Università degli Studi di Ferrara

DOCTORAL COURSE IN ENGINEERING SCIENCES

CYCLE 33

COORDINATOR Prof. Stefano Trillo

PLANNING, FAULT DIAGNOSIS, AND DATA COMPRESSION USING SPARSE SIGNALS TO DESIGN ELECTRICAL SYSTEMS

Scientific/Disciplinary Sector (SDS) ING-INF/04

Candidate	**Supervisor**	**Supervisor**
Dott. Milton Ruiz	Prof. Simani Silvio	Prof. Inga Esteban
_____	_____	_____

Year 2017/2021

Abstract

The design of optimal, scalable and reliable electrical networks is one of the main challenges for researchers due to their large size and complexity. This dissertation proposes an heuristic algorithm for the design of electrical networks using georeferenced coordinates. The design of power transmission networks, due to the fact that it supplies electricity to cities, countries, and even interconnects continents, does not present established georeferenced routes that can be used as restrictions. For example, when planning under the heuristics proposed in this research, there is the possibility that a section of the power transmission network passes through a lagoon, a river or certain places that are difficult to access, hindering the actual deployment of the network. In order to test the proposed heuristics, the design of electric power distribution networks is presented using georeferenced coordinates of parks, houses, and buildings as constraints. The buildings are considered as subscribers of the electric service and have been divided into different strata of electric consumption with random loads. The results show that all buildings are supplied by a low voltage feeder, i.e., 100% coverage. It is shown that the medium and low voltage feeders respect the route of the roads and sidewalks allowing the implementation of subway electrical networks. Finally, the size of the transformers is automatically selected from a list of commercial capacities considering the design standards.

The safety and reliability of smart grids require the implementation of large numbers of meters, actuators, and control systems. One of the main problems is the loss of information in the electrical signals that can be caused by sensors or problems in the transmission of information. Therefore, this dissertation proposes the creation of a unique dictionary based on optimization techniques, finding the atoms that allow the recovery of lost samples in electrical systems in steady state and dynamic state. Signal reconstruction from sparse measurements has been performed by applying three optimization philosophies Basis Pursuit which allow reconstructing signals with 10% of missing samples in times of about 10s. Matching Pursuit allows reconstructing signals from 25% of missing samples in times of about 0.2s. Finally, Orthogonal Matching Pursuit allows reconstructing signals with 30% of missing samples in times of about 1.5s.

This dissertation presents an algorithm for fault detection and classification in transmission lines using the Haar-type wavelet mother transform. Voltage and current signals contain all the information of the power system, therefore when a fault occurs in the electrical power system, these signals present disturbances in their amplitude, phase changes, and presence of harmonics. Mathematically,

all mother wavelets respond with an impulse when there is an abrupt change in the signals, which allows to detect when a fault has occurred(in time domain). Experimental results have shown that with frequencies above 100 kHz it is possible to detect a fault with 100% accuracy by taking only 4 samples of the signal and applying the Haar wavelet. The fault detection times vary between 0.31ms and 1.15ms and the fault classification times vary between 0.94s and 1.31s.

Finally, the massive amount of data to be transported and stored generates a new problem in information compression. In this dissertation a high-performance algorithm is proposed to compress the data of electrical signals. First biorthogonal wavelet level six transform is applied, however after compression, the reconstructed signal will have a different amplitude and it will be shifted when compared to the original one. Then, normalization is used (for amplitude correction between the original signal and reconstructed one) by multiplying the reconstructed signal by the result of the division between the original signal maximum magnitude and the reconstructed signal maximum magnitude. Thirdly, the ripple in the reconstructed signal is eliminated by applying a moving average filter. Finally, the shifting is corrected by finding the difference between the maximum points in a cycle of the original signal and the reconstructed one. After the compression algorithm was performed the best rates are 99.803% for compression rate, RTE 99.9479%, NMSE 0.000434, and Cross-Correlation 0.999925.

Abstract (Italian)

La progettazione di reti elettriche ottimali, scalabili e affidabili è una delle principali sfide per i ricercatori a causa delle loro grandi dimensioni e complessità. Questa tesi propone un algoritmo euristico per la progettazione di reti elettriche utilizzando coordinate georeferenziate. La progettazione delle reti di trasmissione di potere fornisce elettricità a città, paesi e anche interconnette continenti, per questa ragione, non presenta percorsi georeferenziati stabiliti che possano essere utilizzati come restrizioni. Ad esempio, quando si pianifica secondo le euristiche proposte in questa ricerca, esiste la possibilità che una sezione della rete di trasmissione di energia attraversi una laguna, un fiume o certi luoghi di accesso difficile, ostacolando lo sviluppo reale della rete. Allo scopo di testare l'euristica proposta, si presenta il progetto delle reti di distribuzione dell'energia elettrica utilizzando come vincoli le coordinate georeferenziate di parchi, case ed edifici in coercizione. Gli edifici sono considerati abbonati al servizio elettrico e sono stati suddivisi in diversi strati di consumo elettrico con carichi casuali. I risultati mostrano che tutti gli edifici sono forniti da un alimentatore a bassa tensione, ovvero una copertura del 100%. Si evidenzia che le linee di media e bassa tensione rispettano il tracciato delle strade e dei marciapiedi implementando la realizzazione di reti elettriche metropolitane. Infine, la taglia dei trasformatori si seleziona automaticamente da un elenco di capacità commerciali considerando gli standard di progetazzione.

La sicurezza e l'affidabilità delle reti intelligenti richiede l'implementazione di un gran numero di contatori, attuatori e sistemi di controllo. Uno dei problemi principali è la perdita di informazioni nei segnali elettrici che può essere causata da sensori o problemi nella trasmissione delle informazioni. Pertanto, questa tesi propone la creazione di un dizionario unico basato su tecniche di ottimizzazione, trovando gli atomi che consentono il recupero di mostre perse nei sistemi elettrici in stato stazionario e dinamico. La ricostruzione del segnale da misure sparse è stata sviluppata applicando tre filosofie di ottimizzazione Basis Pursuit che consentono di ricostruire segnali con il 10% dei campioni mancanti in tempi di circa 10s. Matching Pursuit permette di ricostruire i segnali dal 25% dei campioni mancanti in tempi di circa 0,2s. In sintesi, Orthogonal Matching Pursuit permette di ricostruire segnali con il 30% dei campioni mancanti in tempi di circa 1,5s.

Questa tesi presenta un algoritmo per il rilevamento dei fallimenti e la classificazione nelle linee di trasmissione utilizzando la Haar-type wavelet mother transform. I segnali di tensione e corrente contengono tutta l'informazione del sistema di alimentazione, quindi quando si verifica un fallimento nel sistema di alimentazione elettrica, questi segnali presentano disturbi nella loro ampiezza, sfasamento e pre-

senza di armoniche. Matematicamente, tutte le wavelets madri rispondono quando esiste un cambiamento improvviso nei segnali che permette riconoscere quando un fallimento succede (nel dominio del tempo).

Risultati esperimentali hanno dimostrato che con frequenze superiori a 100 kHz è possibile rilevare un fallimento con precisione del 100%, prendendo soltanto 4 campioni della segnale ed applicando la Haar wavelet. Il rilevamento di fallimenti varia fra 0.31ms e 1.15ms e i tempi dellla classificazione di fallimenti varia tra 0.94 e 1.31s.

Alla fine, l'enorme quantità di dati da trasportare ed archiviare genera un nuovo problema nella compressione delle informazioni. In questa tesi si propone un algoritmo ad alte prestazioni per comprimere i dati dei segnali elettrici. La prima ondata biortogonale di livello sei viene applicata, tuttavia dopo la compressione, il segnale ricostruito avrà un'ampiezza diversa e sarà spostato rispetto a quello originale. Quindi, si utilizza la normalizzazione (per la correzione dell'ampiezza tra il segnale originale e quello ricostruito) moltiplicando il segnale ricostruito per il risultato della divisione tra l'ampiezza massima del segnale originale e l'ampiezza massima del segnale ricostruito. In terzo luogo, l'ondulazione nel segnale ricostruito si elimina applicando un filtro a movimento medio. Infine, lo spostamento viene corretto trovando la differenza tra i punti massimi in un ciclo del segnale originale e quello ricostruito. Dopo l'esecuzione dell'algoritmo di compressione, i tassi migliori sono 99,803% per il tasso di compressione, RTE 99,9479% NMSE 0,000434 e Cross-Correlation 0,999925.

Acknowledgments

Firstly, I am grateful to my family, thank you to my wife, Andrea, for supporting me and our kids Martin, Mía, and Olivia.

Special thanks to my supervisors, Silvio Simani and Esteban Inga, for the constant assistance and the valuable suggestions that encouraged my personal and academic growth.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Throughout history and especially during the last few decades, the phenomenon of migration from rural to urban areas has increased. United Nations reports indicate that, currently, more than half of the population lives in urban areas, and they estimate that, during the next 50 years, seven out of every ten people will live in the metropolises. This phenomenon presents new challenges regarding governance, citizenship, mobility, internet of things, energy, environment, energy efficiency, and sustainability. This scenario requires an active participation of citizens assisted by digital technologies to shape smart cities that provide services for improving the welfare of all its members without compromising the resources of future generations. The evolution from traditional cities to smart cities has generated new challenges, which focusing on planning scalable electrical networks with high standards in safety, efficiency, autonomy, resilience, and reliability, providing consumers with continuous and high-quality electrical energy. The traditional electrical system is the most complex network in the world, it is an interconnected network of generation units, substations, transmission lines, distribution lines, loads, and users. Traditional electrical systems have evolved over the last few decades into intelligent electrical networks, or what are called the smart grids which are composed of distributed energy resources and energy storages. A smart grid offers the following advantages: allows to link a large number of nonlinear loads, gives support to the growing fleet of electric vehicles, makes possible the advances in information technology, and communications, increasing the complexity of information processing in real time in order to characterize, identify, protect, and diagnose the precise behavior of the system. At the same time, the control and monitoring of power quality requires a large number of sensors, smart meters, and phase measurement units, which are installed at different points of the smart grid. Therefore, the generation, acquisition, transport, and processing of data is possible through the creation of new systems such as: Measurement Data Management System (MDMS), Customer Information System (CIS) and billing, Geographic Information System (GIS), Supervision Control, and Data Acquisition (SCADA), Energy Management System (EMS), Distribution Management System (DMS), Outage Management System (OMS), assets management system, mobile workforce management system, demand response management system, dis-

tributed energy resource management system, business intelligence, data analytics, and telecommunications systems [Daim et al., 2018]. To this end, in order to mitigate the problems of power supply, it is necessary to install sensors in strategic points of the electrical system in order to mitigate the problems related to the electrical supply. For example, in smart homes, all devices have the ability to send and receive information, enabling real-time demand management. This situation causes that the large number of sensors installed generate enormous data rates that must be transmitted, processed and stored by the management systems. The sampling time of the electrical signals, which depends on their nature since there are phenomena that can be generated in the order of nanoseconds to long duration disturbances with times exceeding 60 seconds. As an example, one of the systems with the largest expansion and investment worldwide is that of the United States of America, where investment in the electricity sector to 2040 will be $44 billion dollars. According to data from the United States Energy Information Administration, sustained interruptions of electrical service are 5.8 hours, causing economic losses due to power outages that are estimated to range from $22 billion to $135 billion dollars a year [American Society of Civil Engineers, 2020]. In this sense, the operation of the electrical systems near their stability limits, the presence of equipment sensitive to electromagnetic disturbances and the high economic losses generated by the loss of productivity are factors that encourage the design and planning of scalable electrical networks that allow minimizing the interruption times of the electrical service supported by an adequate failure diagnosis. Therefore, one of the biggest challenges for the researchers and the motivation of the present thesis is the planning of electrical networks that allow minimizing the time required for the detection, isolation, and restoration of the system, as well as compressing the amount of data measured by the sensors in the whole electrical system. In the following section there are terms and definitions related to the literature, followed by the state of the art on fault diagnosis and the corresponding solution strategies. Finally, the content of the thesis is summarized.

1.1 Nomenclature

In order to maintain the adequate operation of a system, the supervision functions are defined to indicate unwanted or not allowed process states with the purpose of taking the appropriate actions to correct them. The terms and definitions on the subject of fault diagnosis were taken from [IEEE Std 1159-2019, 2019] and the IFAC Technical Committee SAFEPROCESS [Isermann and Ballé, 1997].

- **System states, signals, time dependency, and frequency**:
  - **Disturbance**: An unknown and uncontrolled input acting on a system.
  - **Declared voltage**: The nominal voltage or a value of a voltage different from the nominal voltage obtained by agreement between the electricity supplier and a consumer.

- **Duration of a voltage swell**: The interval between the time when the root-mean-square (RMS) voltage rises above the swell threshold and the time it subsequently drops below the swell threshold.
- **Error**: Deviation between a computed or measured value and the true, specified or theoretically correct value.
- **Failure**: An interruption of a systems ability to perform a required function under specified operating conditions
- **Fault**: Unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable, usual, or standard conditions.
- **Interharmonic**: A frequency component of a periodic quantity that is not an integer multiple of the frequency at which the supply system is operating (e.g., 50 Hz, 60 Hz).
- **Impulsive transient**: A sudden nonpower frequency change in the steady-state condition of voltage or current that is unidirectional in polarity.
- **Malfunction**: Intermittent irregularity in fulfillment of a systems desired function.
- **Residual**: Fault indicator, based on deviations between measurements and model-equation-based calculations.
- **Root-mean-square (RMS) variation**: A term often used to express a variation in the rms value of a voltage or current measurement from the nominal.
- **Short-circuit power**: The theoretical value expressed in MVA of the initial symmetrical three-phase short-circuit power at a point on the supply system. It is defined as the product of the initial symmetrical short-circuit current, the nominal system voltage, and the factor $\sqrt{3}$ with the aperiodic component (dc) being neglected.
- **Supply system**: All the lines, switchgear, and transformers operating at various voltages that make up the transmission systems and distribution systems to which customers' installations are connected.
- **Sustained interruption**: A type of long-duration RMS voltage variation where the complete loss of voltage (< 0.1 pu) on one of more phase conductors is for a time greater than 1 minute.
- **Temporary interruption**: A type of short-duration RMS variation where the complete loss of voltage (> 0.1 pu) on one or more phase conductors is for a time period between 3 s and 1 minute.
- **Total interharmonic distortion**: The ratio of the root mean square of the harmonic content, considering interharmonic components up to the 50th order and specifically excluding harmonics, expressed as a

percent of the fundamental. Interharmonic components of order greater than 50 may be included when necessary.

- **Time synchronization**: A method of using a time signal from an accurate time source for time synchronizing equipment.
- **Traveling wave**: The resulting wave when the electric variation in a circuit takes the form of translation of energy along a conductor, such energy being always equally divided between current and potential forms.
- **Voltage change**: A variation of the RMS or peak value of a voltage between two consecutive levels sustained for definite but unspecified durations.
- **Voltage fluctuation**: A series of voltage changes or a cyclical variation of the voltage envelope.
- **Voltage interruption**: The disappearance of the supply voltage on one or more phases. It is usually qualified by an additional term indicating the duration of the interruption (e.g., momentary, temporary, sustained).
- **Waveform distortion**: A steady-state deviation from an ideal sine wave of power frequency principally characterized by the spectral content of the deviation.
- **Symptom**: Change of an observable quantity from normal behavior.

- **Functions and tasks**:
  - **Fault detection**: Determination of faults present in a system and time of incidence.
  - **Fault isolation**: Determination of kind, location and time of detection of a fault by evaluating symptoms. Follows fault detection
  - **Fault identification**: Determination of the size and time-variant behavior of a fault. Follows fault isolation
  - **Fault diagnosis**: Determination of type, size, location and time of detection of a fault by evaluating symptoms. It includes fault detection, isolation, and identification.
  - **Monitoring**: A continuous real-time task of determining the possible conditions of a physical system, recognizing, and indicating anomalies of the behavior.
  - **Supervision**: Monitoring a physical system and taking appropriate actions to maintain the operation in the case of faults.
  - **Protection**: Means by which a potentially dangerous behavior of the system is suppressed if possible, or means by which the consequences of a dangerous behavior are avoided.

- **Models**:
  - **Quantitative model**: Use of static and dynamic relations among system variables and parameters in order to describe the system behavior in quantitative mathematical terms.
  - **Qualitative model**: Use of static and dynamic relations among system variables in order to describe the system behavior in qualitative terms, such as causalities.
  - **Analytical model**: The mathematical relations among the system variables are based on physical laws deriving by the knowledge of the system behavior.
  - **Data-driven model**: The mathematical relations among the system variables are inferred on the basis of a set data coming from the system itself.
- **System properties**:
  - **Reliability**: Ability of a system to perform a required function under stated conditions, within a given scope, during a given period of time.
  - **Safety**: Ability of a system not to cause danger to persons or equipment or the environment.
  - **Availability**: Probability that a system or equipment will operate satisfactorily and effectively at any point of time.
- **Faults, time dependency**:
  - **Abrupt fault**: Fault modeled as step-wise function. It represents bias in the monitored signal.
  - **Incipient fault**: Fault modeled by using ramp signals. It represents drift of the monitored signal.
  - **Intermittent fault**: Combination of pulses with different amplitudes and lengths.

1.2 Electric Distribution Networks Planning

Due to the sustained growth in the demand for electricity supply, the planning of the electrical power system is becoming more complex, involving the expansion of substations, the installation of distributed generation, the deployment, and upgrading of feeders to meet requirements in an optimal, timely, and cost-effective way, while respecting environmental standards.

The first stage is the generation of electrical energy from primary energy sources such as fuels, solar energy, wind energy, geothermal, and hydraulic energy. To transmit the electrical energy generated by the plants, it is necessary to raise

the voltage in order to reduce losses in transmission to the distribution centers, allowing the flow of electrical energy at service voltage levels to the subscribers of the electrical sector.

Electrical network planning must take into account the level of reliability along with costs related to substation location, transformer location, and capacity, appropriate transmission line diameters, and conductor options based on required voltage levels. The main limitation in the planning of electrical networks are the high investment costs related to the updating of the system, as well as the operation and maintenance costs, with the main objective of minimizing the load loss by maximizing the reliability of the system. The solutions to optimization problems presented by researchers depend on the objective function and the constraints posed. The complexity of planning issues depends on the constraints of the problem which include power balance, voltage levels, power flow, active, and reactive power control, penetration limits, substation capacity, radiality, power factor regulation, power losses, total investment, and operation constraints. Several solutions have been proposed by researchers using mathematical optimization and metaheuristic methods. The evolution of traditional power grids to smart grids is possible through the addition of distributed generation, energy storage, and communications among all the actors. An extremely complex scenario is the planning of the energy distribution network because it presents constraints such as voltage drops, operating limits, power balance, distributed generation entry, and the inclusion of geopositioning systems for the laying of the power grid.

1.3 Fault Diagnosis of Electrical Distribution Networks

Due to the complexity of the electrical system and the operation of the network near its stability thresholds, there is always the possibility that there will be disturbances or failures in the Electrical Power System (EPS). An electrical fault is an abnormal condition in a power system that can be caused by degradation of insulation, short circuit or interruption of current flow caused by the operation of circuit breakers. In electrical systems, specifically in overhead line systems, most short circuit failures, typically between 80% and 90%, tend to occur in overhead lines and cables, in substation equipment, and combined bus bars. The information contained in the current and voltage signals can be used to analyze the behavior of the electrical system and make the diagnosis of failures using mathematical models applied to dynamic systems. Each stage that constitutes the electrical system must be capable of detecting, isolating, classifying and locating all types of faults by activating devices under the design of protection schemes in order to minimize the interruption of energy supply to consumers.

