3}$ is involved in the abnormal behaviour in Fig. 2.20; and that the active power variables $P_{69,1}$, $P_{69,2}$, and $P_{69,3}$ gathered on November 02, 2018 are involved in the abnormal behaviour in Fig. 2.21. Further evaluation of the network topology shows that constraint violations with SPE statistics are due to a high energy consumption within the coverage area of the IEDs involved in the occurrence.

2.6.4 Discussion

The results of Section 2.6.3 show that the multilayer PCA strategy presented in Section 2.6 allows for detection and isolation of different types of abnormalities in energy production and consumption patterns in parallel, taking into consideration the time ranges of distinct phenomena or problems of interest while respecting the temporal hierarchy of decisions (see Fig. 2.1). As an outcome, multilayer PCA is more flexible and scalable than traditional multiway PCA when it comes to the handling of distinct time scales, variables, and layers of interest.

The data organization in multilayer PCA enables to apply the same data set to identify different types of abnormal behaviours more effectively. This procedure provides an appropriate sampling rate so that a specific set of variables can be used over the length of time associated with a given problem of interest with good performance. On the top

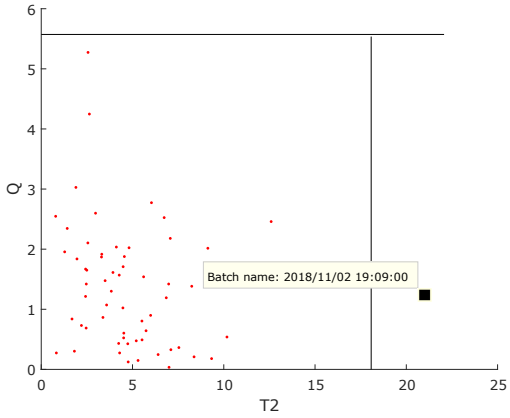

Figure 2.16: Results – SPE vs. T^2 at layer 1: minutes between 07:00 PM and 07:59 PM on November 02, 2018

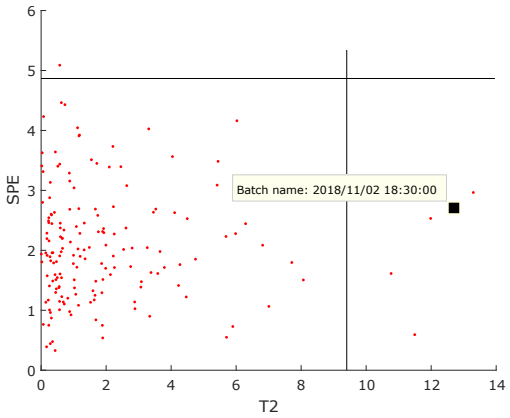

Figure 2.17: Results – SPE vs. T^2 at layer 2: hours from October 29, 2018 to November 04, 2018

of this, the results obtained with the data organization procedure show the importance of selecting an appropriate sampling according to the length of time of the layer so that different types of abnormal behaviours can be detected with an adequate representation over time. If this procedure was not used, the results detected at the longest layers would contain redundant information.

In the case study described in Section 2.6.3, variations in the normal operating conditions are attributed to the uncertain, intermittent nature of the electricity production from solar PV panels, whereas other perturbations are attributed to abnormal energy consumption patterns. Consequently, events detected with T^2 statistics are related to the energy generated by the solar PV panels, whereas events detected with SPE statistics are related to the energy consumption. As a matter of a fact, T^2 measures the distance of the projected data to the centre of the model and is expected to present larger values associated with variations in the normal operating conditions, whereas SPE measures the

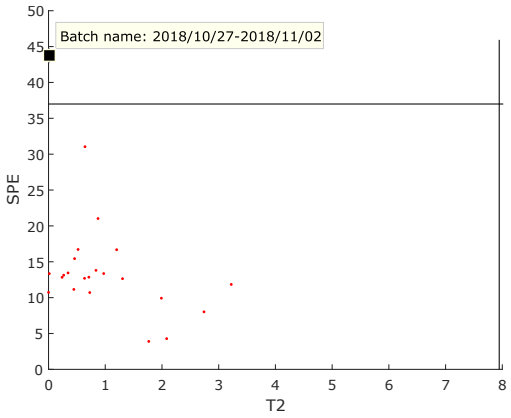

Figure 2.18: Results – SPE vs. T^2 at layer 3: afternoons of the weeks in the 2018 Autumn term

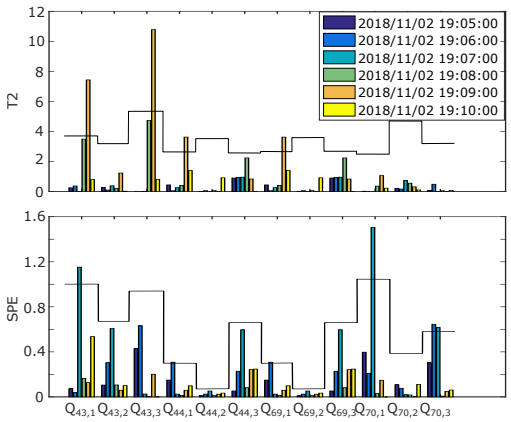

Figure 2.19: Results – Contribution analysis of T^2 (top) and SPE (bottom) at layer 1: load behaviour

distance of the observation to the projection subspace and is expected to present larger values associated with changes in the correlation structure of the observations. In any case, it is recommended to proceed with a further comprehensive investigation of any results detected with any of the T^2 or SPE statistics, as they produce violation of constraints related to major concerns about power system operation in different ways.

However, it is noteworthy that any types of abnormal behaviours can be identified only if they produce changes in the measured quantities that last enough to be recorded. This principle applies to parts of the grid that are directly or indirectly monitored by IEDs. To this extent, the results also show that it is possible to zoom in and out of abnormal energy production and consumption patterns through this multilayer PCA and thereby associate distinct problems of interest over time whenever they are identified at different layers. In addition, it can be noticed that, the shorter duration of the layer and the higher sampling rate of the data, the more principal components are needed to express the same

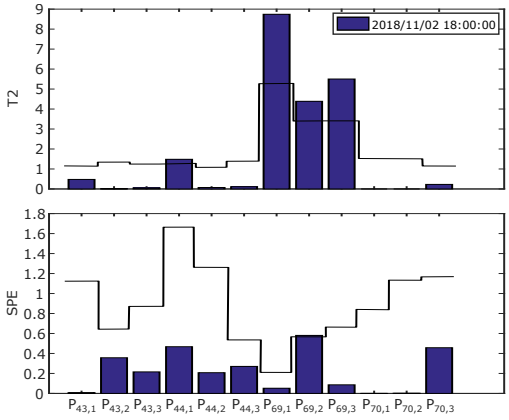

Figure 2.20: Results – Contribution analysis of T^2 (top) and SPE (bottom) at layer 2: hourly power changes

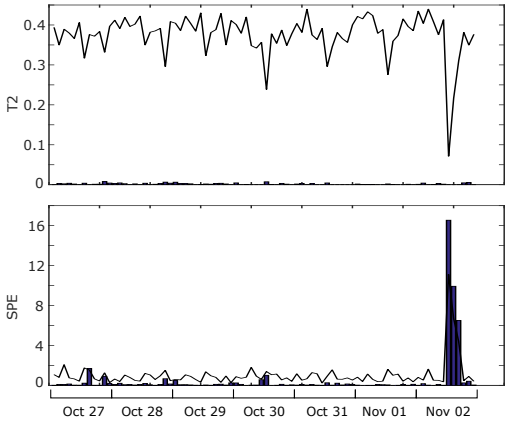

Figure 2.21: Results – Contribution analysis of T^2 (top) and SPE (bottom) at layer 3: daily power changes

information. In fact, the analysis requested $r_1 = 3$ hourly, $r_2 = 2$ weekly, and $r_3 = 1$ biannually.

Furthermore, the procedure of isolation allows to identify the most probable substations, lines, and electrical quantities responsible for the abnormal behaviour, as it computes the influence of each variable – referred to an electrical quantity of a substation – in the calculated values of the T^2 and SPE statistics. Consequently, information about the network topology and its electrical parameters and energy appliances are helpful for the correct identification of probable locations and causes of the abnormal behaviour.

2.7 Final Remarks

This chapter presented data-driven methods for event detection in power systems, with applications to modelling and monitoring tasks. The presented methodologies are applicable to any power transmission and distribution networks supervised by WAMS, IED, PMU, and/or SM infrastructures and are particularly relevant to identify minor events that would be missed otherwise when screening for information in large amounts of data. Despite that, these methods do not aim at replacing conventional power protection systems based on well-established methods and physics principles.

Table 2.2: Results – PCA-based event isolation with a 5 s window

Event	T^2 statistics	SPE statistics
1	UTPan (100.0%)	UTPan (100.0%)
2	UTPan (100.0%)	UTPan (67.2%)
		McDonald (32.8%)
3	McDonald (100.0%)	McDonald (62.8%)
		Waco (16.6%)
		Harris (12.8%)
		Austin (7.8%)
4	UTPan (100.0%)	UTPan (100.0%)
5	UTPan (100.0%)	UTPan (100.0%)
6	McDonald (100.0%)	McDonald (100.0%)
7	McDonald (74.3%)	McDonald (61.5%)
	Waco (15.5%)	UTPan (38.5%)
	UTPan (10.2%)	
8	Waco (100.0%)	Waco (43.8%)
		Austin (39.3%)
		UT3 (16.9%)
9	UTPan (100.0%)	UTPan (100.0%)
10	UTPan (100.0%)	Waco (40.6%)
		McDonald (38.1%)
		UTPan (21.3%)
11	UTPan (100.0%)	McDonald (29.6%)
		UTPan (70.4%)
12	UTPan (59.5%)	UTPan (42.2%)
	Harris (25.0%)	McDonald (28.3%)
	Waco (15.5%)	Waco (24.0%)
		Austin (5.5%)
13	UTPan (100.0%)	Waco (57.1%)
		McDonald (16.3%)
		UTPan (26.6%)
14	UTPan (100.0%)	McDonald (46.5%)
		UTPan (34.9%)
		Waco (18.6%)
15	McDonald (100.0%)	McDonald (76.9%)
		UTPan (23.1%)
16	UTPan (100.0%)	UTPan (63.1%)
		UT3 (36.9%)

Table 2.3: Case study – Types of events detected with signal-based methods

time (s)	Location	Category
435	McDonald	Impulse, multiple
445	Waco	Step change
565	McDonald	Impulse, single
645	McDonald	Impulse, single
945	McDonald	Impulse, single
1180	McDonald	Step change
1250	McDonald	Impulse, single
1530	McDonald	Impulse, single
1600	McDonald	Step change
1670	McDonald	Impulse, single
1795	McDonald	Impulse, single
1875	Waco	Step change
2320	McDonald	Impulse, single
2355	Waco	Impulse, single
2700	McDonald	Impulse, single
2715	McDonald	Impulse, single
2735	McDonald	Impulse, multiple
2770	Waco	Step change
2910	McDonald	Impulse, single
3575	McDonald	Impulse, single

Table 2.4: Results – Event detection over a 10 s window without noise

time (s)	FFT	Yule-Walker	Min-Max	Derivative	PCA	SVM
435	Yes	No	Yes	No	Yes	Yes
445	No	Yes	No	No	Yes	No
565	No	No	Yes	Yes	Yes	Yes
645	No	No	No	Yes	Yes	Yes
945	Yes	No	Yes	No	Yes	Yes
1180	Yes	No	Yes	No	Yes	Yes
1250	Yes	No	Yes	No	Yes	Yes
1530	No	No	Yes	Yes	Yes	Yes
1600	Yes	Yes	No	No	Yes	Yes
1670	No	No	Yes	No	Yes	Yes
1795	No	No	No	Yes	Yes	Yes
1875	No	Yes	Yes	No	Yes	No
2320	No	No	No	Yes	Yes	Yes
2355	No	No	No	Yes	Yes	Yes
2700	Yes	No	Yes	No	Yes	Yes
2715	Yes	No	No	No	Yes	Yes
2735	Yes	No	No	No	Yes	Yes
2770	No	Yes	Yes	No	Yes	No
2910	No	No	No	Yes	Yes	Yes
3575	No	No	Yes	Yes	Yes	Yes

Table 2.5: Results – Event detection over a 10 s window with noise

time (s)	FFT	Yule-Walker	Min-Max	Derivative	PCA	SVM
435	No	No	No	No	No	Yes
445	No	Yes	No	No	Yes	No
565	No	No	No	No	No	Yes
645	No	No	No	No	No	Yes
945	No	No	No	No	Yes	Yes
1180	No	No	No	No	No	Yes
1250	No	No	No	No	Yes	Yes
1530	No	No	No	No	Yes	Yes
1600	Yes	No	No	No	Yes	No
1670	No	No	No	No	Yes	Yes
1795	No	No	No	No	No	Yes
1875	No	Yes	Yes	No	No	No
2320	No	No	No	No	Yes	Yes
2355	No	No	No	Yes	Yes	No
2700	No	No	No	No	Yes	Yes
2715	No	No	No	No	Yes	Yes
2735	No	No	No	No	Yes	Yes
2770	No	Yes	Yes	No	No	No
2910	No	No	No	No	Yes	Yes
3575	No	No	No	No	No	Yes

Table 2.6: Case study – Types of events detected with distinct PCA techniques

#	Location	Category
1	Edinburg	Impulse, multiple
2	Edinburg	Transient
3	Fort Davis	Impulse, single
4	Edinburg	Impulse, multiple
5	Edinburg	Transient
6	Fort Davis	Impulse, single
7	Fort Davis	Impulse, single
8	Waco	Transient
9	Edinburg	Transient
10	Edinburg	Impulse, multiple
11	Edinburg	Impulse, multiple
12	Edinburg	Impulse, multiple
13	Edinburg	Transient
14	Edinburg	Transient
15	Fort Davis	Impulse, single
16	Edinburg	Impulse, multiple

Table 2.7: Results – PCA-based event detection over a 10 s window

	Kaiser			Scree plot			Variance			VRE			Detectability			
#	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	r	Q
1	3	Y	Y	4	Y	Y	3	Y	Y	3	Y	Y	2	Y	2	Y
2	2	N	Y	3	N	Y	3	N	Y	1	N	Y	4	Y	1	Y
3	2	Y	Y	2	Y	Y	3	N	Y	2	Y	Y	4	Y	1	Y
4	2	Y	Y	2	Y	Y	3	Y	Y	1	Y	Y	2	Y	2	Y
5	3	Y	N	3	Y	N	3	Y	N	2	Y	Y	3	Y	1	Y
6	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	N	2	Y
7	2	N	Y	3	N	Y	3	N	Y	2	N	Y	2	N	2	Y
8	2	N	Y	3	Y	N	2	N	Y	2	N	Y	3	Y	3	N
9	3	Y	Y	2	N	Y	3	Y	Y	1	N	Y	3	Y	1	Y
10	2	Y	Y	2	Y	Y	2	Y	Y	1	N	Y	2	Y	2	Y
11	2	Y	Y	5	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
12	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
13	2	Y	Y	3	Y	N	3	Y	N	2	Y	Y	3	Y	1	Y
14	2	Y	Y	2	Y	Y	3	Y	Y	1	Y	Y	2	Y	2	Y
15	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
16	2	Y	Y	2	Y	Y	2	Y	Y	1	N	Y	2	Y	2	Y

Table 2.8: Results – PCA-based event detection over a 1 min window

	Kaiser			Scree plot			Variance			VRE			Detectability			
#	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	Q	r	T2	r	Q
1	3	Y	Y	4	Y	N	3	Y	Y	2	Y	Y	2	Y	2	Y
2	2	Y	Y	5	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
3	2	Y	Y	5	Y	N	3	Y	Y	1	Y	Y	2	Y	2	Y
4	2	Y	Y	2	Y	Y	3	Y	Y	1	Y	Y	2	Y	2	Y
5	2	Y	Y	2	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
6	2	Y	N	2	Y	N	3	Y	N	1	N	Y	2	Y	1	Y
7	2	Y	Y	3	Y	Y	3	Y	Y	2	Y	Y	4	Y	1	Y
8	2	N	N	3	N	N	3	N	N	1	N	N	2	N	2	N
9	3	Y	Y	2	N	Y	3	Y	Y	1	N	Y	3	Y	3	Y
10	2	N	Y	2	N	Y	3	Y	Y	1	N	Y	3	Y	3	Y
11	2	Y	Y	3	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
12	3	Y	Y	5	Y	Y	3	Y	Y	1	N	Y	3	Y	3	Y
13	2	N	Y	2	N	Y	3	Y	N	2	N	Y	3	Y	1	Y
14	2	Y	Y	3	Y	Y	2	Y	Y	2	Y	Y	2	Y	2	Y
15	2	Y	Y	3	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y
16	2	Y	Y	3	Y	Y	3	Y	Y	1	N	Y	2	Y	2	Y

Table 2.9: Case study – Layers and problems of interest

k	Problem of interest
1	Load behaviour
2	Hourly power changes
3	Daily power changes

Table 2.10: Case study – Layers, time lengths, and input data

k	Time lengths				Input			
	f_k	θ_k	σ_k	τ_k	data	n_k	m_k	p_k
1	1 min^{-1}	1 min	1 min	1 h	Q only	60	12	1
2	1 h^{-1}	1 h	1 h	1 week	P only	84	12	1
3	1 day^{-1}	1 week	1 week	6 months	P only	26	12	7

Chapter 3

Fault Location in Power Systems

This chapter describes state-of-the-art fault location techniques and presents the data-driven methods for fault location in power systems proposed in the articles derived from this thesis, further tested and demonstrated in different case studies.

3.1 Definitions

In this chapter, faults refer to short-circuits of different types and at different locations of the grid. In the event of a fault, fault location methods shall be able to determine where it occurred in the grid fast and accurately to enable restoration of service and ensure quality and continuity of supply at acceptable levels.

3.2 State-of-the art methods

Fault location techniques used in power systems can be classified as impedance-based, travelling-wave-based, knowledge-based, and hybrid approaches, described as follows. For a thorough explanation about specific deductions and implementations of these methods, refer to [81].

- Impedance-based methods rely on fundamental-frequency voltage and current measurements together with information about the network topology and electrical parameters to estimate the point of fault.
- Travelling-wave methods rely on high-frequency components and accurate timing information to estimate the point of fault with use of propagating wave principles.
- Knowledge-based methods rely on large amounts of historical data for extraction of the underlying knowledge of specific fault conditions and can be categorized as quantitative or qualitative, as described in the previous chapter.

- Hybrid approaches consist of a combination of the strategies previously described.

Among the existing techniques, impedance-based methods are the cheapest and easiest to implement, relying on fundamental-frequency voltage and current measurements together with information about the network parameters to estimate the faulted section of the network as a function of the fault distance and impedance. To this extent, there are several methods based on iterative calculations (e.g., [72, 59]) or relying on fault distance estimations (e.g., [8, 107]). However, they may be inaccurate over a range of scenarios, including varying loads, distributed generation, and complex network topologies [89]. On the other hand, travelling-wave methods are independent from the system conditions, but are inadequate to heterogeneous distribution networks and prohibitive in some settings due to the high sampling rates requested for good accuracy, the high number of equipment (transformers, capacitors, voltage regulators) connected to the network, and lateral branches that may produce several reflected waves [9]. These stumbling blocks make travelling-wave methods impractical to accurately locate faults. Alternatively, knowledge-based methods may help to increase the accuracy of fault location for their inner flexibility, but may be impractical when the quality of training or testing data is poor and/or the adaptation to different fault conditions is computationally too expensive. Consequently, the selection of the most suitable fault location strategy depends on the information available about the network topology, electrical parameters, and operating conditions plus measurement devices and investment budget.

3.2.1 Literature review

Fault location techniques in power systems have been extensively studied and improved for decades. In recent years, however, this topic has re-gained importance for the deregulation of the energy sector and the increasing penetration level of new technologies in the grid, such as distributed energy resources, inverter-interfaced energy appliances, and smart grid technologies. These structural changes alter the one-directional current flow and the behaviour of the grid under short-circuits, for which traditional fault location methods are no longer valid.

In this scenario, a number of methodologies have been developed to overcome the limitations of traditional fault location methods, relying on impedance-based algorithms (e.g., [11]), travelling waves (e.g., [28]), knowledge-based techniques (e.g., [50]), and hybrid approaches (e.g., [59]) updated to new situations. The increasing digitalization of electric power infrastructures has also contributed to the task, with the deployment of WAMS infrastructures combined with advanced communication and data processing capabilities [58, 36]. However, the usage of fault location methods based on computational intelligence [60] is limited due to the high complexity of implementation [76, 50] and may be impractical in some cases. As an option, dimensionality reduction techniques are easily implementable and reduce the amount of information required for the monitoring of the grid, but accuracy is limited to the correlations identifiable in the data. Moreover, most distribution networks do not have an advanced monitoring infrastructure because of the high implementation

cost of multiple measurement devices. As a result, existing applications at distribution level are limited to sparse estimation [36, 79], which may produce inaccurate results over a range of scenarios.

Regardless of the digital capabilities of the grid, other pieces of information may be used to improve efficiency and accuracy of fault location tasks. In particular, the incorporation of expert knowledge and historical data of previous occurrences allow for prior beliefs to be updated when new evidence is available, which may be helpful to reduce the search area and select candidate locations.

Among the probabilistic methods previously proposed, [105] investigated the contributions of fault events to system instability considering the stochastic nature of faults. In [77], Monte Carlo simulations were performed to determine the most vulnerable areas of the grid against specific fault events, using historical data to compute the trend of possible fault locations and types. In [108], a Bayesian approach was presented for joint outage identification and state estimation using the minimum probability of error. Likewise, when the pre-fault and/or post-fault conditions of the network are known or can be estimated, this information can be used to calculate the posterior probability of a specific fault in the grid. This calculation can be used to minimise the overall risk of a fault in the grid and reduce the impact of faults in vulnerable areas.

Bayesian approaches were also presented in [99, 94, 88, 29, 32, 47], among others. Notably, [99] relied on failure rates of grid components, voltage and current measurements, and information about digging activity to compute the posterior probability distribution of faulted sections, whereas [94] used field technician subjective opinion. Information about the status of the grid was used in [88] for line failure detection after a cyber-physical attack on the grid, in which some lines are disconnected and some measurements become unavailable; and in [29] for sparse estimation and subsequent fault location. Meanwhile, [32] presented an analytical method for modelling and evaluation of the impact of protection system failures on bulk power system reliability, considering different failure modes and the interactions among components; and [47] indicated that outages in the power grid are correlated with the geographical position of lines. Despite these efforts, a probabilistic approach capable of minimizing the overall risk of faults in power networks over a range of scenarios is still missing.

3.2.2 Contributions

The following articles derived from this thesis explore contributions to the state-of-the-art methods for fault location in power systems based on the incorporation of expert knowledge, historical data of previous occurrences, and/or simulations of representative scenarios.

- Souto, L. & Broda, J. (2020). Network-Informed Bayesian Approach to Minimize the Risk of Faults in Power Distribution Networks. Submitted to: International Journal of Electrical Power and Energy Systems.

- Souto, L. (2020). Fault Behaviour of Power Distribution Networks with Distributed Generation and Uncertainties. In 2020 IEEE PES Transmission & Distribution Conference and Exhibition Latin America (T&D-LA), Montevideo, 2020, pp. 1-5.
- Souto, L., Melendez, J. & Herraiz, S. (2020). Fault Location in Power Distribution Networks using Arbitrary Similarity Criteria in the Principal Component Subspace. In 2020 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 2020, pp. 1-5.
- Souto, L., Melendez, J. & Herraiz, S. (2020). Fault Location in Low Voltage Smart Grids Based on Similarity Criteria in the Principal Component Subspace. In 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 2020, pp. 1-5.

Section 3.3 presents PCA-based methods for feature selection and subsequent fault location based on similarity criteria, whereas Section 3.4 presents a procedure to estimate the fault behaviour of power systems with uncertainties based on extensive simulations. Finally, Section 3.5 presents a Bayesian approach to minimise the overall risk of faults in power distribution networks, which allows for updates in prior beliefs when new evidence is available, using information about the network topology, electrical parameters, and short-circuit currents. The case studies presented in the next sections are focused on MV and LV distribution networks, for their inner characteristics and limitations of existing methods.

3.3 Fault location based on similarity criteria in the principal component subspace

This section presents fault location methods relying on equivalent PCA models of the network obtained with the procedure described in Section 2.2.6. To compute the similarity criteria, the method compares a given testing scenario with a set of reference scenarios representing distinct operating conditions of the power network under consideration (i.e., including standard operation and short-circuits). Two calculation procedures are considered in the next sections: a weighed similarity criterion and an arbitrary similarity criterion.

3.3.1 Weighed similarity criterion

To compute a weighed similarity criterion, consider the set of training scenario $\mathcal{K}$ (with standard operation or faults at all network buses) and a generic testing scenario denoted by $k_{test} \notin \mathcal{K}$, possibly containing a fault or some deviation from the standard operating conditions, and let $\mathbf{V}_{k_{test}}$ and $\mathbf{\Lambda}_{k_{test}}$ be its eigenvector matrix and eigenvalue matrix, respectively. The choice of an appropriate value of r to reduce $\mathbf{V}_{k_{test}}$ into $\mathbf{P}_{k_{test}}$, $\forall k \in \mathcal{K}$,

is based on the calculation of a weighed cosine sum $\varphi_{k,k_{test}}$ of the dot product of $\mathbf{v_{j,k}}$ and $\mathbf{v_{j,k_{test}}}$ weighed by the normalized variance $\bar{\lambda}_{j,j,k_{test}}$, $j = 1, \cdots, r$, as in (3.1). This procedure also allows to identify the correct network configuration and operating condition of the testing scenario, since $\varphi_{k,k_{test}}$ should be close to 1.0 if $\mathbf{v_{j,k}}$ and $\mathbf{v_{j,k_{test}}}$, $j = 1, \cdots, r$, are similar.

$$\varphi_{k,k_{test}} = \sum_{j=1}^{r} |\mathbf{v_{j,k}} \times \mathbf{v_{j,k_{test}}}| \ \bar{\lambda}_{j,j,k_{test}} \tag{3.1}$$

with

$$\bar{\lambda}_{j,j,k_{test}} = \frac{\lambda_{j,j,k_{test}}}{\sum_{j=1}^{m} \lambda_{j,j,k_{test}}} \tag{3.2}$$

Once r is defined, if a fault occurred, further investigation is conducted to identify its possible locations. In this case, fault location and classification are performed by comparing the statistical model given by $\mathbf{P}_{k_{test}}$ and $\mathbf{\Lambda}_{k_{test}}$ with those computed for each reference scenario $k \in \mathcal{K}$. The results calculated with (3.1) are ranked in descending order. Finally, the fault buses $\bar{k}$ of the training scenarios with the highest values of (3.1) are identified as the most probable locations of the fault in the testing scenario.

3.3.2 Arbitrary similarity criteria

In this case, consider a generic testing scenario denoted by k_{test} with a few observations of an unknown operating condition and let $\mathbf{V}_{k_{test}}$ and $\mathbf{\Lambda}_{k_{test}}$ be its $m \times m$ eigenvector matrix and eigenvalue matrix, respectively. Under the assumption that the operating conditions of some scenarios $\bar{k}$ of all reference scenarios $\mathcal{K}$ are nearly the same as k_{test}, then $\mathbf{V}_{\bar{k}}$ and $\mathbf{V}_{k_{test}}$ are expected to be very similar. However, in the event of a fault whose impedance is different from the reference scenarios, the principal components whose loadings are not negligible in the variables involved in the fault are expected to deviate more than the others.

As a matter of a fact, the equivalent impedance $\bar{Z}_{eq}$ seen at the substation bus in the event of a fault is given by (3.3), where $\bar{Z}_{us}$ is the upstream equivalent impedance, $\bar{Z}_F$ is the fault impedance, and $\bar{Z}_{ds}$ is the downstream equivalent impedance.

$$\bar{Z}_{eq} = \bar{Z}_{us} + \frac{\bar{Z}_F \bar{Z}_{ds}}{\bar{Z}_F + \bar{Z}_{ds}} \tag{3.3}$$

In the event of a fault with $\bar{Z}_{F,test} \neq \bar{Z}_F$, where $\bar{Z}_F$ is the fault impedance of a reference scenario and $\bar{Z}_{F,test}$ is the fault impedance of k_{test}, the difference between the equivalent impedance calculated with (3.3) is given by (3.4).

$$\bar{Z}_{eq} - \bar{Z}'_{eq} = \frac{\left[\bar{Z}_F\left(\bar{Z}_{F,test} + \bar{Z}_{ds}\right) - \bar{Z}_{F,test}\left(\bar{Z}_F + \bar{Z}_{ds}\right)\right]\bar{Z}_{ds}}{\left(\bar{Z}_F + \bar{Z}_{ds}\right)\left(\bar{Z}_{F,test} + \bar{Z}_{ds}\right)} \tag{3.4}$$

Assuming that the fault impedance is very small in comparison with the downstream equivalent impedance, $\bar{Z}_F + \bar{Z}_{ds} \approx \bar{Z}_{ds}$ and $\bar{Z}_{F,test} + \bar{Z}_{ds} \approx \bar{Z}_{ds}$. Then (3.4) can be simplified as (3.5)

$$\bar{Z}_{eq} - \bar{Z}'_{eq} = \bar{Z}_F - \bar{Z}_{F,test} \tag{3.5}$$

Thereby, the principal components whose loadings are not negligible in the variables involved in the fault should not be selected to perform fault location. Thus, in the arbitrary criterion, dimensionality reduction is achieved by retaining $r < m$ principal components or columns of $\mathbf{V}_k$ and $\mathbf{V}_{k_{test}}$ which represent the major trends of the data set, excluding those with significant loadings in the variables involved in the fault. As a result, $\mathbf{V}_k$ is reduced with some loss of information to an $m \times r$ matrix $\mathbf{P}_k$ given by (3.6). It is noteworthy that the assembling of $\mathbf{P}_k$ is arbitrary and $\mathbf{p_i}_k \neq \mathbf{v_i}_k$ may hold, $i = 1, \cdots, r$.