Taking the example of the transmission system in England and Wales, about 300 short-circuit faults are registered per year. Failures that appear correspond to the four main types: line-to-ground faults (67%), line-to-line faults (25%), 5% cor-

respond to three-phase symmetrical faults, and finally 3% are Line-to-Line Ground faults. Approximately 77% of the breakdowns recorded in the England/Wales transmission system are caused by weather phenomena due to storms, snow, saline pollution in the insulators, and strong winds that cause the collapse of the transmission lines, thus showing a total of 11 types of faults that can occur in the transmission lines [Tleis, 2019].

In the event of an electrical system fault, intermittent voltage spikes, and power quality irregularities, blackouts or brownouts occur if demand exceeds capacity during particular periods of time. The consequences for the end user are the loss of reliability in the supply of electricity, which imposes economic losses in homes and businesses. The economic loss is caused by damage to equipment manufactured with sensitive electronic circuits that are affected by voltage spikes, deterioration of food that needs electricity to maintain in controlled temperature conditions, wasted, and unproductive time for workers with temporarily paralyzed activities in factories or businesses, additional costs incurred by increased dependence and use of backup generators.

Table 1.1 presents the estimated cost impact by the economic sector in billions of dollars [American Society of Civil Engineers, 2020].

Table 1.1: Cumulative Impacts by Sector ($ billions).

Economic Sector	2012	Cumulative, 2012–20	Cumulative, 2012–40
Residential	6	71	354
Commercial/other	6	74	402
Industrial	4	52	239
Transportation	0.03	0.38	3.82

One of the fundamental pillars for the development of smart cities is electricity. The significant increase in non-linear loads that are part of the electrical grids reduces the quality of the energy supplied. With the purpose of complying with quality standards, it is necessary to monitor electrical networks in order to detect, classify, and locate disturbances at any particular point of the electrical system [Arauz, 2019] [Basavaraj, 2016] [Jia et al., 2018]. Monitoring requires a significant number of sensors and technological advantages in microelectronics, mechanical systems and wireless communications have made it possible to have a massive number of sensors at a relatively low cost. However, a new challenge facing communications networks is the ability to process, transport, and store the enormous amount of data without losing important information [Maldonado, 2018] [Baimel et al., 2016].

1.4 Outline of the Thesis

The contents of the thesis are briefly summarized in this section, in order to provide an overview of the work.

- **Chapter 2**: the state of the art is presented on topics related to: *planning* of electrical networks, showing the possible objective functions, restrictions of the problem, and variables. Current procedures used to *detect, classify, and locate* faults in electrical systems are described. Finally, the *compression* techniques used in electrical signals are shown.

- **Chapter 3**: this chapter describes the formulation of the problem and the mathematical models, heuristics, and algorithms developed to give solution to the planning of electrical networks, diagnosis of electrical failures and data compression.

- **Chapter 4**: the chapter presents the results obtained in the planning of electrical networks, fault diagnosis and signal compression, and makes the comparison with the results of other studies.

- **Chapter 5**: in this chapter, the comments of the results obtained and future studies are shown.

Chapter 2

Power System and Fault Modeling

This chapter shows the research related to the planning of electrical networks, detailing objective functions, variables and constraints proposed by researchers in scientific publications. Section 2, describes the methods of detection, classification and location of faults in electrical networks. Finally, the theory of compressed sensing is presented and the results of research applied to electrical systems are described.

2.1 Optimum electrical distribution network planning

Power network planning optimizes the decision variables according to the objectives and the desired pressure. The complexity of electrical network planning increases with the dimensionality of the problem. Generally, the decision variables in planning problems are location, size and investment. There are optimization problems that involve more than one objective function to be solved and are called multiobjective optimization. For example, it is required to minimize energy losses, investment cost and pollution emission, as well as to maximize the reliability level. The result is a set of solutions from which the best one must be chosen. Optimization problems are generically presented as:

$$Minimizing\ or\ Maximizing\ f(x)$$
$$subject\ to\ A(x)\ \leq\ b$$

Where $f(x)$ is the objective function of the problem, x represents the decision variable and $A(x)\ \leq\ b$ describes a set of constraints expressed by a set of equations or inequations.

The decision variables can be continuous or discrete. Continuous variables represent the amount of voltage level at each bus, the energy flow of feeders, distributed generation, energy injected from substations and restricted load. The

binary variables represent for example: the installation of new feeders and distribution substations, the change in the type of conductors and the improvement of the capacity of the distribution substation.

Conventional decision options take into account the routing of feeders for the primary and secondary systems, determining the capacity of the transformers, as well as the size of the substations. The appropriate conductor and cable diameters, the type of overhead or subway installation are determined. In recent years, distributed generation, integration of micro-grids, energy procurement, allocation of protective devices, penetration of electric vehicles, energy storage, demand response and the ability to generate reactive power by providing economically viable solutions and reliable services to consumers have been involved. The main objective function involved in the expansion problem is the investment and operating cost. The investment cost is related to the actualization and improvement of the system. Operating costs are related to the cost of maintenance, energy losses during the whole period of operation, cost of energy not supplied and cost of distributed energy generation that would be minimized to improve the reliability level of the network. Table 2.1 presents the main objective functions used in the research.

Table 2.1: Objective functions used in the distribution network expansion planning problem [Vahidinasab et al., 2017].

Objective functions	References
Emission	[Ahmadigorji et al., 2017, Zeng et al., 2014, Hung et al., 2014]
Reliability	[Jeddi et al., 2019, Arasteh et al., 2019, Ahmadi et al., 2019]
Feeder loss	[Ruiz et al., 2020, Xie et al., 2018, Banol Arias et al., 2018]
Voltage Drop	[Aghaei et al., 2014, El-Zonkoly, 2013, Niknam et al., 2012]
DG Investment	[Banol Arias et al., 2018, Munoz-Delgado et al., 2016, Arasteh et al., 2016]
Transformer Loss	[Munoz-Delgado et al., 2016, Mazhari and Monsef, 2015, Vélez M. et al., 2014]
Purchased Energy	[Gholizadeh-Roshanagh et al., 2020, Arasteh et al., 2016, Yao et al., 2014]
DG Operation Cost	[Zakernezhad et al., 2021, Jooshaki et al., 2019]
Feeder Investment	[Mehrtash et al., 2020, Gholizadeh-Roshanagh and Zare, 2019]
Feeder Operation Cost	[Xie et al., 2018, Sipoli et al., 2014, Aghaei et al., 2014]
Substation Investment	[Kaewmamuang et al., 2019, Abedi et al., 2019]
Substation Operation Cost	[de Quevedo et al., 2019, Xie et al., 2018, Banol Arias et al., 2018]

The size, type and location of distributed generation to be installed is determined based on geographic conditions, fuel resources and the feasibility of using renewable energy resources such as wind turbines, photovoltaic panels, biomass, *etc.* [Verma et al., 2020, Vahidinasab et al., 2017]. Table 2.2 shows the main variables used in electrical network planning. The main technical constraints or operating constraints incorporated in the objective functions of the Distribution Network Expansion Planning Problem (DNEP) are shown below:

Table 2.2: Commonly used variables in the distribution expansion planning problem [Verma et al., 2020].

Variable	Constraint Equation
Voltage limit for nodes and busses	$V_i^{min} \leq V_i \leq V_i^{max}$
Power balance	$P_{DG,i} + P_{DG,i}^{cur} - \sum_{i=1}^{N_B}[\frac{(\lvert V_i\rvert - \lvert V_j\rvert)^2}{\lvert Z_{ij}\rvert}] * pf - P_{L,i} - \sum_{i=1}^{N_B} P_{b,f,ss}(V_i, \theta_i) = 0$
Branch current flow	$I_f^{min} \leq I_f \leq I_f^{max}$
Radiality of the distribution network	$\sum_{i \epsilon \Psi_{NW2}} N_{Br2} = N_{LP2} - \sum_{S2 \epsilon \Psi_{SS2}} N_{SS2}$
Branch thermal limits	$\sqrt{[P_{b,f,ss}(V_{ij}, \theta)]^2 + [Q_{b,f,ss}(V_{ij}, \theta)]^2} \leq S_b^{max}$
Power flow capacity	$\lvert S_{b,f,ss}\rvert \leq S_{b,f,ss}^{max} b \ \epsilon \ \Psi_{B1}, f \ \Psi_{F1}, ss \ \epsilon \ \Psi_{N1}, \Psi_{N2}$
Power factor	$Q = P * tan(\Phi)$
Uni-directional flow	$S_{Purch}^1 \geq 0$
Reliability	$CRS_z^{min} \leq CRS_z \leq CRS_z^{max}$

2.2 Methods of Diagnosis and Classification of Faults

Fault diagnosis in electric power systems is one of the most challenging and particularly important problems in protection design and operation. Diagnostic methods aim to detect faults as early as possible and isolate them from the system. Subsequently, classification, and localization are carried out. These methods must be reliable, selective and with high response speeds, ensuring maximum service continuity. In inefficient electrical systems, a fault can propagate and cause the total collapse of the system, known as blackout. To meet increasingly demanding standards, researchers are combining advanced signal processing techniques, artificial intelligence, global positioning systems, and communications systems to improve power system performance. The dynamics of the power system is recorded periodically by measuring current and voltage variables. It is important to have fast systems in the sampling of variables, processing, and control, with high selectivity rates and ability to differentiate between the different types of faults that the system can tolerate, for example the operating speed of protection systems such as high-speed relays operate in less than 20 ms. Depending on the sampling rate of the monitoring equipment, it is possible to determine the type of disturbance, for example to detect a transient impulse, it is necessary to acquire samples in the

order of nanoseconds, which implies the generation of large amounts of data.

Although the current and voltage signals have all the information about the system, it is impossible to frame in a set of rules and criteria all the faults that exist in the electrical power systems. There are several techniques to extract information from the signals, for example, it is possible to extract information about the fundamental frequency component and the harmonic components from the time signals.

It is important to consider the performance of the algorithm with respect to the amount of data to be processed, as the amount of data increases the results are more accurate in the detection, classification and localization of faults.

The models are specific to the process or system, however, fault diagnosis techniques are classified into different categories [Aleem et al., 2014].

In the quantitative method, models are based on physical or fundamental engineering principles and provide the most accurate production estimators when they are well formulated. Quantitative methods are usually complex and require computational effort, the effectiveness of the procedure is limited by the availability of sensor information and it is difficult to build dynamic models. There is a limited modeling approach as linear models are mainly used, these models cannot explain the cause of the fault in the diagnosis. If a fault is not modeled specifically, there is no guarantee of its detection. Challenging efforts to derive explicit mathematical models would make it less likely for increasingly complex systems, however, it can provide important transient resolution for fault diagnosis. The qualitative method is simple to develop and apply, since they are causal in nature and can provide an explanation in the diagnosis and they do not require extensive knowledge or analytical information to make the diagnosis. These methods are specific to the process or system, then it is difficult to tabulate a set of rules for complex systems. They rely heavily on the heuristic knowledge and experience of the developer, thus introducing a variance bias in qualitative reasoning and explanation. Qualitative methods can provide fast and efficient diagnosis for non-critical processes. Methodology is useful when systems expertise exists, but analytical mode is lacking. The process history-based method is suitable when training data are large or are simple to create or collect. Mapping functions are well understood and have been developed extensively. There is extensive research in the development of pattern recognition (machine learning) algorithms. These methods demonstrate considerable robustness to system noise under dynamic conditions. In the case of supervised learning, simulation of training data with steady-state and dynamic information is needed.

These models pose difficulties in classifying an anomaly outside the established measurements, it is also difficult to classify multiple faults simultaneously. Process history-based methods demonstrate robustness to system noise and are applicable in many pattern recognition problems. These methods are also suitable for systems where it is difficult to obtain parameters, especially when parameters change with system dynamics.

The methodology of fault diagnosis based on process history begins with the measurement of parameters such as voltages, currents and line impedance followed

by the extraction of characteristics. Generally, in case of faults in power systems, the measurement is of electrical parameters to detect and classify the fault. Feature extraction is an essential phase in the pattern recognition problem that forms the basis of classifier performance. For a reliable classifier, it is important that the feature extraction technique reduces the processing data and preserves important features.

Nodal transformations such as the Clarke transform and Park relations are commonly used in three-phase AC machine-oriented control. The Clarke transform converts the time-domain components of a three-phase system (of an abc frame) into two components of an orthogonal stationary frame ($\alpha\beta$). Park's transform converts the two components of the frame ($\alpha\beta$) to an orthogonal rotating reference frame. Implementing these two transforms consecutively simplifies the calculations by converting the AC voltage and current waveform into DC signals. In [Gupta and Tripathy, 2015], the author uses superimposed sequence components based on integrated impedance (SSCII). The algorithm is tested on a 50Hz 400kV system and analyzes the current and voltage profiles at the beginning and end of the transmission lines. This method is reliable for low and high resistance faults and the algorithm is suitable for high speed communication channels. This method can detect faults in less than 20ms. The Fourier transform obtains the frequency characteristics of the signal; the fundamental frequency of power systems is 50 or 60 Hz, when performing the Fourier transform additional frequencies are presented. A system is under conditions of instability due to transients, lightning or faults at different points of the system. Based on the literature, the Fourier transform is applied in half or full cycle, the difference is the time to detect a fault, for example if the fault occurs at the beginning of the positive edge using the full cycle, the algorithm will take (1/50) or (1/60) seconds and if it is applied in each half cycle the algorithm will take (1/25) or (1/30) seconds depending on the fundamental frequency of the system. In [Raza et al., 2020], the detection method proposed by the author is based on Fuzzy-neuro, it is tested in a 50Hz 220kV system and analyzes the fault current and several voltage samples. This method is based on back-propagation and fuzzy control theory, the high harmonic components are eliminated by FFT and the fault detection is performed in less than 10 ms.

The wavelet transform is one of the most widely used methods in the extraction of features for fault detection. The main feature of wavelets is the extraction of information from current and voltage signals using a multiple resolution and frequency decomposition. The frequency bands used vary between 100 Hz and 1 MHz. In [Silva et al., 2006], the fault is detected through a method based on discrete wavelet transform and artificial neural networks. It is tested on a 60Hz 230kV system and analyzes current and voltage signals. This method normalizes the voltage and current signals to values in the range of -1 to 1 and implements the Db4 wavelet. This method can obtain 100% accuracy for fault detection and 99.83% accuracy for fault classification.

In [Cheng Hong, 2000], the detection method is based on the Daubechies level 5 wavelet transform and an artificial neural network, the strategy is tested on a

50Hz 500kV system and analyzes the voltage and current waveforms, then 3960 fault cases are considered for classification and adaptive response training. The results present an accuracy of 99.7% for fault detection in single lines and 92% in parallel lines, and with respect to fault classification, the result is an accuracy of 99.65%. In [Yadav and Swetapadma, 2015], the fault is detected by a method based on linear discriminant analysis (LDA) and wavelet transform. The algorithm is tested on a 50Hz 400kV system and analyzes only current signals. This strategy uses the wavelet transform to process the current samples up to three levels, then the relays are set up to 90% of the transmission line. This method can obtain 100% accuracy for both fault detection and classification. The best results obtained in previous work with similar methodologies are summarized in Table 2.3.

Table 2.3: Fault detection and classification best results in literature review.

Fault detection		
Technique	**Accuracy**	**Elapsed time**
Wavelet and Fuzzy-neuro based [Wang and Keerthipala, 1998]	High but not 100%	10 ms
Discrete Wavelet transform and Artificial Neural Networks [Abdullah, 2018]	41-100%	Not specified
Wavelet transform and a self-organized artificial neural network, [Cheng Hong, 2000]	99.7%	Not specified
Superimposed sequence components based integrated impedance [Gupta and Tripathy, 2015]	High but not specified%	20ms
Fault classification		
Technique	**Accuracy**	**Elapsed time**
Back-propagation network classifier [Aggarwal et al., 1999]	99%	not specified
Fuzzy-logic and WT based method [Youssef, 2004]	99%	10ms
Fuzzy-logic [Das and Reddy, 2005]	97%	not specified
Field-programmable gate array (FPGA) with WT [Valsan and Swarup, 2009]	100%	6ms

Among the methods based on artificial intelligence, whose principle of operation is based on the learning or training of the algorithm from events and measurements recorded in the system at different stages are Artificial Neural Networks (ANN), Fuzzy Logic (FL), Expert Systems (ES), Genetic Algorithms (GA) and Support

Vector Machines (SVM).

Table 2.4 presents different methods of fault localization.

Table 2.4: Comparison of different fault location finding methods.

Algorithm	Input	Features	Results
Artificial neural network [Mazon et al., 2000].	Pre-fault current and voltage samples.	Training time varies from 5 s to 2.5 min to accomplish the mentioned error level.	The maximum error is 0.7% while 0.12% is the minimum error.
Neuro-fuzzy systems and wavelet transform [Reddy, 2007].	Current and voltage magnitudes.	Back-propagation is used for learning and 228 various faults created for analysis.	–
Artificial neural network and wavelet transform [Reddy, 2008].	Current and voltage samples.	Daubechies is employed and dissolved up to three levels, 10 kH sampling frequency.	The maximum error is 0.06% while 0.68% is the minimum error.
Support vector machine and scaled conjugate gradient [Gayathri, 2015].	Positive sequence voltage and line currents.	5 kHz sampling rate, 2e-004s is time to locate the fault.	The maximum error is 1.852 km while 7.874e-3 km is minimum.
Fast Fourier transform with traveling-wave theory [Mamiş et al., 2013].	Current samples measured from one end.	25.6 kHz sampling frequency and 512 samples, to reduce FFT leakage Hanning window is employed.	Fault location error is 0.12%.
Impedance based method [Roostaee et al., 2017]	Current and voltage samples.	DIGSILENT is used to simulate the test system, 10 kHz as sampling frequency with simulation time 0.2 s.	Achieved FL error is less than 1%.

Depending on the available circuit data, different fault location algorithms can be implemented, *e.g.*, two-ended synchronized method, two-ended unsynchronized method, two-ended current unsynchronized method.

Single-ended impedance-based fault location algorithms estimate the location of the fault on a transmission line from one end. To locate all types of faults, phase-to-ground voltages and currents must be measured on each phase. The voltage and current waveforms captured during a fault by an Intelligent Electronic Device

(IED) at one end of the line are used to determine the apparent impedance between the IED and the short-circuit fault location. The advantages of using single-ended algorithms are that they are simple to implement, produce reasonable location estimates, and require data from only one end of a line. Depending on the circuit data available, different fault location algorithms can be implemented, *e.g.*, single reactance method, fault location method without using source impedances, and fault location methods using source impedance. Two-ended impedance-based algorithms use voltage and current waveform data captured at both ends of a transmission line to estimate the location of a fault. Additional measurements from the remote end of a transmission line are used to eliminate any reactance errors caused by fault resistance, load current, or system inhomogeneity. A communication channel transfers data from one IED device to another.