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{p_{1,k}} & \cdots & \mathbf{p_{r,k}} \end{bmatrix} \tag{3.6}$$

with $\mathbf{p_i} = \begin{bmatrix} p_{1,i} & \cdots & p_{m,i} \end{bmatrix}^T$, $i = \{1, \cdots, m\}$

In the event of short-circuit operation in k_{test}, further investigation is required to locate the fault over all candidate scenarios $\bar{k}$ of $\mathcal{K}$ representing a given fault type and grid setting. The selection of r principal components is arbitrary and depends on the following aspects:

1. Network configuration: The selection of principal components may consider a single PMU or more PMUs, depending on the status of the switches in the network. The status of the switch is determined by calculating the dot product between $\mathbf{p_{i,\bar{k}}}$ and $\mathbf{p_{i,k_{test}}}$, which should be very close to the unity for all $i = 1, \cdots, m$ if the switch connecting the PMU is off. If the switch mode is off, then investigation proceeds with PMU data recorded at the substation where the fault is detected only; otherwise, if the switch mode is on, then investigation proceeds with PMU data recorded at all substations involved in the fault.

2. Fault type: the selection of r principal components is made separately for each reference scenario $\bar{k}$ with short-circuit operation and is based on the loadings of the principal components $\mathbf{p_{i,\bar{k}}}$ of $\mathbf{P}_{\bar{k}}$, $i = 1, \cdots, m$. The principal components whose relative weighs or loadings are minimum in the variables involved in the fault are selected to perform fault location, as they are expected to be very similar to the testing scenario k_{test} due to the same operating conditions.

The similarity between $\bar{k}$ and k_{test} (assembled with the principal components used in $\bar{k}$) can be computed through the dot product between the principal components $\mathbf{p_i}$ of those matrices, $i = 1, \cdots, m$, which is equal to the unity if $\mathbf{p_{i,\bar{k}}}$ and $\mathbf{p_{i,k_{test}}}$ are identical. This procedure allows to identify the operating condition of k_{test}, including the presence of faults and phases involved, as well as the grid setting of the network in the presence of switches.

Finally, the overall similarity criteria $\varphi_{\bar{k},k_{test}}$ is calculated in (3.7) as the sum of the dot product of all principal components in the resulting matrices $\mathbf{P}_{\bar{k}}$ and $\mathbf{P}_{k_{test}}$. Since r may be different for distinct reference scenarios, the sum should be averaged over r to ensure consistency of results. The results calculated with (3.7) are ranked in descending order such that the reference scenario with the highest $\varphi_{\bar{k},k_{test}}$ is the most probable location of the fault in k_{test}. If there is a tie between two or more buses, the similarity criteria proceeds with investigation of all the probable locations.

$$\varphi_{\bar{k},k_{test}} = \frac{\sum_{i=1}^{r} |\mathbf{p_{i,\bar{k}}} \cdot \mathbf{p_{i,k_{test}}}|}{r} \tag{3.7}$$

3.3.3 Case study: the RESOLVD project network

This section presents the results obtained with the weighed similarity criteria and the arbitrary similarity criteria. It presents the real-based power distribution network in which short-circuit simulations were performed and reports the fault location results obtained with different criteria at different buses of the network.

The short-circuit simulations were performed in MATLAB in a real-based LV power distribution network within the project RESOLVD, illustrated in Fig. 3.1. It represents a LV network located in Catalonia, Spain, which consist of primary distribution feeders with branches connecting the substation node to the customers (i.e., local energy producers or consumers).

Fig. 3.1 displays 2 distribution substations (SS-2 with a 250 kVA and SS-1 with a 630 kVA transformer Dyn11, 400 V secondary), 28 feeders drawn to scale modelled as short R-L lines (with $\frac{R}{X} = 5.4$ for overhead lines and $\frac{R}{X} = 2.7$ for underground cables), 1 switch, distributed generation from 4 solar photovoltaic (PV) modules (10 kW each), and 20 different energy consumers. Among the consumers, there are 1 industrial three-phase customer with 70 kW of contracted power ($\bar{S}_{40,A-B-C}$) and 19 residential single-phase customers with less than 10 kW of contracted power ($\bar{S}_{10,A}$, $\bar{S}_{10,C}$, $\bar{S}_{20,A}$, $\bar{S}_{20,C}$, $\bar{S}_{30,A}$, $\bar{S}_{50,C}$, $\bar{S}_{60,A}$, $\bar{S}_{60,C}$, $\bar{S}_{70,B}$, $\bar{S}_{80,B}$, $\bar{S}_{90,B}$, $\bar{S}_{100,C}$, $\bar{S}_{110,A}$, $\bar{S}_{120,C}$, $\bar{S}_{130,A}$, $\bar{S}_{140,B}$, $\bar{S}_{150,A}$, $\bar{S}_{150,C}$, $\bar{S}_{160,B}$). In total, the length of the primary distribution feeder connecting SS-1 to SS-2 is 325 m and the length of the longest lateral branch is 95 m.

PMUs are installed at both substation nodes and sample phase voltage and line current phasor quantities from which the equivalent impedance is calculated, which provides $m = 2 \times 3 = 6$ variables in total in all scenarios. A statistical model is built for each PMU in separate, as they are supposed to represent systems that are possibly independent from

each other. This choice provides 2 distinct statistical models per scenario with $m = 3$ variables each. The statistical models are built with impedance magnitudes only to suit the linear algorithm in use. This choice is not expected to make a negative impact on the accuracy of fault location, despite the non-linear behaviour of the faults, since the network behaviour and loads are mainly resistive and the faults are purely resistive.

Historical data are obtained from simulations of standard operation and permanent short-circuits with typical hourly values of PV generation and load consumption profiles over a year. Considering a specific hour of the day, this choice provides $n = 365$ observations per reference scenario. Although these time intervals are chosen due to the real PV generation and load consumption profiles available for the simulations, a shorter or longer time interval can be used to build the statistical models without loss of generality. In addition, the reference fault scenarios include different types of faults applied at the substation and load buses with fault resistance $\bar{Z}_F = 0.001\Omega$ between phases and/or phases and ground. The fault impedance model is illustrated in Fig. 3.2. Both switch statuses on and off are considered in all training and testing scenarios.

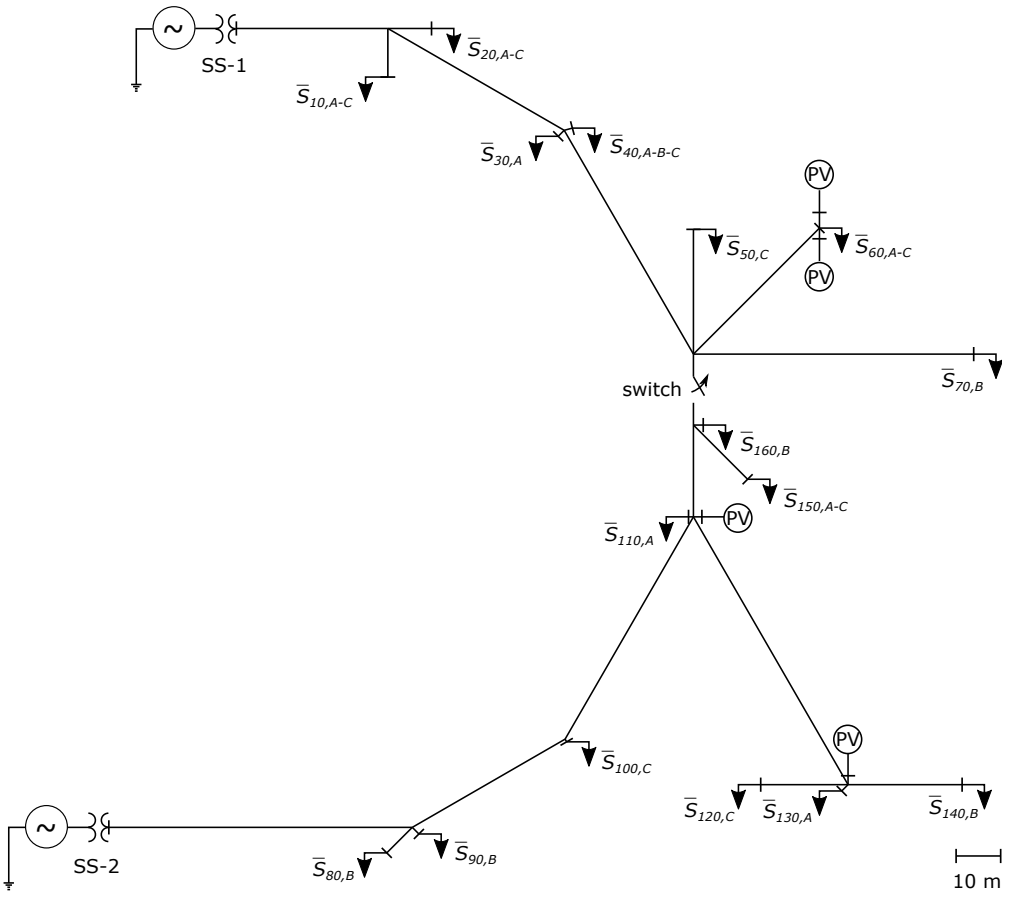

Figure 3.1: Case study – Single-line diagram of the LV distribution network

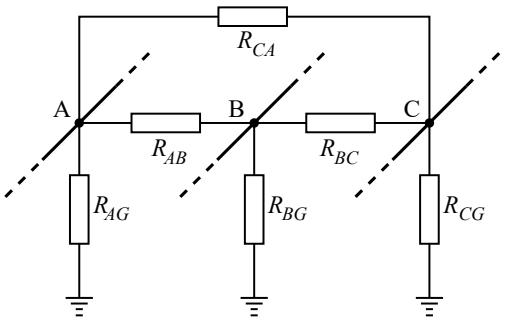

Figure 3.2: Case study – Fault impedance equivalent circuit

3.3.4 Results: weighed similarity criteria

The testing scenarios of this section consist of variations in the PV generation under normal operation, reduced by 25%, 50%, 75%, and 100% of the standard operation profiles, and three-phase (ABC) symmetrical faults with fault resistance $\bar{Z}_F = 1\Omega$ simulated at the load buses at midday, considering the same PV generation and load consumption profiles as the reference scenarios. In total, 32 fault scenarios are tested, since buses 10, 20, 60, and 150 contain two households each (i.e., 16 faulty buses times 2 switch modes), besides 4 scenarios with variations in PV generation profiles under normal operation.

The deviations computed with (3.1) under normal operation with variations in the PV generation profiles are displayed in Table 3.1, considering standard operation and faults at the substation buses. Meanwhile, the fault location results obtained with the switch off and on are displayed in Tables 3.2 and 3.3, respectively. These tables describe the fault scenario k_{test} (i.e., faulty bus) in the first column and the results obtained with the similarity criteria in the principal component subspace in the next columns, including the correct identification of the grid setting in the second column (regarding the switch mode and part of the network where the fault is); the faulty bus $\bar{k} \in \mathcal{K}$ determined by the sum of dot products calculated with (3.1) for all candidate buses, given information about the right network setting, in the third column; the distance error between k_{test} and $\bar{k}$ in the fourth column; and the r principal components used to compute (3.1) for all candidate scenarios in the fifth column.

3.3.5 Results: arbitrary similarity criteria

The testing scenarios of this section consist of permanent three-phase (ABC), phase-ground (AG), phase-phase (BC), and phase-phase-ground (BCG) faults simulated at the load buses, considering the same PV generation and load consumption profiles as the reference scenarios. Different values of fault resistance are considered as a function of the fault type: $\bar{Z}_F = 0.1\Omega$ and $\bar{Z}_F = 1.0\Omega$ between phases (i.e., R_{AB}, R_{BC}, R_{CA} in Fig. 3.2) and/or phases and ground (i.e., R_{AG}, R_{BG}, R_{CG} in Fig. 3.2), whenever applicable. In total, 256 fault scenarios are tested per hour (i.e., 16 faulty buses times 4 fault types times

Table 3.1: Results – Deviations computed with the weighed criterion with variations in the PV generation

Testing scenario	Reference scenario					
	Std. operation		Fault SS-1		Fault SS-2	
Switch mode	off	on	off	on	off	on
75% PV switch off	0.8254	0.6126	0.0350	0.5056	0.0381	0.1426
75% PV switch on	0.7567	0.9996	0.6860	0.8864	0.1442	0.3411
50% PV switch off	0.8244	0.6355	0.0725	0.5271	0.0264	0.1179
50% PV switch on	0.7557	0.9995	0.6865	0.8856	0.1408	0.3435
25% PV switch off	0.7954	0.6385	0.1065	0.5340	0.0146	0.0865
25% PV switch on	0.7558	0.9995	0.6863	0.8852	0.1398	0.3435
0% PV switch off	0.7944	0.6587	0.1472	0.5491	0.0040	0.0651
0% PV switch on	0.7512	0.9991	0.6908	0.8834	0.1340	0.3505

2 fault resistance values times 2 switch modes). In a desktop setting, calculation times are between 0.025 s and 0.050 s for all testing scenarios, whereas the file size of all reference scenarios together for each fault type is 20 kB.

The fault location results computed under different fault operating conditions at midday are displayed in Tables 3.4 to 3.11, obtained for three-phase, phase-ground, phase-phase, and double-phase-ground faults with $R_F = 0.1\Omega$ and $R_F = 1.0\Omega$ at different buses (i.e., fault scenarios k_{test}) with the switch off and on. The faulty bus k_{test} is displayed in the first column and the results obtained with the similarity criteria in the principal component subspace are shown in the following columns, including the calculated $\bar{k}$ determined by considering the sum of dot products calculated with (3.7) for all candidate buses, given information about the right network setting, and the cable length between k_{test} and $\bar{k}$. As the results vary from simulation to simulation, the cable length between k_{test} and $\bar{k}$ is regarded as a good indicator of how far the calculated point of fault $\bar{k}$ is from the actual point of fault k_{test}.

3.3.6 Discussion

Overall, the results of 3.3.3 indicate that the multivariate statistical strategy is capable of distinguishing between faults, standard operation, and variations in the standard operating conditions correctly. Additionally, in the event of a fault, the methodology identifies switch status, the fault type, and the part of the network where it occurred correctly in all cases and the true location of the fault with good accuracy in most cases, regardless of the fault type, location, and impedance of k_{test}.

Considering the weighed similarity criterion, the fault location is identified correctly in 2 out of 16 scenarios when the switch is off and in 4 out of 16 scenarios when the switch is

Table 3.2: Results – Three-phase faults computed with the weighed criterion, switch off

Faulty bus	Grid setting	Distance errors		r
(k_{test})	(Right/wrong)	Bus ($\bar{k}$)	$\bar{k} - k_{test}$ [m]	
10	Right	10	0.0	2
20	Right	10	20.6	2
30	Right	40	3.8	4
40	Right	10	58.5	2
50	Right	10	140.7	4
60	Right	40	98.1	2
70	Right	40	121.0	2
80	Right	120	194.2	3
90	Right	160	122.1	2
100	Right	120	146.9	4
110	Right	120	90.9	4
120	Right	160	94.2	2
130	Right	150	109.2	4
140	Right	160	117.8	2
150	Right	120	89.6	2
160	Right	160	0.0	2

on, whereas the maximum distance error is 194.2 m when the switch is off (faulty bus 80 identified as 120 in Table 3.2) and 174.7 m when the switch is on (faulty bus 70 identified as 20 in Table 3.3). Despite the correct identification of the network setting, these errors are approximately the double of the length of the longest lateral branch and respectively stand for 59.8% and 53.8% of the total length from one substation to the other. Nevertheless, the actual faulty bus is among the first ranked results in almost all scenarios with both switch modes on and off when only the right network setting is considered, which evinces the importance of identifying the grid setting correctly before performing fault location with this methodology. Consequently, the maximum errors of this fault location procedure remain in the same part of the network delimited by the switch where the point of fault is. The average errors of all scenarios listed in Tables 3.2 and 3.3 stand for 27.0% and 23.0% of the total length from one substation to the other, which is less than the length of the longest lateral branch.

Considering the arbitrary similarity criterion, the calculated location of the fault is accurate in most cases, as the results displayed in Tables 3.4 to 3.11 represent the exact point of fault or another bus in the same branch of the network in most cases. On the top of this, the calculation times and storage space requirements show the feasibility of an online implementation of the method for detection and location of different faults in a small LV network. However, the results indicate diverse effects on the fault location

Table 3.3: Results – Three-phase faults computed with the weighed criterion, switch on

Faulty bus	Grid setting	Distance errors		r
(k_{test})	(Right/wrong)	Bus ($\bar{k}$)	$\bar{k} - k_{test}$ [m]	
10	Right	10	0.0	2
20	Right	20	0.0	2
30	Right	20	57.5	4
40	Right	20	57.5	2
50	Right	20	139.7	4
60	Right	20	151.8	3
70	Right	20	174.7	2
80	Right	110	107.2	2
90	Right	110	101.2	2
100	Right	110	60.5	4
110	Right	110	0.0	4
120	Right	150	109.2	2
130	Right	130	0.0	4
140	Right	150	132.8	2
150	Right	100	79.8	2
160	Right	110	24.7	3

errors with an increase in the fault resistance. In many cases, the errors remain roughly the same; in others, however, they increase or decrease considerably. This is attributed to the small impedance of the line sections (few mΩ), to the statistical models of the network (main features) under different operating conditions, and also to the loadings of the variables involved in the fault in the principal components subspace (not negligible in a few reference scenarios). To this extent, the results obtained with fault resistance values greater than 1 Ω are expected to be considerably less accurate than the results calculated with small fault resistance values due to the equivalent resistance seen at the substation bus.

It is noteworthy that the ranking of results calculated with (3.1) and (3.7) may be inaccurate over a range of scenarios due to the differences between the training and testing scenarios in use, such as variations in the standard operating conditions, different faults, inaccurate representations of the network provided by the statistical models in use, poor quality of data, among others. Moreover, the equivalent impedance seen at the substations is almost the same for faults at different buses in the same part of the network and a fixed grid configuration, which poses a challenge for distinguishing between buses with similar load behaviours. As an option, the method could be improved by including more reference scenarios in the training data sets, with different faults, time scales, operating conditions, additional information about the network topology, more measurement devices at different

Table 3.4: Results – Three-phase faults computed with the arbitrary criterion, switch off

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	10	0.0	40	58.5
20	10	20.6	40	57.5
30	40	3.8	10	58.5
40	40	0.0	40	0.0
50	40	75.4	40	75.4
60	50	68.0	40	87.5
70	40	110.5	40	110.5
80	80	0.0	80	0.0
90	80	10.2	80	10.2
100	100	0.0	100	0.0
110	160	24.7	120	90.9
120	120	0.0	120	0.0
130	120	21.5	120	21.5
140	120	45.1	120	45.1
150	160	19.3	160	19.3
160	160	0.0	160	0.0

locations, data from different sources, etc.

3.4 Fault behaviour of power systems with uncertainties

This section evaluates the fault behaviour of power systems with uncertainties. It describes the initial assumptions, the algorithms used in short-circuit calculations, and the modelling of uncertain parameters.

3.4.1 Assumptions

This analysis is valid for any power transmission and distribution systems, but especially convenient for power distribution networks with variable loads and distributed energy resources along their feeders. It requests information about the network topology and electrical parameters of the feeders, stored in the utility's database, estimations of the operating status of loads, distributed energy resources, and fault impedance parameters, and uncertainties associated with the voltage and current phasor quantities measured at

Table 3.5: Results – Three-phase faults computed with the arbitrary criterion, switch on

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	20	20.6	10	0.0
20	20	0.0	20	0.0
30	30	0.0	20	57.5
40	40	0.0	20	57.5
50	50	0.0	50	0.0
60	50	68.0	40	87.5
70	50	90.9	40	110.5
80	80	0.0	80	0.0
90	90	0.0	80	10.2
100	110	60.2	90	42.5
110	110	0.0	130	72.2
120	140	45.1	130	21.5
130	130	0.0	130	0.0
140	140	0.0	130	27.5
150	160	19.3	150	0.0
160	160	0.0	150	19.3

the substation bus. It is based on phase components to deal with unbalanced networks, which implies that all phasor quantities and admittance matrices are expressed in terms of circuit phases. Essentially, the algorithm computes the equivalent admittance matrix of the network, then calculates the fault current at each line section under investigation to determine the fault behaviour of the network over a range of scenarios.

3.4.2 Algorithm

The methodology consists of an optimization via simulations approach conducted in OpenDSS. Thereby, the algorithm implemented in this software is briefly introduced in the following paragraphs.

First, the algorithm computes the nodal admittance matrix of the network, denoted by $\bar{Y}$, representing a specific operating condition of the network previously defined in the input data. This procedure includes all loads as admittances, as well as the equivalent admittance of all generation buses. Thus, all generation buses (i.e., slack bus and connection points of distributed generation units) are converted to their Thevenin equivalents. Then, the equivalent system is described by (3.8), where $\dot{I}$ denotes the vector of current injections into the network and $\dot{V}$ is the vector of node voltage to the ground.

Table 3.6: Results – Phase-ground faults computed with the arbitrary criterion, switch off

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	10	0.0	20	20.6
20	10	20.6	20	0.0
30	40	3.8	40	3.8
40	40	0.0	50	75.4
50	50	0.0	50	0.0
60	40	87.5	50	68.0
70	50	90.9	50	90.9
80	80	0.0	80	0.0
90	90	0.0	90	0.0
100	100	0.0	100	0.0
110	160	24.7	100	60.2
120	140	45.1	100	153.3
130	120	21.5	100	135.7
140	140	0.0	100	159.2
150	160	19.3	100	96.0
160	160	0.0	100	81.1

$$\dot{I} = \bar{Y}\dot{V} \tag{3.8}$$

Next, a direct solution of $\bar{Y}$ is calculated with source injections and generator injections. This procedure allows for computation of the resulting open-circuit voltage vector $\dot{V}_{OC}$ and short-circuit admittance matrix $\bar{Y}_{SC}$, from which the short-circuit current vector $\dot{I}_{SC}$ can be calculated. The short-circuit currents are then computed at each bus using the Thevenin equivalent model as in (3.9).

$$\dot{I}_{SC} = \bar{Y}_{SC}\dot{V}_{OC} \tag{3.9}$$

The short-circuit currents and fault distances representing a specific operating condition of the network are then saved. The algorithm proceeds with an update in the estimations of uncertain parameters, described as follows in Section 3.4.3, until all operating scenarios of interest are evaluated. A flowchart of the proposed algorithm is shown in Fig. 3.3.

Table 3.7: Results – Phase-ground faults computed with the arbitrary criterion, switch on

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	10	0.0	20	20.6
20	20	0.0	20	0.0
30	40	3.8	30	0.0
40	40	0.0	50	75.4
50	50	0.0	40	75.4
60	50	68.0	60	0.0
70	70	0.0	50	90.9
80	90	10.2	90	10.2
90	90	0.0	90	0.0
100	100	0.0	100	0.0
110	140	96.9	110	0.0
120	160	111.8	140	45.1
130	160	94.2	140	27.5
140	160	117.8	140	0.0
150	160	19.3	110	39.7
160	110	24.7	140	117.8

3.4.3 Estimations and uncertainties

The algorithm previously described may be computed over a range of scenarios representing variations of the fault operating conditions of the network. Thus, consider that the exact operating status of loads, distributed energy resources, and fault impedance parameters are unknown, but can be estimated over a range of values. In addition, assume that the phasor quantities may present inaccuracies caused by measurement errors and/or noise. Such uncertainties are described in the following paragraphs.

The loads of the network may be a combination of constant-impedance, constant-current, and constant-power models (that is, ZIP loads). The load estimation depends on its corresponding electrical behaviour, represented by (3.10) for a constant-impedance model, by (3.11) for a constant-current model, and by (3.12) for a constant-power model.

$$\bar{Y_L} = \frac{\dot{I}_L}{\dot{V}_L} = constant \longrightarrow \begin{cases} \bar{Y_L} = constant \\ \dot{I}_L \propto \dot{V}_L \end{cases} \tag{3.10}$$

$$\dot{I}_L = \bar{Y_L}\dot{V}_L = constant \longrightarrow \begin{cases} \bar{Y_L} \propto \dot{V}_L^{-1} \\ \dot{I}_L = constant \end{cases} \tag{3.11}$$

Table 3.8: Results – Phase-phase-ground faults computed with the arbitrary criterion, switch off

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	20	20.6	20	20.6
20	20	0.0	10	20.6
30	60	87.5	10	58.5
40	50	75.4	10	58.5
50	50	0.0	50	0.0
60	50	68.0	60	0.0
70	50	90.9	50	90.9
80	80	0.0	80	0.0
90	80	10.2	80	10.2
100	100	0.0	100	0.0
110	120	96.9	100	60.2
120	140	45.1	120	0.0
130	140	27.5	140	27.5
140	140	0.0	140	0.0
150	140	132.7	120	126.7
160	140	117.8	120	111.8

$$\bar{S}_{load} = \bar{Y}_L \left|\dot{V}_L\right|^2 = constant \longrightarrow \begin{cases} \bar{Y}_L \propto \dot{V}_L^{-2} \\ \dot{I}_L \propto \dot{V}_L^{-1} \end{cases} \tag{3.12}$$

In (3.10), $\bar{Y}_L$ is known and constant; in (3.11), $\dot{I}_L$ is known and $\bar{Y}_L$ may assume a range of values proportionally to $\dot{V}_L^{-1}$; in (3.12), the apparent power $\bar{S}_{load}$ is known and $\bar{Y}_L$ may assume a range of values proportionally to $\dot{V}_L^{-2}$. In turn, $\dot{I}_L$ is proportional to $\dot{V}_L$ in (3.10) and proportional to $\dot{V}_L^{-1}$ in (3.12). Considering that $\dot{V}_L$ is lower and upper bounded, that is, $\underline{\dot{V}_L} \leq \dot{V}_L \leq \overline{\dot{V}_L}$, the minimum and maximum values of $\bar{Y}_L$ and $\dot{I}_L$ in (3.10) to (3.12) can be determined.

In turn, the voltage and current phasor quantities at the connection points of distributed generation units depend on the type of generator and might be obtained from IEDs or estimated otherwise. For distributed generation units interfaced with power electronics, the output current is assumed to be constant and nearly equivalent to the rated current [17]. Thereby, considering a lower and upper bound for $\dot{I}_{DG}$ and $\dot{V}_{DG}$, $\underline{\dot{I}_{DG}} \leq \dot{I}_{DG} \leq \overline{\dot{I}_{DG}}$ and $\underline{\dot{V}_{DG}} \leq \dot{V}_{DG} \leq \overline{\dot{V}_{DG}}$ hold. In particular, for P-V generators, $\dot{V}_{DG}$ is fixed and thereby $\dot{I}_{DG}$ is proportional to the active power $\bar{P}_{DG}$ supplied to the grid, according to (3.13).

Table 3.9: Results – Phase-phase-ground faults computed with the arbitrary criterion, switch on

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	20	20.6	20	20.6
20	20	0.0	20	0.0
30	30	0.0	50	75.4
40	30	3.8	50	75.4
50	40	75.4	60	68.0
60	60	0.0	20	141.2
70	40	110.5	50	90.9
80	90	10.2	90	10.2
90	90	0.0	90	0.0
100	100	0.0	100	0.0
110	100	60.2	100	60.2
120	100	153.3	100	153.3
130	100	135.7	100	135.7
140	100	159.2	140	0.0
150	100	96.0	150	0.0
160	100	81.1	100	81.1

$$\bar{P}_{DG} = Real\left\{\dot{V}_{DG} \times \dot{I}_{DG*}\right\} \longrightarrow \bar{P}_{DG} \propto \left|\dot{I}_{DG}\right| \tag{3.13}$$

Although the fault admittance is assumed to be infinite in Section 3.4.2, a range of finite pre-specified values can be used to calculate the short-circuit currents at a given point of fault denoted by d. Considering a lower bound $\underline{\bar{Y}_F}$ and upper bound $\overline{\bar{Y}_F}$, $\underline{\bar{Y}_F} \leq \bar{Y}_F \leq \overline{\bar{Y}_F}$ holds. Alternatively, the same fault impedance $\overline{\bar{Y}_F}$ can be used to compute the short-circuit currents at different locations along the feeders such that $0 \leq d \leq d_{max}$, where d_{max} is the length of the feeder.

In addition, the voltage and current phasor quantities at the substation bus, denoted by $\dot{V}_S$ and $\dot{I}_S$, may be inaccurate due to measurement errors and/or noise, with uncertainties denoted by $\Delta\dot{V}_S$, $\Delta\dot{I}_S$. Thereby, the actual voltage and current phasor quantities are in the ranges $\dot{V}_S - \Delta\dot{V}_S \leq \dot{V}_S \leq \dot{V}_S + \Delta\dot{V}_S$ and $\dot{I}_S - \Delta\dot{I}_S \leq \dot{I}_S \leq \dot{I}_S + \Delta\dot{I}_S$. This produces an offset in the short-circuit currents calculated with (3.9).

As a result, the short-circuit currents calculated deterministically with (3.9) change as a function of $\bar{Y}_L$, $\dot{I}_L$, $\dot{V}_L$, $\bar{Y}_{DG}$, $\dot{I}_{DG}$, $\dot{V}_{DG}$, $\dot{V}_S$, $\dot{I}_S$, $\bar{Y}_F$, and d. To this extent, a few possible fault scenarios can be investigated so that a range of short-circuit currents can be determined and associated with a specific point of fault. Additionally, if historical data of fault events are available, the uncertainties aforementioned can be associated with specific

Table 3.10: Results – Phase-phase faults computed with the arbitrary criterion, switch off

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	10	0.0	10	0.0
20	10	20.6	10	20.6
30	40	3.8	70	110.5
40	40	0.0	70	110.5
50	60	68.0	50	0.0
60	40	87.5	10	142.2
70	50	90.9	10	155.2
80	80	0.0	120	194.3
90	80	10.2	120	188.3
100	110	60.2	120	153.3
110	110	0.0	120	90.9
120	120	0.0	120	0.0
130	120	21.5	120	21.5
140	110	96.9	140	0.0
150	160	19.3	160	19.3
160	150	19.3	160	0.0

probability distribution functions such that the conditional probability of occurrence can be determined for each fault scenario under evaluation.

3.4.4 Case study: Modified IEEE 4-bus DY-balanced system

This section presents the results obtained in the study of the uncertain fault behaviour of a distribution network. It describes the test network in which short-circuits were simulated in the OpenDSS software tool and displays the corresponding short-circuit current magnitudes along the feeders over a range of scenarios.

Three-phase short-circuit simulations were performed in a modified version of the IEEE 4-bus DY-balanced system, whose topology is displayed in Fig.3.4. This distribution system has 5 buses and is composed of a 12.47/4.16 kV distribution substation (buses 2-3) connecting the stiff source (slack bus 1) and the primary distribution feeder to two lateral branches (buses 4 and 5), where the distributed generation DG1 and the load L1 are installed. The total length of the primary feeder is 2000 ft, whereas the length of each lateral branch is 250 ft.

The external grid is represented as an infinite bus with 200 MVA short-circuit power.

Table 3.11: Results – Phase-phase faults computed with the arbitrary criterion, switch on

Faulty bus	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
(k_{test})	Bus ($\bar{k}$)	Cable length [m]	Bus ($\bar{k}$)	Cable length [m]
10	20	20.6	10	0.0
20	20	0.0	10	20.6
30	60	87.6	10	58.5
40	60	87.6	10	58.5
50	60	68.0	50	0.0
60	70	103.0	60	0.0
70	60	103.0	50	90.9
80	80	0.0	80	0.0
90	80	0.0	80	10.2
100	100	0.0	150	96.0
110	120	96.9	160	24.7
120	120	0.0	120	0.0
130	120	21.5	130	0.0
140	120	45.1	120	45.1
150	150	0.0	150	0.0
160	150	19.3	150	19.3

The load was modelled as 100% constant power, 100% constant impedance, and 100% constant current in distinct scenarios, with a total consumption between 4.2 MW and 6.6 MW (power factor of 0.9). In turn, the distributed generation was modelled as a P-V equivalent source with $\left|\dot{V}_{DG}\right| = 1.02$ p.u. and injected power varying from 1.0 MW to 2.0 MW.

The poles and conductors used in the simulation cases, typical of medium voltage distribution level, are displayed in Fig. 3.5. The line sections are all composed of 336.4 MCM ACSR Linnet phase conductors and 4/0 Penguin ground wires.

The faults were simulated in OpenDSS in fault study mode along the lateral branches where DG1 and L1 are connected, with fault admittances introduced every 25 ft along these laterals. This provided 21 simulated faults for every possible combination of $\bar{P}_{DG}$ and $\bar{P}_L$ (i.e., 10 simulations along each branch and 1 simulation at the connection node per combination of $\bar{P}_{DG}$, $\bar{P}_L$).

The fault scenarios evaluated consist of:

1. $P_{DG} = 1.5$ MW fixed and load modelled as 100% constant power, 100% constant impedance, and 100% constant current with $P_L \in [4.2, 6.6]$ MW, 0.6 MW step;

2. $P_L = 5.4$ MW fixed in 100% constant power, 100% constant impedance, and 100%

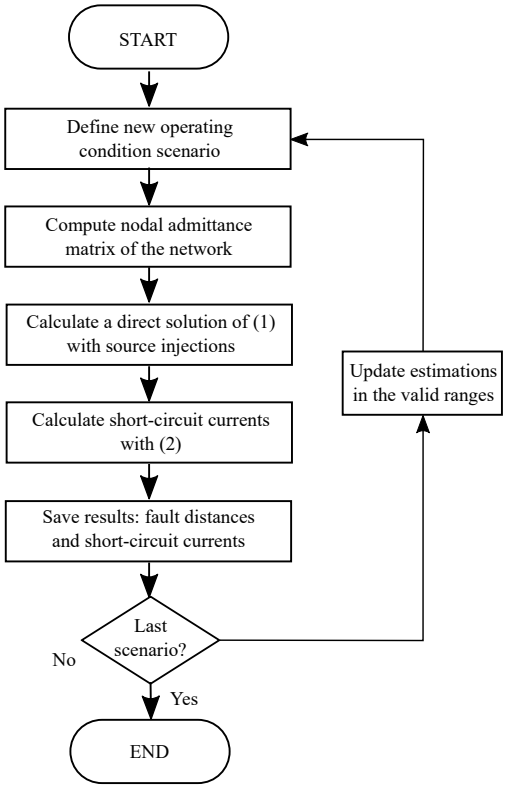

Figure 3.3: Fault behaviour with uncertainties – Flow diagram of the algorithm

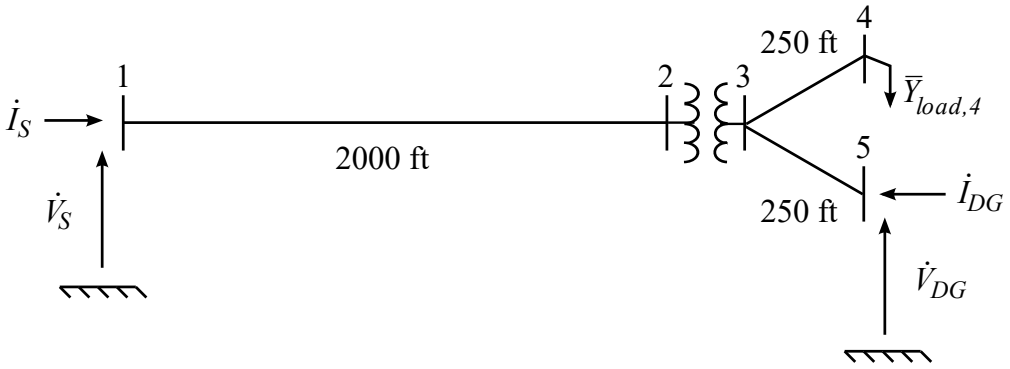

Figure 3.4: Case study – Modified IEEE 4-bus DY-balanced system

constant current models, and distributed generation with $P_{DG} \in [1.0, 2.0]$ MW, 0.25 MW step;

3. $P_{DG} = 1.0$ MW and $P_L = 6.6$ MW versus $P_{DG} = 2.0$ MW and $P_L = 4.2$ MW fixed in 100% constant power, 100% constant impedance, and 100% constant current models.

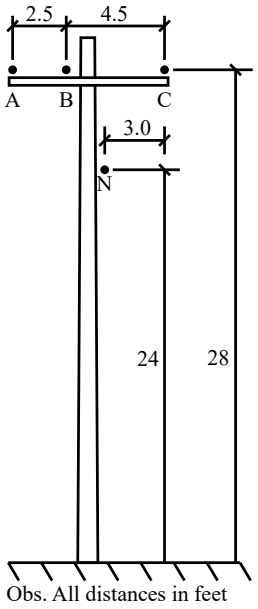

Figure 3.5: Case study – Arrangement of distribution pole and conductors

3.4.5 Results

This section presents the short-circuit currents obtained over a range of scenarios. For clarity, the results are illustrated graphically along the lateral branches $3-4$ and $3-5$ of Fig. 3.4 in all graphs of Fig. 3.6 to 3.8, which show three-phase symmetrical short-circuit currents at a single phase. Solid lines show the fault currents along the branch $3-4$, whereas dashed lines show the fault currents along the branch $3-5$.

Scenario 1 The results of the proposed method for the first scenario (with varying power consumption and load types) are illustrated graphically in Fig. 3.6 with a constant impedance, constant current, and constant power load model (graphs of Fig. 3.6 from the top to the bottom). A comparison between solid and dashed lines of the same color shows that the differences between the short-circuit currents along the branches $3-4$ and $3-5$ increase as P_{DG} increases. In turn, a comparison between solid and dashed lines of different colors shows that P_{DG} introduces an offset in the short-circuit currents simulated. Moreover, the graphs show that, for the same fault distances, the grid setting with a constant-impedance load presents the highest short-circuit current magnitudes, followed by the constant-current load and the constant-power load. However, the short-circuit levels do not vary significantly with the load model in use. This is explained by the changes made in the equivalent representation of the grid calculated with (3.9), as the current injections increase as P_{DG} increases.

Scenario 2 The results of the proposed method for the second scenario (varying injected power from distributed generation unit with different load types) are presented in Fig. 3.7 for a constant impedance, constant current, and constant power load model (graphs of Fig. 3.7 from the top to the bottom). A comparison between solid and dashed lines of the same colour shows that the differences between the short-circuit currents along the branches $3-4$ and $3-5$ increase slightly as P_L decreases. In turn, a comparison between solid and dashed lines of different colours shows that P_L introduces a small offset in the short-circuit currents simulated. In addition, the graphs show that, for the same fault distances, the grid setting with a constant-impedance load presents the highest short-circuit current magnitudes, followed by the constant-current load and the constant-power load. However, the short-circuit levels do not vary significantly with the load model in use and remain roughly the same for different values of P_L. Furthermore, the changes in the fault behaviour of the grid presented hereby are much smaller than those verified with varying P_{DG}. This is attributed to the changes made in the Thevenin equivalent calculated with (3.9); the power injected by the generator changes the current injections of the grid considerably, whereas the load admittance is much smaller than the fault admittance and does not produce significant changes in the equivalent admittance at the point of fault.

Scenario 3 The results of the proposed method for the third scenario (minimum and maximum fault currents with different load types) are displayed in Fig 3.8. In tjis case, a comparison between solid and dashed lines of the same colour shows that the differences between the short-circuit currents along the branches $3-4$ and $3-5$ increase as P_{DG} increases and P_L decreases. In turn, a comparison between solid and dashed lines of different colours shows that the offset in the short-circuit currents decreases with the distance along the branch $3-4$ and increases with the distance along the branch $3-5$.

Interpretation of the results The results aforementioned can be interpreted in two ways: looking at a fixed short-circuit current and looking at a fixed distance. The former shows which points of fault may present the same fault current, whereas the latter shows which fault currents may occur at a given point of fault.

For illustration, consider the simulation results plotted in Fig. 3.6 in two distinct situations: a 12 kA fixed short-circuit current and a 200 ft fixed fault distance. The resulting fault distances and short-circuit currents obtained with different operating conditions are summarized in Table 3.12 and Table 3.13, respectively. Note that they may be extended to the graphs of Fig. 3.7 and Fig. 3.8 and take into consideration different short-circuit currents and fault distances.

3.4.6 Discussion

The results of 3.4.4 show that a range of possible scenarios has to be evaluated when uncertainties are taken into consideration in fault location tasks, given a fixed short-circuit current and/or fault distance. Additionally, if historical data of fault events are available,

Table 3.12: Results – Possible fault distances calculated with a 12 kA short-circuit current in Fig. 3.6

Setting	Lateral	P_{DG} (MW)	fault distance (ft)
constant-impedance load	$3-4$	1.0	46
		1.25	68
		1.5	90
		1.75	111
		2.0	132
	$3-5$	1.0	48
		1.25	74
		1.5	100
		1.75	128
		2.0	157
constant-current load	$3-4$	1.0	42
		1.25	65
		1.5	87
		1.75	109
		2.0	130
	$3-5$	1.0	45
		1.25	71
		1.5	97
		1.75	126
		2.0	155
constant-power load	$3-4$	1.0	39
		1.25	62
		1.5	84
		1.75	107
		2.0	128
	$3-5$	1.0	42
		1.25	68
		1.5	94
		1.75	123
		2.0	153

Table 3.13: Results – Possible short-circuit currents calculated with a 200 ft fault distance in Fig. 3.6

Setting	Lateral	P_{DG} (MW)	fault current (kA)
constant-impedance load	$3-4$	1.0	10.77
		1.25	10.92
		1.5	11.07
		1.75	11.22
		2.0	11.36
	$3-5$	1.0	10.86
		1.25	11.04
		1.5	11.23
		1.75	11.42
		2.0	11.62
constant-current load	$3-4$	1.0	10.87
		1.25	11.02
		1.5	11.17
		1.75	11.33
		2.0	11.48
	$3-5$	1.0	10.96
		1.25	11.14
		1.5	11.33
		1.75	11.53
		2.0	11.74
constant-power load	$3-4$	1.0	10.72
		1.25	10.87
		1.5	11.02
		1.75	11.17
		2.0	11.32
	$3-5$	1.0	10.81
		1.25	11.00
		1.5	11.20
		1.75	11.39
		2.0	11.59

these uncertainties can be associated with specific probability distribution functions. For instance, if a short-circuit is more likely to occur when P_L is the highest and P_{DG} is the lowest (blue lines of Fig. 3.8), then this operating condition shall be prioritized in the search for faults. In this case, for a given short-circuit current, the closest fault distance to node 3 shall be investigated first; for a given point of fault, the smallest short-circuit current shall be considered.

Thereby, the fault analysis presented in 3.4.4 is able to provide an expansive, but accurate estimation of the points of fault and short-circuit currents, considering a range of values for loads, distributed generation, and fault distance and impedance parameters. Furthermore, the results obtained over a range of fault scenarios are comprehensive enough to provide a good overview of the possible fault conditions of the grid.

3.5 Data-driven Bayesian approach to minimize the risk of faults

This section presents a data-driven Bayesian method to minimise the risk of faults and support fault location in power distribution networks. The goal is to determine which grid node or region is the most vulnerable to specific types of faults. The algorithm relies on short-circuit currents and information about the network topology and electrical parameters, which allows for prior beliefs to be used and updated when new evidences are available. The following sections describe the initial assumptions and calculations of probabilities and risks.

3.5.1 Initial considerations

The fault location algorithm presented in this section is valid for any power transmission and distribution networks, but especially relevant to power distribution networks due to the limited availability of measurements along their lateral branches. Typically, power transmission systems are interconnected systems supplied by large power plants, whereas the topology of power distribution networks is radial and consists of a primary feeder connecting the distribution substation to few lateral branches, with IEDs installed at the substation (mainly) and distributed generation buses.

The algorithm incorporates information about the network topology and electrical parameters and simulated data of fault events, stored at the utility's database. Input data may also include post-fault voltage and current phasor measurements provided by IEDs with time synchronization capabilities, if available, used to identify fault occurrences.

The algorithm relies on short-circuit simulations to calculate the maximum fault currents at each bus and investigates all buses to determine a range of candidate fault scenarios, as the actual fault impedance and distance are unknown. It relies on phase components to handle unbalanced power generation and consumption profiles, which implies

that all electrical measurements and parameters in this text are expressed in terms of circuit phases. The results of the short-circuit simulations are used to define classes of faults, based on the fault current magnitudes computed at each bus.

The fault statistics represent specific network topologies and operating conditions, which are used to calculate the *a priori* and conditional probabilities. In the event of a feeder reconfiguration (which may happen constantly in distribution systems) or change in the network topology (e.g., new components), the classes of faults are expected to change, since they are associated with specific current magnitudes. Likewise, substantial changes in operating conditions (e.g., DG and loads) may affect the classes of faults. As a consequence, the methodology performs best with usage of representative data of the network operating conditions, under the assumption that the grid configuration is known.

Once the candidate fault parameters are computed, the algorithm proceeds with the calculation of conditional *a posteriori* probabilities as a function of the fault parameters of all candidate scenarios, adapted from [99]. Afterwards, the total risk is computed for each candidate scenario, considering costs related to the out-of-service areas and repair crew services. As the fault location algorithm aims at minimizing the total risk, the fault scenario which presents the minimum total risk shall be the first to be investigated.

Thereby, the fault location algorithm presented in this section can be divided in three parts: estimation of the point of fault, calculation of conditional *a posteriori* probability, and calculation of risks.

3.5.2 Estimation of the point of fault

Impedance-based fault location methods rely on voltage and current measurements to estimate the fault impedance $\bar{Z}_F$ and the fault distance x. As this problem typically presents multiple feasible solutions combining distinct pairs $\left(x, \bar{Z}_F\right)$, using iterative algorithms to pinpoint the fault is a challenging task. Likewise, estimating the fault distance using protection devices and further estimating the fault impedance may be impractical, depending on the information available.

The present algorithm overcomes this limitation through short-circuit simulations that enable to determine the fault behaviour of the network. For a given fault type, the short-circuit currents calculated at each bus are associated with specific classes of faults S_i, $i = 1, \cdots, N_C$, where N_C is the number of classes. If each class of faults is associated with a different bus of the network, then N_C is equivalent to the number of buses considered in the analysis; else if each class of faults is associated with a range of short-circuit currents, then N_C is equivalent to the number of areas with similar short-circuit characteristics.

As the probability of x falling into one or another of these categories is different, the algorithm relies on a probabilistic Bayesian approach to estimate the point of fault. Thus, it requires information about the *a priori* probability of x being in the class of faults S_i, denoted by $p\left(S_i\right)$, and the conditional probability or likelihood of x being in S_i, denoted by $P\left(x|S_i\right)$. This can be estimated or obtained with historical data of fault events stored

at the utility's database, whenever available, under the assumption that the event classes are independent from each other.

3.5.3 Calculation of *a posteriori* probability

Having determined the candidate fault scenarios, the algorithm proceeds with the calculation of posterior probabilities as a function of the fault parameters of all candidate scenarios. Thus, it requires information about $p(S_i)$ and $P(x|S_i)$, $i = 1, \cdots, N_C$. The posterior probability $P(S_i|x)$ that x belongs to the class S_i is then calculated using (3.14).

$$P(S_i|x) = \frac{p(S_i)\, P(x|S_i)}{\sum_{j=1}^{N_C} p(S_j)\, P(x|S_j)} \tag{3.14}$$

3.5.4 Calculation of risks

The hypotheses $\mathcal{H} = \{\mathcal{H}_0, \cdots, \mathcal{H}_{N_C-1}\}$ in the Bayesian method are given by (3.15) [21] as follows.

$$\mathcal{H}_i : x \sim P(S_i|x), \quad i = 0, \cdots, N_C - 1 \tag{3.15}$$

Thereby, the conditional risk of choosing $\mathcal{H}_i$ for a given x is given by (3.16), where λ_{ij} is the relative cost of classifying x in S_i when it actually belongs to S_j.

$$R(\mathcal{H}_i|x) = \sum_{j=1}^{N_C} \lambda_{ij} P(S_j|x) \tag{3.16}$$

with

$$\begin{cases} \lambda_{ij} = 0, \ i = j \\ \lambda_{ij} \geq 0, \ i \neq j \end{cases} \tag{3.17}$$

To minimize the overall risk of a fault at a specific location being wrongly classified, the Bayes decision rule chooses the hypothesis $\mathcal{H}_j$ which presents the minimum calculated value of (3.18) as follows.

$$\text{Decide}\mathcal{H}_j : R(\mathcal{H}_j|x) = \min\{R(\mathcal{H}_i|x), \ i = 0, \cdots, N_C - 1\} \tag{3.18}$$

In (3.17), λ_{ij} is computed with historical data (or simulated data of representative scenarios), considering costs related to the out-of-service areas (power not supplied) and repair crew services (logistics and maintenance costs), according to (3.19).

$$\lambda_{ij} = \lambda_{ij,o} + \lambda_{ij,m} \tag{3.19}$$

with

$$\lambda_{ij,o} = \left| \frac{\phi_j - \phi_i}{\sum_i \phi_i} \right| \tag{3.20}$$

$$\lambda_{ij,m} = \left| \frac{d_j - d_i}{\max_i d_i} \right| \tag{3.21}$$

where ϕ denotes the power not supplied and d denotes the distance to be walked by the repair crew.

From (3.19), it holds that every λ_{ij} is finite and bounded by a minimum cost $\underline{\lambda} = 0$ and a maximum cost $\overline{\lambda}$ given by (3.22) such that $\underline{\lambda} \leq \lambda_{ij} \leq \overline{\lambda}$ always holds.

$$\overline{\lambda} = \max_{i,j} \lambda_{ij,o} + \max_{i,j} \lambda_{ij,m} \tag{3.22}$$

Since (3.18) is continuous and the feasible region is closed and bounded, the Bolzano-Weierstrass theorem guarantees the existence of an optimal solution for (3.18) (which may not be unique).

3.5.5 Case study: IEEE 37-bus and 123-bus test systems

This section describes the test networks in which short-circuits were simulated, the analysis, and the fault location risks calculated with the proposed data-driven Bayesian algorithm over a range of scenarios.

The proposed data-driven Bayesian algorithm was tested in the IEEE 37-bus system, whose topology is displayed in Fig. 3.9, and in the IEEE 123-bus system, whose topology is displayed in Fig. 3.10. Both diagrams display the feeders drawn to scale and the feeders' thickness proportional to the currents flowing through the feeders in standard operation. The IEEE 37-bus system represents an actual 4.8-kV network in California, delta configured, with all line segments underground, two single-phase open-delta voltage regulators, spot loads, and very unbalanced, whereas the IEEE 123-bus system represents a typical power distribution network with 7 switches possibly open or closed and 2^7 possibilities of reconfiguration.

The analysis was carried out using the OpenDSS software tool in fault study mode [31]. IEEE circuit data available in the open-source examples implemented in this platform were used to test the method. First, short-circuit simulations (three-phase faults in the IEEE 37-bus system and phase-ground faults in the phase A of the IEEE 123-bus system) are used to identify the parts of the network with similar fault behaviours. Then, the classes of faults are defined such that each one corresponds to a terminal bus of the network, for a better visualization of the results. Thus, the buses are sorted decreasingly according to the calculated fault currents from S_1 – which corresponds to the bus with the highest short-circuit current – to S_{N_C} – which corresponds to the bus with the lowest short-circuit

current. This arrangement is convenient to distinguish between classes upstream the fault and downstream the fault, as explained as follows.

Next, the fault probabilities are defined as a function of the measured currents for each class. At first, the *a priori* probability of x being in a specific class of faults is assumed to be the same for all classes of faults, as there are no historical data of past fault events available for these networks. Then, the conditional probability of x being in a given class of fault S_i is determined as a function of the fault current. Considering that the expected class is S_j, the conditional probability is equivalent to a maximum $\overline{P}$ when $S_i = S_j$, null in the classes upstream (i.e., when $i > j$), and reduced proportionally to the distance between classes i and j in the classes downstream (i.e., when $i > j$), such that (3.23) holds with $\sum_{i=1}^{N_C} P\left(x|S_i\right) = 1$.

$$P\left(x|S_i\right) = \begin{cases} \overline{P}, i = j \\ 0, i < j \\ \frac{N_C - i}{N_C}\overline{P}, i > j \end{cases} \tag{3.23}$$

For illustration, consider the IEEE 37-bus system illustrated in Fig. 3.9. In the event of a fault at bus 701 (j^{th} class), the probability defined by (3.23) would be equivalent to $\overline{P}$ at bus 701, null at bus 799, and reduced to $\frac{37-i}{37}\overline{P}$ in the buses downstream 701, $i = j + 1, \cdots, N_C$.

3.5.6 Results

The calculated risks at a few selected nodes of the IEEE 37-bus system are displayed in Tables 3.14 to 3.20, whereas the calculated risks at a few selected nodes of the IEEE 123-bus system in different configurations are displayed in Tables 3.21 and 3.22. The first columns display the network buses, whereas the other columns show the calculated risks corresponding to faults at different buses (i.e., representing different classes of faults). The nodes selected are terminal nodes containing loads. As a result, the risk is evaluated in 14 different classes with 14 three-phase faults simulated in these classes in the IEEE 37-bus system; and in 18 different classes with 18 single-phase faults at phase A simulated in these classes in the IEEE 123-bus system.

In Tables 3.14 to 3.22, all entries are per unit values; the bold values correspond to the minimum calculated risk among all buses for a fault at a specific class, whereas the diagonal elements represent the risk calculated for a specific bus (row) in the exact class associated with that bus (column). To verify the method correctness, Table 3.14 represents the reference scenario of the IEEE 37-bus system which all other scenarios are compared to.

Ideal scenario The ideal scenario does not contain any forced errors and is used as the reference scenario for comparative purposes. The network operating conditions correspond to the standard operation without any uncertainties related to renewable energy

Table 3.14: Results – Risks in the ideal scenario

Bus	712	742	728	729	718	731	732	725	722	735	740	724	741	736
712	0.21	0.25	0.30	0.26	0.23	0.30	0.31	0.29	0.30	0.43	0.54	0.36	0.58	0.58
742	0.21	0.24	0.29	0.25	0.22	0.29	0.30	0.28	0.29	0.42	0.53	0.35	0.57	0.57
728	0.17	**0.16**	**0.12**	0.16	0.18	0.19	0.20	0.19	0.17	0.35	0.46	0.28	0.50	0.50
729	0.18	0.18	0.14	**0.09**	0.13	0.12	0.12	0.09	0.13	0.26	0.37	0.16	0.38	0.38
718	**0.17**	0.19	0.20	0.16	**0.13**	0.19	0.20	0.18	0.20	0.32	0.44	0.25	0.48	0.47
731	0.22	0.21	0.15	0.10	0.13	**0.05**	0.06	0.06	0.04	0.17	0.28	0.09	0.32	0.32
732	0.24	0.22	0.16	0.10	0.14	0.06	**0.05**	0.05	0.06	0.18	0.29	0.08	0.30	0.30
725	0.22	0.21	0.15	0.10	0.13	0.07	0.07	**0.04**	0.08	0.20	0.31	0.10	0.33	0.32
722	0.22	0.20	0.14	0.12	0.14	0.06	0.08	0.08	**0.02**	0.19	0.31	0.12	0.35	0.34
735	0.38	0.36	0.29	0.24	0.26	0.17	0.18	0.19	0.18	**0.00**	0.12	0.10	0.15	0.15
740	0.49	0.47	0.40	0.35	0.38	0.29	0.29	0.30	0.29	0.11	**0.00**	0.21	0.04	0.03
724	0.31	0.29	0.22	0.16	0.20	0.11	0.10	0.11	0.12	0.10	0.21	**0.00**	0.23	0.22
741	0.53	0.51	0.44	0.37	0.41	0.32	0.31	0.31	0.33	0.15	0.04	0.23	**0.00**	0.01
736	0.52	0.50	0.43	0.37	0.40	0.31	0.30	0.31	0.32	0.15	0.03	0.22	0.01	**0.00**

production or varying energy consumption and the repair crew is initially located at bus 799 (reference to compute the walking distance). In addition, for a specific class of faults, the *a priori* probability is assumed to be the same at any of the buses of the network and the conditional probabilities are calculated according to (3.23).

The results of the proposed method for the ideal scenario are presented in Table 3.14. Note that the most vulnerable bus lies within the expected class of faults in 11 out of 14 scenarios (diagonal elements of Table 3.14) and in an adjacent class in 2 out of 14 scenarios.

Influence of variable loads The presence of variable loads is expected to change the calculation of the risks and change the overall result. Thus, the present analysis aims at quantifying the influence of variable loads in the results.