The k-Nearest Neighbors (k-NN) algorithm is a non-parametric classification method that can achieve high classification accuracy on problems with non-normal and unknown distributions. For a particular sample, k nearest points are found between the data and the sample. Generally, the Euclidean distance is used, where the components of a point are used to compare with the components of another point. The basis of the k-NN algorithm is a data matrix consisting of N rows and M columns. The parameters N and M are the number of data points and the dimension of each data point, respectively. Using the data matrix, a query point is provided and the k closest points within this data matrix that are closest to this query point are searched.

Nearest neighbor classifiers are based on learning by analogy, *i.e.*, comparing a given test sequence with training sequences. The training sequences are described by n attributes. Each sequence represents a point in a n-dimensional space. Thus, all training sequences are stored in an n-dimensional pattern space. An unknown sequence, the k nearest neighbor classifier searches the pattern space for the k training sequences closest to the unknown sequence. These k training sequences are the "k nearest neighbors" of the unknown sequence. The closeness is defined in terms of the Euclidean distance. The training complexity of k-NN is O(N), *i.e.*, linear in the size of the training set in both time and space. The Euclidean distance between two points or sequences, let's say $X_1 = (x_{11}, x_{12}, ..., x_{1n})$ and $X_2 = (x_{21}, x_{22}, ..., x_{2n})$, is shown in Eq. 2.1:

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (X_{1i} - X_{2i})^2} \tag{2.1}$$

Methods based on traveling waves are based on the reflection and transmission of the signals generated in the electrical system in case of failure. They offer high accuracy in the location of the fault and are not affected by constructive parameters of the network and their implementation is complex and costly, since it requires the additional installation of acquisition equipment as well as a great accuracy in the synchronization at both ends of the line. These methods can present problems in distinguishing between reflected waves at the fault point and at the lateral branches of the line, as well as being affected by the presence of high noise levels in the input

signals. The main methods for detecting traveling waveheads include the wavelet transform, mathematical morphology, and the Hilbert-Huang transform.

Fault location methods using traveling waves are independent of the network configuration and the devices installed in the network. These techniques are very accurate, but require a high sampling frequency and are more expensive to implement than impedance-based techniques. In steady state, the line parameters are homogeneous and the voltages and currents can be calculated by means of a linear relationship. When a fault occurs along a transmission line, voltage and current transients will travel to the line terminals. These transients will continue to bounce back and forth between the fault point and the two faulted line terminals until the post-fault steady state is reached. The evolution of the terminal bus transients can be constructed using the well-known lattice diagram method [Fei et al., 2019].

Considering a single-phase lossless transmission line of length l, connected between buses A and B, with characteristic impedance Z_c and traveling wave velocity of v. If a fault occurs at a distance x from bus A, this would appear as an abrupt injection at the fault point. This injection will travel as a surge along the line in both directions between the fault point and two terminals until the subsequent fault occurs.

The differential equations representing the voltage and current in Eq. 2.2 and Eq. 2.3, respectively, in the transmission lines are functions of distance and time, which are described below:

$$\frac{\partial e(x,t)}{\partial x} = -Ri(x,t) - L\frac{\partial i(x,t)}{\partial t} \tag{2.2}$$

$$\frac{\partial i(x,t)}{\partial x} = -Gv(x,t) - C\frac{\partial e(x,t)}{\partial t} \tag{2.3}$$

where: R is the resistance in units per length, L is the inductance, G is the conductance and C is the capacitance. All these parameters are characteristics of the transmission lines.

For short transmission lines, the model used is the lumped concentrated parameter model where G and C are omitted. The location of the fault can be calculated by solving the differential equations, representing the voltage and current in Eq. 2.4 and Eq. 2.5, respectively:

$$\frac{\partial e(x,t)}{\partial x} = -L\frac{\partial i(x,t)}{\partial t} \tag{2.4}$$

$$\frac{\partial i(x,t)}{\partial x} = -C\frac{\partial e(x,t)}{\partial t} \tag{2.5}$$

The solution of the differential equations for voltage and current in Eq. 2.6 and Eq. 2.7, respectively:

$$e(x,t) = e_f(x-vt) + e_r(x-vt) \tag{2.6}$$

$$i(x,t) = \frac{1}{Z_c}e_f(x-vt) - \frac{1}{Z_c}e_r(x-vt) \tag{2.7}$$

where:

$$Z_c = \sqrt{\frac{L}{C}}$$

is the impedance of the transmission line and:

$$v = \sqrt{\frac{1}{LC}}$$

is the propagation velocity.

The Traveling Waves (TW) produced by a perturbation at location x travel in different directions toward the ends of the transmission line, with velocities close to the speed of light. The ends of the transmission line represent a discontinuity or impedance change in which part of the wave energy will be reflected in the perturbation. The remaining energy will travel to other elements of the electrical system or transmission lines. The lattice diagram in Figure 2.1 illustrates the multiple waves generated at the ends of the lines. The wave amplitudes are represented by reflection coefficients kA and kB, which are determined by the characteristic impedance relationships at the discontinuities. τ_A and τ_B represent the travel time from the fault to the discontinuity.

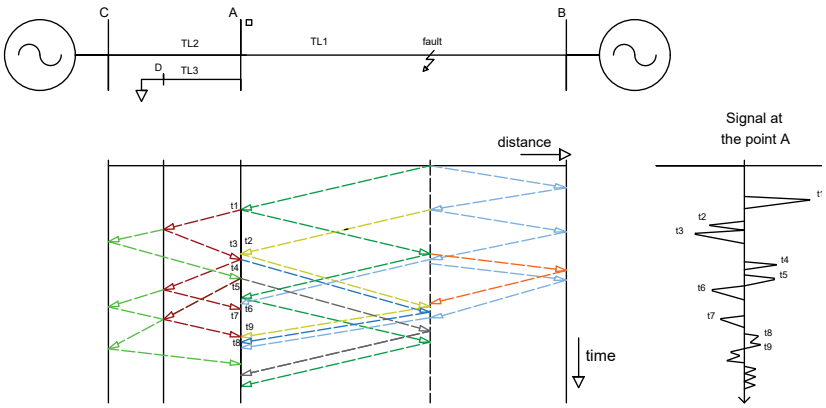


Figure 2.1: Ladder diagram of TW propagation and the form of the signal at the point measurement.

With GPS technology, τ_a and τ_b can be accurately determined. By knowing the length (l) of the line and the arrival time difference (τ_a – τ_b), the distance (x) to the fault from substation A can be calculated by means of Eq. 2.8:

$$x = \frac{l - C(\tau_a - \tau_b)}{2} \tag{2.8}$$

where: C is the propagation speed of a wave 299,79 m/s. The construction of the lattice diagram becomes computationally difficult if the attenuation and

distortion of the signals as they propagate along the line are taken into account. On the other hand, the time-frequency resolution of the transient signals can be used to determine the travel times of these transients between the fault point and the line terminals. In three-phase transmission lines, if losses are taken into account, there are three modes of propagation, so that for the analysis of the traveling wave effect, the phase values must be converted into modal values.

Chapter 3

Planning and Fault Diagnosis Design

Chapter 3 presents the models that have been developed to plan electrical networks using georeferenced coordinates of users, transformers, routes and substations. Additionally, the models and algorithms have been exploited for the diagnosis of electrical faults taking into account the detection, classification and localization of faults. A recovery algorithm in case of data loss and finally a data compression algorithm are presented.

3.1 Scalable Electrical Networks Planning

The present research is based on the planning of the electrical distribution networks from the data generated by the geographic information system known as Open Street Map (OSM). The information acquired are the positions and the number of buildings, streets, parks, etc.

The route of the feeders must maintain direct relation with the georeferenced system and the distance between the transformer and the end user must respect the voltage drop levels established by the standards and regulations.

The design of the distribution network considers as a restriction that the subscribers are supplied by a single transformer that provides low voltage electricity supply to each user. To determine the location of the transformers, a set of possible candidate sites is generated based on the geographical characteristics of the area where the network will be implemented. The transformers are fed by the primary distribution network from the medium voltage substation. The maximum capacity of the transformers to supply energy to a defined number of users is restricted by power and cost.

The ampacity of the cables and conductors are calculated according to the voltage level, current, type of material and cross section. The cost of implementing the distribution network only considers the unit cost of the conductor per km.

When current flows through a conductor, the Joule effect causes energy losses by transformation into heat. Therefore, the voltage of a load bus or PQ bus is

different from the voltage of the generator bus or PV bus. Eq. 3.1 shows the voltage V_i on each PQ_i bus which must be within the specified range:

$$V_{min} \leq V_i \leq V_{max} \tag{3.1}$$

In distribution networks, the conductor resistance is very large in comparison with the conductor reactance, therefore the losses caused by the conductor reactance can be neglected and for this reason they are not analyzed in the present research.

The main objective function involved in the planning problem is the cost of implementing the electrical network. The objective is to find the shortest distances and the positions to locate the transformers, so the objective function is presented in Eq. 3.2:

$$min \sum_{j=1}^{M} \sum_{i=1}^{N} X_i Y_j \tag{3.2}$$

where: M is the number of candidate sites, N is the number of users, X_i is the locations of each user, and Y_i is the locations of each transformer.

The constraints are the following: According to Eq. 3.3 only one transformer provides energy to one user. Eq. 3.4 is the restriction that limits the maximum number of users that can be connected to a transformer. Eq. 3.5 is used to enable or disable the location of the transformers.

Subject to:

$$X_i * Y_j = 1 \qquad \forall\ X_i / \{X_i \in only\ Y_j\} \tag{3.3}$$

$$\sum_{i=1}^{N} (X_i * P_i) \leq Cmax_j \qquad \forall\ X_i / \{X_i \in coverage\ area\ Y_j\} \tag{3.4}$$

$$Y_j \in 0, 1 \qquad \forall\ Y_j \tag{3.5}$$

Using the geo-referenced coordinates of a city, buildings are considered as subscribers of the electric service, therefore, it is imperative to supply electric power to each user through a transformer.

For determining the shortest route between the subscribers and the transformers, the combinatorial optimization heuristic known as Steiner Tree is used. Sets of subscribers are established according to locations or strata, each user presenting random power consumptions. The k-medoids clustering algorithms and the Delaunay-Voronoi triangulation are used to establish the groups.

The Haversine formula is used to calculate the distances to be covered by the feeders, which takes into account the curvature of the earth. The routing of cables and conductors of the medium and low voltage network is based on the number of users, the coordinates of the buildings and the coordinates of the streets. The optimal location of the transformers starts from the establishment of

candidate locations that are proposed at the intersections of the streets, which is a combinatorial problem known as NP-Complete called Set-Cover whose solution is presented below.

Algorithm 1 consists of the following phases. Step 1: Data acquisition from open street map, building and street coordinates are extracted. Step 2: Constructing sets of subscribers, intersections and candidate locations for transformer placement. Step 3: Calculation of subscriber group sizes. Step 4: creation of the sets. Step 5: Generation and assignment of random loads for each subscriber.

Algorithm 1: Reconstruction and clustering

Step 1: Acquire data from OSM
$x^u, y^u, m_{con}, x^{in}, y^{in}$
Step 2: Build sets
$\Omega_u = [x^u, y^u],\ \Omega_{in} = [x^{in}, y^{in}],\ \Omega_s = [x^{in}, y^{in}]$
$Se = [x^{se}, y^{se}]$
Step 3: Calculate the size of group
$clust = round(N/2^{nc})$
Step 4: Build cluster
$[idx, m^{xy} = kmedoids(clust, \Omega_u)]$
$tri = dalaunay(m^{xy}(:,1), m^{xy}(:,Z)$
$voronoy(m^{xy}(:,1), m^{xy}(:,Z), tri);$
Step 5: Allocation of random loads
$m^U_{idx} = [(1:N), idx]$
$for\ C_1 = 1:N$
    $m^U_{idx}(C_1, 3) = random(idx(C_1))$
$endfor$

In order to cover all the subscribers of the electric sector, a minimum expansion tree is generated using Algorithm 2, taking into account that the maximum distances must comply with the voltage drop regulations. Once all the users are connected, step 2 is to solve the SetCover problem using Greedy's algorithm, which

selects the number of transformers needed to cover the power required by the users.

Algorithm 2: Placement of transformation centers

Step 1: PRIM modified
  Define: C_{max}, D_{ct}
  $X = [x^u, x^s, x^{se}, mean(x^u)]$
  $Y = [y^u, y^s, y^{se}, mean(y^u)]$
  $for\ C_2 = 1 : M$
    $x_a = X_N + C_2$
    $y_a = Y_N + C_2$
    $gp = primmodified(\Omega_u, x_a, y_a, D_{ct}^{min}, C_t^{max})$
    $u_p^{mn}(C_2, gp) = 1$
  $endfor$
Step 2: Greedy algorithm
  $greedy(u_p^{mn})$
  Return: $solC, solL$

Algorithm 3 selects the capacity of each transformer according to the number of users to which it must supply the electrical service. With the matrix of active sites and users, it calculates, for each transformer, the number of users it must supply energy to and the power of the users. With the power of each transformer, the capacities of the transformers to be installed are identified in the database.

Algorithm 3: Sizing

Step 1: Define: C_{max}, D_{ct}
Step 2: Redundance Error:
  $for\ C_3 = 2 : N$
    $aux = find(solC(C_{3,:} == 1)$
    $if\ length(aux) > 1$
      $for\ C_4 = 2 : length(aux)$
        $solC(i, aux(m)) = 0$
      $endfor$
    $endif$
  $endfor$
Step 3: Power demand
  $for\ C_5 = 1 : M_{C_T}$
    $ind = find(solC(C_{5,:} == 1)$
    $S_{pt} = 0$
    $for\ C_6 : length(ind)$
      $S_{pt} = S_{pt} + m_{idx}^u(ind(C_6, 3))$
      $P_T(C_5) = S_{pt}$
    $endfor$
  $endfor$

Algorithm 4 determines the voltage drops in each circuit by applying the formulas for the current flowing through the conductor, the resistance of the conductor

and the distance.

Algorithm 4: Drop voltage in circuits

Step 1: Define: T_j^{xy}
Step 2: Calculate
    $d_{ui}^{sj}; \forall\ U \mid U \in cobert(T_j)$
Step 3: Assignments
    Define conductor: ρ_c, S_c
    Define network parameters: Φ_c, v_L
Step 2: Calculate
    $d_{U_i S_j} = haversine(U_i\ ,\ T_j)$
    $for\ C_6 = 1 : length(P_{T_j})$
        Section resistance: $R_c(C_6) = d_{U_i S_j} * \rho_c / S_c$
        Current by line: $I = P_T(C_6)/(v_L)$
        $\Delta V = R_C(C_6) * I(C_6))$
    $endfor$

Table 3.1: Used variables

SETS	
Ω_u	Set of users in georeferenced area.
Ω_{in}	Set of intersections in georeferenced area.
Ω_s	Set of candidate sites for the location of transformers.
Ω_Z	Set of active locations.
VARIABLES	
$x^u,\ y^u$	Longitude and latitude of the i-th user.
$x^{in},\ y^{in}$	Longitude and latitude of the i-th intersection.
m_{con}	Connectivity matrix with georeferenced intersections.
S_e	Location of the electrical distribution substation
idx	Indices assigned to each user based on their cluster
m^{xy}	Location of medians obtained when applying K-medoids
m^u_{idx}	User matrix and assigned cluster index
tri	Matrix with the set of triangles formed in the Delaunay triangulation
d_{eT}	Distances between transformers
gp	Graph obtained by applying the Prim algorithm
m_p^{MN}	Preliminary matrix of candidate sites and users
$solC$	Distribution Transformers and users with coverage
$solL$	Labels of the coverage matrix
T_i^{xy}	Coordinates of the i-th active transformer
$d_{u_i Z_j}$	Distance from user i to the transformer j
R_c	Electric resistance of the conductor
ΔV_u	Voltage loss per user
P_Z	Powers of transformers
C_1, C_2, C_3, C_4, C_5	Counters
S_{pt}	Adder
aux	Auxiliary variables
PARAMETERS	
N	Number of users
M	Number of candidate sites
N^{in}	Number of intersections
$x^{se},\ y^{se}$	Longitude and latitude of the distribution substation
n_c	Number of regions for clustering
$clust$	Number of users belonging to the same cluster
C_{max}	Transformers maximum rating
D_{ct}	Minimum distance between transformers
M_{Tac}	Number of active transformers
ρ	Electrical resistivity of the employed conductor
S_c	Cross section of the employed driver
V_L	Line voltage in secondary circuit
Φ	Power factor

3.2 Compressive Sensing Techniques

The detection of an event in electrical power systems is the basis for various applications in the fields of protection, power quality analysis, and system control. The equipment designed to detect and evaluate disturbances and distortions in an electrical system must be as selective as possible so that the alarm is given only for relevant events and not false alarms. Smart grids generate a huge amount of data obtained by the uniform sampling performed by the sensors as they apply the Nyquist formulation and the higher the resolution required, the greater the amount of data generated. For example, there is equipment such as the SEL-T401L, which is an Ultra High-Speed Line Relay with sampling rates of 1 MHz at 18 bits of resolution that generates around 7 MB per second, which represents around 200 TB per year. Compression techniques based on Fourier, DCT, wavelets use an orthogonal point of the signal and keep the coefficients where the signal energy is concentrated.

Compressed sensing is the mathematical framework that optimizes the number of samples needed, the samples contain all the information of the original signal recovering high-quality signals.

Compressed Sensing (CS) is an alternative technique to Shannon/Nyquist sampling. When the original signal X is sparse on an orthogonal basis Ψ. A stochastic stationary observation matrix Φ can be used to compress the number of samples based on a spatial transformation. The number of samples obtained are smaller compared to the original signal. The values obtained keep the structure of the original signal x is accurately reconstructed by solving the numerical optimization problem.

In compressed sensing, a measurement is a linear functional applied to a signal

$$\langle x, f \rangle \tag{3.6}$$

The compressed sensor makes multiple linear measurements. This can best be represented by the action of a sensing matrix Φ on the signal x given by:

$$y = \Phi x \tag{3.7}$$

where: $\Phi \in \mathbb{C}^{MxN}$ represents M different measurements made on the signal x by the sensing process. Each row of Φ represents one linear measurement.

The vector $y \in \mathbb{C}^M$ is known as measurement vector. $\mathbb{C}^N$ forms the signal space while $\mathbb{C}^M$ forms the measurement space.

It is assumed that the signal x is K-sparse or K-compressible in D and $K \ll N$. The objective is to recover x from y given that Φ and D are known. First recover the sparse representation α from y and then computing $x = D\alpha$

If $M \geq N$ then the problem is a straightforward least squares problem.

The more interesting case is when $K < M \ll N$ *i.e.* the number of measurements is much lower than the dimension of the ambient signal space while more than the sparsity level of signal namely K. Note that when α is determined, finding x is straightforward. The simplified problem given as: recover x from y with

$y = \Phi x$ where $x \in \mathbb{C}^N$ itself is assumed to be K-sparse or K-compressible and $\Phi \in \mathbb{C}^{MxN}$ is the sensing matrix.

A generic sparse acquisition problem could be represented by the following linear matrix expression:

$$y = \Phi x + n \;=\; \Phi D\alpha = (\Phi D)\alpha \tag{3.8}$$

where: x is the $N \times 1$ dimension original signal; Φ is the $M \times N$ dimension ($M \ll N$) observation matrix; n is the observation noise; there are only K ($K \ll N$) non-zero elements in α, which is an N-dimensional K-sparse vector; $\Theta = \Phi\Psi$ is called the sensing matrix. CS recovers the sparse vector S from the observation vector y, thereby accurately reconstructing the original signal $\hat{X} = \Phi\hat{S}$.