The results of the proposed method with variable loads, modelled as a random variable multiplied to the nominal power, are presented in Table 3.15. It can be noticed that the most vulnerable bus lies within the expected class of faults in 10 out of 14 scenarios (diagonal elements of Table 3.15) and in an adjacent class in 2 out of 14 scenarios.

Influence of distributed generation The presence of distributed generation is expected to increase the short-circuit currents of the grid. As a result, the classes of fault S_j will be different from those obtained without distributed generation.

To study the impact of distributed generation on the overall performance of the method, a 10 kW, 4.8 kV, P-V type, single-phase, delta-connected distributed genera-

Table 3.15: Results – Risks with variable loads

Bus	712	742	728	729	718	731	732	725	722	735	740	724	741	736
712	0.18	0.20	0.27	0.27	0.24	0.33	0.32	0.31	0.32	0.49	0.60	0.41	0.61	0.61
742	0.17	0.19	0.26	0.25	0.22	0.32	0.31	0.30	0.31	0.48	0.59	0.40	0.59	0.60
728	0.15	0.13	0.09	0.09	0.12	0.12	0.12	0.11	0.12	0.26	0.37	0.18	0.41	0.40
729	0.14	**0.12**	**0.09**	**0.07**	**0.11**	0.11	0.10	0.09	0.10	0.27	0.38	0.19	0.38	0.38
718	**0.14**	0.14	0.15	0.14	0.11	0.20	0.19	0.18	0.19	0.36	0.47	0.28	0.47	0.48
731	0.20	0.18	0.11	0.10	0.13	**0.04**	0.06	0.07	0.04	0.17	0.29	0.08	0.33	0.32
732	0.20	0.17	0.10	0.09	0.12	0.05	**0.04**	0.05	0.03	0.18	0.29	0.10	0.31	0.30
725	0.18	0.15	0.09	0.08	0.11	0.06	0.05	**0.03**	0.04	0.21	0.32	0.13	0.33	0.33
722	0.19	0.16	0.10	0.08	0.12	0.05	0.04	0.04	**0.02**	0.19	0.30	0.11	0.31	0.31
735	0.36	0.33	0.25	0.23	0.26	0.17	0.17	0.19	0.18	**0.00**	0.12	0.09	0.17	0.15
740	0.47	0.44	0.36	0.34	0.37	0.28	0.28	0.29	0.28	0.12	**0.00**	0.21	0.05	0.03
724	0.28	0.25	0.18	0.16	0.19	0.10	0.11	0.13	0.11	0.09	0.21	**0.00**	0.26	0.24
741	0.48	0.45	0.38	0.35	0.38	0.31	0.30	0.31	0.29	0.17	0.05	0.26	**0.00**	0.01
736	0.48	0.45	0.38	0.35	0.38	0.30	0.29	0.30	0.29	0.15	0.03	0.24	0.01	**0.00**

tion was connected to buses 720, 737, and 742 of the IEEE-37 bus network (displayed in Fig. 3.9 without distributed generation). The generators are set to 1.05 p.u. and provide adequate reactive power support to the grid between -1000 and 1000 kVAr. In this scenario, as the presence of distributed generation is known and used in the short-circuit simulations, the classes of faults are different from those obtained without distributed generation.

The results of the proposed method calculated with distributed generation are presented in Table 3.16. In this case, it can be noticed that the most vulnerable bus lies within the expected class of faults in 12 out of 14 scenarios (diagonal elements of Table 3.16) and in an adjacent class in 1 out of 14 scenarios.

Influence of fault impedance The fault impedance is expected to decrease the short-circuit currents of the grid with respect to the ideal scenario. As a result, the classes of fault S_j identified by the method will be different from those obtained in the ideal scenario.

To study the impact of a different fault impedance on the overall performance of the method, a small fault impedance equivalent to 0.05 Ω was added to the terminal line lengths of the IEEE-37 bus network (displayed in Fig. 3.9 without the equivalent extra length). The results of the proposed method are presented in Table 3.17. A comparison with the columns of Table 3.14 shows that the fault classes of buses 712, 724, 740, 728, and 729 are identified wrongly in Table 3.17 (as 742, 741, 741, 729, and 718, respectively). In this case, the most vulnerable bus lies within the expected class of faults in 8 out of

Table 3.16: Results – Risks with distributed generation

Bus	712	742	728	729	718	731	732	725	722	735	740	741	724	736
712	0.22	0.26	0.30	0.27	0.24	0.30	0.31	0.29	0.29	0.43	0.54	0.58	0.36	0.58
742	0.21	0.24	0.29	0.26	0.23	0.29	0.31	0.28	0.28	0.42	0.53	0.57	0.35	0.57
728	0.18	**0.16**	**0.11**	0.16	0.19	0.20	0.21	0.19	0.17	0.35	0.46	0.50	0.28	0.50
729	0.18	0.18	0.15	**0.09**	0.14	0.12	0.12	0.10	0.12	0.26	0.37	0.38	0.16	0.38
718	**0.18**	0.19	0.19	0.16	**0.13**	0.19	0.21	0.19	0.19	0.32	0.44	0.48	0.25	0.47
731	0.22	0.21	0.16	0.11	0.13	**0.05**	0.06	0.06	0.03	0.17	0.28	0.32	0.09	0.32
732	0.23	0.22	0.17	0.11	0.14	0.06	**0.05**	0.06	0.05	0.18	0.29	0.30	0.08	0.30
725	0.21	0.21	0.16	0.10	0.14	0.07	0.07	**0.04**	0.07	0.20	0.32	0.33	0.10	0.32
722	0.22	0.21	0.15	0.12	0.15	0.07	0.08	0.08	**0.01**	0.19	0.31	0.35	0.12	0.34
735	0.37	0.36	0.30	0.24	0.26	0.17	0.18	0.18	0.19	**0.00**	0.12	0.15	0.10	0.15
740	0.48	0.47	0.41	0.35	0.37	0.28	0.29	0.29	0.30	0.11	**0.00**	0.04	0.21	0.03
741	0.52	0.51	0.45	0.37	0.41	0.32	0.30	0.31	0.34	0.15	0.04	**0.00**	0.23	0.01
724	0.30	0.29	0.23	0.16	0.20	0.11	0.10	0.11	0.12	0.10	0.21	0.23	**0.00**	0.22
736	0.51	0.50	0.44	0.37	0.40	0.31	0.30	0.30	0.33	0.15	0.03	0.01	0.22	**0.00**

14 scenarios (diagonal elements of Table 3.16) and in an adjacent class in 4 out of 14 scenarios.

Influence of walking distance The distance to be walked by the repair crew patrol is also expected to change the calculation of the risks and consequently the overall result. Thereby, the present analysis aims at quantifying the effect of the walking distance on the results, by locating the repair crew in a different initial position.

The results of the proposed method with a different distance to be walked by the repair crew, re-located to bus 775 as its initial position, are presented in Table 3.18. It can be noticed that the most vulnerable bus lies within the expected class of faults in 12 out of 14 scenarios (diagonal elements of Table 3.18) and in an adjacent class in 1 out of 14 scenarios.

Influence of fault probabilities The fault probabilities assigned to the classes of faults are also expected to make an impact on the calculation of the risks. Thus, this section evaluates the influence of an alternative estimation of the *a priori* probability and conditional probability in the overall results.

The results obtained with a different *a priori* probability are presented in Table 3.19, whereas the results obtained with a different conditional probability are presented in Table 3.20. The former analysis considers that the *a priori* probability in the expected class of faults is two times bigger than in the other classes of faults (i.e., $p\left(S_j\right) = 2 \times p\left(S_i\right), \forall i \neq j$).

Table 3.17: Results – Risks with different fault impedance

Bus	712	742	728	729	718	731	732	725	722	735	740	724	741	736
712	0.25	0.25	0.26	0.23	0.23	0.30	0.31	0.29	0.30	0.43	0.58	0.58	0.58	0.58
742	0.24	0.24	0.25	0.22	0.22	0.29	0.30	0.28	0.29	0.42	0.57	0.57	0.57	0.57
728	**0.16**	**0.16**	0.16	0.18	0.18	0.19	0.20	0.19	0.17	0.35	0.50	0.50	0.50	0.50
729	0.18	0.18	**0.09**	0.13	0.13	0.12	0.12	0.09	0.13	0.26	0.38	0.38	0.38	0.38
718	0.19	0.19	0.16	**0.13**	**0.13**	0.19	0.20	0.18	0.20	0.32	0.48	0.48	0.48	0.47
731	0.21	0.21	0.10	0.13	0.13	**0.05**	0.06	0.06	0.04	0.17	0.32	0.32	0.32	0.32
732	0.22	0.22	0.10	0.14	0.14	0.06	**0.05**	0.05	0.06	0.18	0.30	0.30	0.30	0.30
725	0.21	0.21	0.10	0.13	0.13	0.07	0.07	**0.04**	0.08	0.20	0.33	0.33	0.33	0.32
722	0.20	0.20	0.12	0.14	0.14	0.06	0.08	0.08	**0.02**	0.19	0.35	0.35	0.35	0.34
735	0.36	0.36	0.24	0.26	0.26	0.17	0.18	0.19	0.18	**0.00**	0.15	0.15	0.15	0.15
740	0.47	0.47	0.35	0.38	0.38	0.29	0.29	0.30	0.29	0.11	0.04	0.04	0.04	0.03
724	0.29	0.29	0.16	0.20	0.20	0.11	0.10	0.11	0.12	0.10	0.23	0.23	0.23	0.22
741	0.51	0.51	0.37	0.41	0.41	0.32	0.31	0.31	0.33	0.15	**0.00**	**0.00**	**0.00**	0.01
736	0.50	0.50	0.37	0.40	0.40	0.31	0.30	0.31	0.32	0.15	0.01	0.01	0.01	**0.00**

In this case, the results of Table 3.19 show that the most vulnerable bus lies within the expected class of faults in 13 out of 14 scenarios. The latter analysis considers that, for a given class of faults, the conditional probability of the classes downstream (i.e., when $i > j$) is given by an exponential function $\overline{P} \exp j - i$. In this case, the results of Table 3.20 show that the most vulnerable bus lies within the expected class of faults in all scenarios.

Influence of grid reconfiguration Since the fault currents change as a function of the grid configuration, the classes of faults and the calculation of the risks also depend on the grid configuration, with 2^n possibilities for n reconfiguring elements. Thus, this section evaluates the influence of different grid configurations in the overall results.

In this section, the proposed Bayesian algorithm is tested in the IEEE 123-bus system in two different grid settings so that the differences between distinct grid configurations can be compared in terms of accuracy of results: all switches closed (Table 3.21) and all switches closed but switch Sw-2 (Table 3.22). In both settings, the repair crew is initially located at bus $150R$ (reference to compute the walking distance). The network operating conditions correspond to the standard operation without any forced errors or uncertainties (e.g., related to renewable generation or varying energy consumption). For a specific class of faults, the *a priori* probability is assumed to be the same at any of the buses of the network and the conditional probabilities are calculated according to (3.23).

The results of the proposed method are presented in Tables 3.21 and 3.22. Note that the classes of faults are sorted in different ways, as the grid configuration changes the current magnitudes. In Table 3.21, the most vulnerable bus lies within the expected class

Table 3.18: Results – Risks with different walking distance

Bus	712	742	728	729	718	731	732	725	722	735	740	724	741	736
712	0.30	0.35	0.36	0.30	0.28	0.37	0.37	0.35	0.38	0.15	0.02	0.51	0.07	0.06
742	0.31	0.35	0.37	0.31	0.27	0.38	0.38	0.34	0.37	0.17	0.01	0.50	0.05	0.05
728	0.30	**0.31**	**0.22**	0.26	0.41	0.32	0.33	0.62	0.63	0.17	0.34	0.83	0.39	0.38
729	0.31	0.33	0.25	**0.19**	0.35	0.26	0.26	0.52	0.59	0.08	0.24	0.71	0.26	0.26
718	0.40	0.44	0.46	0.38	**0.23**	0.47	0.46	0.23	0.26	0.30	0.14	0.36	0.12	0.12
731	0.36	0.36	0.31	0.30	0.44	**0.19**	0.22	0.71	0.75	0.25	0.41	0.90	0.46	0.45
732	0.37	0.37	0.31	0.29	0.44	0.20	**0.20**	0.69	0.76	0.25	0.42	0.88	0.44	0.43
725	0.59	0.61	0.62	0.52	0.38	0.60	0.58	**0.08**	0.11	0.51	0.35	0.15	0.30	0.31
722	0.60	0.63	0.63	0.56	0.42	0.64	0.63	0.13	**0.05**	0.55	0.39	0.16	0.37	0.38
735	**0.30**	0.32	0.27	0.22	0.32	0.28	0.29	0.47	0.52	**0.01**	0.17	0.66	0.22	0.21
740	0.31	0.35	0.37	0.30	0.27	0.38	0.38	0.34	0.37	0.17	**0.00**	0.50	0.05	0.04
724	0.73	0.76	0.76	0.66	0.53	0.75	0.72	0.22	0.20	0.66	0.49	**0.00**	0.45	0.45
741	0.34	0.38	0.40	0.32	0.27	0.41	0.40	0.30	0.35	0.22	0.05	0.44	**0.00**	0.01
736	0.34	0.38	0.40	0.32	0.27	0.41	0.39	0.31	0.36	0.21	0.04	0.45	0.01	**0.00**

of faults in 11 out of 12 scenarios (diagonal elements), whereas in Table 3.22, the most vulnerable bus lies within the expected class of faults in 10 out of 12 scenarios and in an adjacent class in 1 out of 12 scenarios.

3.5.7 Discussion

The results of 3.5.5 indicate that the data-driven Bayesian algorithm proposed in Section 3.5 is capable of identifying the most vulnerable areas of the grid with good accuracy and robustness for different fault scenarios. Overall, variations in the load estimations, the presence of distributed generation, fault impedance, walking distance, fault probabilities, and grid reconfigurations do not affect the results significantly, as long as the network topology and its components are represented accurately.

The highest costs $\lambda_{ij,o}$ are associated with the pairs of buses where the difference between power consumption is the highest, whereas the highest costs $\lambda_{ij,m}$ are associated with the pairs of buses where the distance to be walked is the longest (using the reference bus as a starting point). Therefore, the minimum cost parameter is expected to result in minimum load shedding with minimum displacement.

The calculations show that the accuracy can be improved when the repair crew patrol is initially at an intermediate coordinate of the network. This shows that some initial positions may be more adequate than others for risk minimization, comparing the risks calculated with different reference points. The accuracy of results is also improved when the calculated risks are negligible in the classes of faults that are far away from the expected

Table 3.19: Results – Risks with change in the *a priori* probability

Bus	712	742	728	729	718	731	732	725	722	735	740	724	741	736
712	0.14	0.16	0.29	0.23	0.16	0.27	0.29	0.26	0.29	0.43	0.54	0.36	0.58	0.58
742	**0.14**	**0.15**	0.27	0.22	0.15	0.27	0.28	0.26	0.28	0.42	0.53	0.35	0.57	0.57
728	0.20	0.20	**0.06**	0.14	0.17	0.19	0.20	0.18	0.16	0.35	0.46	0.28	0.50	0.50
729	0.19	0.19	0.13	**0.04**	0.11	0.10	0.09	0.07	0.12	0.26	0.37	0.16	0.38	0.38
718	0.15	0.15	0.18	0.12	**0.05**	0.17	0.18	0.16	0.19	0.32	0.44	0.25	0.48	0.47
731	0.23	0.23	0.17	0.10	0.15	**0.02**	0.03	0.04	0.03	0.17	0.28	0.09	0.32	0.32
732	0.25	0.24	0.18	0.09	0.16	0.03	**0.02**	0.03	0.05	0.18	0.29	0.08	0.30	0.30
725	0.23	0.22	0.16	0.07	0.14	0.05	0.04	**0.01**	0.06	0.20	0.32	0.10	0.33	0.32
722	0.24	0.23	0.15	0.12	0.17	0.04	0.06	0.06	**0.01**	0.19	0.31	0.12	0.35	0.34
735	0.39	0.38	0.32	0.25	0.30	0.17	0.18	0.20	0.19	**0.00**	0.12	0.10	0.15	0.15
740	0.51	0.50	0.44	0.36	0.41	0.28	0.29	0.31	0.30	0.12	**0.00**	0.21	0.04	0.03
724	0.32	0.31	0.25	0.16	0.23	0.10	0.08	0.10	0.12	0.10	0.21	**0.00**	0.23	0.22
741	0.54	0.53	0.47	0.38	0.45	0.32	0.30	0.32	0.34	0.15	0.04	0.23	**0.00**	0.01
736	0.54	0.53	0.47	0.37	0.44	0.31	0.30	0.32	0.34	0.15	0.03	0.22	0.01	**0.00**

one. However, this might not be the case and has to be adjusted to each network under consideration.

In terms of accuracy, the worst overall results are those calculated with a different fault impedance in the IEEE 37-bus system. This is expected due to the significant change in the short-circuit current magnitude, which leads to a wrong identification of the most probable class of fault in a few buses. This inaccuracy can be overcome by considering a different fault impedance in the definition of the reference classes of faults. Nevertheless, the results are still acceptable, as the most probable class of fault is identified correctly in most cases. In turn, the most troublesome buses are the closest to the distribution substation. This is expected because there are more buses connected downstream with a non-zero conditional probability and relatively high $\lambda_{ij,o}$ and $\lambda_{ij,m}$.

Additionally, the analysis of distinct configurations shows that the classes of faults may change considerably as a function of the network setting. Nonetheless, the accuracy of the results is not affected by changes in the network configuration and distributed generation, as far as the classes of faults and probabilities are calculated for the correct setting. Were more buses (and classes of faults) of the original networks to be considered, the accuracy of the method could be enhanced or worsened, depending on the costs and probabilities associated with each scenario.

Table 3.20: Results – Risks with change in the conditional probability

Bus	712	742	728	729	718	731	732	725	722	735	740	724	741	736
712	**0.01**	0.06	0.29	0.19	0.15	0.26	0.28	0.26	0.31	0.44	0.52	0.39	0.58	0.58
742	0.02	**0.05**	0.28	0.19	0.14	0.26	0.27	0.25	0.29	0.43	0.51	0.38	0.57	0.57
728	0.26	0.25	**0.02**	0.13	0.17	0.18	0.20	0.17	0.18	0.36	0.44	0.31	0.50	0.50
729	0.20	0.20	0.14	**0.02**	0.12	0.09	0.08	0.07	0.13	0.27	0.35	0.19	0.38	0.38
718	0.10	0.12	0.19	0.09	**0.04**	0.16	0.17	0.16	0.20	0.34	0.42	0.28	0.48	0.47
731	0.26	0.27	0.19	0.10	0.16	**0.00**	0.03	0.04	0.05	0.18	0.26	0.12	0.32	0.32
732	0.28	0.28	0.21	0.09	0.17	0.02	**0.01**	0.03	0.06	0.19	0.27	0.11	0.30	0.30
725	0.25	0.26	0.18	0.07	0.16	0.03	0.03	**0.01**	0.08	0.21	0.29	0.13	0.33	0.32
722	0.28	0.28	0.17	0.12	0.18	0.03	0.05	0.06	**0.02**	0.20	0.29	0.15	0.35	0.34
735	0.42	0.43	0.36	0.26	0.32	0.17	0.18	0.20	0.17	**0.01**	0.11	0.11	0.15	0.15
740	0.54	0.55	0.47	0.38	0.44	0.28	0.30	0.32	0.29	0.10	**0.03**	0.19	0.04	0.03
724	0.35	0.36	0.28	0.17	0.25	0.09	0.08	0.11	0.12	0.11	0.19	**0.03**	0.23	0.22
741	0.58	0.59	0.51	0.39	0.46	0.32	0.31	0.33	0.32	0.14	0.06	0.20	**0.00**	0.01
736	0.57	0.58	0.51	0.39	0.47	0.31	0.30	0.33	0.32	0.14	0.06	0.19	0.01	**0.00**

3.6 Final Remarks

This chapter presented data-driven approaches to support fault location in power systems in different ways, with applications to modelling, uncertainty quantification, and probabilistic assessment of faults. The presented methodologies are applicable to any power transmission and distribution networks with distinct characteristics, but particularly relevant to power distribution networks for their topology and inner characteristics. They are particularly helpful to provide the utility companies with additional pieces of information and evidences in the search for different types of faults.

Table 3.21: Results – Risks with all switches closed

Bus	48	94	79	10	20	88	37	46	33	71	111	114
48	**0.22**	0.30	0.27	0.35	0.24	0.44	0.15	0.29	0.11	0.45	0.68	0.75
94	0.35	0.32	0.47	0.54	0.42	0.34	0.46	0.53	0.36	0.73	0.36	0.42
79	0.24	0.28	**0.16**	0.26	0.18	0.60	0.29	0.11	0.11	0.60	0.48	0.55
10	0.41	0.43	0.29	**0.24**	0.40	0.81	0.50	0.16	0.34	0.83	0.32	0.32
20	0.23	0.27	0.17	0.26	**0.16**	0.58	0.28	0.13	0.09	0.58	0.50	0.57
88	0.56	0.54	0.68	0.73	0.582	**0.17**	0.48	0.82	0.65	0.99	0.65	0.71
37	0.31	0.39	0.35	0.42	0.30	0.36	**0.07**	0.41	0.24	0.58	0.80	0.87
46	0.31	0.34	0.21	0.25	0.26	0.68	0.37	**0.03**	0.21	0.70	0.39	0.45
33	0.22	**0.27**	0.21	0.30	0.19	0.52	0.22	0.19	**0.02**	0.51	0.58	0.65
71	0.63	0.71	0.68	0.76	0.64	0.87	0.57	0.67	0.50	**0.01**	1.09	1.15
111	0.52	0.45	0.50	0.46	0.56	0.65	0.75	0.41	0.59	1.08	**0.00**	0.06
114	0.58	0.51	0.55	0.49	0.62	0.71	0.82	0.48	0.65	1.14	0.06	**0.00**

Table 3.22: Results – Risks with all switches closed except Sw-2

Bus	48	10	20	37	46	33	79	94	111	88	114	71
48	0.27	0.35	0.45	0.29	0.46	0.44	0.37	0.83	0.54	0.42	0.32	0.80
10	0.26	**0.24**	0.39	0.25	0.38	0.54	0.15	0.61	0.32	0.19	0.10	0.58
20	0.52	0.51	0.41	0.64	0.79	0.94	0.65	1.12	0.83	0.70	0.60	1.09
37	**0.23**	0.25	**0.35**	**0.20**	0.37	0.50	0.25	0.71	0.42	0.29	0.20	0.68
46	0.41	0.39	0.47	0.29	**0.16**	0.34	0.44	0.44	0.67	0.54	0.44	0.36
33	0.56	0.61	0.67	0.47	0.36	**0.21**	0.73	0.73	0.96	0.83	0.73	0.65
79	0.32	0.30	0.44	0.30	0.42	0.58	**0.07**	0.54	0.25	0.12	0.02	0.50
94	0.75	0.70	0.80	0.61	0.45	0.61	0.500	**0.03**	0.32	0.45	0.55	0.06
111	0.54	0.52	0.65	0.50	0.61	0.76	0.24	0.30	**0.01**	0.13	0.23	0.32
88	0.42	0.40	0.54	0.39	0.51	0.66	0.15	0.42	0.13	**0.00**	0.10	0.39
114	0.33	0.32	0.46	0.31	0.43	0.59	0.08	0.52	0.23	0.10	**0.00**	0.48
71	0.69	0.64	0.74	0.55	0.40	0.56	0.45	0.09	0.32	0.39	0.483	**0.00**

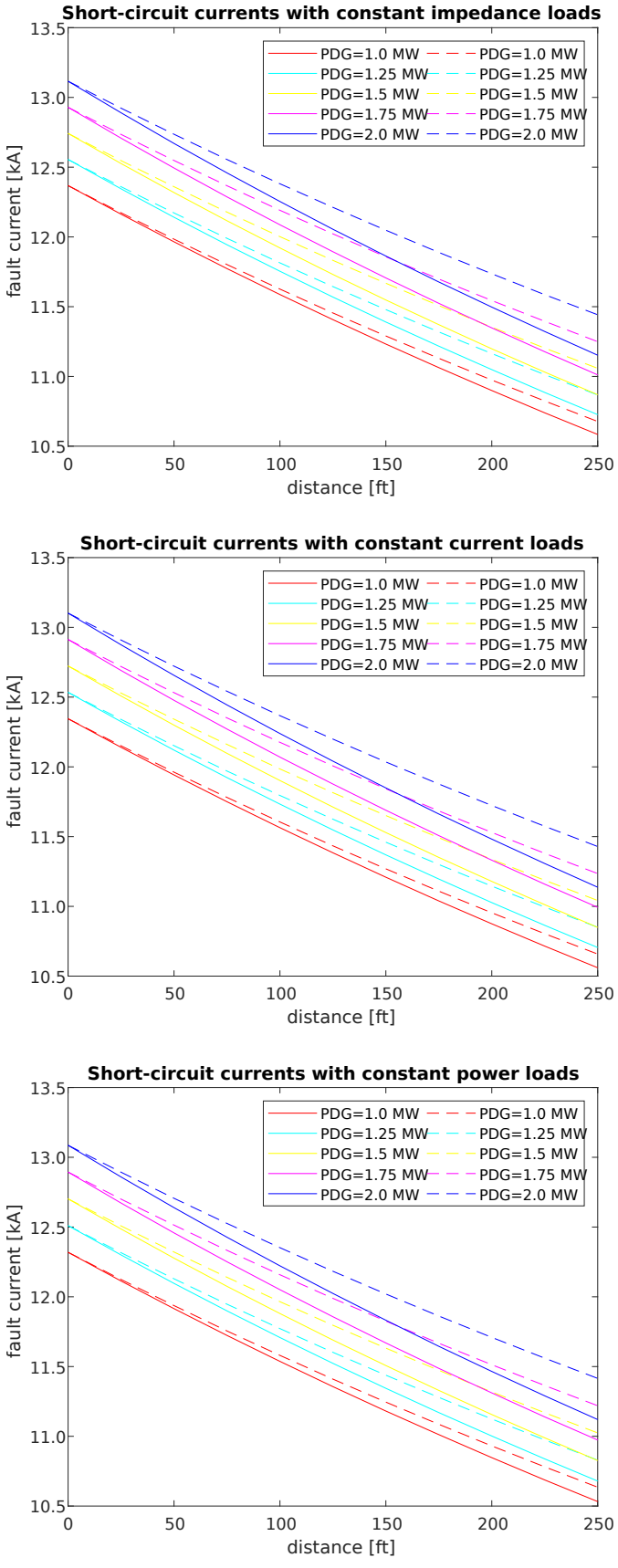

Figure 3.6: Results – Short-circuit currents at phase A with $P_L = 5.4$ MW (constant impedance, current, and power load models) and varying P_{DG}

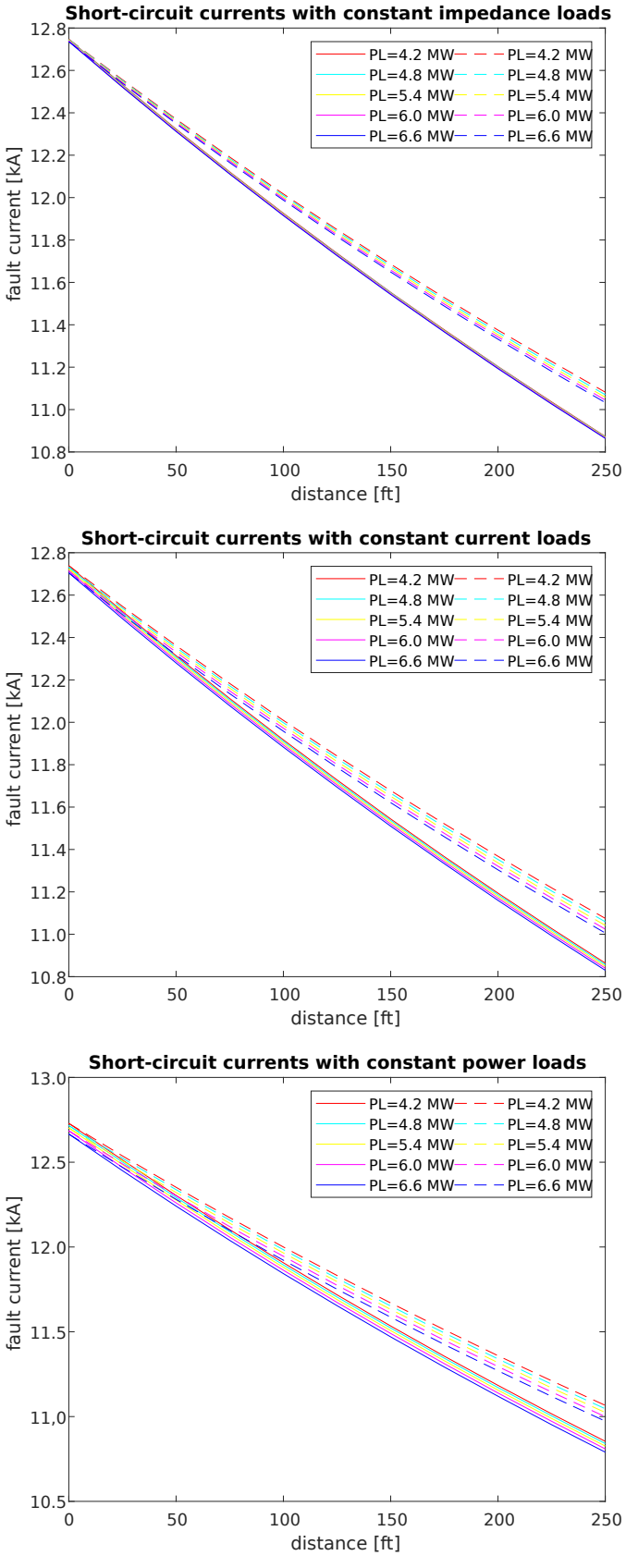

Figure 3.7: Results – Short-circuit currents at phase A with varying P_L (constant impedance, current, and power load models) and $P_{DG} = 1.5$ MW