To represent a signal with compression detection, it is necessary to obtain the dictionary matrix, D can be constructed by means of different elementary waveforms generated from a variety of basic functions, as the Short-time Fourier Transform (STFT), Wavelet Transform (WT), Discrete Cosine Transform (DCT), Hilbert Transform (HT), Gabor Transform (GT), Wigner Distribution Function (WDF), S Transform (ST), Gabor-Wigner Transform (GWT), Hilbert–Haung Transform (HHT), and hybrid transform based methods.

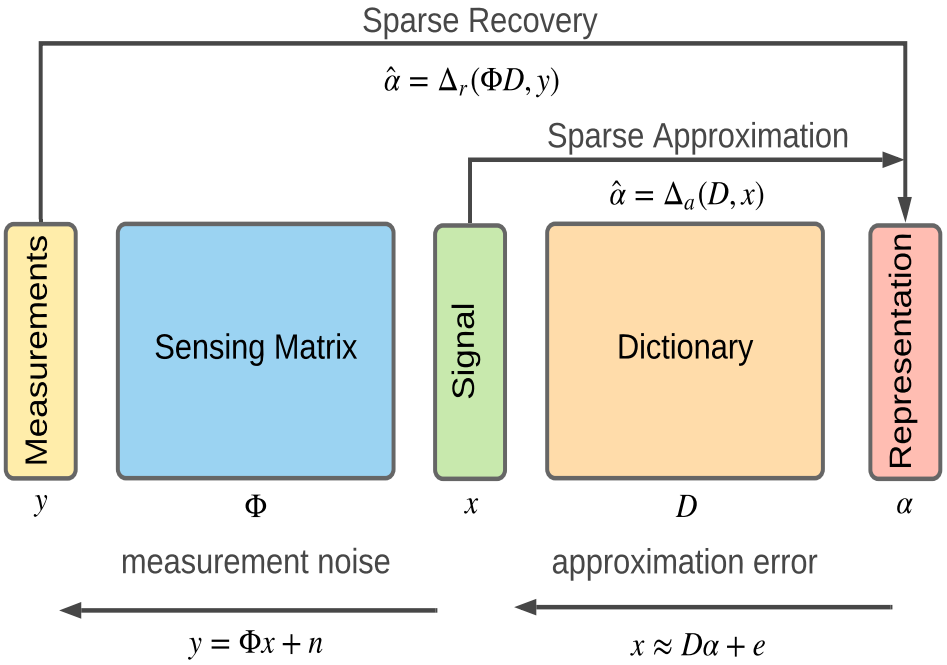


Figure 3.1: Framework of compressed sensing [Ruiz and Montalvo, 2020]

.

If the k-column sub matrices of Φ are badly conditioned, then, it is possible that some sparse signals get mapped to very similar measurement vectors. Thus, it is numerically unstable to recover the signal. Moreover, if noise is present, stability further degrades [Sahoo and Makur, 2015]. The works of Candes and Tao showed

that the geometry of sparse signals should be preserved under the action of a sensing matrix. In particular, the distance between two sparse signals should not change by much during sensing [Candes and Tao, 2006]. They quantified this idea in the form of a restricted isometric constant of a matrix Φ as the smallest number δ_k for which the following holds:

$$(1-\delta_k)||x||_2^2 \leq ||\Phi x||_2^2 \leq (1+\delta_k)||x||_2^2 \; \forall \; x : ||x||_0 \leq k \tag{3.9}$$

When $\delta_k < 1$ then, the inequalities imply that every collection of k columns from Φ is non-singular. Since is needed, every collection of 2k columns is needed to be non-singular, $\delta_{2k} < 1$ is needed, which is the minimum requirement for recovery of k sparse signals.

Moreover, if $\delta_{2k} \ll 1$, then, sensing operator very nearly maintains the l_2 distance between any two k sparse signals. As a consequence, it is possible to invert the sensing process in a reliable way.

It is now known that many randomly generated matrices have excellent RIP (Restricted Isometry Property) behavior. One can show that if $\delta_{2k} \leq 0.1$, then with $M = O(kln^{\alpha}N)$ measurements, one can recover x with high probability.

A simple formulation of the problem is as follows: minimize $||x||_0$ subject to $y = \Phi x$ has infinite solutions, since it entails a combinatorial explosion in the search space.

The algorithms can be broadly classified into the following categories: Greedy pursuits.- These algorithms attempt to build the approximation of the signal iteratively by making locally optimal choices at each step. Examples of such algorithms include Orthogonal Matching Pursuit (OMP), stage-wise OMP, regularized OMP, Compressive SaMpling Pursuit (CoSaMP), and and Iterative Hard Thresholding (IHT). Convex relaxation.- These techniques relax the l_0 norm minimization problem into a suitable problem which is a convex optimization problem. This relaxation is valid for a large class of signals of interest. Once the problem has been formulated as a convex optimization problem, a number of solutions are available, *e.g.* interior point methods, projected gradient methods and iterative thresholding. Combinatorial algorithms.- These methods are based on research in group testing and are specifically suited for situations where highly structured measurements of the signal are taken. This class includes algorithms like Fourier sampling, chaining pursuit, and HHS pursuit.

Minimum Norm Solution

ℓ_0 *norm.*- In the CS context, the solution is explicitly required to be sparse. A natural formulation for the solution recovery might be:

$$\hat{x} = min||x||_0 \qquad s.t \quad ||\Phi\hat{x}||_0 \leq N \tag{3.10}$$

ℓ_1 *norm.*- given a measurement matrix A satisfying the RIP, highly sparse solutions can be obtained by convex optimization. In literature, this algorithm became famous under the denomination of Basis Pursuit (BP) [Donoho, 2006, Narayanan et al., 2018]:

$$\hat{x} \;=\; \min_{x} ||x||_1 \qquad s.t \quad ||\Phi\hat{x} - y||_2 \leq N \tag{3.11}$$

The sparse recovery problem can be interpreted as a convex optimization problem, and can be efficiently solved via linear programming techniques based on the canonical simplex method or the more recent interior point method.

ℓ_2 *norm.*- also known as Least Squares (LS) solution, minimizes the residual energy, thus ensuring denoising and measurement data fitting [Narayanan et al., 2018]:

$$\hat{x} \;=\; \min_{x} ||x||_2 \qquad s.t \quad y = \Phi x \tag{3.12}$$

Matching Pursuit (MP)

A sparse approximation algorithm which finds the "best matching" projections of multidimensional data onto the span of an over-complete dictionary D. The basic idea is to approximately represent a signal f from Hilbert space H as a weighted sum of finitely many functions (called atoms) taken from D [Carta et al., 2018].

Orthogonal Matching Pursuit (OMP)

This is an approximation method that seeks to recover the sparse vector α through the identification of the vector of indices $\jmath$ composed by $i = 1, \ldots, N$ so that the columns ϕ_i of Θ are chosen adequately to minimize the cardinality of the error ε of the signal compressed in the approximation $\Theta\hat{\alpha}$. In OMP, the residue is always orthogonal to the amplitude of the already selected atoms α_i. The found vector $\jmath$ therefore contains the index $\hat{\alpha}$. The optimization problem that is solved is:

$$\min_{\jmath}\{|\jmath| : y = \sum_{i\epsilon j} \phi_i\alpha_i \quad s.t \;\; ||y - \Theta\hat{\alpha}||_2 \;< \varepsilon \;\text{ and }\; \hat{\alpha} \;>\; 0 \tag{3.13}$$

OMP is an ambitious algorithm since it starts by determining the largest atoms of $\hat{\alpha}$; and it is also based on MP. Nevertheless, this work does not expose the characteristics of this algorithm since it was not implemented due to the limitations in the number of iterations [Wang et al., 2016, Li et al., 2018, Sahoo and Makur, 2015].

Basis Pursuit (BP)

It exploits the convex relaxation by replacing $l_0 -$"$norm$" with $l_1 - norm$. For real signals, it can be implemented as a linear programming approach. For complex signals, it can be implemented as a second order cone program.

$$\hat{\alpha} \;=\; arg \;\min_{\alpha\in\mathbb{C}^D} ||\alpha||_1 \; s.t ||x - D\alpha||_2 \;\leq \epsilon \tag{3.14}$$

Efficient solvers are available to solve BP problems using convex optimization techniques. They are usually polynomial time and involve sophisticated algorithms for implementation. Their advantage is that they allow a global and particular

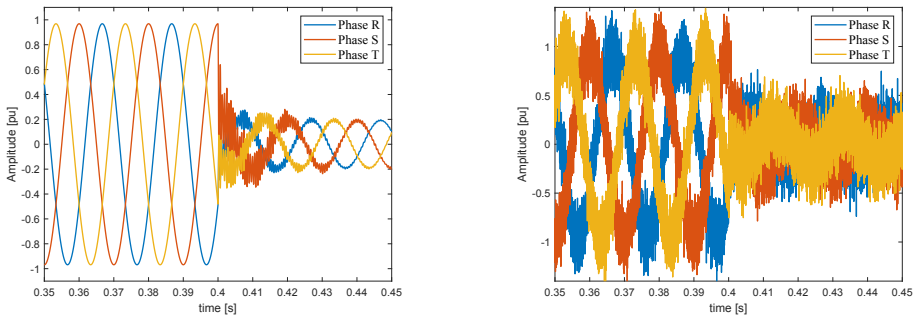
solution, but the disadvantage is that convex optimization methods are yet very expensive computationally.

3.3 Fault Diagnosis

Electrical faults are electromagnetic phenomena that occur in current and voltage signals and are recorded by smart meters or by phasor measurement units. Based on the literature reviewed, it has been shown that fault diagnosis is performed using ideal signals, *i.e.*, signals without noise and with all the samples. In practice, the data acquired by the sensors are affected by errors modelled by noise processes and there is a probability of loss of readings or samples.

The algorithm reads the signals from a database, the first step adds white Gaussian noise to the signal. The second step is the application of filters for noise elimination. Two types of filters have been designed, the first one is the design of a low-pass Finite Impulse Response filter (FIR), based on the frequencies found with the Fourier transform; the cutoff frequency has been placed at 1000 hertz.

Figure 3.2 shows a faulty voltage signal measured from a bus and the same signal signal affected by noise.



(a) Noise-free signal. **(b)** Signal with noise.

Figure 3.2: Fault voltage signals

The second one is a moving average filter, that is a common method used to smooth noisy data. This example uses the filter function to calculate averages over a vector of data. A moving average filter slides a window of length *windowSize* along the data, calculating averages of the data contained in each window. The window size is variable and depends on the number of samples per cycle. The following difference equation defines a moving average filter of a vector x:

$$y(n) = \frac{1}{windowSize}(x(n) + x(n-1) + ... + x(n-(windowSize-1))) \quad (3.15)$$

When a fault occurs, the current and voltage signals are affected by increasing the signal frequencies. The signal of the electric network in steady state has a

frequency of 50 or 60 Hz. When a fault occurs there is evidence of the creation of additional frequencies to the fundamental one as shown in Figure 3.3. Additional frequencies corresponding to the failure are presented.

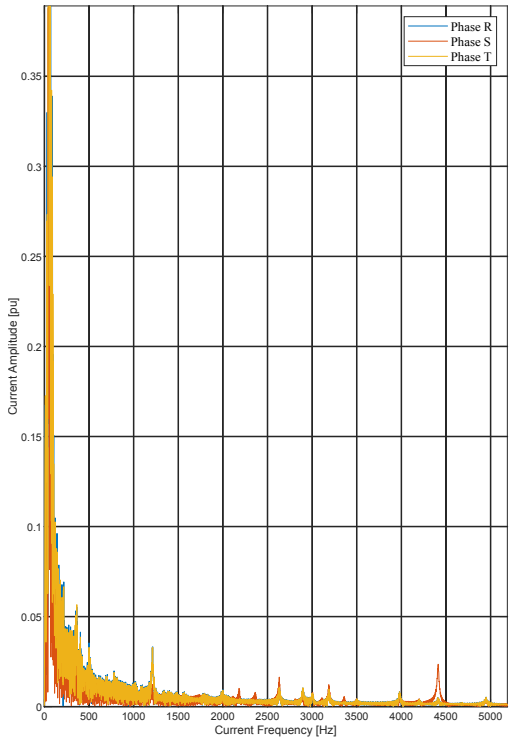


Figure 3.3: Frequency spectrum of signal with noise

Compressed detection is an alternative technique to Nyquist-Shannon sampling, for the reconstruction of a sparse signal x in R^N that can be recovered well with only few components of $M * N$ base matrix. To this end, x should be sparse, *i.e.*, the majority has k nonzero elements where $k \epsilon N$ is used to recover a very low signal from a small number of measurements. There is a high probability to reconstruct a signal with loss of information based on a set of random linear projections using nonlinear reconstruction algorithms. The signal can be either time or frequency domain and the number of random samples needed to recover the original signal is much smaller than the samples of the original signal.

There is a second problem related to the loss of information, with samples of the zero amplitude signal, *i.e.*, there may be lost data and the signals become scattered. In order to solve the data loss, the creation of a dictionary is proposed. The dictionary contains the most important information of all the signals either in steady state or in faulty conditions. The design of the dictionary is carried out by means of optimization techniques to calculate the number of volumes that compose the dictionary. Different optimization techniques were used, and the difference is

based on the optimization philosophy. The techniques for finding a dictionary and subsequent reconstruction of the signals are Least Squares, Matching Pursuit and Orthogonal Matching Pursuit. Step 1 is the signal reading. In step 2, the signal is filtered by applying the FIR filter and moving average removing the signal noise. Step 3 is to generate an orthogonal basis of the signals, the orthogonal basis is generated by applying the Discrete Cosine Transform (DCT). The step 4 performs the calculation of the number of samples needed to recover the signal by applying the techniques of Least Squares, Matching Pursuit or Orthogonal Matching Pursuit. In step 5, once the optimal number of atoms is obtained, the signal recovery is performed. The recovery of the signal starting from a signal with loss of information and dispersed.

The number of samples per cycle is stored in the variable N. Then the Psi matrix is created by calculating the discrete cosine transform creating an N-by-N matrix. The sensing matrix called as Phi is formed from the inverse of the Psi matrix. The DCT of the signal is calculated by multiplying the transpose of Psi by the signal. To determine the number of samples needed to reconstruct the original signal. k is calculated which is the number of atoms needed from the original signal, the value of k is calculated by progressively increasing the number of atoms, as presented in the FOR loop. Then the optimization is searched using for example Matching Pursuit, etc. Solver is the result of the optimization and is applied to the original signal. Finally, the error in the signal recovery is calculated.

Algorithm 5: Signal Recovery

```
Step 1: Number of samples per cycle.
        N = ns;
Step 2: Discrete cosine transform
        Psi = dctmtx(N)';
Step 3: Sensing matrix
        Phi = inv(Psi);
Step 4: Discrete cosine transform
        s_dct = Psi' * signal;
Step 5: Signal Reconstruction
        for k = 1 : 10 : N
            solver = MatchingPursuit(Phi, k);
            solution = solver.solve(signal);
            recoveryerror = (norm(mp_diff)/norm(x)) * 100;
        end
```

Once the optimum value of k has been obtained, the number of samples needed to recover the original signal is calculated. The second step is to create sparse signal of the original signal with random samples. Finally, by applying the optimization algorithms, the number of minimum samples of the signals as a function of the error is calculated.

Algorithm 6 shows the steps to perform an ultra-fast detection of electrical faults. Step 1: Once the distribution network is designed, the network simulation

is performed in CIMDYST. Once the network is designed, the electrical faults are generated and stored in a database. The first step is the acquisition of the data to be analyzed, the signals are 6, 3 represent the voltages and 3 represent the currents in each bus, according to the voltage and current values of time. Based on experimentation, it has been verified that at least four samples of the current or voltage signal are necessary, and, the Haar level 1 wavelet transform is applied to the samples, identifying an abrupt change in the signal caused by a fault. In step 3, the wavelet coefficients are extracted. Application of the level 1 wavelet transform compresses the signal to half the original signal data. In step 4, the data of the coefficients are interpolated until the original signal dimension is obtained. Finally, in step 5, the maximum value of the coefficients is determined, and the indices of the maximum value indicate the time of occurrence of the failure.

Algorithm 6: Fault detection

Step 1: Acquire data from .csv file
    $OS = load('filename.csv')$
Step 2: Wavelet coefficients calculations
    $for\ i = 1 : length(OS)$
    $[c, l] = wavedec(OS(1:4), level, wavelet\_type)$
Step 3: Extract detail coefficients
    $d1 = detcoef(c, l, level)$;
Step 4: Interpolation using FFT method
    $FD = interpft(d1, 2 * length(d1))$;
Step 5: Finding of Maximum values and comparison with established limits
    $MV = max(abs(FD))$
    $if\ MV > threshold$
    $TF = i + 1$
    $AF\ MV$
    end
    end

The installation of a large number of sensors with very high sampling frequencies makes it possible to increase the reliability of electrical systems. One of the challenges for researchers is to work with large amounts of information. In this section, the algorithms proposed for data compression of electrical signals are presented. There are three parameters to measure the fidelity of the reconstructed signals with respect to the original: Retained Energy Percentage (RTE), Normalized Mean Square Error (NMSE) and Correlation (COR).

Algorithm 7 represents the initial process of signal processing, once the noise has been removed and the signal has been reconstructed, in case of missing samples, signal processing proceeds at each stage. First, the data are acquired and stored in OS representing the Original Signal. The number of signal samples is stored in ROS (Row Original Signal) and COS (Column Original Signal) stores the number of signals. To calculate the number of samples per cycle, the zero crossing of

the signals is calculated and it is determined how many samples per cycle each electrical current or voltage signal has and the results are stored as indices in A and B. The difference between the calculated indices A and B represents the number of samples of each cycle, so it is necessary to start the algorithm when the system is in steady state allowing to find important characteristics of the signals such as the maxima that represent the peaks, these data are stored in C and D. Finally, the number of samples per cycle of the signal is stored in E and is the size of the signal between the indices A and B. The described process of algorithm 7 must be performed by all 3 phases.

Algorithm 7: Signal characteristics extraction

Step 1: Acquire data from .csv file
$OS = load('filename.csv')$
Step 2: Zero crossing
$[ROS, COS] = size(OS)$
$for\ i = 1 : COS$
$A(i) = OS(:, i) < 0\ and\ OS(: +1, i) > 0$
$B(i) = OS(:, i) > 0\ and\ OS(: +1, i) < 0$
$endfor$
Step 3: Amplitude and coordinates extraction in steady state
$for\ i = 1 : COS$
$[C(i), D(i)] = max(OS(A(i) : B(i), i))$
$endfor$
Step 4: Calculation of number of samples per cycle
$for\ i = 1 : COS$
$E(i) = length(OS(A(i) : B(i), i))$
$endfor$
Return: $A,\ B,\ C,\ D,\ E$

Smart grids are made up of a large number of sensor and actuator devices with real-time communications, which implies an enormous generation of data. Currently, there is equipment capable of sampling at 1 Mhz, generating TB of information annually. A solution to the problem of information transmission and storage is presented in Algorithm 8, where the proposed methodology for data compression using wavelets. The first step consists of taking the filtered and corrected signal in case of missing data and applying the bior1 wavelet transform level 6. The application of the transform compresses the original signal as a function of the level, but affects certain characteristics of the signal such as amplitude variation, ripple addition and phase shift when compared to the original signal. The second step of the proposed algorithm is to calculate the zero crossing allowing to determine the number of samples per cycle. To correct the amplitude, step 3 is proposed, which finds the maximum values of each signal in current or voltage of the steady-state signals, taking as reference the first cycle of the signal. The obtained amplitude value allows normalizing the amplitude of the reconstructed

signal. The new signal is stored in SCN (Signal Compressed Normalized). The ripple elimination is performed using the moving average filter, the parameters of the moving average filter is to determine the number of samples of the signal to be taken by the filter, the value is stored in the variable N, according to the values of N are equal to 0.1% of the number of samples of the compressed signal. The result is stored in SCNA (Compressed Normalized Average Signal). Finally, the signal displacement is corrected by calculating the difference in time of the maxima between the original and the reconstructed signal. The calculation of the time difference allows to shift the reconstructed signal by correcting the deviation. The result is stored in SCNAS matrix (Compressed Normalized Average Shifting Signal).

Algorithm 8: Wavelet compression of an electrical signal

Step 1: Wavelet Compression Signal (SC)
  $for\ i = 1 : COS$
    $[c, l] = wavedec(OS, level, wavelet\_type)$
    $SC(:, i) = appcoef(c, l, wavelet\_type)$
  $endfor$
Step 2: Zero crossing (compressed signal)
  $[RSC, CSC] = size(SC)$
  $for\ i = 1 : CSC$
    $F(i) = SC(:, i) < 0\ and\ SC(: +1, i) > 0$
    $G(i) = SC(:, i) > 0\ and\ SC(: +1, i) < 0$
  $endfor$
Step 4: Normalization, moving average, shifting correction (SCNAS)
  $for\ i = 1 : CSC$

$SCN(:, i) = SC(:, i). * (max(OS(A(i) : B(i), i))/(max(SC(F(i) : G(i), i))$

    $SCNA(:, i) = 1/N \sum_{i=1}^{N} SCN(1 : N, i)$
    $[H, I] = max(SCNA(F(i) : G(i), i))$
    $DT = I - D(i)$
    $SCNAS(:, i) = padarray(SCNA(:, i), DT, 0, 'pre');$
  $endfor$
Return: $SCNAS$

Algorithm 9 describes the proposed steps to eliminate repeated signals by allowing the data to be compressed. The first step determines the number of samples and the number of signals, the values are stored in J and K respectively. OSP (Original Signal Period) represents one cycle of the original signal, SCP (Signal Compressed Period) represents one cycle of the compressed wavelet signal. Each cycle of the original signal is compared one by one with all the cycles of the compressed signal. If the RTE is high, between 0.9999 and 1.0009 when comparing the two signals, it indicates that both signals are identical and then zeros are placed in that signal cycle in the SCW matrix. This process creates a vector, which rep-

resents an index that will later allow the reconstruction. Conversely, if the RTE value is below 0.9999, the signal values are placed for that cycle in the SCW (Signal Compressed Window) matrix. By doing this process, equal signals (high RTE) are eliminated, allowing a higher compression of the signals.

Algorithm 9: Windowing for elimination of repeated signals

```
1: cont = 1
2: [J, K] = size(SCNAS)
3: for i = 1 : K
4:     for ii = 1 : floor(J/E(i))
5:         OSP = OS(A(i) : B(i), i)
6:         SCP = SCNAS(A(i) + (ii − 1) ∗ E(i) : B(i) + (ii ∗ E(i), i)
7:         RTE = ∑(OSP)^2 / ∑(SCP)^2
8:         if RTE > 0.9999 && RTE < 1.00009
9:             Ind(ii, i) = cont
10:            SCW(A(i) + (ii − 1) ∗ E(i) : B(i) + (ii ∗ E(i), i) = 0
11:            cont = cont + 1
12:        else
13:            Ind(ii, i) = cont
14:            SCW(A(i) + (ii − 1) ∗ E(i) : B(i) + (ii ∗ E(i), i) = SCP
15:            cont = cont + 1
16:        endif
17:    endfor
18: endfor
19: Return: SCW, Ind
```

Algorithm 10 presents the signal reconstruction. The first step is to calculate the number of signals and the number of samples per signal, these data are stored in the variables L and M. A FOR loop from 1 to M is created, allowing the reconstruction of all signals. The second FOR loop from 1 to $E(i)$ indicates the actual size of each signal and up to which point the signals should be reconstructed. With the third FOR loop from 1 to the maximum number of indices per signal, the indices represent the number of repeated periodic signals and the order of repetition. The variable ni stores the value of the index. The variable SC stores the cycles of the SCW signal. If the value of the variable iii is equal to the index ni, the signal cycle stored in SC is placed. In other words, each cycle of the SCW signal is revised and the original signal is reconstructed based on the indices. Finally, the RTE, Energy Recovery, NMSE, and COR are calculated to verify the relationship indices between the original signal and the reconstructed one.

Algorithm 10: Reconstruction

$[L, M] = size(SCW)$
$for\ i = 1 : M$
  $for\ ii = 1 : (ROS/E(i))$
    $for\ iii = 1 : max(Ind(:, i))$
      $ni = Ind(ii, i)$
      $SC = SCW(E(i) * ni, i)$
      $if\ iii == ni$
        $SR(E(i) * ni, i)) = SC$
      end
    end
  end
  $$RTE(:, i) = \frac{\sum_{n=0}^{ROS} OS[n]^2}{\sum_{n=0}^{ROS} SR[n]^2}$$
  $$Energy\ Recovery(\%)(:, i) = \frac{100 \cdot ||SR||}{||OS||}$$
  $$NMSE(:, i) = \frac{||OS - SR||^2}{||OS||^2}$$
  $$COR(:, i) = \frac{OS^T \cdot SR}{OS^T \cdot OS}$$
end
Return: $RTE, Energy\ Recovery, NMSE, COR$

Chapter 4

Simulations and Results

This chapter shows the simulations of the different proposed algorithms and the results obtained for both power network planning and fault diagnosis. First of all, results obtained from the deployment of electrical distribution networks using georeferenced coordinates are presented. In addition, to evaluate the results obtained, simulations of the network are carried out to verify that the results are within the limits regulated by the standards. Next, the experiments related to the diagnosis of electrical faults based on sparse signal processing techniques are presented, allowing the detection and classification of faults even if there is loss of formation. Finally, results on the compression of data related to electrical measurements are presented.

4.1 Planning of Electrical Distribution Networks

The planning algorithms require georeferenced coordinates of the location of streets and buildings. The results of the Borgo Punta neighborhood, located in the city of Ferrara - Italy, are presented below. As shown in Figure 4.1, the georeferenced coordinates are 11.6352 W 44.8391 S and 11.6504 E 44.8485 N. The georeferenced data of buildings and streets have been obtained from OSM, which is a free, collaborative platform that allows map editing. The coordinates corresponding to latitude and longitude are stored in an OSM file extension.

The scenario accounts for approximately 629 buildings over an area of 1,182 square miles. The buildings have been divided into five clusters. Each cluster represents an electricity consumption stratum, *i.e.*, there are different consumption profiles randomly assigned to users depending on location. Figure 4.1 shows the Delaunay-Voronoi tessellation. The model developed assigns different power levels: the orange region is zone 1, the power consumption of the users is 4 kW $\pm$ 20%; the blue region is zone 2, the users' consumption powers are 5 kW $\pm$ 20%; the red region is zone 3 the users' consumption powers are 6 kW $\pm$ 20%; the red region is zone 3, the users' consumption powers are 6 kW; the green region is zone 4 the users' consumption powers are 7 kW $\pm$ 20%; and the purple region is zone 5, the users' consumption powers are 8 kW $\pm$ 20%.

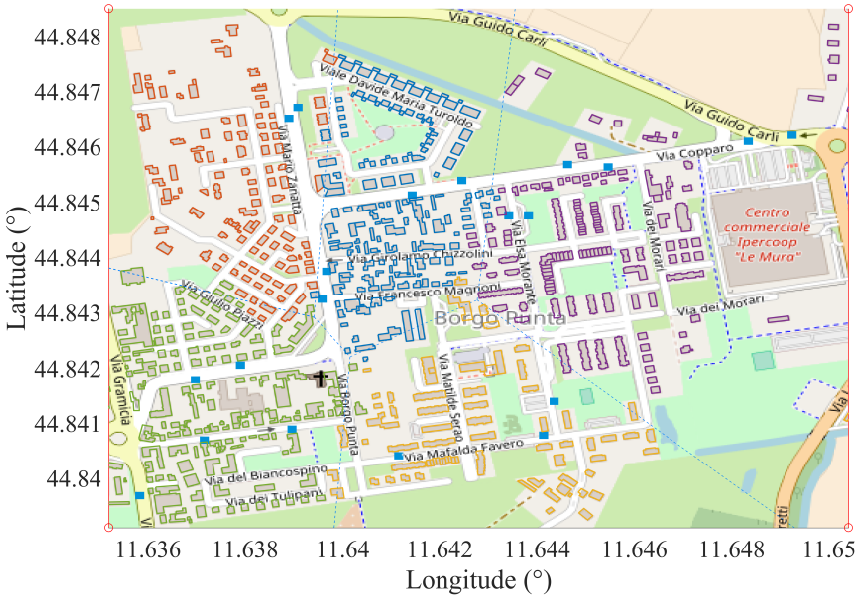


Figure 4.1: Scenario for deployment of underground electrical distribution grids.

On the other hand, each building has 4 coordinates in georeferenced space, corresponding to latitude and longitude. The algorithm distributes the users between transformers and inspection wells. The distances are calculated using the Haversine formula. Next, the Prim algorithm is used to generate the Minimum Spanning Tree (MST) of a medium voltage subway radial network deployed along the streets, ensuring a minimum total length of branches. This algorithm considers the active nodes and identifies the shortest route between a user and the nearest transformer, while respecting the voltage and power ratings of the transformers. In addition, the voltage drop on the lines is required to be less than 3%.

Table 4.1, shows the results obtained after applying the optimization methodology in the scenario of 100 meters of spacing and 150 KVA of more power, with data from the application of Algorithm 2 with which the powers are defined together with the number of transformation units required by the scenario.

Table 4.1: Scenario with spacing parameters 100 meters and power 150 KVA.

Parameter	Value
Users	629
Active sites	27
Coverage[%]	100
MV Network Length [km]	12.5
Capacity [KVA]/ Quantity	5/5; 10/2; 25/9; 37.5/3; 50/2; 75/6

Figure 4.2 shows the first scenario after optimization, the spacing distance between transformer units is 100 meters and their maximum power is 150 KVA.

For this case, the model requires 27 active sites.

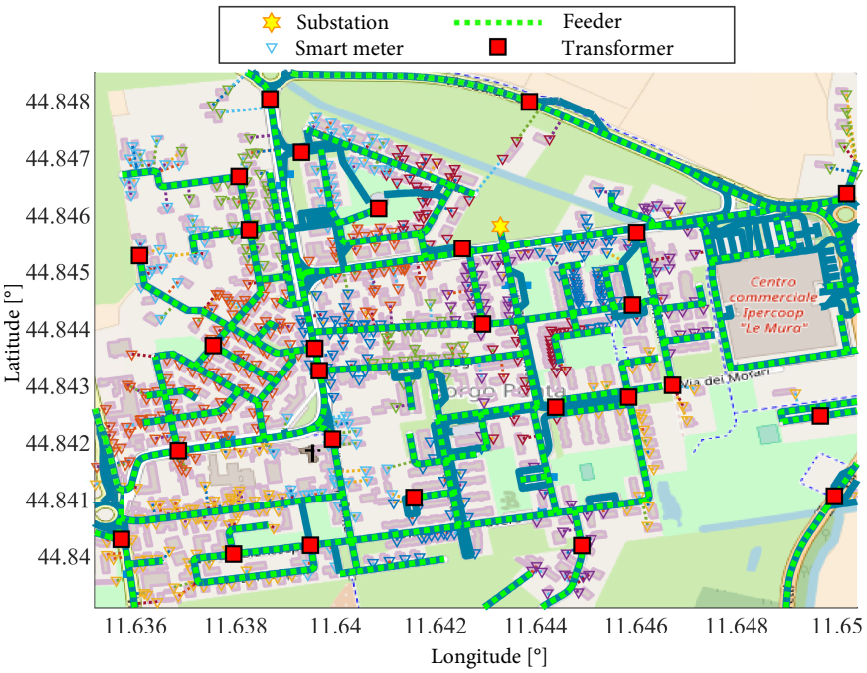


Figure 4.2: Optimized distribution network in the simulation scenario: spacing between transformer stations 100 meters and maximum transformer capacity of 150 KVA.

Table 4.2 shows the results in this scenario, and similarly, the powers together with the number of transformers required. It can be seen how the optimization leads to decrease the length of the medium voltage network, which is beneficial for the distribution system.

Table 4.2: Scenario with spacing parameters 80 meters and power 350 KVA.

Parameter	Value
Users	629
Active sites	18
Coverage [%]	100
MV Network Length [km]	8.37
Capacity [KVA]/ Quantity	5/4; 10/3; 25/3; 50/3; 75/3; 100/1; 125/1

Figure 4.3 shows the optimized distribution network in the second simulation scenario, which considered a spacing of 80 meters and a maximum power of the 350 KVA transformers. For this case, the model requires 18 active sites.

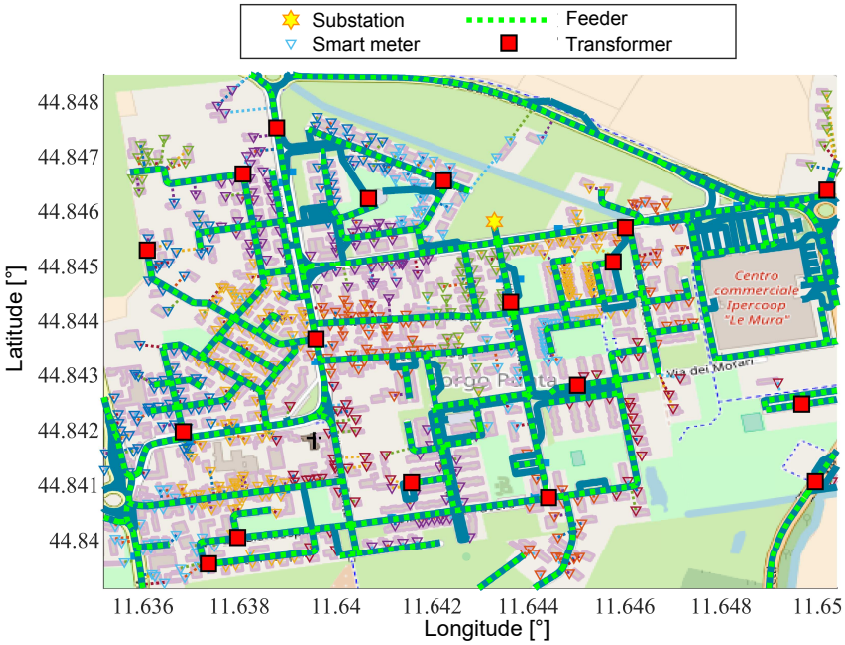


Figure 4.3: Optimized distribution network in the simulation scenario: spacing between transformer stations 80 meters and maximum transformer capacity of 350 KVA.

It can be seen that the spacing distances of the transformation centers have a well-know effect on the time required to reach the optimum response, maintaining a proportional relationship; the greater the distance, the longer the execution time and vice versa, and the same happens with the powers; as shown in Figure 4.4, the higher the capacity of the assigned transformation units, the longer the time required for optimization. Finally, the computer equipment used in the development of the experiment has the following characteristics: Processor Intel (R) Xeon (R) E-2176M CPU @ 2.70GHz and 64 GB RAM.

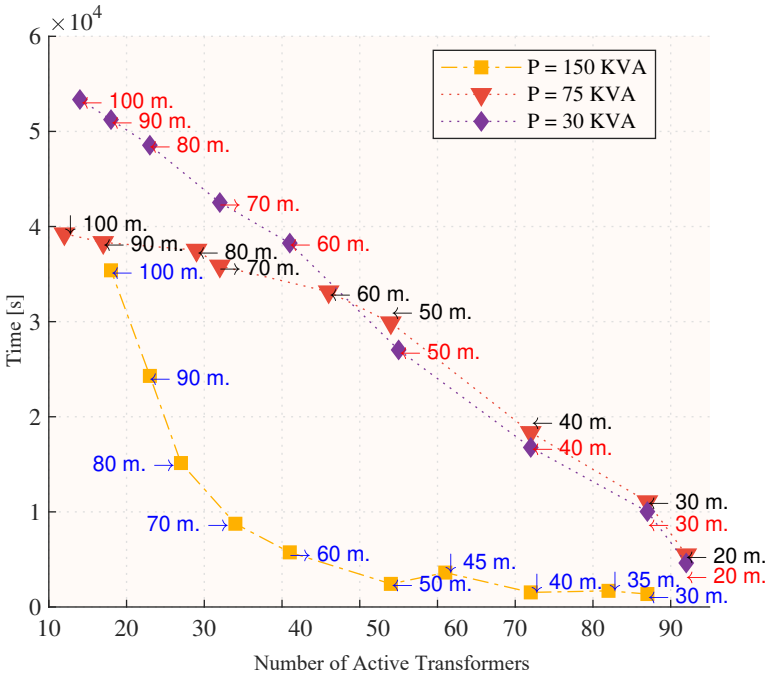


Figure 4.4: Model execution time as a function of the number of active transformer stations for the various simulation scenarios.

AC models are realistic representations of energy flow in power lines. Due to their nonlinearity, solving the power flow problem is a computationally expensive task. In this paper, a standard AC power flow model is used, which decomposes the network to improve the performance when solving large-scale power flow problems. To this end, a power system is divided into parts in which the problem is estimated independently to reduce the execution time. In particular, the power flow model implemented in CYMDIST utilizes the algorithm unbalanced voltage drops with a 3% tolerance to solve the AC power flow problem. Figure 4.5 shows the single-line circuit of the power distribution system and the power flow.