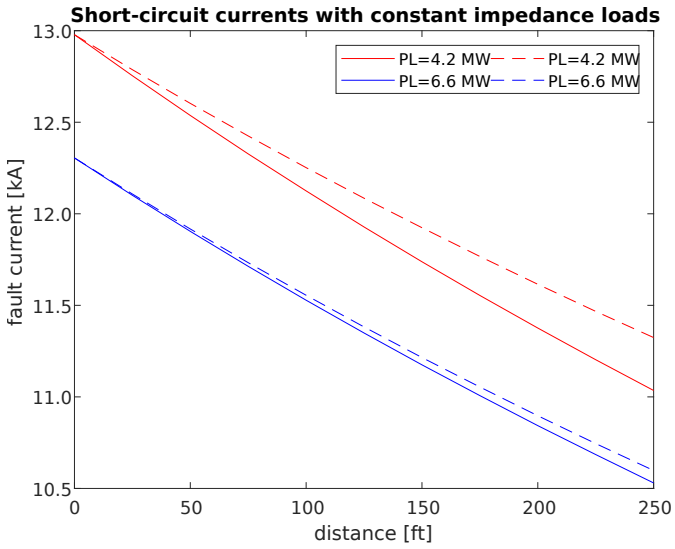

Figure 3.8: Results – Minimum (blue) and maximum (red) short-circuit currents

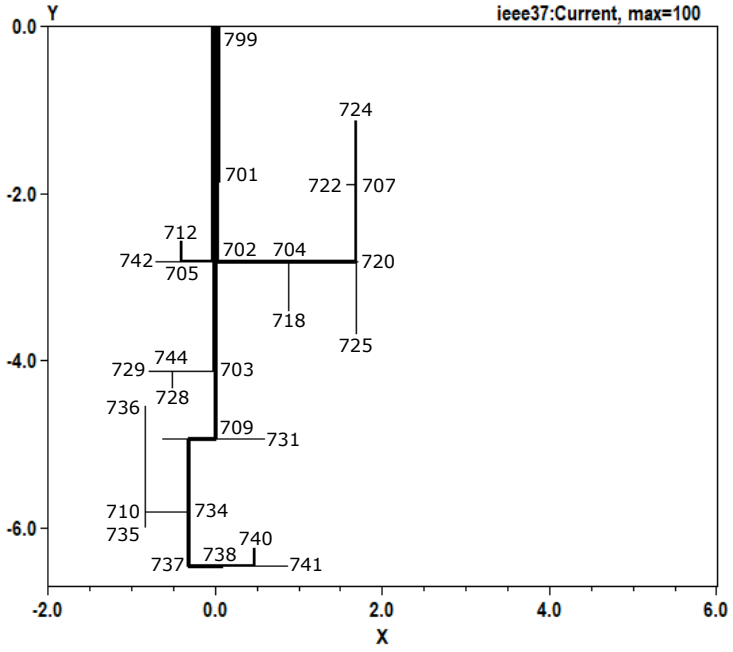

Figure 3.9: Case study – IEEE 37-bus feeder representation

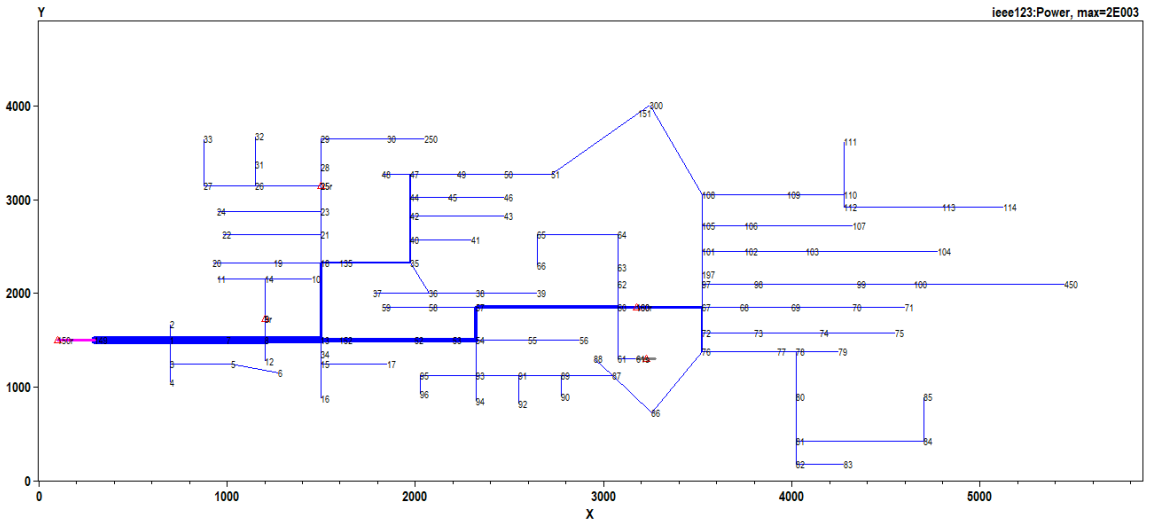

Figure 3.10: Case study – IEEE 123-bus feeder representation

Chapter 4

Resilience Assessment in Power Systems

This chapter describes strategies to improve power system resilience to external HILP events. It presents definitions, state-of-the-art techniques, and the data-driven methods for enhancements in power system resilience proposed in this article, further demonstrated in case studies.

4.1 Definitions

In this chapter, resilience refers to the ability to anticipate, respond, and recover from external HILP events, such as natural disasters, extreme weather conditions, and man-made attacks. Improving power system resilience has become increasingly important because the frequency and intensity of natural disasters is expected to rise in the future due to climate change [37]. In this ever-changing scenario, power systems shall be designed and operated to withstand and recover from such extreme events, maintaining quality of supply and minimizing performance degradation.

There are different definitions of the main attributes of a resilient power system. The main features of resilience according to [48], applicable to any critical infrastructures, are robustness, resourcefulness, rapid recovery, and adaptability, described as follows.

- Robustness: keep operating and withstand HILP events
- Resourcefulness: effectively manage a disaster as it evolves, identify and prioritize options to control and mitigate the damage
- Rapid recovery: contingency plans, emergency operation, return to normal operation as fast as possible after a disaster

- Adaptability: introduce new strategies to improve robustness, resourcefulness, and rapid recovery before the next disaster

In turn, the main features of a resilient critical infrastructure according to [64] are resistance, reliability, redundancy, and response and recovery, described as follows.

- Resistance: prevent damage or disruption by providing strength or protection to resist the hazards
- Reliability: ensure that the system and its components are properly designed to operate under a range of conditions
- Redundancy: enable normal operation to be switched with use of backup installations or spare capacity
- Response and recovery: contingency plans, emergency operation, and return to normal operation after a disaster

On the one hand, the definition provided by [48] is particularly suitable for existing infrastructures, as adaptability is a key aspect to improve their resilience based on lessons learned from previous disasters. On the other hand, the definition provided by [64] is more relevant for designing critical infrastructures in general, as it describes the physical characteristics of a resilient system. Thereby, the complementary usage of both definitions is recommended.

4.2 State-of-the art methods

Measures to improve power system resilience can be divided into planning, response, and restoration actions, depending on the time remaining for the expected occurrence [45], combining hardening and operational strategies [70] along with suitable metrics for performance assessment [57]. To this extent, distinct measures can be prioritized according to resilience- and cost-efficiency indicators, considering vulnerability and adaptation studies along with inputs from past events and/or simulations [66]. In this context, the next sections provide an overview of relevant metrics for power system resilience assessment, followed by a description of hardening and operational strategies commonly used and a literature review on power system resilience approaches.

4.2.1 Resilience metrics

Resilience metrics may be categorized into qualitative and quantitative metrics. The former usually rely on conceptual frameworks derived from the definition of resilience, such as the conceptual resilience trapezoid associated to an event [70], and/or semi-quantitative indicators, which assess specific resilience attributes within a per unit scale. In turn,

the latter typically rely on performance indicators, which are preferred to make a clear comparison between the pre- and post-event resilience levels of a power system [57].

The conceptual resilience trapezoid associated to an event is introduced in the next section, followed by some quantitative metrics commonly used in power system resilience. It is noteworthy that some of these quantitative metrics compute the difference between the real and the ideal performance to indicate the impact of an event on the system without applying any recovery actions, whereas others consider the impact of recovery actions on the post-event resilience level of the system.

Conceptual resilience trapezoid

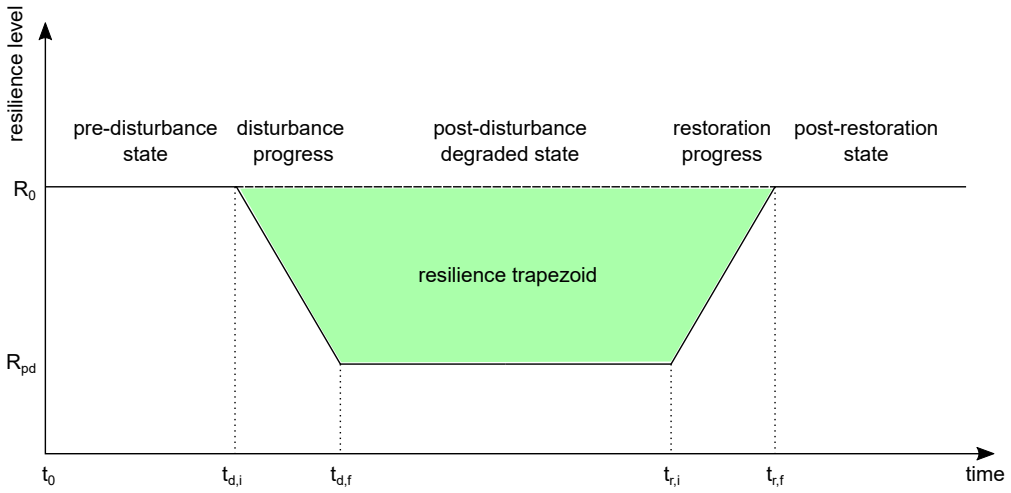

Figure 4.1: Definitions – Conceptual resilience trapezoid associated to an event (Adapted from [70])

The conceptual resilience trapezoid of a generic system, illustrated in Fig. 4.1, represents the evolving states of the system associated to the occurrence of a generic event over time. Following the notation shown in Fig. 4.1, the system is in the pre-disturbance state between the initial time t_0 and $t_{d,i}$, when the event hits the system and the disturbance progress starts; the disturbance progress ends in $t_{d,f}$ and leads the system to its post-disturbance degraded state, which lasts from $t_{d,f}$ to $t_{r,i}$, when the restoration starts; the restoration progress then ends in $t_{r,f}$ and eventually leads the system to its post-restoration state. The pre-disturbance and post-disturbance resilience levels of the system are denoted by R_0 and R_{pd}, respectively.

Considering that the green trapezoid illustrated in Fig. 4.1 represents the power system resilience levels "as is", enhancements due to the incorporation of preventive actions will reduce the painted area by increasing R_{pd} and/or reducing $t_{r,i} - t_{d,f}$ and $t_{r,f} - t_{r,i}$. Thereby, resilience enhancements may be quantified with this conceptual framework by calculating the difference between the trapezoid area of the system "as is" and with resilience enhancements.

Total number of customers without electricity

The total number of customers without electricity NC is given by (4.1) as follows, where $n(t)$ denotes the customers without electricity on day t, $t = 1, \cdots, \tau$ and τ is the time horizon of the analysis:

$$NC = \sum_{t=1}^{\tau} n(t). \tag{4.1}$$

This quantitative metric does not incorporate any recovery actions.

Loss of load probability

The loss of load probability $LOLP$ is given by (4.2) as follows, where OC is the outage capacity and RC is the reserve capacity of the system. It is calculated as a function of the effective system capacity and load demand:

$$LOLP = \mathcal{P}(OC > RC). \tag{4.2}$$

This quantitative metric does not incorporate any recovery actions.

Loss of load expectation

The loss of load expectation $LOLE$ is given by (4.3), where τ is the time horizon of the analysis and $LOLP$ is given by (4.2):

$$LOLE = \tau \times LOLP. \tag{4.3}$$

It is the expected time in the horizon τ in which the effective system capacity dos not meet the load demand. This quantitative metric does not incorporate any recovery actions, as the $LOLP$.

Loss of load frequency

The loss of load frequency $LOLF$ is the cumulative frequency of system interruption, given by (4.4):

$$LOLF = F(OC > RC). \tag{4.4}$$

This quantitative metric does not incorporate any recovery actions, as the $LOLP$.

Expected energy not supplied

The expected energy not supplied $EENS$ represents the expected value of lost energy, given by (4.5), over the horizon τ (assumed to be an integer):

$$EENS = \sum_{t=1}^{\tau} \left[OC\left(t\right) - RC\left(t\right) \right]. \tag{4.5}$$

This quantitative metric does not incorporate any recovery actions.

Value of loads lost

The value of loads lost $VOLL$ is the cost of the expected value of lost energy, given by (4.6), where η_{EENS} is the unit cost of loads lost:

$$VOLL = \eta_{EENS} \times EENS. \tag{4.6}$$

This quantitative metric does not incorporate any recovery actions, as the $EENS$.

Lost revenue impact

The lost revenue impact LRI represents the total revenue lost by the utility company, given by (4.7), where w_i is a weighting factor for the i^{th} load, PLN_i is the peak load in normal operation, PLR_i is the peak load restored in the t^{th} restoration stage, and Δ_t is the duration of the t^{th} restoration stage:

$$LRI = \sum_{t} \sum_{i} w_i \left(PLN_i - PLR_{i,t} \right) \Delta_t. \tag{4.7}$$

Note that this quantitative metric incorporates the effects of recovery actions.

Total restoration cost

The total restoration cost TR aggregates the total resources used during the recovery process, including labour, replacement, and generator operating costs. This quantitative metric incorporates the effects of recovery actions on the accumulated cost.

Recovery resilience

The recovery resilience represents the total impact of the event on the utility company, given by (4.8) using a weighting factor W to express the relation between supplied energy and recovery resources:

Table 4.1: Definitions – Resilience strategies and implementation times as a function of the event occurrence

Resilience strategy	Implementation time
Planning	From years to a few days ahead
Response	From day ahead to the end of the occurrence
Restoration	Post-occurrence

$$RR = \frac{LRI + W \times TR}{\sum_t \sum_i w_i PLN_i - \Delta_t}. \tag{4.8}$$

Note that this quantitative metric incorporates the effects of recovery actions, as it relies on the LRI and TR.

Severity risk index

The severity risk index SRI_t, defined in [92], is calculated with information about the impact $I_{k,t}$ and the probability of occurrence $\mathcal{P}_{k,t}$ at the time instant t over a range of scenarios $k = 1, \cdots, K$, as in (4.9):

$$SRI_t = \sum_{k=1}^{K} \mathcal{P}_{k,t} \times I_{k,t}. \tag{4.9}$$

It does not incorporate the effect of recovery actions.

4.2.2 Resilience strategies

Improvements in power system resilience may combine hardening and operational strategies. The former are used for resilience-based planning of power systems, whereas the latter are used for resilience-based response and restoration actions. The implementation of hardening and/or operational strategies shall consider the time remaining for an event to occur, based on forecasts. For reference, the time scales associated with planning, response, and restoration actions for improvements in power system resilience are listed in Table 4.1.

The following sections list distinct resilience-based planning, response, and restoration strategies commonly applied at transmission and distribution level. For more details, refer to [70, 67, 57].

Resilience-based planning

Resilience-based planning of power systems may include the strategies listed as follows.

- Hardening of lines, towers, poles, foundations, and other grid components
- Replacing overhead lines with underground cables
- Relocating grid components to areas less prone to severe events
- Increasing reserve capacity and redundancy of equipment
- Pruning of vegetation
- Implementing FACTS and switches for grid reconfiguration
- Updating load assignments and determining priority loads at different zones
- Upgrading the system for defensive islanding and microgrid operation
- Adding DER into distribution networks
- Risk assessment and management for evaluating and preparing for HILP events
- Accurate estimation of the location and severity of HILP events
- Improving emergency and preparedness plans
- Increasing situation awareness through advanced monitoring and prediction tools

Resilience-based response

Resilience-based response of power systems may include the strategies listed as follows.

- Resilience-based unit commitment
- Transmission line switching
- Defensive islanding and microgrid operation

Resilience-based restoration

Resilience-based restoration of power systems may include the strategies listed as follows.

- Optimal switching of transmission and distribution lines
- Optimal resource allocation
- Microgrid service restoration
- Using DG and black-start units

Table 4.2: Definitions – Differences between typical outages and outages caused by natural disasters

Typical outage	Outage caused by natural disaster
Low impact, high probability	High impact, low probability
Few faults (component failures)	Multiple faults (catastrophic damage)
No spatio-temporal correlation	Spatio-temporal correlation
Most generation units remain in service	Generation units may be out of service
Supported by contingency analysis tools	Unforeseen event
Network remains intact	Network damaged/collapsed
Only involve power grid infrastructure	Interdependent infrastructures
Quick restoration	Long restoration

4.2.3 Literature review

Forecast models used to estimate typical power outages and blackouts caused by natural disasters may be statistical – based on power system and environmental data and parameters, data fitting failure models, or fitting goodness – or simulation based [96]. For comparison, the differences between typical outages and outages caused by natural disasters are summarized in Table 4.2, adapted from [96] and [70].

Grid resilience modelling from the aspects of hazard, fragility, and probabilistic risk assessment under extreme weather conditions has been addressed in [67, 68, 73, 35, 98], among others. Notably, [67] performed a stochastic evaluation to quantify the effect of adverse weather conditions on the resilience of critical electric power infrastructure, considering distinct regional wind profiles and the impact of human response on the implementation of actions to mitigate damages and restore the system. In this context, [98] modelled hurricane exposure using fragility curves and [73] applied a probabilistic metric to predict the impact of future extreme events on grid resilience and evaluate the improvements of different planning and restoration actions. In turn, [68] and [35] assessed resilience of critical power transmission system components based on the expected energy not supplied under strong winds and earthquakes, respectively. In other references, the impact of extreme weather conditions on power systems resilience is represented by an uncertainty variable, as in [86, 102, 34], as well as the effect of high penetration of renewable energy on the decision of hardening grid components, as in [104, 16]. However, these procedures cannot be generalized to all types of extreme weather events and natural disasters, as they present different behaviours and rely on distinct Earth systems and climate models and data.

Alternatively, the objective of enhancing grid resilience with minimum cost was previously studied in [63, 52, 12], among others. Notably, [63] presented an optimization framework based on conditional value at risk, defined by the expected energy not supplied, aimed at re-designing distribution substations with minimum cost and limited risk

exposure to generic HILP events, which might not be robust for long-term resilience enhancements. In [52], investment portfolios that offer the highest resilience enhancements against potential risks caused by generic natural hazards were identified using an optimization via simulation approach, whereas [12] considered the duration-dependent value of resilience to integrate the value of lost load in resilient planning and operation decisions.

It is important to emphasize that resilience assessment also varies as a function of the grid components affected. To this extent, among all power system components, overhead power lines are particularly vulnerable to outages caused by different external events, for their manufacturing and assembling, insofar as they transport electricity from generation sites to consumers across long distances. As a consequence, their failure rates are expected to be high in comparison with other grid components and can put the overall system operation at a higher risk of collapsing under extreme weather events [95]. As an option, they could be hardened or replaced with underground lines, which present low failure rates and cheap operational costs in comparison with overhead lines. However, this choice takes its toll with high installation costs and long restoration times in the event of an outage. In turn, defensive island operation strategies, as proposed in [69], are not feasible in distribution networks with radial configuration and one-directional centralized supply.

To this extent, electric power substations out of service are also troublesome, since they connect large power plants to a number of consumers. Among HILP weather events, floods are especially problematic for substations, as they are installed on or near to the ground level and must be shut down in case of flooding [20]. Moreover, depending on the hydrological characteristics of a region and the severity of the episode, flood clearance may take several days and lead to long interruption times, huge amounts of energy not supplied, and additional costs for operation in reserve mode, repair of damaged components, and penalty for loads lost. In this scenario, considering hydrological models and a range of realistic flood scenarios to assess the impact of potential flood events on the grid is relevant to correlate the locations affected and minimize the performance degradation in the worst-case scenarios. Moreover, this representation is more suitable to capture the dynamic behaviour of floods than the constant probability taken from flood plans.

Apart from the grid components, other aspects concerning the energy service are also relevant for an adequate resilience assessment. In particular, electricity distribution networks in low-income neighbourhoods are particularly vulnerable to HILP risks of different nature – either for inadequate risk assessment for infrastructure planning, or for the impossibility to pay for the repair costs associated with the damages to infrastructure. As a matter of a fact, they usually present a high number of irregular connections, as well as thefts of electricity and grid components combined with inefficient energy appliances [84]. As a consequence, they are less reliable than in high-income neighbourhoods and may be overloaded at higher ratings than the maximum acceptable limits that ensure a safe operation. To make matters worse, they are often provided with a limited service by the utility company and lack alternatives for energy provision [1]. It is noteworthy that extreme events affect households and external grid components in different ways. Thus, distinct approaches to improve power system resilience in low-income neighbourhoods should be considered.

4.2.4 Contributions

The following articles derived from this thesis explore contributions to the state-of-the-art methods for power system resilience enhancements to extreme events.

- Souto, L., Yip, J., Cunha, V. C., Austgen, B., Wu, W.Y., Kutanoglu, E., Hasenbein, J., Yang, Z.L., King, C., & Santoso, S. (2020). Power System Resilience to Floods: Modelling, Impact Assessment, and Long-Term Mitigation Strategies. Submitted to: International Journal of Electrical Power and Energy Systems.
- Souto, L. & Santoso, S. (2020). Overhead versus Underground: Designing Power Lines for Resilient, Cost-Effective Power Distribution Networks under Windstorms. In Resilience Week 2020, Salt Lake City, UT, 2020, pp. 1-5.
- Souto, L. & Santoso, S. (2020). Evaluation of Power System Resilience Enhancements in Low-Income Neighborhoods. In 2020 IEEE PES Transmission & Distribution Conference and Exhibition Latin America (T&D-LA), Montevideo, 2020, pp. 1-5.

In the next sections, resilience assessment is based on resilience metrics and cost indicators. Section 4.3 evaluates resilience improvements in low-income neighbourhoods, using information about demand profiles and candidate solutions. Next, Section 4.4 is focused on designing power lines with resilience to wind storms, using information about the network topology and operation and candidate solutions. Finally, Section 4.5 investigates power system resilience to floods, using information about hydrological models, flood scenarios, and network topology within a robust optimization framework for hardening transmission substations subject to a budget limit. The case studies are focused on long-term planning of transmission and distribution systems in different contexts.

4.3 Evaluation of power system resilience improvements in low-income neighbourhoods

This section is concerned about extreme events affecting low-income neighbourhoods. The sections as follows present the problem description and formulation, including impact assessment and mitigation strategies.

4.3.1 Problem description

Electricity infrastructures are often in trouble due to natural hazards, such as floods and landslides, extreme weather conditions, such as heavy rain and wind storms, and human-made attacks, such as vandalism and theft of energy or electrical equipment. In low-income households and communities, the consequences of such extreme events are further

amplified due to the network topology and irregularities combined with social poverty [13], as explained as follows.

Power distribution networks in low-income neighbourhoods are typically radial, with households connected to the main distribution feeder at low voltage level every few meters. The line rating tends to be relatively low because of the low energy intensity of low-income households, but is usually higher than expected due to irregular users connected to the grid. As a consequence, power distribution networks in low-income neighbourhoods are less reliable than in high-income neighbourhoods and prone to outages under extreme operating conditions.

The impact of an outage in low-income neighbourhoods involves both the infrastructure and service and also depends on the financial means available to restore the out-of-service loads. Consequently, the supply interruption times may be very long or even permanent due to the impossibility of reconnecting to the grid, whereas the maintenance cost and the value of load lost may be much higher than the investment cost in mitigation strategies towards a resilient grid. This motivates the search for feasible solutions able to improve power system resilience of low-income neighbourhoods to extreme events. Ultimately, it will guarantee that the energy needs of the whole society are met adequately and also help to keep low-income communities safe.

4.3.2 Problem formulation

The impact of extreme events on electricity infrastructures is reflected on damaged equipment and on the amount of energy not supplied. Extreme events may affect individual households in vulnerable areas, as well as distribution lines, poles, transformers, and other auxiliary equipment not protected. To this extent, the damages to individual grid components and to the physical infrastructure may be repairable or request full replacement, depending on the impact of the extreme event in consideration on the grid components (i.e., from minor to major). In turn, the interruption duration depends on the impact on the energy infrastructure and on the restoration time and may last for a few hours (for example, when caused by the removal of grid components) or forever in case of permanent disconnection from the grid (for example, when caused by non-replaceable destruction of property or loss of life). The amount of energy not supplied is equivalent to the expected energy demand not supplied over time. In case of permanent disconnection from the grid, the energy not supplied is infinite (worst-case scenario).

Thereby, the methodology computes the costs of repair or replacement of damaged infrastructures of each component $i = 1, \cdots, N_C$ affected, denoted by $\mu_{re,i}$, and the the value of loads lost of each load $i = 1, \cdots, N_L$ affected, denoted by $\mu_{out,i}$, over the time horizon τ. Next, the sum of these costs is compared with the cost of implementation of mitigation alternatives for each component $i = 1, \cdots, N_C$ affected, denoted by $\mu_{inst,i}$. This comparison enables to determine if investing on resilience enhancements is more advantageous than doing nothing for impact mitigation. In this context, some particularities are described as follows.

Equation (4.10) describes $\mu_{re,i}$ as a function of the damage level $d_{i,t} \in [0,1]$ over time, $t = 1, \cdots, \tau$, considering a maximum cost $\overline{\mu_{re,i}}$ over time for each component $i = 1, \cdots, N_C$.

$$\mu_{re,i} = \sum_{t=1}^{\tau} d_{i,t} \times \overline{\mu_{re,i}}, \text{ for} i = 1, \cdots, N_C. \tag{4.10}$$

The decision of implementing a specific mitigation strategy is formulated as a long-term cost minimization function over τ, considering the costs associated with the installation of mitigation options, repair of damaged components, and amount of energy not supplied. Hence, let $\mu_{inst,i}$, $\mu_{re,i}$, and $\mu_{out,i}$ be the costs of implementation of mitigation alternatives, maintenance (repair or replacement) of damaged components, and amount of energy not supplied associated with the i^{th} component, $i = 1, \cdots, N_C$, or load, $i = 1, \cdots, N_L$. For the utility company, the decision is justified if the implementation cost is lower than the maintenance and outage costs "as is", that is, if (4.11) holds for some $i = 1, \cdots, N_C$:

$$\mu_{inst,i} \leq \mu_{out,i} + \mu_{re,i}. \tag{4.11}$$

In turn, for the community and household owners, the decision of implementing a specific mitigation alternative is justified if the implementation cost is lower than the maintenance and outage costs "as is", considering the costs that are not subsidized by the government. Thereby, (4.11) is turned into (4.12) in the case of low-income community and households, with $i = 1, \cdots, N_L$ and $s_{inst,i}$, $s_{out,i}$, $s_{re,i} \in [0,1]$ denoting the per unit costs that are not subsidized by the government with regards to implementation, energy consumption, and maintenance, respectively:

$$s_{inst,i}\mu_{inst,i} \leq s_{out,i}\mu_{out,i} + s_{re,i}\mu_{re,i}. \tag{4.12}$$

As the occurrence of a natural disaster, extreme weather event, or man-made attack is often uncertain, a probabilistic modelling may be assumed to account for failure of individual grid components, as in [68]. This approach allows for a probabilistic evaluation of outages triggered by specific probabilistic conditions. In this situation, (4.11) and (4.12) become (4.13) and (4.14), where $\mathcal{P}(x)$ is the probability of the value x of a specific variable of interest (e.g., wind speed, precipitation, electric current) to occur and lead to an outage, as follows:

$$\mu_{inst,i} \leq \mathcal{P}(x)\left(\mu_{out,i} + \mu_{re,i}\right), \tag{4.13}$$

and

$$s_{inst,i}\mu_{inst,i} \leq \mathcal{P}(x)\left(s_{out,i}\mu_{out,i} + s_{re,i}\mu_{re,i}\right). \tag{4.14}$$

It is noteworthy that $\mathcal{P}(x)$ may be obtained from fragility curves of specific grid components, historical data, expectations, among others.