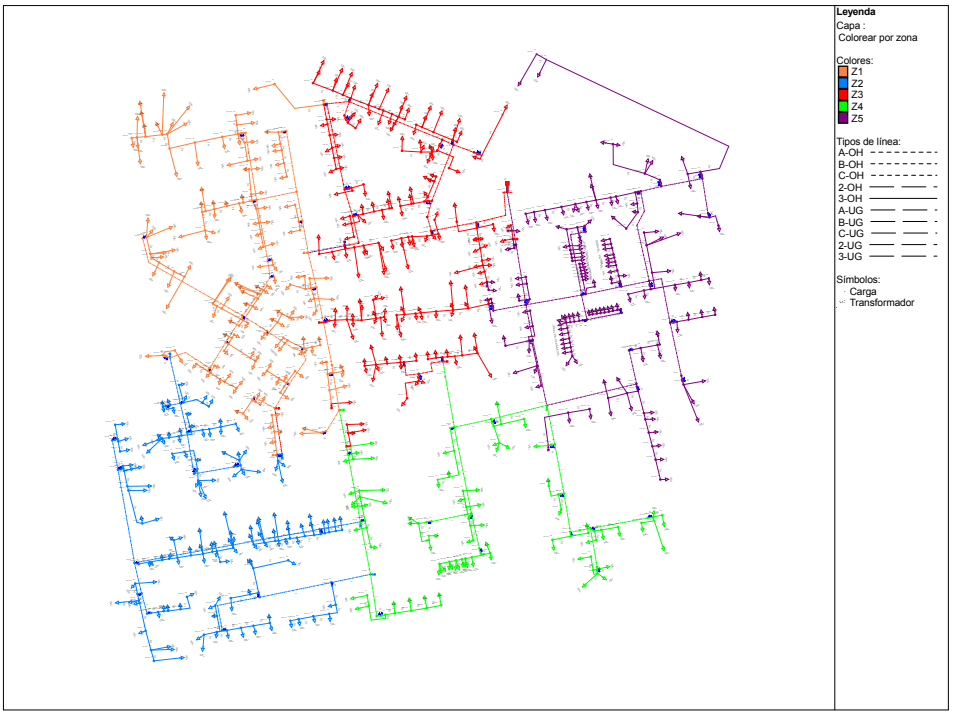


Figure 4.5: Power flow model

The information on the most representative feeders in each zone is described. The length of the main medium voltage feeder in zone 1 is 1.23 km; the secondary low voltage feeder has a length of 147 m and feeds 21 users, with a maximum voltage drop of 4.81 V on phase A. The length of the main medium voltage feeder in zone 2 is 1.53 km; the secondary low voltage feeder has a length of 152m and feeds 20 users, with a maximum voltage drop of 9.3 V in phase A. The length of the main medium voltage feeder in zone 3 is 0.72 km; the secondary low voltage feeder has a length of 178m and feeds 16 users, with a maximum voltage drop of 7.9 V on phase A. The length of the main medium voltage feeder in zone 4 is 1.22 km; the secondary low voltage feeder is 72m long and feeds 13 users, with a maximum voltage drop of 9.7 V on phase A. The length of the main medium voltage feeder in zone 5 is 0.52 km; the secondary low voltage feeder has a length of 80 m and feeds 15 users, with a maximum voltage drop of 6.8 V on phase A. Figure 4.6 shows the phase voltage profiles in the five zones considered in the simulations.

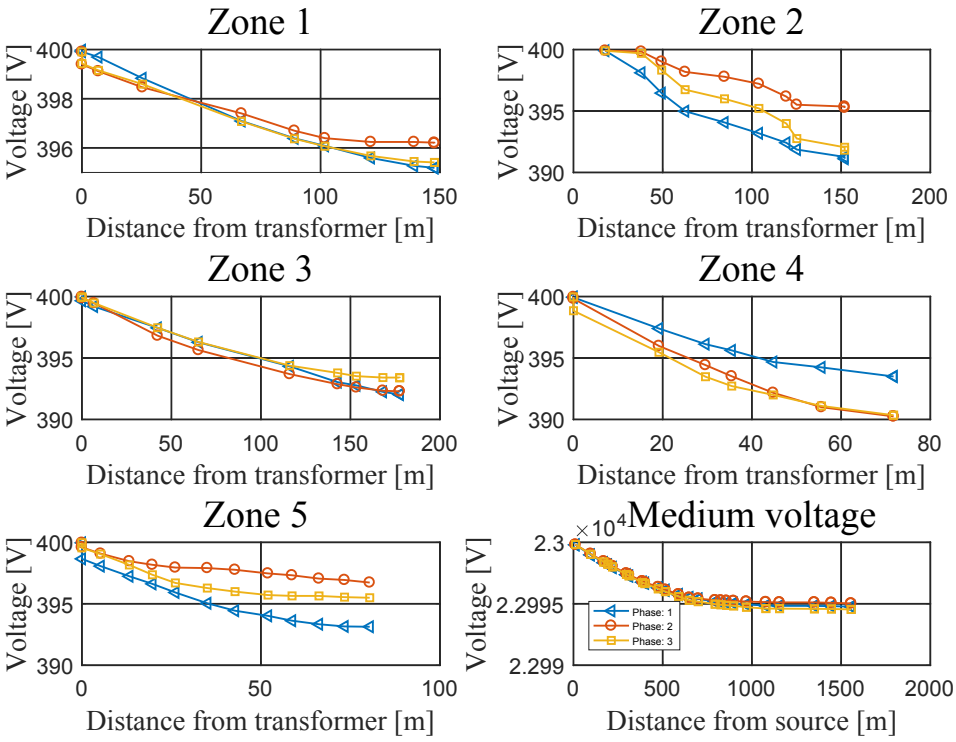


Figure 4.6: Voltage profiles maximum lengths: a) Zone 1: 21 users with 148 m, b) Zone 2: 20 users with 154 m, c) Zone 3: 26 users with 186 m, d) Zone 4: 13 users with 76 m, e) Zone 5: 15 users with 80 m.

In the final analysis, none of the feeders should exceed the maximum length from the low voltage side of the transformer to the user's meter. The medium voltage network should be deployed to connect as many transformers as required to cover 100% of the users. The proposed model uses N transformers to provide coverage to M users with loads ranging from 4 KW to 8 KW. It is important to note that the power transmitted by the low voltage network is the sum of the power consumed from the transformers by the downstream users. The installed load is 3810 KVA, distributed in 1272, 1268 and 1270 for phases A, B and C, respectively. The average rated voltage is 23000 VLL and 13279 VLN, and the rated currents are 97.6 A, 94.8 A and 94.5 A for phases A, B and C, respectively. Based on the results obtained with the power flow analysis in the distribution network, it can be concluded that there is a load balance between phases A, B and C, respectively.

Table 4.3, presents a summary of the power flow analysis, including the power source, the installed load in each zone, the conductor ampacity and losses in lines, conductors and transformers.

Table 4.3: Summary of the power flow in the power distribution grid.

Total summary	**Active Power [kW]**	**Reactive Power [kVAR]**	**FP[%]**
Sources (Balance power)	3804,47	206,03	99,85
Total production	3804,47	206,03	99,85
Load Zone 1	629,98	87,65	99,05
Load Zone 2	634,97	14,57	99,97
Load Zone 3	930,95	14,42	99,99
Load Zone 4	447,99	-0,04	100
Load Zone 5	1092,93	81,4	99,72
Total loads	3736,82	198	99,86
Cable capacitance	0	158,08	0
Line capacitance	0	0	0
Total shunt capacitance	0	158.08	0
Losses in lines	30,93	24,69	78,16
Losses in cables	1,7	1,33	78,77
Losses in transformers	35,02	140,09	24,25
Total losses	67,65	166,1	37,72

4.2 Electrical faults signals restoring based on compressed sensing techniques

4.2.1 Basis Pursuit Results

Figure 4.7 shows the number of samples required to reconstruct the 90% signal with respect to the original one. The scenario with the lowest number requires 441 random samples of the original signal (Figure 4.7 (c)) and the time required for the solution of the algorithm is approximately 10.71 seconds, which is also the lowest time compared to the other samples analyzed in the other literals of the same figure.

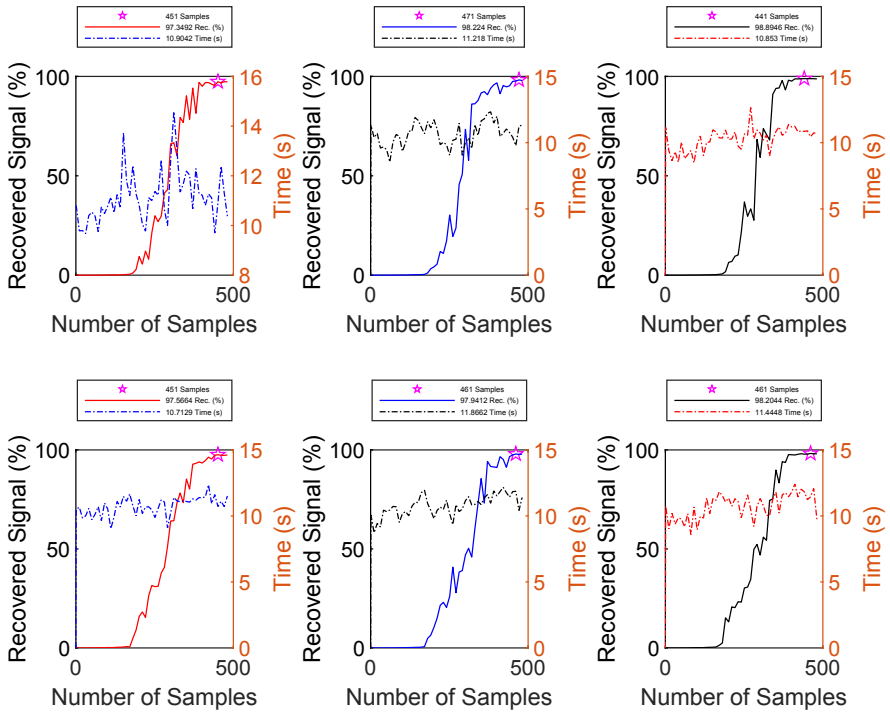


Figure 4.7: Number of samples m needed to recover the signal vs processing time.

Figure 4.8 shows the reconstruction of the signal at a different point of the system based on random samples from 50% to 100%, so the value of k would be 431. It is important to mention that, the signals are obtained from 60% of all the information, where the orange signals eliminate the noise produced when the samples are diffused.

4.2.2 Matching Pursuit Results

To solve the signal restoration problem using the matching pursuit model, it is first necessary to establish the number of atoms in the signal known as k. The atoms in the signal are the minimum number of samples that represent the complete

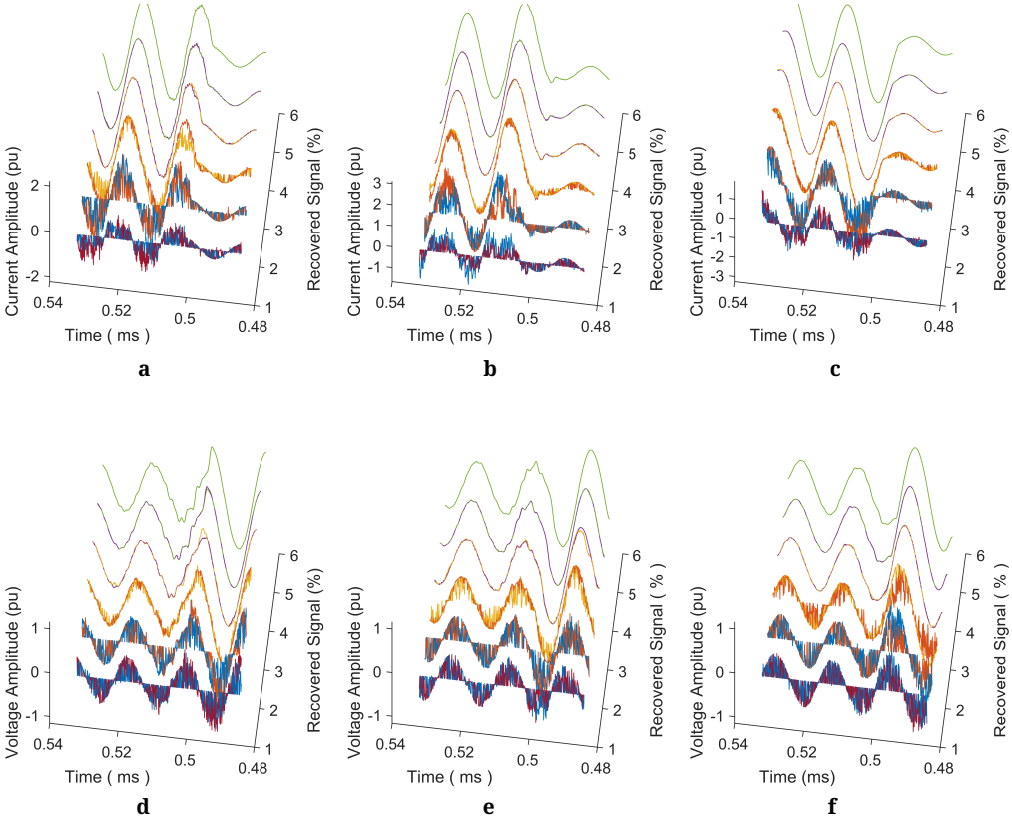


Figure 4.8: Signal reconstruction using m values vs processing time.

signal. Figure 4.9 shows the optimum number of k "signal atoms" of the signal under fault conditions. Therefore, the number of random samples needed is 171 for a noise-free fault signal. As can be seen, the running time of the algorithm to find the optimal value of k is less than 0.2 seconds.

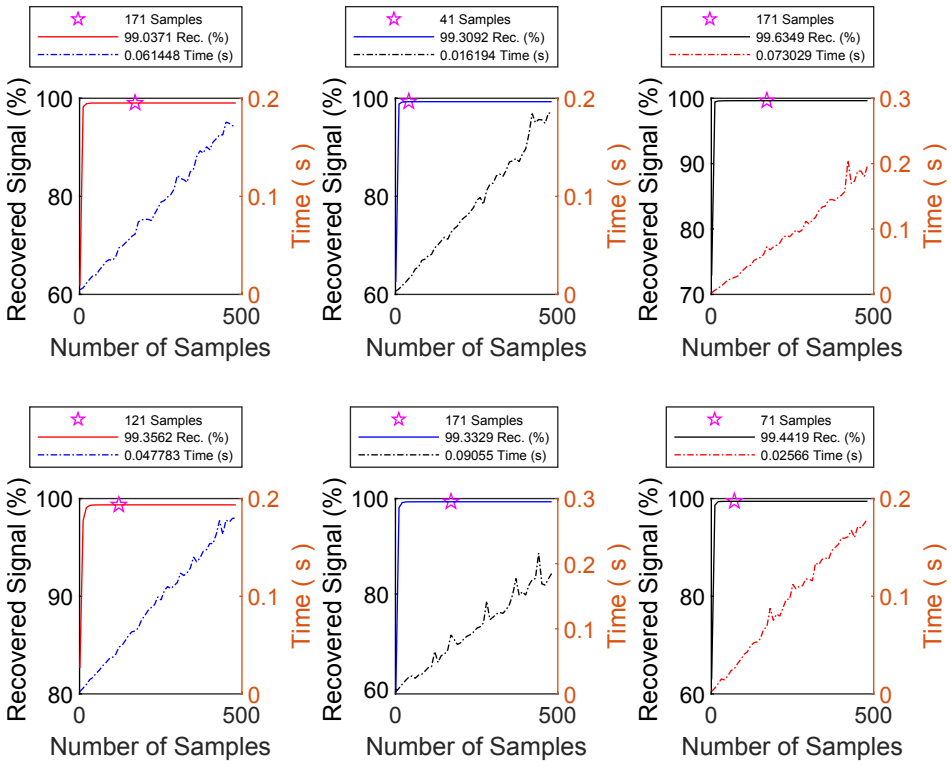


Figure 4.9: Calculation of the number of k atoms vs. processing time.

On the other hand, the number of samples needed to reconstruct the signal has been calculated. Figure 4.10 shows the number of samples needed to reconstruct the 90% signal as a function of the original one. The scenario with the highest number is shown in Figure 4.10 (a). It requires 391 random samples of the original signal which is equivalent to 80% of the original signal data and the machine time required to process the algorithm is 0.1726 seconds.

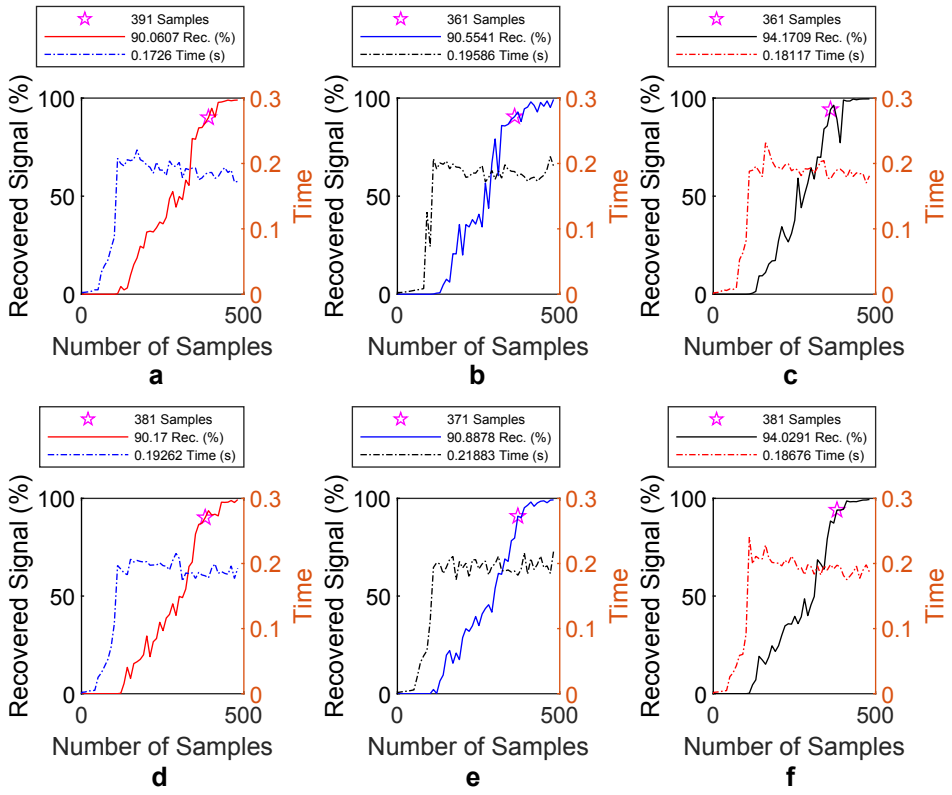


Figure 4.10: Number of m samples necessary to recover the original signal vs. processing time.

Finally, once the number of samples has been calculated, the optimum values for reconstructing the signals can be obtained. In this case, $k = 171$ which corresponds to 35% of the base signal and $m = 391$ corresponding to 80% of the base signal. Moreover, the reconstruction tests of the signals produced by the electrical faults in bus 14 are developed.

Figure 4.11 shows the reconstruction of the signal at a different point in the system based on random samples from 50% to 100%, so the value of k would be 171. It is important to mention that the signals are obtained from the 80% of all the information, where the orange signals eliminate the noise that is produced when the samples are extended. The computation time to run this algorithm to find the optimal k value is less than 0.2 seconds.