4.3.3 Case studies

In this section, two distinct scenarios are taken into consideration to evaluate whether the decision of improving grid resilience against extreme events is feasible in low-income neighbourhoods, inspired in Brazil. Notably, the following mitigation strategies are taken into consideration:

1. deploying intelligent electronic devices to detect and eliminate thefts of energy, and
2. installing off-grid PV systems or independent generators at household or community level for autonomous operation from the external grid against supply interruption.

These choices are made because the first strategy contemplates upgrades on the power distribution network for hardening infrastructure, whereas the second strategy includes distinct off-grid generators at household or community level to provide autonomous operation from the external grid. These situations are evaluated in distinct settings to determine whether the decision of improving grid resilience against extreme events is feasible in low-income neighbourhoods.

Initial considerations

The value of loads lost is equivalent to $\mu_{out,i} = \eta_{out,i} \times e_{out,i}$, with unit cost per energy $\eta_{out,i} = 0.75$ R$/kWh given in the Brazilian Real currency (R$) and energy consumption $e_{out,i}$ of an individual household between 30 and 220 kWh/month [4]. Even though the government usually subsidizes the electricity bills of low-income households usually (partially or totally), the value of loads lost is still supposed to be paid in full for the distribution utility company.

Regarding electrical components, the unit costs ($\eta_{inst,i}$, $\eta_{re,i}$) in use for overhead power lines are equivalent to 80 R$/m and $\mu_{inst,i} = \mu_{re,i} = 80 \times l_i$ in R$, with l_i representing the line length in m. The investment cost of an intelligent electronic device for smart metering is equal to R$ 100 for a single-phase device and R$ 300 for a three-phase device, whereas their respective maintenance costs correspond to R$ 20 and R$ 50 per year. The installation costs of off-grid PV panels with batteries are given in Table 4.3.

Deploying IEDs to detect and eliminate thefts of energy

In this situation, the feasibility of deploying IEDs to monitor electric variables and detect energy thefts is evaluated in two settings:

1. along the distribution lines, and
2. at the connection points of the households with the external grid through metered supply.

Table 4.3: Case study – Off-grid PV system cost parameters

Power (kW)	$\mu_{inst,i}$ (R$)
1	2500
2	5000
5	7500
6	10000
10	12500

The feasibility of deploying IEDs hereby evaluated regards irregular customers possibly connected to the grid as energy thefts, but the electricity is accounted as commercial losses. In this case, if the distribution utility decides to issue an electricity bill to the people of the community, it must pay taxes to the government in advance, risking not receiving payment from these people afterwards. As an outcome, these unpaid bills lead to commercial losses, as the customers might not be able to afford their electricity bills.

In this case, $\mu_{out,i}$, $i = 1, \cdots, N_L$, corresponds to the value of loads that are supplied irregularly and not accounted by the utility, whereas the number of metering devices is denoted by N_{IED}. They are supposed to be three-phase devices along the distribution lines and single-phase at the connection points of the households with the external grid.

Considering that the implementation costs are paid in full by the utility company, (4.11) can be re-written as (4.15) as follows:

$$\sum_{i=1}^{N_{IED}} \mu_{inst,i} > \sum_{i=1}^{N_L} \mu_{out,i} + \sum_{i=1}^{N_{IED}} \mu_{re,i}. \tag{4.15}$$

Consider a distribution feeder with N_L irregular customers every 5 m, with $\mu_{out,i}$, $i = 1, \cdots, N_L$ varying from $0.75 \times 30 \times 12 \times \tau_{out}$ to $0.75 \times 220 \times 12 \times \tau_{out}$ R$ per household, where τ_{out} is the duration of the irregular supply in years. Then, consider the following settings:

1. $N_{IED} = 1$ along the distribution lines every 5 m and
2. $N_{IED} = 1$ at the connection point of the households with the external grid.

It holds that $\mu_{re,i} = 50 \times \tau_{out}$ R$ in the first setting and $\mu_{re,i} = 20 \times \tau_{out}$ R$ in the second setting, $i = 1, \cdots, N_{IED}$. The outage duration τ_{out} that compensates for the mitigation alternatives aforementioned can be calculated as a function of the costs. The results are displayed in Table 4.4 according to N_L, $e_{out,i}$, and the installation setting in consideration.

If τ_{out} is shorter than the lifespan of the metering infrastructure, then its implementation is recommended. Assuming that it remains in operation for at least 36 months for the purpose of this analysis, then its implementation is justified in all settings of Table 4.4.

Table 4.4: Results – Outage duration (months)

Setting	Household		Distribution	
$e_{out,i}$ (kWh/month)	30	220	30	220
$N_L = 1$	12.5	1.8	11.25	1.8
$N_L = 3$	4.15	0.6	4.2	0.6
$N_L = 5$	2.5	0.36	2.6	0.4

Installing off-grid PV systems at household or community level against supply interruption

This section evaluates the installation of off-grid PV systems (that is, PV panels combined with batteries) at household or community level for autonomous operation from the external grid against man-made attacks. In this case, the affected lines are located after the point of connection to the customers and belong to the household owners.

The installation costs given in Table 4.3 are considered to calculate the implementation cost of off-grid PV systems, whereas the cost associated with replacement of destroyed or stolen lines are given by the installation cost of a new overhead line, i.e., $\mu_{re,i} = 80 \times l_i$ R$. In addition, the costs associated with the power outage are reflected in the operation costs associated, for example, with independent generators (e.g., diesel) and non-electrified appliances (e.g., wood ovens and candles). These costs are assumed to be at least three times more expensive than the electricity cost, i.e., $\eta_{out,i} = 2.25$ R$/kWh, whereas the emergency energy consumption is cut to half over the outage duration (that is, ranging from 15 to 110 kWh/month).

As $\mu_{re,i}$ is a function of the line length, the total line length replaced can be written as a function of the individual line length $l_{unit,i}$ multiplied with the frequency of occurrence, denoted by f_d, over the time horizon τ, with $\mathcal{P}(x) = f_d$. In other words, $l_i = l_{unit,i} f_d \tau$. Thereby, the time horizon when the repair and outage costs compensate for the cost of implementation of the mitigation option, given by (4.16), can be obtained from (4.14):

$$\tau \geq \frac{s_{inst,i}\mu_{inst,i}}{\left(\eta_{out,i} \times e_{out,i} + \eta_{re,i} \times l_{unit,i}\right) f_d}. \tag{4.16}$$

Assume $l_{unit,i} = 2$ m for a single household, an outage duration $\tau_{out} = 48$ h per occurrence, and the emergency energy consumption $e_{out,i}$ equivalent to half of the power supplied by each corresponding off-grid system during τ_{out}. Considering $s_{inst,i} = 1$ (no installation subsidized) and $s_{inst,i} = 0.5$ (50 % subsidized), with $s_{re,i} = 1$ (that is, no replacement subsidized) and $s_{out,i} = 0$ in normal operation for simplification (that is, energy consumption 100 % subsidized in normal operation), the results of (4.16) are summarized in Tables 4.5 and 4.6 as a function of f_d, with $f_d = 1, 2, 3$ per year, and the peak power provided by each off-grid system.

Table 4.5: Results – Return horizon without subsidization (years)

Power (kW)	τ (years)		
f_d (year^{-1})	1	2	3
1	15.4	7.2	5.2
2	30.4	15.2	10.2
5	44.6	22.3	14.9
6	58.7	29.4	19.6
10	70.8	35.4	23.6

Table 4.6: Results – Return horizon with 50% subsidization (years)

Power (kW)	τ (years)		
f_d (year^{-1})	1	2	3
1	7.7	3.6	2.6
2	15.2	7.6	5.1
5	22.3	11.2	7.5
6	29.4	14.7	9.8
10	35.4	17.7	11.8

4.3.4 Discussion

The results of Section 4.3.3 indicate that the feasibility of enhancing grid resilience in low-income neighbourhoods can be interpreted in different ways. To this extent, for a specific extreme event, the probability of occurrence, impact, and possible mitigation alternatives make an impact on the decision process.

In the first strategy evaluated, the results are influenced by the implementation costs and energy consumption from irregular customers. In some cases, the return of investment is almost immediate, while in others it may take longer for this to happen. By and large, the return of investment is expected to occur for all settings in less than 2 months when $e_{out,i} = 220$ kWh/month and in roughly a year when $e_{out,i} = 30$ kWh/month.

In the second strategy, the results are also influenced by the implementation costs not subsidized and the expected frequency of occurrence. In this case, however, the return of investment is more likely to occur when the energy consumption is lower, as the implementation cost of the corresponding off-grid PV system is cheaper in these cases. Nonetheless, the results show that the subsidization provided by the government may lead to a faster return of investment.

In the second situation, the usage of batteries for other purposes than off-grid electricity generation is disregarded. In this context, the people using those off-grid PV systems

should be properly instructed and assisted from the beginning of the implementation until the lifetime end of the system.

Overall, the main findings suggest that the costs associated with maintenance and energy consumption may compensate for the investment costs associated with alternatives for enhancing grid resilience in low-income neighbourhoods. Nevertheless, the results indicate that upgrades in the electricity network infrastructure are more cost-effective than island-mode operation, as the return of the investment is significantly shorter in all simulated settings. However, the results are heavily dependent from the amount of electricity supplied.

The results previously obtained were calculated with a deterministic analysis (that is, using constant demand profiles and frequencies of occurrence over time) that does not consider the temporal value of money. Despite that, the methodology is general enough to handle non-deterministic values and time-varying costs. In this case, the constant demand profiles, event frequencies, and costs involved in the analysis must be replaced with varying ones.

4.4 Designing power lines for resilient, cost-effective power systems under wind storms

This section presents a procedure to evaluate the pros and cons of choosing between overhead and underground power lines for resilient, cost-effective power networks under wind storms. Among all natural disasters, strong winds represent a particular class of HILP events that is highly likely to damage distribution towers and poles and pull down overhead lines in vulnerable areas.

The methodology assesses the costs and risks associated with installation, operation, and repair of underground and overhead power lines and also takes into account the penalty costs for the energy not supplied. The decision process is formulated as a long-term cost minimization function over the time horizon of the analysis τ, considering the costs associated with individual lines together with their probability of failure. Thus, consider a distribution grid with E lines and N poles (in case of overhead lines) and let $\mu_{inst,i}$, $\mu_{op,i}$, $\mu_{re,i}$ be the costs of installation, operation, and repair of the i^{th} corridor over τ, $i = 1, \cdots, E$. In addition, let $\mu_{out,i}$ be the penalty for the energy not supplied due to an outage of the same i^{th} corridor over τ.

The impact of wind storms on the system resilience is assessed through fragility modelling of individual poles and lines, whenever applicable, as in [68]. This procedure allows for a probabilistic evaluation of system outages as a function of the wind speed w. The fragility curves are given by (4.17) for a single pole and (4.18) for an overhead line, considering a critical speed w_{crit} above which the probability of failure increases and a collapse speed w_{coll} from which the failure is certain:

$$P_T(w) = \begin{cases} 0, \text{ if } w < w_{crit} \\ P_{T,hw}, \text{ if } w_{crit} \leq w < w_{coll} \\ 1, \text{ if } w \geq w_{coll} \end{cases}, \tag{4.17}$$

and

$$P_L(w) = \begin{cases} P_{L,std}, \ if \ w < w_{crit} \\ P_{L,hw}, \ if \ w_{crit} \leq w < w_{coll} \\ 1, \ if \ w \geq w_{coll} \end{cases}. \tag{4.18}$$

The outage of each pole is supposed to be independent of the condition of adjacent poles in the i^{th} corridor. Likewise, the failure of a power line is independent from the failure of poles in the i^{th} corridor. Considering that N_i poles are connected in series along the i^{th} corridor, the failure of a single component – whose failure function is denoted by $F_{L,i}$ for a line and equivalent to (4.18) and by $F_{T,j}$ for a pole and equivalent to (4.17), $j = 1, \cdots, N_i$ – will trip the entire corridor i. Thereby, the joint probability of failure as a function of the wind speed, denoted by $P_i(w)$, is such that (4.19) holds:

$$P_i(w) = 1 - P\left[(F_{L,i} = 0) \cap (F_{T,1} = 0) \cap \cdots \cap (F_{T,N_i} = 0)\right], \tag{4.19}$$

with

$$\begin{cases} P\left[(F_{L,i} = 0)\right] = 1 - P_L(w) \\ P\left[(F_{T,j} = 0)\right] = 1 - P_T(w), \ j = 1, \cdots, N_i \end{cases}. \tag{4.20}$$

For underground lines, (4.18) is replaced with $P_L(w) = P_{L,std}$ constant and (4.17) does not apply. Thereby, (4.19) is equivalent to $P_i(w) = P_{L,std}$, $\forall\ w$.

The wind speed profile over time is defined as a stochastic function, as in [103], following the Weibull probability distribution function $f(w)$ (4.21), where c is a scale parameter and k is a shape parameter:

$$f(w) = k\frac{w^{k-1}}{c^k}\exp\left(-\left(\frac{w}{c}\right)^k\right). \tag{4.21}$$

Considering that the cost estimates are made every hour t and that w is time-dependent, i.e., $w = w_t$, the accumulated cost λ_i of a single corridor over τ is given by (4.22):

$$\lambda_i = \mu_{inst,i} + \sum_{t=1}^{\tau}\left[\mu_{op,i,t} + P_i(w_t)\left(\mu_{re,i}(w_t) + \mu_{out,i,t}\right)\right]. \tag{4.22}$$

In (4.22), $\mu_{inst,i}$ is proportional to the line length, denoted by d_L, whereas $\mu_{op,i}$ is proportional to the contracted power of all customers and the line length divided by the total length of the network. The penalty applied for the energy not supplied is calculated

with the contracted power of each affected customer (robust scenario), which is assumed to be time-independent and weather-independent. However, $\mu_{out,i}$ is time-varying due to the out-of-service loads changing over time. In turn, $\mu_{re,i}$ is assumed to be a function of w to reflect that the damage and time to repair increase with the wind speed, according to [68].

Alternatively, the value of lost load resulting from the failure of the i^{th} corridor can be denoted by $VoLL_i$ and calculated in separate with (4.23), as follows:

$$VoLL_i = \sum_{t=1}^{\tau} P_i\left(w_t\right) \mu_{out,i,t}. \tag{4.23}$$

4.4.1 Case study: the RESOLVD project network modified

The methodology is demonstrated in a modified version of the real-based rural distribution network illustrated in Fig. 3.1. In the present analysis, the network under consideration has 2 distribution substations, 24 feeders, 3 switches, and 20 different energy consumers (among them, 1 industrial, three-phase and 19 residential, single-phase). The sizing of substation transformers is 500 kVA at SS-1 and 250 kVA at SS-2, whereas the contracted power is 100 kW for the industrial customer (power factor 0.96) and 10 kW for each single-phase residential customer (power factor 0.92). Note that the PV panels of the original network are omitted from the grid.

The switches connecting the distribution substation with the external grid are normally closed, whereas the switch connecting the two parts of the distribution network is normally open. In the event of a component failure, they enable grid reconfiguration and disconnection whenever needed in order to minimize the total load shedding.

The following situations are considered in the specification of this grid:

1. underground lines,
2. overhead lines without hardening, and
3. overhead lines with additional hardening.

The costs and failure rates of the power lines in use are approximations from [97] and [78], whereas the cost of energy not supplied, given by the electricity price multiplied by the load lost, is an approximation from [56]. In addition, the duration of an outage, denoted by τ_{out}, is assumed to be longer in underground lines than in overhead lines, in accordance to [97]. The costs are assumed to be constant along the year and higher for industrial customers than for residential customers.

For comparison, two time horizons are considered to compute the costs associated with operation, repair, and out-of-service loads of the grid, $\tau = 10$ years and $\tau = 20$ years. In addition, penalty costs equivalent to $\mu_{out,i} = 7.0$ \$/kWh and $\mu_{out,i} = 35.0$ \$/kWh are analysed over $\tau = 10$ years. The other costs in use for individual overhead and

Table 4.7: Case study – Cost parameters of distinct power lines

	Light Overhead	Hardened Overhead	Underground
$\mu_{inst,i}$ (\$/mi)	$100,000$	$200,000$	$400,000$
$\mu_{op,i}$ (\$/kWh)	0.10	0.10	0.05
$\mu_{re,i}$ (\$/mi)	$1,000$	$2,000$	$10,000$
τ_{out} (h)	12	12	72

Table 4.8: Case study – Line parameters of the grid

Number	Length [mi]	N_i	Shedding [kW]
1	21	2	50
2	28	3	10
3	36	4	0
4	42	4	0
5	25	3	0
6	36	4	0
7	15	2	0
8	42	4	30

underground lines are summarized in Table 4.7. The lengths of the lines highlighted in Fig. 3.1 and load shedding resulting from a failure along them are listed in Table 4.8.

The fragility curves are given by (4.24) and (4.25) for an overhead corridor without hardening, by (4.24) and (4.26) for an overhead corridor with additional hardening along the lines, and by (4.27) for an underground corridor:

$$P_T(w) = \begin{cases} 0, \text{ if } w < 40 \text{ mph} \\ \frac{w-40}{60-40} \text{ h}^{-1}, \text{ if } 40 \leq w < 60 \text{ mph} \\ 1, \text{ if } w \geq 60 \text{ mph} \end{cases}, \tag{4.24}$$

and

$$P_{L1}(w) = \begin{cases} \frac{1.25 \times d_L}{8760} \text{ h}^{-1}, \text{ if } w < 25 \text{ mph} \\ \frac{w-25}{40-25}\left(1 - \frac{1.25 \times d_L}{8760}\right) + \frac{1.25 \times d_L}{8760} \text{ h}^{-1}, \text{ if } 25 \leq w < 40 \text{ mph} \\ 1, \text{ if } w \geq 40 \text{ mph} \end{cases}, \tag{4.25}$$

and

Table 4.9: Results – Value of load lost over $\tau = 10$ years, $\mu_{out,i} = 7.0$ \$/kWh (\$)

i	Light Overhead	Hardened Overhead	Underground
1	29,343	4,142	7
2	5,869	895	2
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	17,610	2,825	9

$$P_{L2}(w) = \begin{cases} \frac{0.75 \times d_L}{8760} \text{ h}^{-1}, \text{ if } w < 40 \text{ mph} \\ \frac{w-40}{60-40}\left(1 - \frac{0.75 \times d_L}{8760}\right) + \frac{0.75 \times d_L}{8760} \text{ h}^{-1}, \text{ if } 40 \leq w < 60 \text{ mph} \\ 1, \text{ if } w \geq 60 \text{ mph} \end{cases}, \tag{4.26}$$

and

$$P_{L3}(w) = \frac{0.25 \times d_L}{8760} \text{ h}^{-1}, \ \forall \ w. \tag{4.27}$$

The probability distribution function of the wind speed is calculated with (4.21) using $c = 10$ and $k = 1.2$. The wind speed over time is drawn randomly from (4.21) using MATLAB. A one-week interval is considered to calculate the costs over the outage times defined in Table 4.7.

4.4.2 Results

This section presents the results calculated for the scenarios previously described. The value of load lost and the accumulated costs calculated for overhead lines without hardening, with hardening, and underground lines are shown in Tables 4.9 and 4.10 for $\tau = 10$ years, in Tables 4.11 and 4.12 for $\tau = 20$ years with $\mu_{out,i} = 7.0$ \$/kWh, and in Tables 4.13 and 4.14 for $\tau = 10$ years with $\mu_{out,i} = 35.0$ \$/kWh. Additionally, Fig. 4.2 illustrates the chosen lines in terms of accumulated costs according to the scenarios enumerated in Section 4.4.1.

4.4.3 Discussion

Overall, it can be noticed in Section 4.4.1 that the overhead corridors without hardening present the highest values of load lost when a failure causes load shedding, whereas un-

Table 4.10: Results – Accumulated costs over $\tau = 10$ years, $\mu_{out,i} = 7.0$ \$/kWh (\$)

Number	Light Overhead	Hardened Overhead	Underground
1	31,807	8,601	8,517
2	9,155	6,845	11,347
3	4,224	7,653	14,472
4	4,928	8,929	17,018
5	2,933	5,312	10,129
6	4,224	7,653	14,472
7	1,760	3,185	6,078
8	22,539	11,754	17,027

Table 4.11: Results – Value of load lost over $\tau = 20$ years, $\mu_{out,i} = 7.0$ \$/kWh (\$)

Number	Light Overhead	Hardened Overhead	Underground
1	57,181	8,319	15
2	11,437	1,735	4
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	34,316	5,348	18

Table 4.12: Results – Accumulated costs over $\tau = 20$ years, $\mu_{out,i} = 7.0$ \$/kWh (\$)

Number	Light Overhead	Hardened Overhead	Underground
1	60,002	13,037	8,634
2	15,199	8,031	11,494
3	4,837	8,099	14,774
4	5,643	9,449	17,236
5	3,359	5,622	10,260
6	4,837	8,099	14,774
7	2,015	3,373	6,156
8	39,959	14,797	17,255

Table 4.13: Results – Value of load lost over $\tau = 10$ years, $\mu_{out,i} = 35.0$ \$/kWh (\$)

i	Light Overhead	Hardened Overhead	Underground
1	146,720	20,710	39
2	29,345	4,476	10
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	88,049	14,122	47

Table 4.14: Results – Accumulated costs over $\tau = 10$ years, $\mu_{out,i} = 35.0$ \$/kWh (\$)

Number	Light Overhead	Hardened Overhead	Underground
1	149,184	25,168	8,548
2	32,631	10,425	11,356
3	4,224	7,653	14,587
4	4,929	8,930	17,018
5	2,933	5,312	10,129
6	4,224	7,653	14,587
7	1,760	3,185	6,078
8	92,979	23,052	17,065

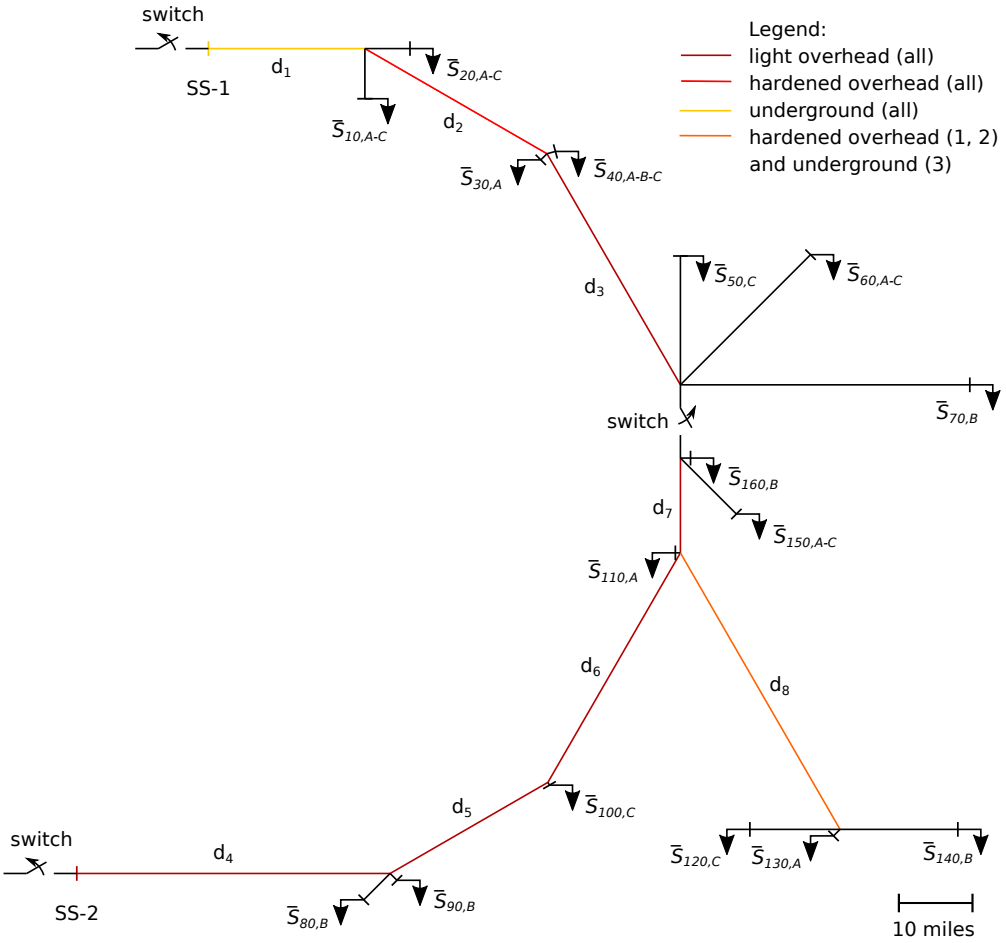

Figure 4.2: Results – Single-line diagram of the radial distribution network showing the lines chosen in terms of accumulated costs

derground lines present the lowest values of load lost when a failure causes load shedding in all simulations. This evinces the high impact of wind storms on overhead corridors in terms of damage and load lost.

Conversely, the costs calculated with (4.22) are the lowest for overhead lines without hardening and the highest for underground cables for all corridors when when a component failure produces no load shedding (that is, all lines except 1, 2, and 8 in all scenarios). In turn, for line number 1, (4.22) is the lowest with use of underground cables for all scenarios due to the relatively high load shedding; for line number 2, (4.22) is the lowest with use of overhead lines with hardening for all scenarios due to the relatively low load shedding; and for line number 8, (4.22) is the lowest with use of overhead lines with hardening when $\mu_{out,i} = 7.0$ US$/kWh and the lowest with use of underground cables when the penalty is increased to $\mu_{out,i} = 35.0$ US$/kWh due to the costs associated with load shedding in different scenarios.

To conclude, the results indicate that the accumulated costs associated with operation, repair, and penalties for the energy not supplied may compensate the installation costs in a long time horizon. In some cases, depending on the length of the line and load shedding caused by a failure, the installation of more reliable, but expensive power lines may be the best option when it comes to both resilience increase and cost reduction in the long run. Furthermore, the results suggest that a high penalty for the energy not supplied may be an effective policy mechanism to incentive resilience-, cost-effective power system planning.

4.5 Power system resilience to floods

The goal of improving the long-term power system resilience in flood-prone areas takes into consideration the hydrological models and the location of electrical equipment, as well as financial and system-related studies. This information is then used to assess the impact of floods on the grid, based on distinct flood scenarios in a specific area. Next, distinct alternatives for mitigation are evaluated in terms of resilience enhancements and costs associated. Ideally, the choice of the best mitigation alternative will improve grid resiliency with minimum cost, considering all possible scenarios to find the best feasible solution.

The following subsections describe the hydrological model, flood modelling, grid representation, impact assessment, and mitigation alternatives within an optimization problem formulation.

4.5.1 Hydrological modelling

This subsection presents an overview of the hydrological model used in the study of flood scenarios. The WRF-Hydro model [41] consists of a land surface model, a terrain routing module, and channel and reservoir routing modules. The inputs are meteorological forcing variables (forecasts or observations), which demonstrate weather conditions dynamically over time. The major output for this study is streamflow.

The base flow parametrization relies on a conceptual groundwater reservoir (that is, a bucket with a conceptual depth and volumetric capacity). It is recharged with spatially-aggregated drainage from the soil profile and discharged into the stream network as channel inflow. The streamflow provides information about the height of the water level in the drainage areas, which corresponds to the HAND values provided by the NWM/WRF-Hydro model [55]. The HAND values are then used to determine whether a substation site is flooded.

It is noteworthy that groundwater mechanisms are not very important during extreme flash floods. In such events, a large amount of the rainfall is not able to infiltrate into soil (infiltration excess) and/or soil gets saturated and cannot hold more water (saturation excess). Infiltration excess and saturation excess produce surface run-off (and shallow

subsurface flow), which is essential for flash flood. Conversely, base flow is only part of streamflow that comes from groundwater discharge.

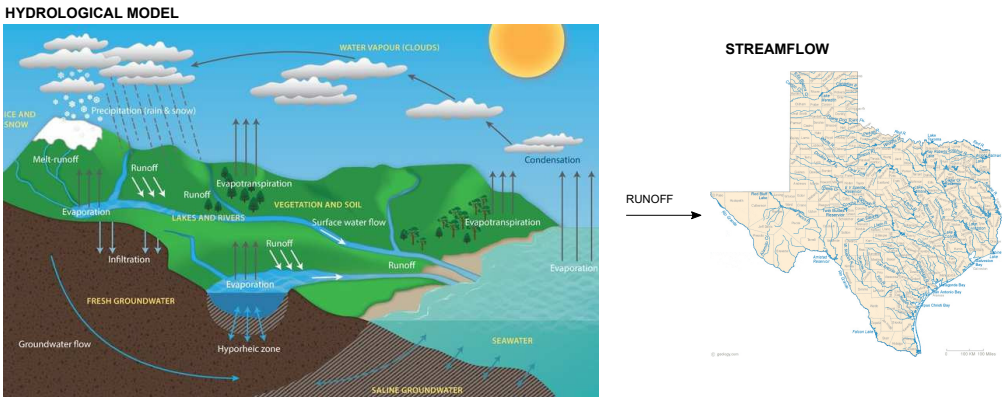

Figure 4.3: Resilience to floods – General framework of the modelling chain with hydrological model (left) and streamflow (right) (Adapted from [80])

A general framework of the modelling chain is shown in Fig. 4.3. It represents the inputs from weather forecasts and the run-off from the hydrological model used to obtain the streamflow.