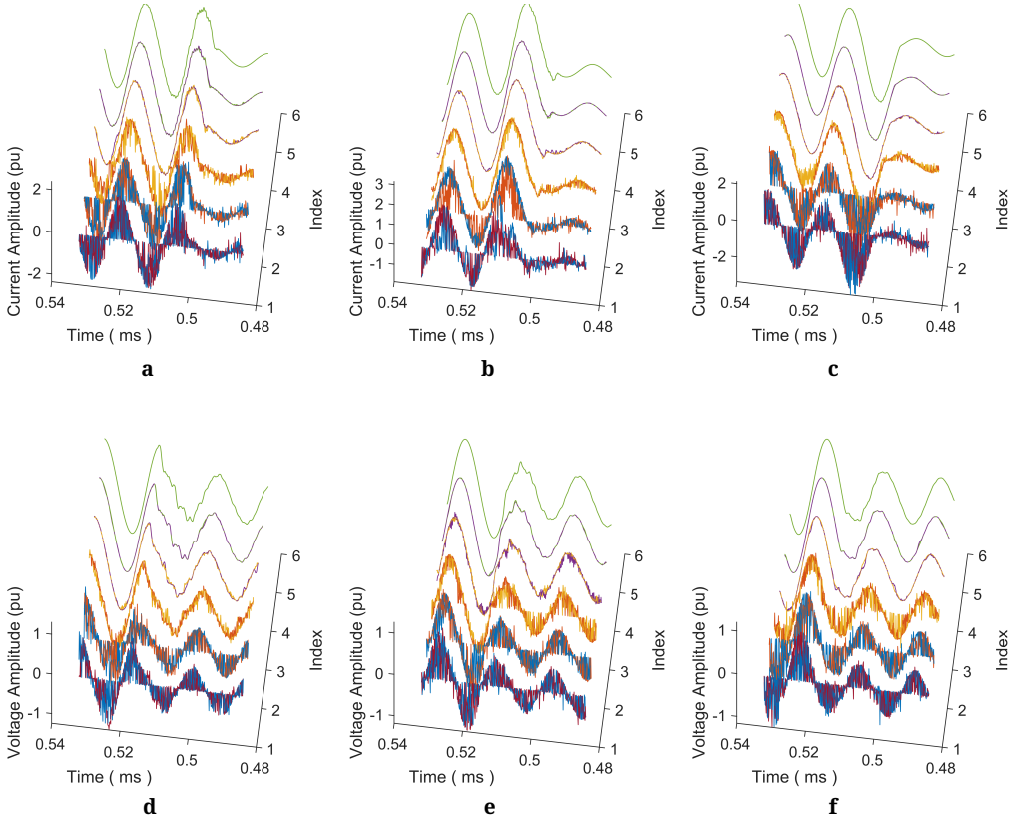


Figure 4.11: Signal reconstruction using k and m values vs. processing time.

4.2.3 Orthogonal Matching Pursuit Results

To solve the signal restoration problem using the orthogonal matching pursuit model, it is necessary to establish the number of signal atoms known as k. The signal atoms are the signal samples that represent the complete signal. Figure 4.12 shows the optimal number of k atoms of the fault signal, it is shown that the number of random samples needed by the atoms is 431 (Figure 4.12 (d)) for a signal under noise-free fault conditions. The computation time of the algorithm is less than 1.71 seconds.

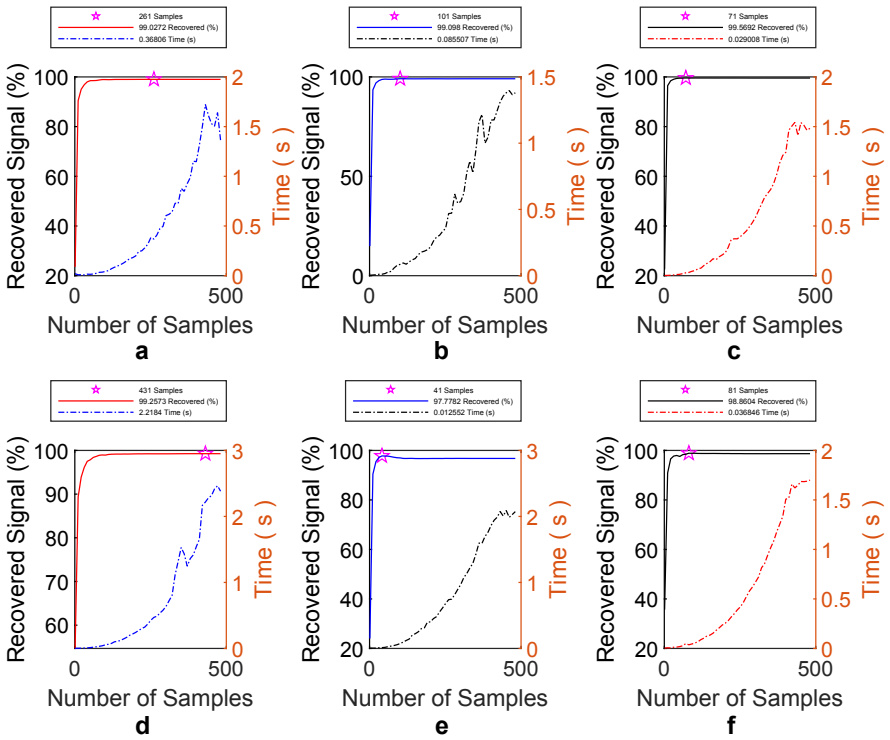


Figure 4.12: Calculation of the number of k atoms vs. processing time.

Figure 4.13 shows the number of samples needed to reconstruct the 90% signal with respect to the original one. The scenario with the highest number requires 341 random samples of the original signal (Figure 4.13 (a) and (e)), however the signal in Figure 4.13 (e) requires a shorter time for the solution of the algorithm, which is 0.182 seconds.

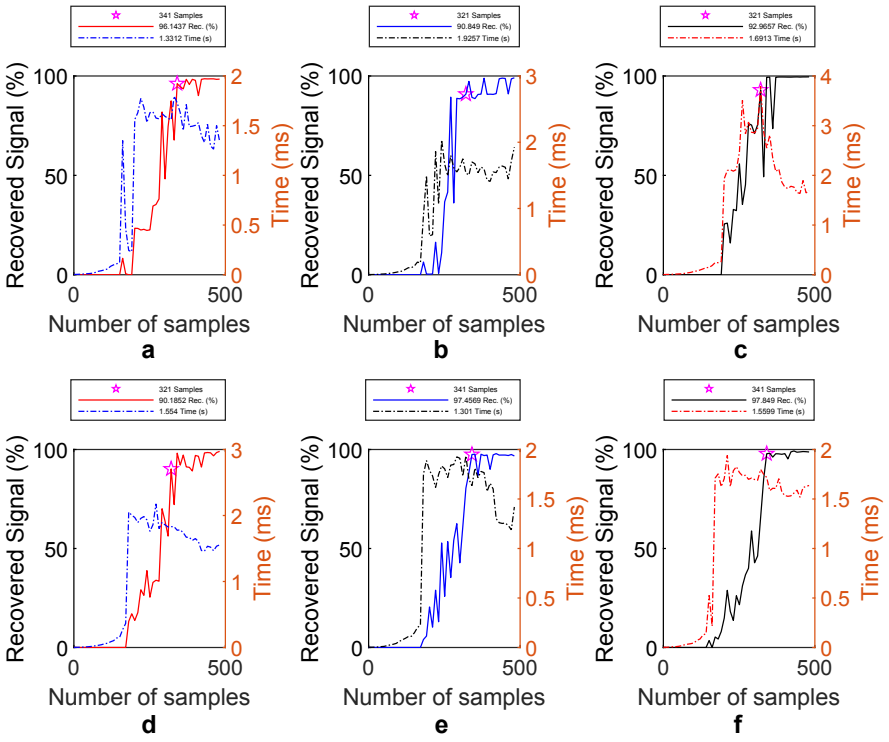


Figure 4.13: Number of m samples needed to recover the signal vs processing time.

Figure 4.14 shows the reconstruction of the signal at a different point of the system based on random samples from 50% to 100%, so the value of k would be 431, as described above. It is important to mention that, the reconstructed signals are obtained from 60% of all the information, where the orange signals eliminate the noise produced when the samples are scattered.

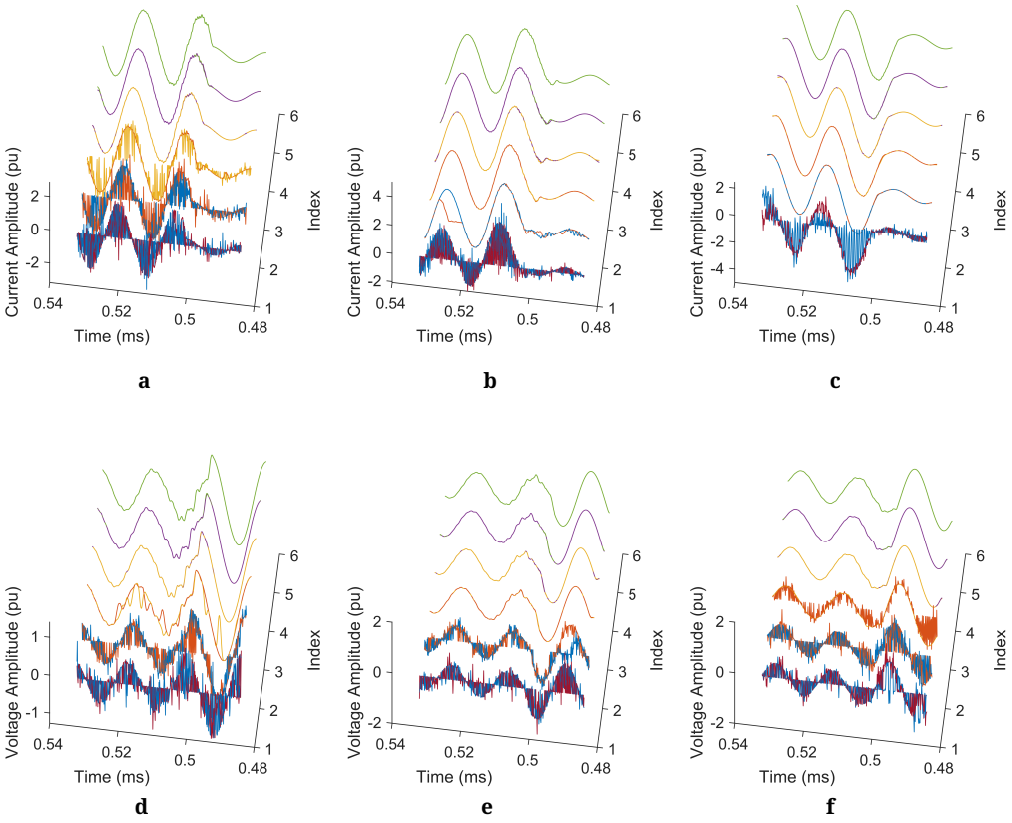


Figure 4.14: Signal reconstruction using k and m values vs. processing time.

The following table presents the summary of the different basis pursuit, matching pursuit and orthogonal matching pursuit techniques used for signal reconstruction based on compressed sensing. The number of k and m samples needed to reconstruct the signals is described, as well as the error percentage and the recovery time.

Table 4.4: Reconstruction signals techniques

k samples	**BP** [%]	**MP**[%]	**OMP**[%]
Current	N/A	35.4	54.1
Voltage	N/A	35.4	89.2
m samples	**BP** [%]	**MP** [%]	**OMP** [%]
Current	90.9	76.8	70.6
Voltage	93.4	80.9	70.6
Error	**BP** [%]	**MP** [%]	**OMP** [%]
Current	1.105	4.451	3.856
Voltage	2.433	4.784	2.151
Time recover	**BP** [s]	**MP** [s]	**OMP** [s]
Current	10.853	0.182	1.331
Voltage	10.712	0.191	1.559

4.3 Fault Diagnosis

4.3.1 Electrical fault detection and classification

Once the power distribution network has been planned, it can be implemented in a power network analysis program such as CYMDIST. At present, power network analysis equipment has the capability to record up to six currents and six voltages at a sampling rate of 1 MHz with 18 bits of resolution, and to record up to six currents and six voltages at a sampling rate of 1 MHz with 18 bits of resolution as SEL-T401L with response times between 1 ms to 5 ms. In the present research, it has been verified that frequencies higher than 100 kHz can record transients as lightning impulses with a typical duration between 50 ns to 1 ms. For the detection of faults in electrical systems, the sampling frequency established in the present investigation is 25 kHz, allowing the recording of disturbances in the order of 40 us. With the established frequency, the number of samples per cycle is 500 for an electrical network frequency of 50 Hz.

To the current and voltage signals Gaussian white noise processes are added, high signal to noise ratios are sought. In the present investigation we have worked with current and voltage signals with power at 0 dB and noise ranging from 20 dB to 50 dB.

To detect faults using the wavelet transform, it was verified in simulation that a minimum of four samples are required to identify a fault, the fault detection time is 160 us corresponding to the four samples required. To test the effectiveness of

the algorithm, the model was tested on the 11 types of faults with resistances ranging from 0 Ohms to 20 Ohms.

In Figure 4.15 (a), the three-phase current signals are shown in bus-bar 4, it is observed that there is a phase and magnitude disturbance. The frequency spectrum that is the result of the electrical fault is shown in Figure 4.15 (b). In this figure, the frequency with the highest amplitude is the fundamental of 50 Hz, however new frequency components were included due to the electrical fault. Figure 4.15 (c) shows discrete cosine transform. In Figure 4.15 (d), the voltage signal has phase disturbances and decreases in magnitude. The frequency spectrum that is the result of an electrical fault is shown in Figure 4.15 (e). In this figure also the frequency with the highest amplitude is the fundamental 50 Hz, however, new frequency components were included due to the electrical fault. In Figure 4.15 (f), the discrete cosine transform is shown. It can be observed that the signal concentrates in few data allowing to create a sparse matrix.

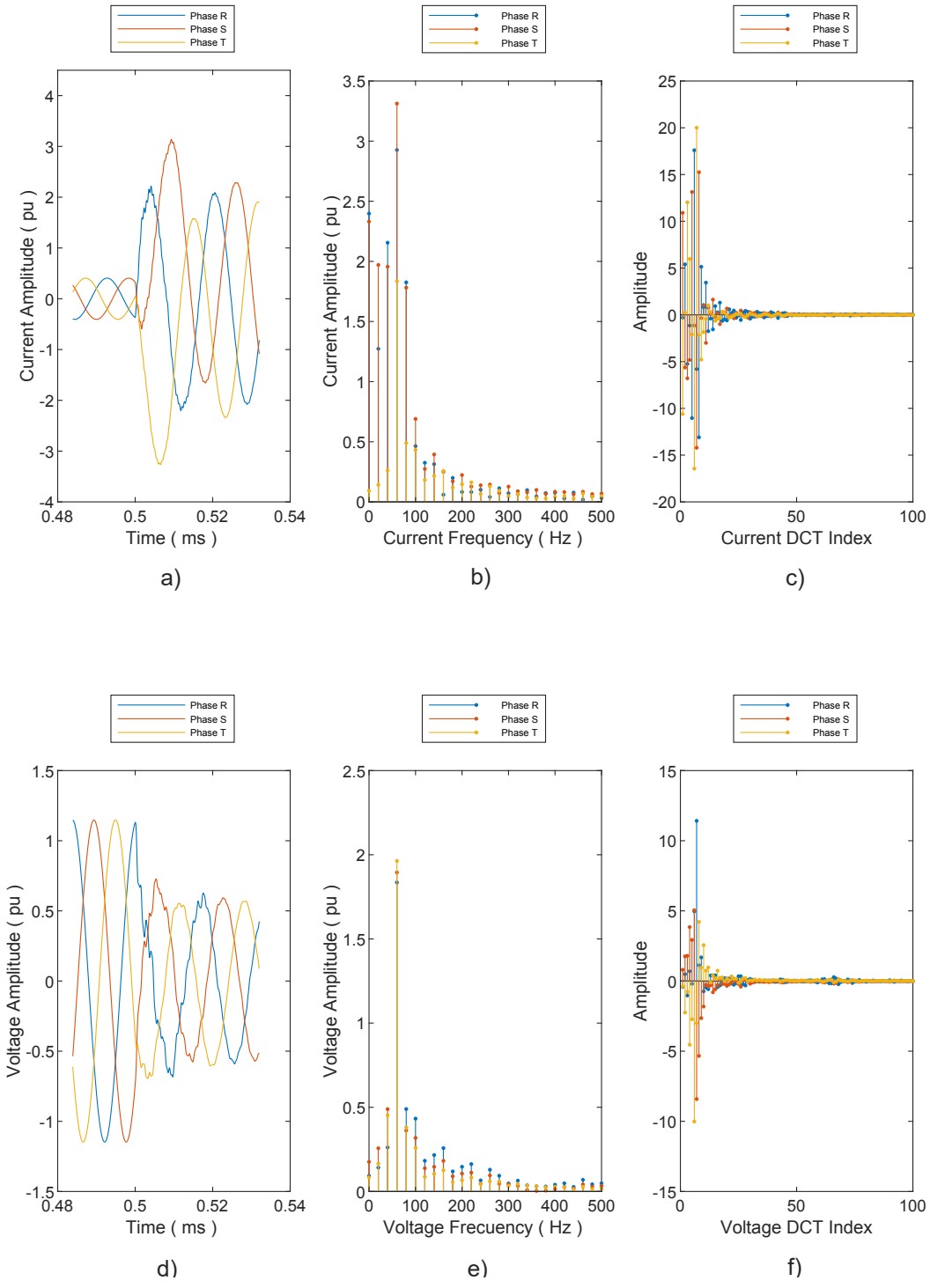


Figure 4.15: Time signals, frequency spectrum, DCT index.

Figure 4.16 shows the voltage signals in the three phases R, S and T. It can be seen that the fault occurs in only one phase because the voltage in phase R reduces its magnitude. The S and T phases have fault disturbances on the R phase; it is evident that the voltage magnitudes in these two phases increase due to the system response to maintain the power supplied. It is important to consider that the fault detection method must be able to distinguish a fault from a possible disturbance since the protection equipment must interrupting the circuit only the faulty phase.

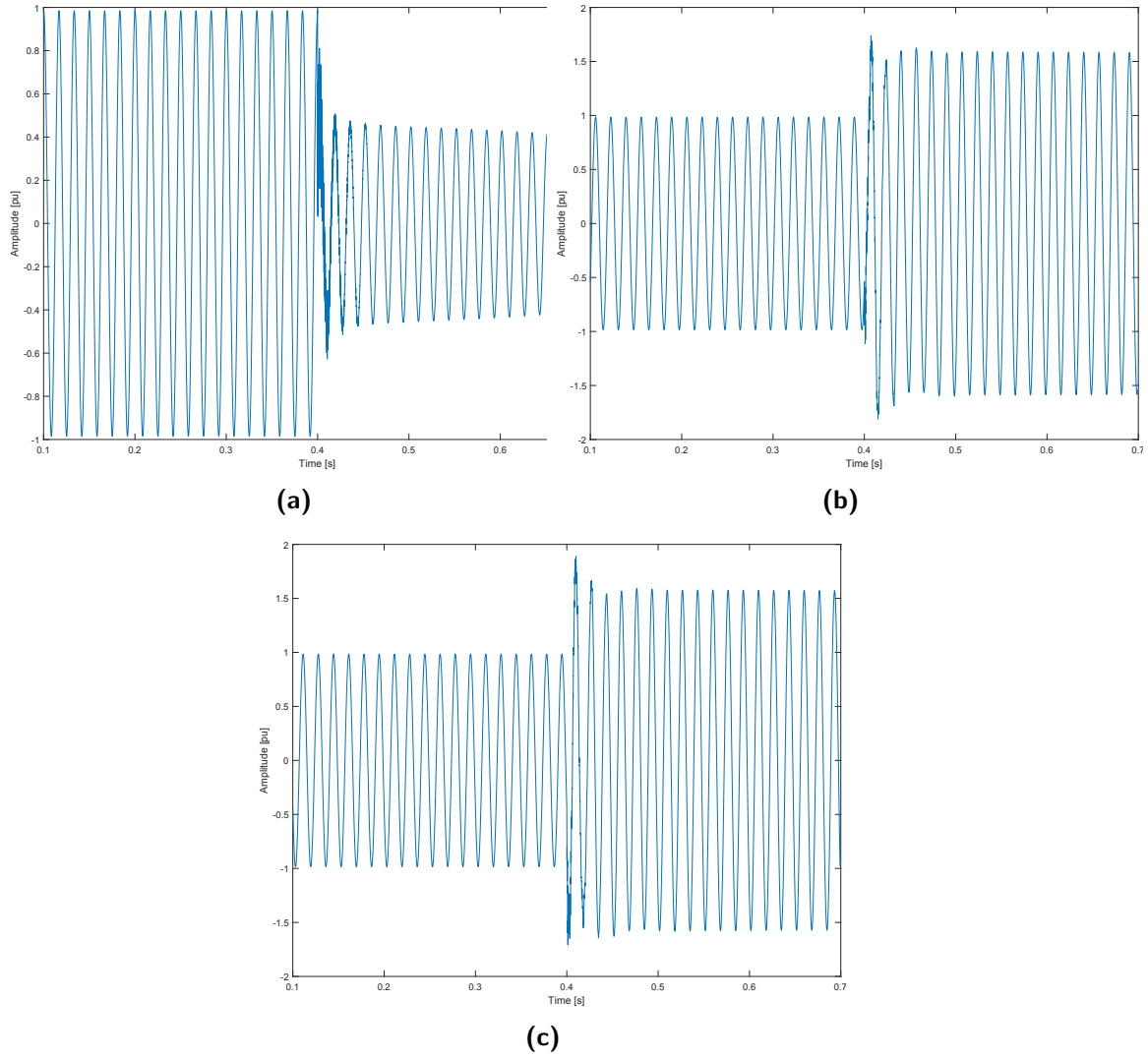


Figure 4.16: (a) Single Phase Fault at Phase R, (b) Phase S without fault, and (c) Phase T without fault.