4.5.2 Flood modelling

Floods are regarded as a function of the height of the water level over time where the individual grid components are located. The outage duration is given by (4.28) as follows, as the affected substations must be shut down and remain disconnected from the grid until the flood is cleared and the damage is repaired:

$$\tau_{out,i,f} = \tau_{flood,i,f} + \tau_{re,i,f}. \tag{4.28}$$

Over a short-term horizon, the probability of an outage caused by flood can be estimated as a function of the probability of adverse weather conditions to produce flooding that reaches the elevation of electrical equipment at a specific location, given its hydrological model. In this scenario, the probability of an adverse weather condition to occur can be obtained with weather forecasts. In turn, the probability of an outage under adverse weather conditions can be estimated with data fitting models (e.g., [96]), simulations of probabilistic outage scenarios (e.g., [68]), and/or using historical data to compute the outage rate under adverse weather conditions in a specific region (e.g., [6]).

Over a long-term horizon, however, the probability of adverse weather conditions to produce flooding at a specific time and location cannot be estimated with available forecasts. In this context, the usage of data from past events and simulations of realistic flood scenarios is recommended to determine the out-of-service areas. As an outcome, the

probabilistic assumption is replaced with “flooded” or “not flooded” for each scenario in consideration.

Scientists expect climate change to increase the intensity and frequency of occurrence of extreme weather events [37]. Moreover, in the coming decades some types of infrastructure, including electric substations, are expected to become more vulnerable to floods due to high rainfall and a rise in sea level. Although these aspects may also be reflected in the duration of an outage and in the damage inflicted to specific components, they are uncertain and may lead to non-realistic assumptions in future projections. To prevent this, the usage of realistic flood scenarios is recommended under the assumption that they may occur with some periodicity over a long-term horizon.

4.5.3 Grid representation

The grid is represented in three situations illustrated in Fig. 4.4: (a) before flooding, (b) flooding “as is”, and (c) flooding mitigated by resilience planning. It is composed originally of N_S transmission substations, N_G generation units, and N_L loads in service. Before flooding, the grid components are part of an interconnected system represented as a connected graph (Fig. 4.4(a)). In the event of a flood f, the remaining part of the operating grid is composed of $N_S - N'_{S,f}$ transmission substations, $N_G - N'_{G,f}$ generators, and $N_L - N'_{L,f}$ transmission loads “as is”; and of $N_S - N'_{S,f} + N''_{S,f}$ transmission substations, $N_G - N'_{G,f} + N''_{G,f}$ generators, and $N_L - N'_{L,f} + N''_{L,f}$ transmission loads with resilience planning.

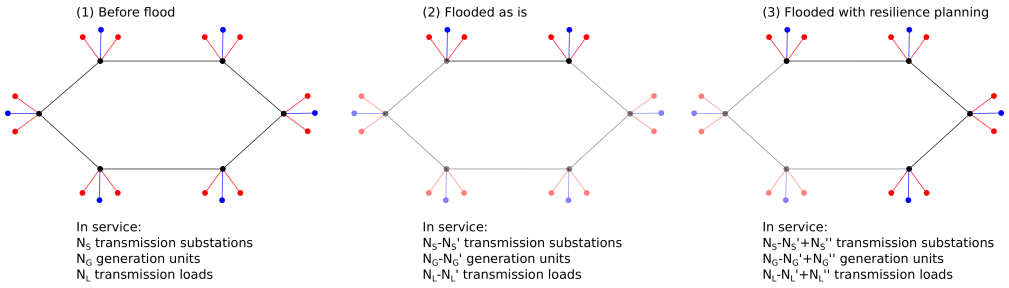

Figure 4.4: Resilience to floods – Equivalent graph representation of a generic interconnected power system (a) before flooding, (b) flooding “as is”, and (c) flooding mitigated by resilience planning, in which the nodes represent substations (large black dots), generation units (blue lozenges), and loads (small red dots) and the edges represent existing paths connecting those nodes

Long-term improvements in power system resilience to floods are targeted at individual grid components located in vulnerable areas. Thus, the present analysis is focused on transmission substations, since these components are usually installed at or nearly above the ground level and belong to the transmission utility company. Transmission substations are typically connected to other transmission and distribution substations, generation units, and loads through lines or power transformers. Hence, system effects shall also be

investigated, as an outage at a transmission substation may affect other grid components connected at lower voltage levels. It is noteworthy that other substations connected at equivalent or higher voltage levels will not be affected as long as the interconnected system is designed in compliance with the $N-1$ reliability criteria [65]. In the event of an outage with loss of generation, the system is supposed to operate in contingency reserve mode.

The search for a re-design to improve grid resilience shall guarantee continuity and quality of service at acceptable levels over τ (assumed to be the ending time for the purpose of this analysis). The system planner is assumed to have information about feasible mitigation solutions, which may contemplate the hardening of existing infrastructure and the replacement of some of the existing components from the original out-of-service area.

4.5.4 Impact assessment

In this subsection, the impact of a flood scenario $f \in \mathcal{F}$ on the grid is regarded as a function of the out-of-service loads, damaged equipment, and operation in contingency reserve mode. The following paragraphs present definitions and mathematical formulations.

The expected energy not supplied at the i^{th} transmission substation until τ ($EENS_{i,f,\tau}$) is calculated with (4.29) and quantified as a function of (4.28) and $p_{load,i,t}$, $\forall i \in \mathcal{S}'_f$, with t assuming discrete values as follows:

$$EENS_{i,f,\tau} = \sum_{t=1}^{\tau} p_{load,i,t} \times \tau_{out,i,f}. \tag{4.29}$$

In this analysis, $\mu_{out,i,f}$ is defined as a penalty cost imposed on the grid due to the outage of some of its loads $\forall l_i \in \mathcal{L}'_f$ and extra costs for operation in reserve mode $\forall g_i \in \mathcal{G}'_f$. It is calculated as a function of the generation capacity and loads connected to the flooded substation, as in (4.30):

$$\mu_{out,i,f,\tau} = \begin{cases} \eta_{l,out} \times EENS_{i,f,\tau}, \text{ if } p_{load,i,t} \geq p_{gen,i,t} \\ \eta_{l,out} \times EENS_{i,f,\tau} + \eta_{g,out} \times \\ \quad (p_{gen,i,t} - p_{load,i,t}) \times \tau_{out,i,f}, \text{ if } p_{load,i,t} < p_{gen,i,t} \end{cases}. \tag{4.30}$$

In turn, the damage inflicted to the transmission substations may vary considerably, requiring repair or full replacement of their components. The accumulated costs over τ are given by $\mu_{re,i,f,\tau}$, $\forall i \in \mathcal{S}'_f$.

The accumulated costs associated with flooded substations "as is" over τ are given by the sum of $\mu_{out,i,f,\tau}$ and $\mu_{re,i,f,\tau}$, as in (4.31), $\forall i \in \mathcal{S}'_f$:

$$\lambda_{i,f,\tau} = \mu_{out,i,f,\tau} + \mu_{re,i,f,\tau}. \tag{4.31}$$

Equation (4.32) represents a vector of the total accumulated costs (4.31) "as is" when considering the f^{th} scenario of $\mathcal{F}$:

$$\lambda_{f,\tau} = \left[\lambda_{1,f,\tau}, \cdots, \lambda_{N'_{S,f},f,\tau}\right]^T. \tag{4.32}$$

Alternatively, considering only the potential decrease in the expected energy not supplied, (4.32) shall be replaced with (4.33), given by the maximum (4.29) "as is" when considering the f^{th} scenario of $\mathcal{F}$ over τ:

$$\mathbf{EENS}_{f,\tau} = \left[EENS_{1,f,\tau}, \cdots, EENS_{N'_{S,f},f,\tau}\right]^T. \tag{4.33}$$

Ultimately, the resilience savings of a candidate option for impact mitigation can be quantified by considering the potential decrease in (4.32) or in (4.33) when all loads affected are in service.

4.5.5 Mitigation options

This subsection describes possible strategies for mitigation of the impacts of floods on transmission substations. For a more thorough description of hardening strategies recommended to enhance power system resilience to extreme weather events, see [93].

Hardening options at substations typically consist of elevating equipment or relocating facilities to areas less prone to flooding. In general, elevated structures provide a reliable, cost-effective, environment-friendly solution. However, elevating an existing transmission substation may be a challenging task. In this context, modular substation designs that can be installed on elevated foundations or platforms are appealing.

The cost $\mu_{RP,i}$ associated with hardening of an individual substation $i \in \mathcal{S}'_f$ is calculated as a function of the total area $\left[m^2\right]$ of the substation sites. It is paid in full before $t = 0$ and given by (4.34):

$$\mu_{RP,i} = \eta_{RP,i} \times A_i. \tag{4.34}$$

The total cost vector $\lambda_{\mathbf{RP}}$ is then given by the individual costs (4.34), as in (4.35):

$$\lambda_{\mathbf{RP}} = \left[\mu_{RP,1}, \cdots, \mu_{RP,N'_{S,f}}\right]^T. \tag{4.35}$$

Supposing that $\lambda_{\mathbf{RP}}$ might be subject to a budget limit $\overline{B}$, the set of feasible solutions has to comply with an additional constraint (4.36) as follows:

$$\lambda_{\mathbf{RP}}{}^T \mathbf{s}'' \leq \overline{B}. \tag{4.36}$$

Thereby, considering (4.32), the minimax problem formulation is given by (4.37):

$$z^*_{\lambda} = \min_{\mathcal{S}''_f \subset \mathcal{S}'_f} \max_{f \in \mathcal{F}} \ \lambda^T_{f,\tau} \left[\mathbf{s}' - \mathbf{s}''\right] + \lambda_{\mathbf{RP}}{}^T \mathbf{s}'' \quad . \tag{4.37}$$
$$s.t. (4.36)$$

Alternatively, considering (4.33), the minimax problem formulation is given by (4.38):

$$z^*_{EENS} = \min_{\mathcal{S}''_f \subset \mathcal{S}'_f} \max_{f \in \mathcal{F}} \ \mathbf{EENS}^T_{f,\tau} \left[\mathbf{s}' - \mathbf{s}''\right] \quad . \tag{4.38}$$
$$s.t. (4.36)$$

Ultimately, the solutions of (4.37) and (4.38) are expected to provide quality and continuity of supply at acceptable levels to the entire grid, while reducing power outages and costs associated with loads lost and damaged equipment over τ. As a consequence, the chosen mitigation alternative is expected to be adequate in terms of both improving resilience and reducing costs.

4.5.6 Case study: Hurricane Harvey's rainfall in the coastal area of Texas

In this section, the methodology presented in Section 4.5 is demonstrated at substation sites in the coastal area of Texas, using simulations of realistic flood events and information about the reduced synthetic grid of the region. The map of the coastal area of Texas is shown in Fig. 4.5, with electric power substations primarily associated with electric power transmission represented by blue dots. Electric substation data obtained from [44] include nominal primary and secondary voltage and GPS coordinates. The hydrological model provides the height of the water level at these locations for a specific flood scenario, which is then compared with the corresponding HAND value. This procedure enables to determine whether a specific location is flooded.

Substation sites are typically designed to prevent flood damage at the 100-year flood elevation plus 0.5 meter [20], under the assumption that the probability of flood is accurately known. Thus, the present analysis considers that a substation $i \in \mathcal{S}$ is flooded at a specific time t if the height of the water level at the substation site is 0.5 m greater than the threshold (i.e., HAND value).

It is noteworthy that the return period of the flooding caused by Hurricane Harvey's rainfall is believed to be larger than 100 years [33] Were the intensity and frequency of events like Hurricane Harvey to increase in future projections, the 100-year flood plan elevation design is no longer adequate for the existing substations. However, as the occurrence of future events with specific characteristics cannot be determined accurately, a range of $|\mathcal{F}| = 25$ possible flood scenarios based on Hurricane Harvey's rainfall is considered for the purpose of this analysis.

The duration of floods is obtained from the scenarios simulated in the coastal area of Texas. As the rainfall scenarios are defined by daily rainfall totals, the costs associated

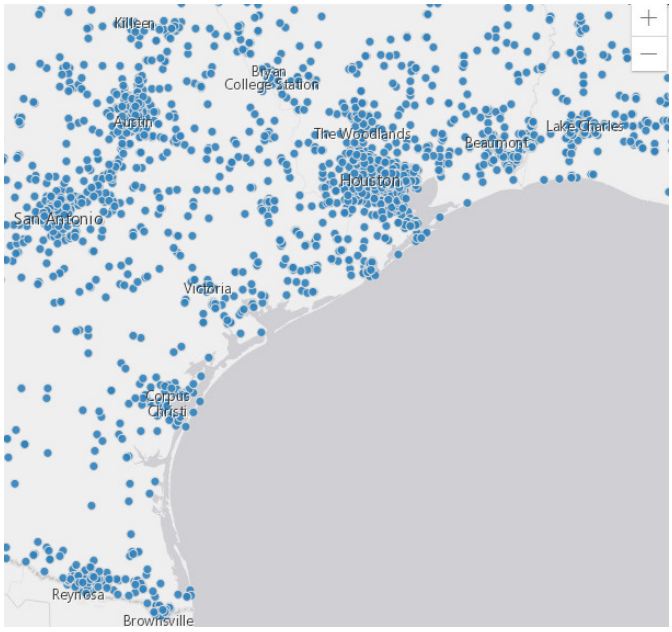

Figure 4.5: Case study – Map of the coastal area of Texas, with electric power substations primarily associated with electric power transmission represented by blue dots (Source: [44])

with the expected energy not supplied are calculated over 24 hours. In addition, the restoration is assumed to be possible in the same day the flood level recedes to below the threshold level, by temporarily allocating mobile substation units. This assumption is likely to lead to estimates of shorter restoration times and lower quantities of expected energy not supplied than we anticipate in real events.

The apparent power supplied by each substation is estimated with use of a reduced equivalent grid obtained from the ACTIVSg2000 synthetic grid, built from public information and a statistical analysis of real power systems in Texas (see [101] for more information). Electric substations from the reduced synthetic grid are matched to the real substation locations by assigning the electrical properties of the synthetic grid substations to the real substations with the minimum Euclidean distance between their coordinates in the GIS data. This procedure enables to connect the generators and loads from the synthetic grid to the real electric substations for a more realistic analysis of the effects of flooding of substations.

In the event of a flood, the maintenance cost depends on the damage to the substations and the time required to restore the power. The associated costs given by $\mu_{re,i,f,\tau}$ may vary from minor repair to full replacement. However, as the actual maintenance costs cannot be determined accurately prior to the occurrence of a flood, a fixed value equivalent to 75% of the full replacement cost is considered per occurrence.

All cost parameters are approximations taken from [7] and [46]. A unit cost of \$2 per

kVA is considered for full replacement of damaged substation components, whereas an unit cost of $100 per MWh is assumed for the expected energy not supplied and $6 per MWh for operation in reserve mode. In addition, an investment budget limit $\overline{B} = \$25$ M is considered.

The installation of elevation structures is regarded as a possible mitigation strategy in all scenarios. The costs associated with elevating structures in the substation area are equal to $200 per m^2. In turn, the substation dimensions adopted for A_i are equivalent to 100 m $\times$60 m when the apparent power is greater than 200 MVA, 60 m $\times$40 m when the apparent power is between 50 MVA and 200 MVA, and 40 m $\times$30 m when the apparent power is between 1 MVA and 50 MVA.

In agreement with long-term power system planning practices [51], the capacity expansion is assumed to be equivalent to 1% per year using the previous year as a reference. For the purpose of this analysis, no new substations are supposed to be built in flood-prone areas. The calculated value of loads lost and the repair/replacement of damaged substations are calculated over a time horizon $\tau = 10$ years. The costs and expected energy not supplied associated with all individual substations possibly affected by a flood are compared to determine the locations possibly presenting the highest values. In total, a range of 25 distinct flood scenarios is evaluated.

In addition, the calculations assume that a severe flood episode is expected to occur at least twice per decade and no more than once a year, which is a conservative assumption in the coastal area of Texas. As the exact dates cannot be determined, two situations are evaluated among the 45 possibilities over a decade: consecutive occurrences of the same flood scenario in the first two years and in the last two years of the time horizon. The results obtained with the smallest and largest amounts of power in the grid in a specific flood scenario reflect the extremes.

4.5.7 Results

Tables 4.15 and 4.16 displays information about the transmission substations flooded in the scenarios evaluated: the first column contains the substation number of the ACTIVSg2000 dataset, the second column shows the number of scenarios in which each substation is flooded out of 25 scenarios, the third and fourth columns show the total power (generation and load, respectively) connected to each substation at $t = 0$, and the fifth column shows the cost of elevating substations.

The results obtained "as is" and with resilience planning to mitigate the worst-case flood scenarios of $\mathcal{F}$ are shown in Tables 4.17 (accumulated costs) and 4.18 (expected energy not supplied). Table 4.17 displays the resulting value of (4.37) "as is" (second column) and with resilience planning (third column) for consecutive occurrences in the last two years of the time horizon (that is, the "latest" occurrences). IN turn, Table 4.18 displays the resulting value of (4.38) "as is" (second column) and with resilience planning (third column) in the "latest" occurrences.

Table 4.15: Case study – Details of flooded transmission substations (part 1 of 2)

substation	floods	$p_{gen,i,1}$ (MW)	$p_{load,i,1}$ (MW)	μ_{RP} (M$)
248	9/25	0.0	7.1	0.6
250	15/25	0.0	117.8	1.2
261	13/25	0.0	79.2	1.2
269	2/25	0.0	13.1	0.6
270	2/25	150.0	0.0	1.2
303	2/25	0.0	5.6	0.6
321	2/25	141.0	63.5	1.2
361	1/25	151.2	0.0	1.2
823	9/25	0.0	18.5	0.6
867	8/25	31.8	0.0	0.6
887	24/25	0.0	8.6	0.6
923	11/25	812.5	221.9	3
931	7/25	211.4	209.0	3
936	9/25	0.0	92.3	1.2
939	11/25	375.8	124.6	3
960	11/25	0.0	74.2	1.2
964	10/25	0.0	210.7	3
969	10/25	0.0	62.2	1.2
970	11/25	0.0	212.5	3
972	15/25	0.0	58.6	1.2
981	15/25	0.0	223.5	3
983	4/25	0.0	5.2	0.6
986	13/25	0.0	97.7	1.2
1003	17/25	0.0	122.5	1.2
1006	3/25	0.0	149.0	1.2

The prioritization lists of substations to be hardened are given in Table 4.19. The first column indicates the evaluation criteria, whereas the second column displays the list of substations chosen according to each evaluation criterion and the third column shows the calculated value of $\lambda_{\mathbf{RP}}{}^{T}\mathbf{s}''$. The results are the same in both the earliest and latest occurrences of Table 4.15 and 4.16 for a given metric, but hardening of substations 1006 and 1077 becomes feasible using (4.37) in the latest occurrences. Note that substations 964, 970, and 1147 are excluded from the list obtained with (4.38) due to their high μ_{RP} and budget constraints. In turn, substations 261, 1089, and 1099 are excluded from the list considering flooding in most scenarios due to their relatively small power capacity and budget constraints.

4.5.8 Discussion

It can be noticed that the results vary considerably from scenario to scenario, as a function of the technical specifications of the substations affected and the outage duration, out-of-service loads, and operation in reserve mode. These factors affect the results calculated with (4.37) and (4.38) directly.

The substation details of Tables 4.15 and 4.16 indicate that some substation sites are more troublesome than others, as the generation and loads connected to them are heavier and they are located in vulnerable areas, prone to severe flood events. To this extent, severe flood episodes are more likely to justify the decision of hardening a substation, as a larger (4.28) results in larger values of (4.29) and (4.31).

It can be noticed that resilience planning reduces the expected energy not supplied in all flood scenarios listed in Table 4.18. Conversely, resilience planning reduces the accumulated cost in the worst-case flood scenarios, but increases the accumulated costs over a range of scenarios listed in Table 4.17. Notably, the accumulated cost ranges given in Table 4.17 show that resilience planning is cheaper than doing nothing in scenarios 13, 14, and 15 for any two flood episodes over time and in scenarios 11 and 12 in the latest occurrences.

The overlapping between some substations in all prioritization lists of Table 4.19 suggest that some of the highest costs are associated with the amount of energy not supplied for long periods of time, which depends on $p_{load,i,t}$ and (4.28). However, the costs associated with operation in reserve mode and repair or replacement of damaged components are higher than the value of loads lost in some cases, especially when the amount of generation connected to a substation is high and/or the value of loads lost is not very high. Consequently, the usage of (4.38) or (4.37) as a decision criterion leads to a different choice of substations to be hardened. For instance, this explains why substation 1110 is among the first substations to be prioritized with (4.37) and does not make to the prioritization list with (4.38). However, this difference can also be attributed to the unit costs and the substation areas used in the analysis.

In addition, note that the substations with the highest values of (4.37) or (4.38) are not necessarily the ones flooded in most scenarios. To this extent, substations installed at the 100-year flood height are flooded more often, but their technical specifications play an important role in the resilience metrics and cost indicators. Thus, comparing the "as is" and "resilience planning" scenarios using (4.37) and (4.38) for each substation site over a range of flood scenarios is recommended to identify the resilience enhancements that shall be prioritized.

Investing in hardening options is advantageous when the implementation cost $\mu_{RP,i}$ is lower than the costs associated with damage to equipment $\mu_{re,i,f,\tau}$ plus the value of loads lost $\mu_{out,i,f,\tau}$ in the f^{th} flood scenario, $i \in \mathcal{S}'_f$. However, the initial budget may not allow for hardening of all substation sites where $\mu_{RP,i} < \mu_{re,i,f,\tau} + \mu_{out,i,f,\tau}$ is true, as occurred in Section 4.5.7. Conversely, some substations where $\mu_{RP,i} > \mu_{re,i,f,\tau} + \mu_{out,i,f,\tau}$ holds might have to be hardened because other substations downstream present high costs associated

with equipment damaged and loads lost. This did not occur in Section 4.5.7 because all affected substations were in the lowest transmission voltage level.

It is noteworthy that mitigating the worst-case flood scenario using (4.37) or (4.38) may increase the accumulated costs over a range of scenarios. To this extent, the benefits of mitigating the worst-case flood scenario are more evident when (4.38) is used as a metric than when (4.37) is used as a metric. In fact, the results shown in Table 4.18 "as is" are higher than those obtained with resilience planning in all flood scenarios, whereas the results shown in Table 4.17 "as is" are lower than those obtained with resilience planning in all flood scenarios but scenarios 10 to 16. Overall, $f = 15$ is the worst-case scenario in terms of expected energy not supplied and accumulated costs "as is", but the hardening of selected substations may change this result, while increasing the impact of other flood scenarios on the grid.

Overall, the results indicate that investing on mitigation alternatives is advantageous not only to improve power system resilience to floods over a range of scenarios, but also to reduce costs and inconveniences associated with loads lost, operation in reserve mode, and damaged equipment. Furthermore, considering that future climate scenarios imply an increase in the frequency of severe flood episodes, the calculations of (4.37) and (4.38) will indicate that hardening is advantageous at more substations, but the budget limit still may prevent it from being fully implemented.

4.6 Final remarks

This chapter presented approaches to enhance power system resilience to HILP events, relying on information from past events and scenario generation. The methodologies performed modelling, impact assessment, and mitigation in different situations that show grid vulnerabilities of different characteristics. Each methodology presented is able to enhance power system resilience to a specific type of extreme event, for which it can be further extended. Furthermore, the usage of information from past events and generation of realistic scenarios improves accuracy of results and boosts confidence in the chosen mitigation strategies.

Table 4.16: Case study – Details of flooded transmission substations (part 2 of 2)

substation	floods	$p_{gen,i,1}$ (MW)	$p_{load,i,1}$ (MW)	μ_{RP} (M$)
1011	10/25	0.0	86.1	1.2
1016	7/25	0.0	68.5	1.2
1027	11/25	0.0	30.8	0.6
1040	2/25	0.0	42.4	0.6
1041	10/25	0.0	107.7	1.2
1051	15/25	0.0	32.6	0.6
1057	11/25	0.0	127.1	1.2
1068	3/25	0.0	16.0	0.6
1069	6/25	0.0	148.5	1.2
1073	13/25	0.0	94.1	1.2
1077	10/25	0.0	93.8	1.2
1079	11/25	0.0	101.2	1.2
1086	13/25	0.0	86.2	1.2
1089	13/25	0.0	63.2	1.2
1091	15/25	0.0	112.2	1.2
1093	12/25	151.2	0.0	1.2
1095	10/25	0.0	196.1	1.2
1096	18/25	137.8	29.2	1.2
1099	13/25	0.0	109.2	1.2
1108	8/25	0.0	14.3	0.6
1110	17/25	1557.5	0.0	3
1111	21/25	0.0	52.2	1.2
1112	21/25	0.0	94.8	1.2
1117	11/25	0.0	111.8	1.2
1121	25/25	0.0	52.0	1.2
1122	19/25	0.0	7.3	0.6
1126	9/25	0.0	105.7	1.2
1135	4/25	0.0	69.4	1.2
1137	8/25	0.0	42.9	0.6
1145	18/25	0.0	41.8	0.6
1147	24/25	288.7	46.9	3
1148	8/25	0.0	35.6	0.6

Table 4.17: Results – Accumulated costs "as is" and with resilience planning in the earliest-latest occurrences

f	$\lambda^T \mathbf{s}'$	$\lambda^T [\mathbf{s}' - \mathbf{s}''] + \lambda_{\mathbf{RP}}{}^T \mathbf{s}''$
1	3.14–3.31	27.74–27.91
2	3.42–3.64	27.44–27.61
3	4.44–4.74	26.98–27.13
4	12.55–13.41	31.65–32.10
5	13.96–14.91	31.65–32.10
6	35.02–37.41	40.62–41.70
7	35.02–37.41	40.62–41.70
8	33.64–35.91	39.24–40.20
9	37.43–39.82	39.79–40.79
10	39.09–41.60	41.00–42.08
11	41.80–44.50	42.42–43.59
12	41.80–44.50	42.42–43.59
13	42.86–45.59	42.20–43.35
14	45.83–48.79	43.13–44.34
15	46.23–49.22	43.53–44.76
16	41.44–44.13	40.64–41.70
17	29.07–30.98	34.76–35.45
18	28.99–30.89	34.68–35.36
19	24.25–25.78	33.67–34.29
20	21.19–22.49	32.66–33.20
21	17.66–18.74	32.21–32.73
22	16.59–17.60	31.69–32.17
23	16.59–17.60	31.69–32.17
24	11.35–12.02	30.25–30.65
25	10.80–11.42	28.29–28.55

Table 4.18: Results – Expected energy not supplied "as is" and with resilience planning in the earliest-latest occurrences

f	$\mathbf{EENS}^T\mathbf{s}'$	$\mathbf{EENS}^T\,[\mathbf{s}' - \mathbf{s}'']$
1	14.87–16.10	3.56–3.85
2	20.54–22.25	4.19–4.54
3	32.84–35.56	3.63–3.93
4	98.55-106.71	38.12–41.28
5	109.33–118.39	38.12–41.28
6	279.65–302.82	122.67–132.84
7	279.65–302.82	122.67–132.84
8	264.03–285.91	115.05–124.58
9	273.38–296.03	117.56–127.30
10	282.71–306.14	126.89–137.40
11	305.65–330.98	126.22–136.68
12	305.65–330.98	126.22–136.68
13	305.30–330.60	124.02–134.30
14	331.65–359.13	131.41–142.30
15	335.58–363.39	137.60–149.00
16	298.97–323.74	107.78–116.71
17	207.13–224.30	67.43–73.01
18	207.00–224.15	69.56–75.32
19	163.41–176.95	63.12–68.35
20	140.54–152.19	54.44–58.95
21	112.51–121.83	52.94–57.32
22	101.10–109.48	42.97–46.53
23	101.10–109.48	42.97–46.53
24	66.32–71.82	25.03–27.11
25	56.22–60.88	13.85–15.00

Table 4.19: Results – Choice of transmission substations to be hardened

Metric	Substations chosen (#)	$\lambda_{\mathbf{RP}}{}^T\mathbf{s}''$
(4.37)	250, 261, 923, 939, 981, 1003, 1057, 1073, 1086, 1091, 1095, 1110, 1111, 1112, 1145	\$24.6 M
(4.38)	250, 261, 923, 939, 970, 981, 1051, 1073, 1086, 1091, 1111, 1112, 1145, 1147	\$24.6 M
floods	250, 887, 972, 981, 986, 1003, 1051, 1073, 1086, 1091, 1110, 1111, 1112, 1121, 1122, 1145, 1147	\$24.6 M

Chapter 5

Conclusions

This thesis has presented different data-driven approaches for event detection, fault location, and resilience assessment in power systems.

The following sections describe the conclusions drawn for each sub-objective.

5.1 Event detection

The case studies presented in Section 2.5.1 and 2.5.2 have considered different signal-based and knowledge-based methods for PMU data event detection, further compared in terms of accuracy of correct event detections with different time windows and noisy data. The results indicate that the statistical knowledge-based event detection techniques present the best performance individually and are more robust to disturbances than signal-based event detection techniques. Therefore, the usage of knowledge-based techniques is recommended for accurate event detection. In particular, the results obtained by using the statistical detectability criterion are the most accurate with both T^2 and SPE statistics (complementary in some cases), which suggests that this is the best approach to select an adequate number of principal components. It can also be noticed that the window size applied to build the PCA model contributes to the task and shall be adjusted according to the events of interest.

In Section 2.3, the results show that the automated PCA-based strategy in use is able to detect voltage sags and swells without requiring any information about the network topology or electrical parameters of its components. However, the assumptions made to build the PCA model cannot be generalized to networks with distinct PMU settings and operating conditions. The resulting observation matrix may contribute to the task and shall be adjusted to different scenarios accordingly.

In turn, the results of Section 2.4.8 show that the methodology presented in Section 2.4 is capable of identifying all detectable events of interest and further isolating the variables involved in the occurrence. It may provide significant dimensionality reduction, which is

advantageous to process large amounts of data, while ensuring event detection in compliance to the operating constraints of the grid. The selection of principal components takes into consideration the statistical detectability of the PCA model with regards to T^2 and SPE statistics (complementary in some cases), which ensures effective detection of all events of interest. Therefore, it is claimed to be more appealing and straightforward than the state-of-the-art techniques to screen for events in power system data.

Finally, the case study featured in Section 2.6.3 is concerned about the detection, isolation, and diagnosis of abnormal energy production and consumption patterns at different time scales with use of OpenLV data. The results indicate that the method presented in Section 2.6 is accurate and efficient, as long as an adequate data set is used to build the PCA model at each individual layer. The proposed strategy is able to identify abnormal operating conditions in power system data based on a multilayer implementation of PCA over different time spans with use of T^2 and SPE statistics (complementary in some cases). This method provides significant dimensionality reduction, which is advantageous to process large amounts of data collected by many IEDs installed at different locations. Moreover, it enables to tackle distinct problems of interest in parallel over time, re-arranging the data to fit the scale and purpose of the analysis whenever required.

5.2 Fault location

The strategies presented in Section 3.3 are capable of locating and classifying faults, identifying the network configuration in the presence of switches, and distinguishing faults from variations in the standard operating conditions. The procedure used to identify the correct grid configuration prior to the location of the fault improves the accuracy of the method, as it reduces the number of candidate scenarios and limits the search to the right part of the network in all testing scenarios. As a result, the errors lie in the same part of the network in all tested scenarios. Nonetheless, it may be inaccurate over a range of scenarios due to the differences between training and testing scenarios, equivalent impedance seen at the substation nodes, and poor quality of data.

The results shown in Section 3.4.4 evince that the fault behaviour analysis provides a comprehensive, but accurate estimation of the points of fault and short-circuit currents, considering a range of values for loads, distributed generation, and fault distance and impedance parameters. The results obtained in a range of fault scenarios are comprehensive enough to provide a good overview of the possible fault conditions of the grid. However, implementation may be computationally too expensive in the presence of a high level of uncertainty from multiple sources and in large-scale power systems with reconfiguration.

In turn, the data-driven Bayesian algorithm presented in Section 3.5.5 is capable of identifying the most vulnerable areas accurately for different faults, regardless of the errors in the load estimations, presence of distributed generation, fault impedance, walking distance, conditional probabilities applied to the simulations, and grid configuration. The results shown in Section 3.5 are not affected significantly by these conditions, as the min-

imum calculated risk lies within the expected class of fault in most cases. Although the method is easily implementable, it relies on probabilistic assumptions and depends on the quality of such assumptions to provide accurate results. Nonetheless, the overall results are acceptable, as the fault location problem is a typical multiple-solution problem.

5.3 Resilience assessment

Overall, the results indicate that enhancements in power system resilience to different types of external high-impact, low-probability events are advantageous to reduce interruption times and costs associated with outage and damaged components.

The case study presented in Section 4.3.3 evaluated the costs and benefits concerning power system resilience enhancements in low-income neighbourhoods, considering distinct events of interest that may lead to power outages. In particular, the feasibility of deploying IEDs to monitor electric variables and detect energy thefts and the installation of off-grid photovoltaic panels combined with batteries at household or community level for autonomous operation from the external grid against man-made attacks are addressed. The decision criteria compares the cost associated with implementation of mitigation alternatives with the value of load lost and the cost to repair damaged components to determine the benefits of different strategies targeted at low-income neighbourhoods. The results of different case studies indicate that the maintenance costs plus the value of load lost may compensate for the investment costs associated with resilience improvements in a long-term horizon, as the return of investment may be relatively long in some cases.

The case study presented in Section 4.4.1 assessed the costs and risks associated with underground and overhead power lines for a resilient, cost-effective power system planning and operation under wind storms. The method evaluated the cost associated with installation, operation, and repair of power lines, as well as the penalties for the energy not supplied (value of lost load), to determine which power line setting is the most appropriate in terms of resilience and costs for different case studies. The results show that there is a trade-off between resilience and costs, as the installation costs are the highest for underground lines, whereas the value of lost load is the highest for overhead lines without hardening in most cases.

On the top of this, the results of Section 4.5.6 evince that floods may be catastrophic to power systems in terms of damage to infrastructure and power outage. To assess the impact of floods on the grid and further define appropriate mitigation strategies, this study used information about hydrological models and the location of electrical equipment over a range of realistic flood scenarios in the coastal area of Texas. Overall, the results indicate that hardening substations at strategic locations may reduce the impacts of severe flood episodes and enhance long-term power resilience. Ultimately, the choice of the most appropriate benchmarking strategy (that is, mitigating or doing nothing) depends on the potential improvements to power system resilience and costs associated with the expected energy not supplied, repair or replacement of damaged grid components, and investment budget for resilience options.

5.4 Further perspectives

Further developments and improvements in event detection, fault location, and resilience assessment towards enhancements in power systems may comprise:

- Improvements in power system modelling: new software tools and customization for realistic scenarios
- Improvements in power system monitoring: implementation of decentralized or distributed control architectures, combined use of distinct metering devices
- New uses of data: usage of distinct electrical quantities, incorporation of historical data and predictions in decision-making processes, integration of multidisciplinary aspects influencing power system operation
- New algorithms: improvements in the state-of-the-art methods, adjustments to specific situations, alternative computing methods
- New problems: created by energy systems in transition, intensified by climate change, and to be known in the future

Bibliography

[1] Energy Equity: Bringing Solar Power to Low-Income Communities. https://e360.yale.edu/features/energy-equity-bringing-solar-power-to-low-income-communities. Accessed: 2020-05-11.

[2] OpenLV | the groundbreaking project that's making local electricity data openly available. http://openlv.net. Accessed: 2020-11-05.

[3] RESOLVD project | renewable penetration levered by efficient low voltage distribution grids. http://resolvd.eu. Accessed: 2020-11-05.

[4] Tarifa Social de Energia Elétrica. https://www.eneldistribuicaosp.com.br/para-sua-casa/tarifa-social-de-energia-eletrica. Accessed: 2020-05-08.

[5] IEEE Standard Electrical Power System Device Function Numbers, Acronyms, and Contact Designations. *IEEE Std C37.2-2008 (Revision of IEEE Std C37.2-1996)*, pages 1–48, 2008.

[6] IEEE standard terms for reporting and analyzing outage occurrences and outage states of electrical transmission facilities. *IEEE Std 859-2018 (Revision of IEEE Std 859-1987)*, pages 1–24, April 2019.

[7] *Transmission Cost Estimation Guide MTEP19*. MISO Energy, 2019.

[8] M. Abad, M. García-Gracia, N. El Halabi, and D. López Andía. Network impulse response based-on fault location method for fault location in power distribution systems. *IET Generation, Transmission Distribution*, 10(15):3962–3970, 2016.

[9] M. A. Aftab, S.M. S. Hussain, I. Ali, and T. S. Ustun. Dynamic protection of power systems with high penetration of renewables: A review of the traveling wave based fault location techniques. *International Journal of Electrical Power & Energy Systems*, 114:105410, 2020.

[10] A. Allen, M. Singh, E. Muljadi, and S. Santoso. PMU data event detection: A user guide for power engineers, 2014. Link to accompanying zipped data file: http://www.nrel.gov/docs/fy15osti/61664-1.zip.

[11] S. F. Alwash, V. K. Ramachandaramurthy, and N. Mithulananthan. Fault-location scheme for power distribution system with distributed generation. *IEEE Transactions on Power Delivery*, 30(3):1187–1195, 2015.

[12] K. Anderson, X. Li, S. Dalvi, S. Ericson, C. Barrows, C. Murphy, and E. Hotchkiss. Integrating the value of electricity resilience in energy planning and operations decisions. *IEEE Systems Journal*, pages 1–11, 2020.

[13] Y. Andersson-Sköld and L. Nyberg. Effective and sustainable flood and landslide risk reduction measures: An investigation of two assessment frameworks. *International Journal of Disaster Risk Science*, 7(4):374–392, Dec 2016.

[14] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao. Load modeling—a review. *IEEE Transactions on Smart Grid*, 9(6):5986–5999, 2018.

[15] N. Ayech, C. Chakour, and M. F. Harkat. New adaptive moving window PCA for process monitoring. *IFAC Proceedings Volumes*, 45(20):606 – 611, 2012. 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes.

[16] A. Bagheri, C. Zhao, F. Qiu, and J. Wang. Resilient transmission hardening planning in a high renewable penetration era. *IEEE Transactions on Power Systems*, 34(2):873–882, March 2019.

[17] P. P. Barker and R. W. De Mello. Determining the impact of distributed generation on power systems—part i: Radial distribution systems. In *2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134)*, volume 3, pages 1645–1656, July 2000.

[18] E. Barocio, B. C. Pal, D. Fabozzi, and N. F. Thornhill. Detection and visualization of power system disturbances using principal component analysis. In *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid*, pages 1–10, Aug 2013.

[19] S. M. Blair, C. D. Booth, G. Williamson, A. Poralis, and V. Turnham. Automatically detecting and correcting errors in power quality monitoring data. *IEEE Transactions on Power Delivery*, 32(2):1005–1013, 2017.

[20] J. M. Boggess, G. W. Becker, and M. K. Mitchell. Storm flood hardening of electrical substations. In *2014 IEEE PES T &D Conference and Exposition*, pages 1–5, 2014.

[21] M.H.J Bollen and I.Y.H. Gu. *Signal Processing of Power Quality Disturbances*. IEEE Press Series on Power Engineering. John Wiley & Sons, 2006.

[22] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao. Characterizing and quantifying noise in PMU data. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5, July 2016.

[23] C. S. Burrus. *Fast Fourier Transforms*. OpenStax CNX, 2012.

[24] L. Cai, N. F. Thornhill, S. Kuenzel, and B. C. Pal. Wide-area monitoring of power systems using principal component analysis and k-nearest neighbor analysis. *IEEE Transactions on Power Systems*, 33(5):4913–4923, 2018.

[25] A. Cataliotti, V. Cosentino, D. Di Cara, P. Russotto, E. Telaretti, and G. Tinè. An innovative measurement approach for load flow analysis in mv smart grids. *IEEE Transactions on Smart Grid*, 7(2):889–896, 2016.

[26] Council of European Energy Regulators. CEER Benchmarking Report 6.1 on the Continuity of Electricity and Gas Supply, July 2018.

[27] T. Dimitrovska, U. Rudež, and R. Mihalič. Real-time application of an indirect power-system contingency screening method based on adaptive PCA. *IEEE Transactions on Power Systems*, 34(6):4665–4673, 2019.

[28] J. Ding, L. Li, Y. Zheng, C. Zhao, H. Chen, and X. Wang. Distributed travelling-wave-based fault location without time synchronisation and wave velocity error. *IET Generation, Transmission & Distribution*, 11(8):2085–2093, 2017.

[29] R. Dobbe, W. van Westering, S. Liu, D. Arnold, D. Callaway, and C. Tomlin. Linear single- and three-phase voltage forecasting and Bayesian state estimation with limited sensing. *IEEE Transactions on Power Systems*, 35(3):1674–1683, 2020.

[30] R. Dunia and S. Joe Qin. A unified geometric approach to process and sensor fault identification and reconstruction: the unidimensional fault case. *Computers & Chemical Engineering*, 22(7):927 – 943, 1998.

[31] Electric Power Research Institute. OpenDSS. Available: `https://www.epri.com/#/pages/sa/opendss?lang=en`.

[32] M. Eliassi, H. Seifi, and M.R. Haghifam. Incorporation of protection system failures into bulk power system reliability assessment by Bayesian networks. *IET Generation, Transmission & Distribution*, 9:1226–1234(8), August 2015.

[33] K. Emanuel. Assessing the present and future probability of Hurricane Harvey's rainfall. *Proceedings of the National Academy of Sciences*, 114(48):12681–12684, 2017.

[34] M. Esfahani, N. Amjady, B. Bagheri, and N. D. Hatziargyriou. Robust resiliency-oriented operation of active distribution networks considering windstorms. *IEEE Transactions on Power Systems*, 35(5):3481–3493, 2020.

[35] S. Espinoza, A. Poulos, H. Rudnick, J. C. de la Llera, M. Panteli, and P. Mancarella. Risk and resilience assessment with component criticality ranking of electric power systems subject to earthquakes. *IEEE Systems Journal*, 14(2):2837–2848, 2020.

[36] G. Feng and A. Abur. Fault location using wide-area measurements and sparse estimation. *IEEE Transactions on Power Systems*, 31(4):2938–2945, July 2016.

[37] C.B. Field, V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2012.

[38] Z. Gao, C. Cecati, and S. X. Ding. A survey of fault diagnosis and fault-tolerant techniques—part i: Fault diagnosis with model-based and signal-based approaches. *IEEE Transactions on Industrial Electronics*, 62(6):3757–3767, June 2015.

[39] Z. Gao, C. Cecati, and S. X. Ding. A survey of fault diagnosis and fault-tolerant techniques—part ii: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Transactions on Industrial Electronics*, 62(6):3768–3774, June 2015.

[40] Y. Ge, A. J. Flueck, D. Kim, J. Ahn, J. Lee, and D. Kwon. Power system real-time event detection and associated data archival reduction based on synchrophasors. *IEEE Transactions on Smart Grid*, 6(4):2088–2097, July 2015.

[41] D.J. Gochis, M. Barlage, R. Cabell, M. Casali, A. Dugger, K. Fitzgerald, M. McAllister, J. McCreight, A. Rafieei-Nasab, L. Read, K. Sampson, D. Yates, and Y. Zhang. The NCAR WRF-Hydro®modeling system technical description, (version 5.1.1), 2020.

[42] W. M. Grady and D. Costello. Implementation and application of an independent Texas synchrophasor network. In *2010 63rd Annual Conference for Protective Relay Engineers*, pages 1–12, March 2010.

[43] Y. Guo, K. Li, D. M. Laverty, and Y. Xue. Synchrophasor-based islanding detection for distributed generation systems using systematic principal component analysis approaches. *IEEE Transactions on Power Delivery*, 30(6):2544–2552, Dec 2015.

[44] Homeland Infrastructure Foundation-Level Data (HIFLD). HIFLD Open GP - Energy. Accessed on 2020-04-24.

[45] G. Huang, J. Wang, C. Chen, J. Qi, and C. Guo. Integration of preventive and emergency responses for power grid resilience enhancement. *IEEE Transactions on Power Systems*, 32(6):4451–4463, Nov 2017.

[46] M. Hummon, P. Denholm, J. Jorgenson, D. Palchak, B. Kirby, and O. Ma. Fundamental drivers of the cost and price of operating reserves, 2013.

[47] T. Iesmantas and R. Alzbutas. Bayesian spatial reliability model for power transmission network lines. *Electric Power Systems Research*, 173:214 – 219, 2019.

[48] A. R. Berkeley III and M. Wallace. A framework for establishing critical infrastructure resilience goals, 2010.

[49] A. R. Jadhav, M. P. R. Sai Kiran, and R. Pachamuthu. Development of a novel IoT enabled power monitoring architecture with real-time data visualization for use in domestic as well as industrial scenarios. *IEEE Transactions on Instrumentation and Measurement*, 2020.

[50] Y. Jiang. Data-driven fault location of electric power distribution systems with distributed generation. *IEEE Transactions on Smart Grid*, 11(1):129–137, 2020.

[51] S. R. Khuntia, B. W. Tuinema, J. L. Rueda, and M. A. M. M. van der Meijden. Time-horizons in the planning and operation of transmission networks: an overview. *IET Generation, Transmission & Distribution*, 10(4):841–848, 2016.

[52] T. Lagos, R. Moreno, A. N. Espinosa, M. Panteli, R. Sacaan, F. Ordonez, H. Rudnick, and P. Mancarella. Identifying optimal portfolios of resilient network investments against natural hazards, with applications to earthquakes. *IEEE Transactions on Power Systems*, 35(2):1411–1421, March 2020.

[53] X. Liang, S. A. Wallace, and D. Nguyen. Rule-based data-driven analytics for wide-area fault detection using synchrophasor data. *IEEE Transactions on Industry Applications*, 53(3):1789–1798, 2017.

[54] X. Liu, D. M. Laverty, R. J. Best, K. Li, D. J. Morrow, and S. McLoone. Principal component analysis of wide-area phasor measurements for islanding detection—a geometric view. *IEEE Transactions on Power Delivery*, 30(2):976–985, April 2015.

[55] Y. Y. Liu, D. R. Maidment, D. G. Tarboton, X. Zheng, and S. Wang. A CyberGIS integration and computation framework for high-resolution continental-scale flood inundation mapping. *JAWRA Journal of the American Water Resources Association*, 54(4):770 – 784, 2018.

[56] London Economics International LLC. Estimating the value of lost load, July 2013.

[57] M. Mahzarnia, M. P. Moghaddam, P. T. Baboli, and P. Siano. A review of the measures to enhance power systems resilience. *IEEE Systems Journal*, pages 1–12, 2020.

[58] M. Majidi, A. Arabali, and M. Etezadi-Amoli. Fault location in distribution networks by compressive sensing. *IEEE Transactions on Power Delivery*, 30(4):1761–1769, Aug 2015.

[59] G. Manassero, S. G. Di Santo, and L. Souto. Heuristic method for fault location in distribution feeders with the presence of distributed generation. *IEEE Transactions on Smart Grid*, 8(6):2849–2858, 2017.

[60] S. E. McHann. Grid analytics: How much data do you really need? In *2013 IEEE Rural Electric Power Conference (REPC)*, pages C3–1–C3–4, April 2013.

[61] A. Meghwani, J. G. Sreenath, S. C. Srivastava, S. Chakrabarti, and A. Sharma. Robust event detection using synchrophasor assisted forecasting-aided state estimation. In *2019 IEEE Power Energy Society General Meeting (PESGM)*, pages 1–5, 2019.

[62] B. Mnassri, E. M. E. Adel, and M. Ouladsine. Unified sufficient conditions for PCA-based fault detectability and isolability. *IFAC Proceedings Volumes*, 45(20):421 – 426, 2012. 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes.

[63] R. Moreno and G. Strbac. Integrating high impact low probability events in smart distribution network security standards through cvar optimisation. In *IET International Conference on Resilience of Transmission and Distribution Networks (RTDN) 2015*, pages 1–6, Sep. 2015.

[64] UK Cabinet Office. Keeping the country running: Natural hazards and infrastructure, Oct 2011.

[65] Organization for Security and Co-operation in Europe. Protecting Electricity Networks from Natural Hazards, 2016.

[66] M. Panteli and P. Mancarella. The grid: Stronger, bigger, smarter?: Presenting a conceptual framework of power system resilience. *IEEE Power and Energy Magazine*, 13(3):58–66, May 2015.

[67] M. Panteli and P. Mancarella. Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events. *IEEE Systems Journal*, 11(3):1733–1742, Sep. 2017.

[68] M. Panteli, C. Pickering, S. Wilkinson, R. Dawson, and P. Mancarella. Power system resilience to extreme weather: Fragility modeling, probabilistic impact assessment, and adaptation measures. *IEEE Transactions on Power Systems*, 32(5):3747–3757, Sep. 2017.

[69] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou. Boosting the power grid resilience to extreme weather events using defensive islanding. *IEEE Transactions on Smart Grid*, 7(6):2913–2922, Nov 2016.

[70] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou. Power systems resilience assessment: Hardening and smart operational enhancement strategies. *Proceedings of the IEEE*, 105(7):1202–1213, July 2017.

[71] D. Peña. *Análisis de datos multivariantes*. McGraw-Hill, 2002.

[72] R. A. F. Pereira, L. G. W. da Silva, M. Kezunovic, and J. R. S. Mantovani. Improved fault location on distribution feeders based on matching during-fault voltage sags. *IEEE Transactions on Power Delivery*, 24(2):852–862, 2009.

[73] S. Poudel, A. Dubey, and A. Bose. Probabilistic quantification of power distribution system operational resilience. In *2019 IEEE Power Energy Society General Meeting (PESGM)*, pages 1–5, Aug 2019.

[74] S. J. Qin and R. Dunia. Determining the number of principal components for best reconstruction. *Journal of Process Control*, 10(2):245 – 250, 2000.

[75] M. Rafferty, X. Liu, D. M. Laverty, and S. McLoone. Real-time multiple event detection and classification using moving window PCA. *IEEE Transactions on Smart Grid*, 7(5):2537–2548, Sep. 2016.

[76] E. A. Reche, J. V. d. Sousa, D. V. Coury, and R. A. S. Fernandes. Data mining-based method to reduce multiple estimation for fault location in radial distribution systems. *IEEE Transactions on Smart Grid*, 10(4):3612–3619, July 2019.

[77] M. Romero, O. J. Murillo, L. Luna, L. Gallego, E. Parra, and H. Torres. Determination of sag disturbing and sag vulnerable zones in a distribution network using stochastic fault simulation. In *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–6, July 2008.

[78] F. Roos and S. Lindahl. Distribution system component failure rates and repair times - an overview. In *Nordic Distribution and Asset Management Conference*, 2004.

[79] I. Rozenberg, Y. Beck, Y. C. Eldar, and Y. Levron. Sparse estimation of faults by compressed sensing with structural constraints. *IEEE Transactions on Power Systems*, 33(6):5935–5944, Nov 2018.

[80] S. Nowicki, N. Gladstone, J. Katuva, Heloise Greeff, A. Manandhar, G. Wekesa and G. Mwania. *The Water Module - Student Resource.* School of Geography and the Environment, University of Oxford, 2018.

[81] M.M. Saha, J.J. Izykowski, and E. Rosolowski. *Fault Location on Power Networks.* Power Systems. Springer London, 2009.

[82] P. W. Sauer and M. A. Pai. *Power System Dynamics and Stability.* Prentice-Hall, 1998.

[83] M. Scholz, M. Fraunholz, and J. Selbig. Nonlinear principal component analysis: Neural network models and applications. In Alexander N. Gorban, Balázs Kégl, Donald C. Wunsch, and Andrei Y. Zinovyev, editors, *Principal Manifolds for Data Visualization and Dimension Reduction*, pages 44–67, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[84] N. Scott, K. McKemey, and S. Batchelor. Energy in low-income urban communities - final technical report, February 2005.

[85] Y. Seyedi, H. Karimi, and S. Grijalva. Irregularity detection in output power of distributed energy resources using PMU data analytics in smart grids. *IEEE Transactions on Industrial Informatics*, 15(4):2222–2232, 2019.

[86] C. Shao, M. Shahidehpour, X. Wang, X. Wang, and B. Wang. Integrated planning of electricity and natural gas transportation systems for enhancing the power grid resilience. *IEEE Transactions on Power Systems*, 32(6):4418–4429, Nov 2017.

[87] J. Shi, N. Yu, E. Keogh, H. K. Chen, and K. Yamashita. Discovering and labeling power system events in synchrophasor data with matrix profile. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1827–1832, 2019.

[88] S. Soltan, P. Mittal, and H. V. Poor. Line failure detection after a cyber-physical attack on the grid using Bayesian regression. *IEEE Transactions on Power Systems*, 34(5):3758–3768, 2019.

[89] L. C. Souto. Fault location in power distribution networks with distributed generation. Master's thesis, Escola Politécnica da Universidade de São Paulo, 2016. doi:10.11606/D.3.2016.tde-28062016-151847.

[90] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer New York, 2014.

[91] D. Topolanek, M. Lehtonen, M. R. Adzman, and P. Toman. Earth fault location based on evaluation of voltage sag at secondary side of medium voltage/low voltage transformers. *IET Generation, Transmission & Distribution*, 9(14):2069–2077, 2015.

[92] D. N. Trakas, N. D. Hatziargyriou, M. Panteli, and P. Mancarella. A severity risk index for high impact low probability events in transmission systems due to extreme weather. In *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6, 2016.

[93] U.S. Department of Energy. Hardening and Resilience: U.S. Energy Industry Response to Recent Hurricane Seasons, August 2013.

[94] B. Uzunoğlu. Locating distribution power system fault employing Bayes theorem with subjective logic. In *2017 2nd International Conference on System Reliability and Safety (ICSRS)*, pages 130–134, Dec 2017.

[95] B. Vaagensmith, T. McJunkin, K. Vedros, J. Reeves, J. Wayment, L. Boire, C. Rieger, and J. Case. An integrated approach to improving power grid reliability: Merging of probabilistic risk assessment with resilience metrics. In *2018 Resilience Week (RWS)*, pages 139–146, Aug 2018.

[96] Y. Wang, C. Chen, J. Wang, and R. Baldick. Research on resilience of power systems under natural disasters—a review. *IEEE Transactions on Power Systems*, 31(2):1604–1613, March 2016.

[97] W.M. Warwick, T.D. Hardy, M.G. Hoffman, and J.S. Homer. Electricity distribution system baseline report, July 2016.

[98] E. B. Watson and A. H. Etemadi. Modeling electrical grid resilience under hurricane wind conditions with increased solar and wind power generation. *IEEE Transactions on Power Systems*, 35(2):929–937, March 2020.

[99] Y. Xiang and J. F. G. Cobben. A Bayesian approach for fault location in medium voltage grids with underground cables. *IEEE Power and Energy Technology Systems Journal*, 2(4):116–124, Dec 2015.

[100] L. Xie, Y. Chen, and P. R. Kumar. Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis. *IEEE Transactions on Power Systems*, 29(6):2784–2794, Nov 2014.

[101] T. Xu, A. B. Birchfield, and T. J. Overbye. Modeling, tuning, and validating system dynamics in synthetic electric grids. *IEEE Transactions on Power Systems*, 33(6):6501–6509, 2018.

[102] M. Yan, Y. He, M. Shahidehpour, X. Ai, Z. Li, and J. Wen. Coordinated regional-district operation of integrated energy systems for resilience enhancement in natural disasters. *IEEE Transactions on Smart Grid*, 10(5):4881–4892, Sep. 2019.

[103] T. Yeh and L. Wang. A study on generator capacity for wind turbines under various tower heights and rated wind speeds using weibull distribution. *IEEE Transactions on Energy Conversion*, 23(2):592–602, June 2008.

[104] W. Yuan, J. Wang, F. Qiu, C. Chen, C. Kang, and B. Zeng. Robust optimization-based resilient distribution network planning against natural disasters. *IEEE Transactions on Smart Grid*, 7(6):2817–2826, Nov 2016.

[105] H. Yuan-Yih and C. Chung-Liang. Probabilistic transient stability studies using the conditional probability approach. *IEEE Transactions on Power Systems*, 3(4):1565–1572, Nov 1988.

[106] H. Zhang and P. Duhamel. On the methods for solving Yule-Walker equations. *IEEE Transactions on Signal Processing*, 40(12):2987–3000, Dec 1992.

[107] L. Zhang, N. Tai, W. Huang, and Y. Wang. Fault distance estimation-based protection scheme for dc microgrids. *The Journal of Engineering*, 2019(16):1199–1203, 2019.

[108] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor. Identification of outages in power systems with uncertain states and optimal sensor locations. *IEEE Journal of Selected Topics in Signal Processing*, 8(6):1140–1153, Dec 2014.

[109] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee. Ensemble-based algorithm for synchrophasor data anomaly detection. *IEEE Transactions on Smart Grid*, 10(3):2979–2988, 2019.

[110] J. Zhu, Z. Ge, and Z. Song. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data. *IEEE Transactions on Industrial Informatics*, 13(4):1877–1885, 2017.