Figure 4.17 shows the response of the wavelet Haar level 1 transformation. It can be seen that in steady state, the wavelet response is unchanged. However, when a fault or disturbance occurs, it is possible to see that how an impulse is generated at the exact moment of the sudden change in the signal. The amplitude of the steady state wavelet response varies between 0.002 and 0.005; when a fault occurs, the amplitude of the impulse varies between 0.08 and 0.18 [pu], and when a

disturbance occurs the amplitude of the impulse varies between 0.08 and 0.18 [pu] and when a disturbance occurs, the amplitude of the pulse varies between 0.045 and 0.055 [pu]. To distinguish a fault from a disturbance, a threshold is fixed and a comparison is made between the obtained amplitudes.

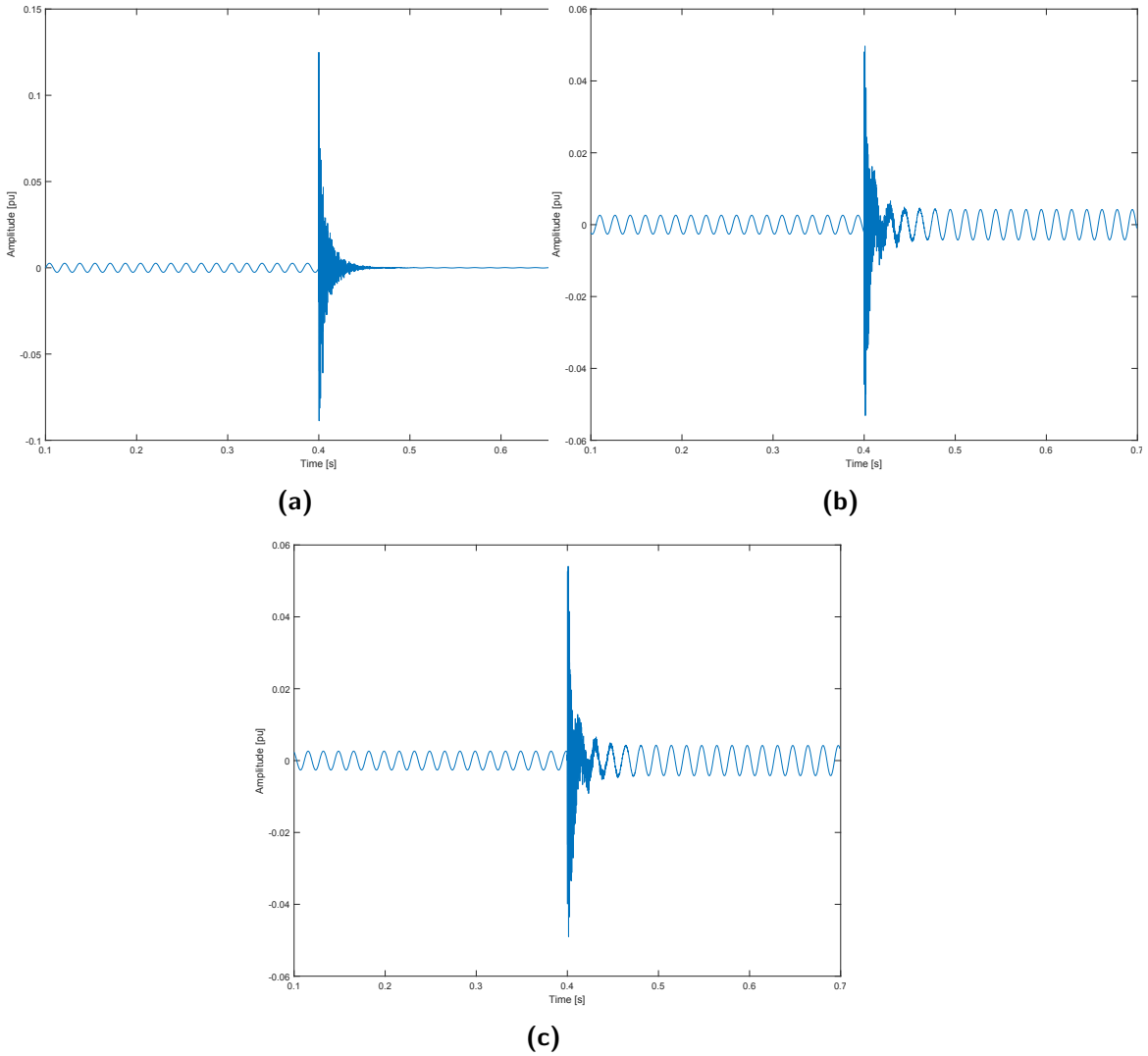


Figure 4.17: (a) Wavelet transform in Phase R, (b) Wavelet transform in Phase S, and (c) Wavelet transform in Phase T.

Figure 4.18 shows the frequency vs. time plot for a single-phase fault. When the voltage or current signals are in steady state, only the fundamental frequency exists as shown in the red boxes. When a fault occurs as in Figure 4.18 (a), several additional frequencies are created in addition to the fundamental, each with different amplitude and with different duration in time, in addition, it can be noted that the fundamental frequency disappears. Figure 4.18 literals (b) and (c) show that for the other two phases, frequencies are added at the same instant that the fault occurs in phase R, but the fundamental frequency does not disappear.

Therefore, it is possible to distinguish between a fault and a disturbance simply by analyzing the frequency response of the signal.

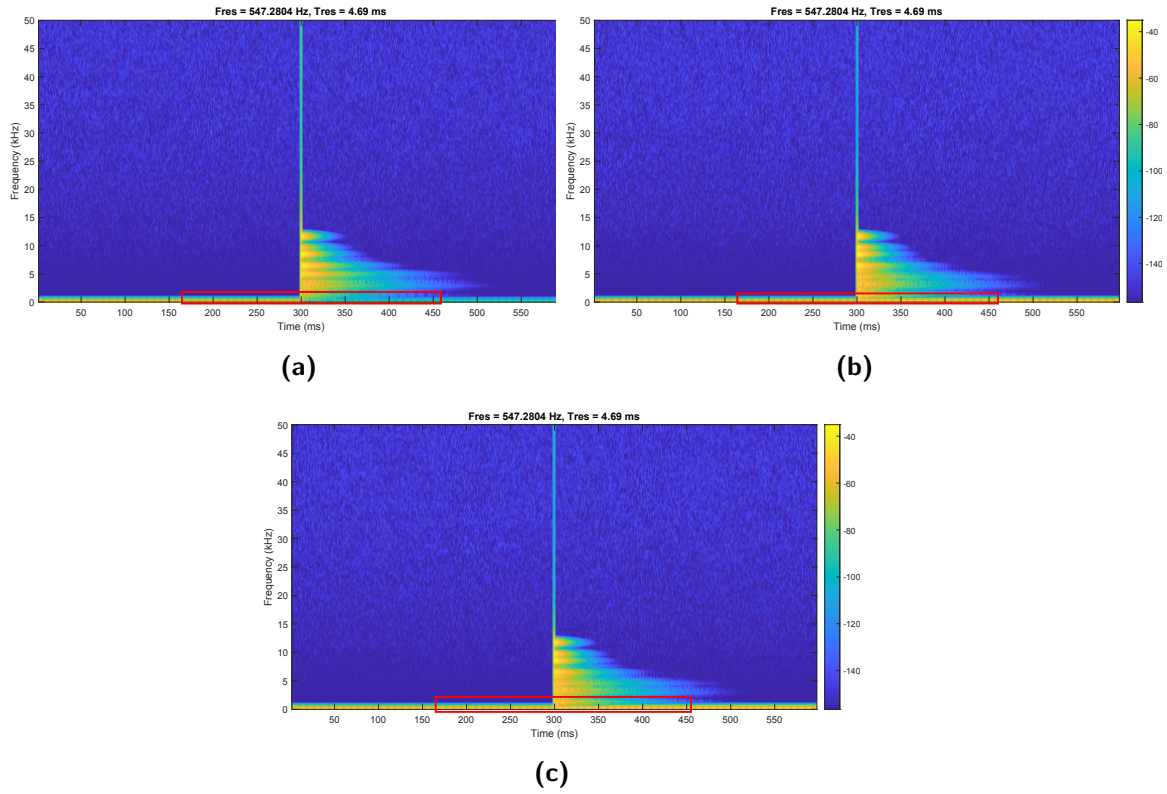


Figure 4.18: Frequency vs time analysis. (a) Phase R (with fault), (b) Phase S (no fault), and (c) Phase T (no fault).

Figure 4.19 shows the voltage signals in three phases R, S and T. It can be seen that the fault occurs in two phases, specifically in phases R and T due to amplitude reduction. The S phase presents disturbances due to faults in the R and T phases; in this phase, the voltage magnitude increases its value due to the response of the system to maintain the supplied power.

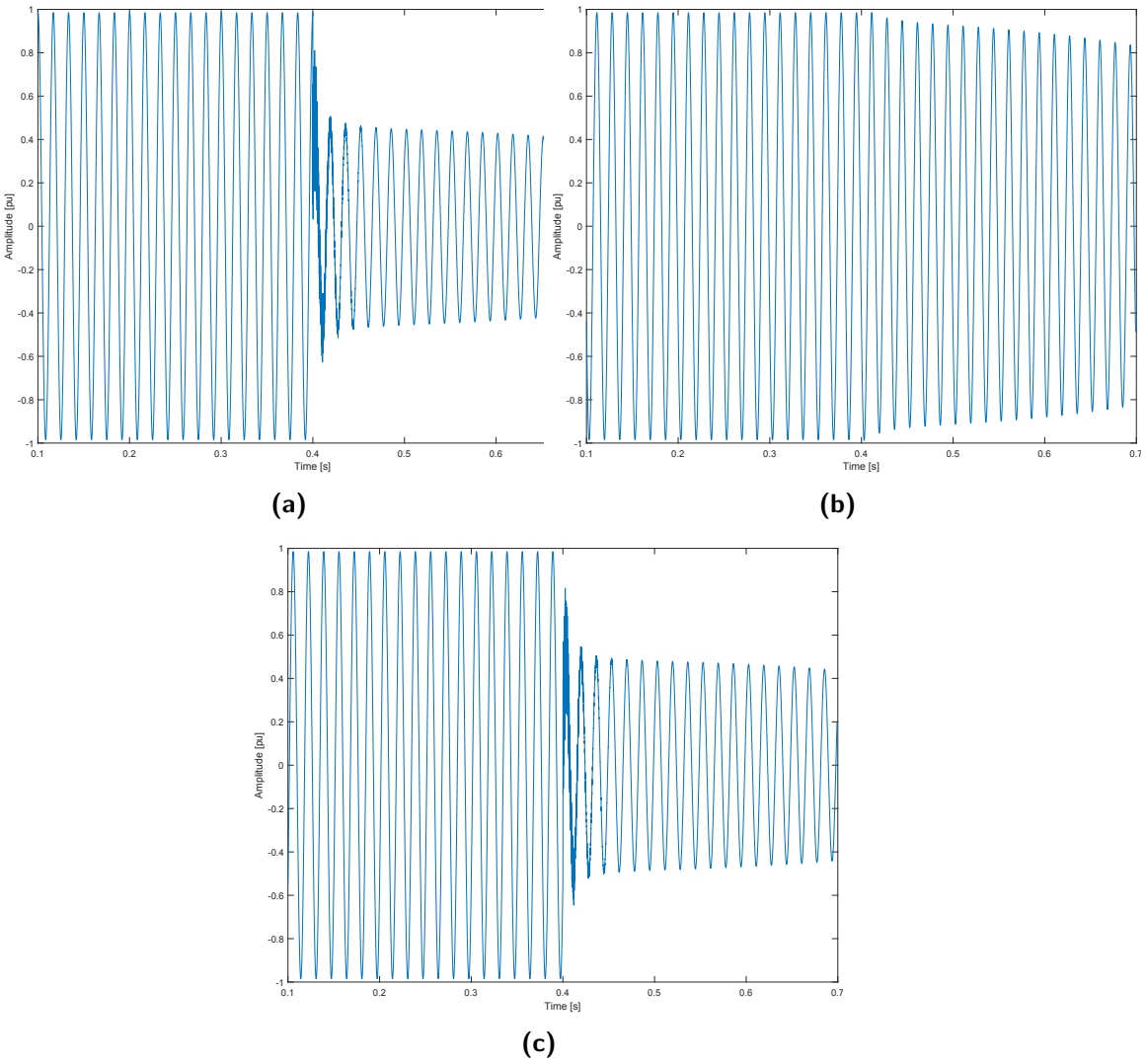


Figure 4.19: (a) Fault in phase R, (b) Phase S without fault, and (c) Fault in phase T.

Figure 4.20 shows the response of the Haar level 1 wavelet transform applied to a two-phase fault. For the R and T phases (fault condition), the amplitude of the wavelet response varies between 0.14 and 0.15 [pu]. While the S phase that does not experience any fault has an amplitude of 2.6 x 10-3 [pu].

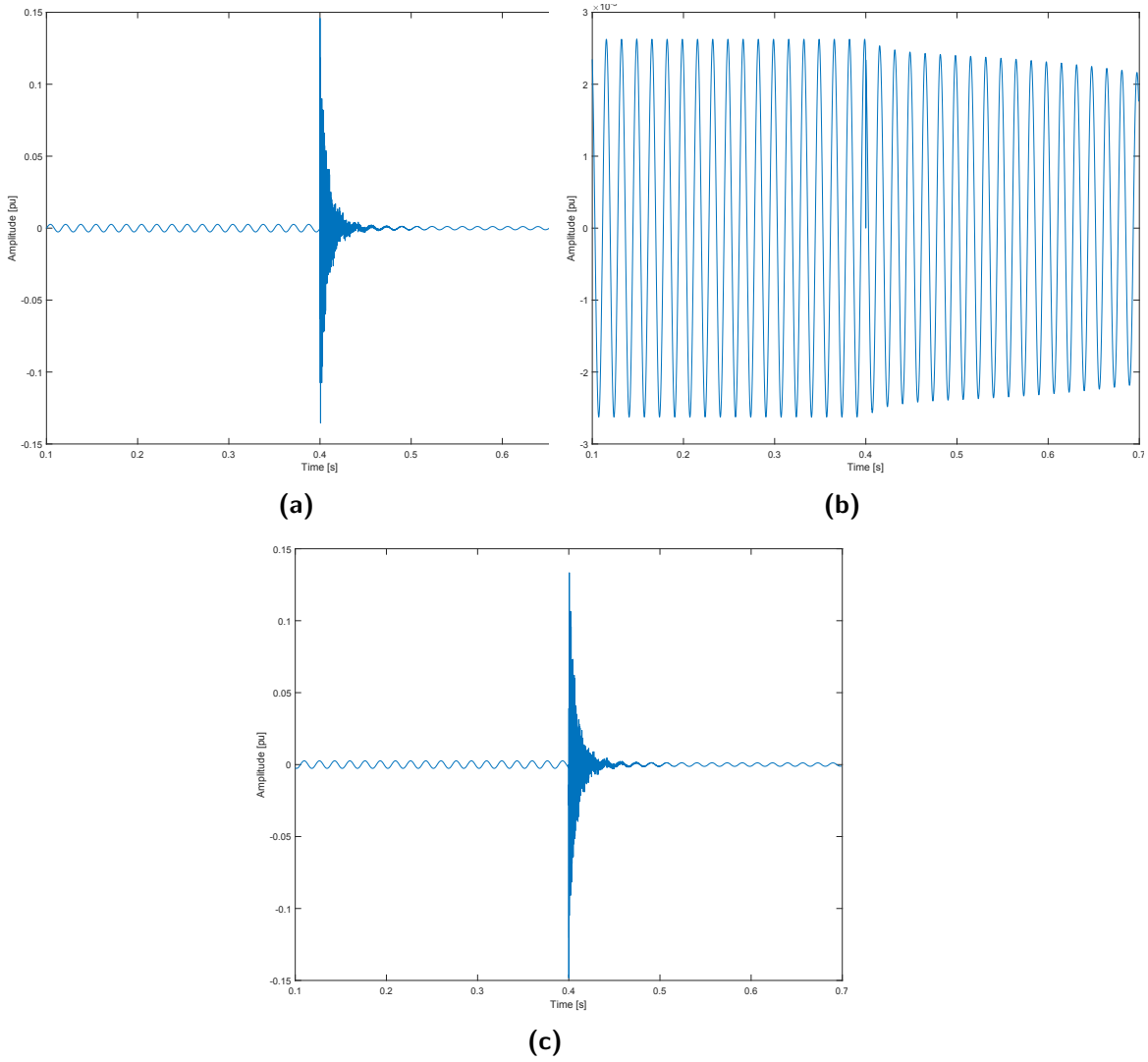


Figure 4.20: (a) Wavelet transform in phase R, (b) Wavelet transform in Phase S, and (c) Wavelet transform in Phase T.

Figure 4.21 shows the frequency vs. time plot in a two-phase fault scenario. When the voltage or current signals are in steady state, only the fundamental frequency exists as shown in the red boxes. When a fault occurs as in Figure 4.21 (a) and (c), several additional frequencies to the fundamental are created, each with different amplitude and with different duration in time. Moreover it can be noted that the fundamental frequency disappears. Figure 4.21 (b) show that for phase S (no fault), the frequencies are added at the same instant that the fault occurs in phases R and T, but the fundamental frequency does not disappear. Therefore, in this scenario, it is also possible to discriminate between a fault and a disturbance by simply analyzing the frequency response of the signal.

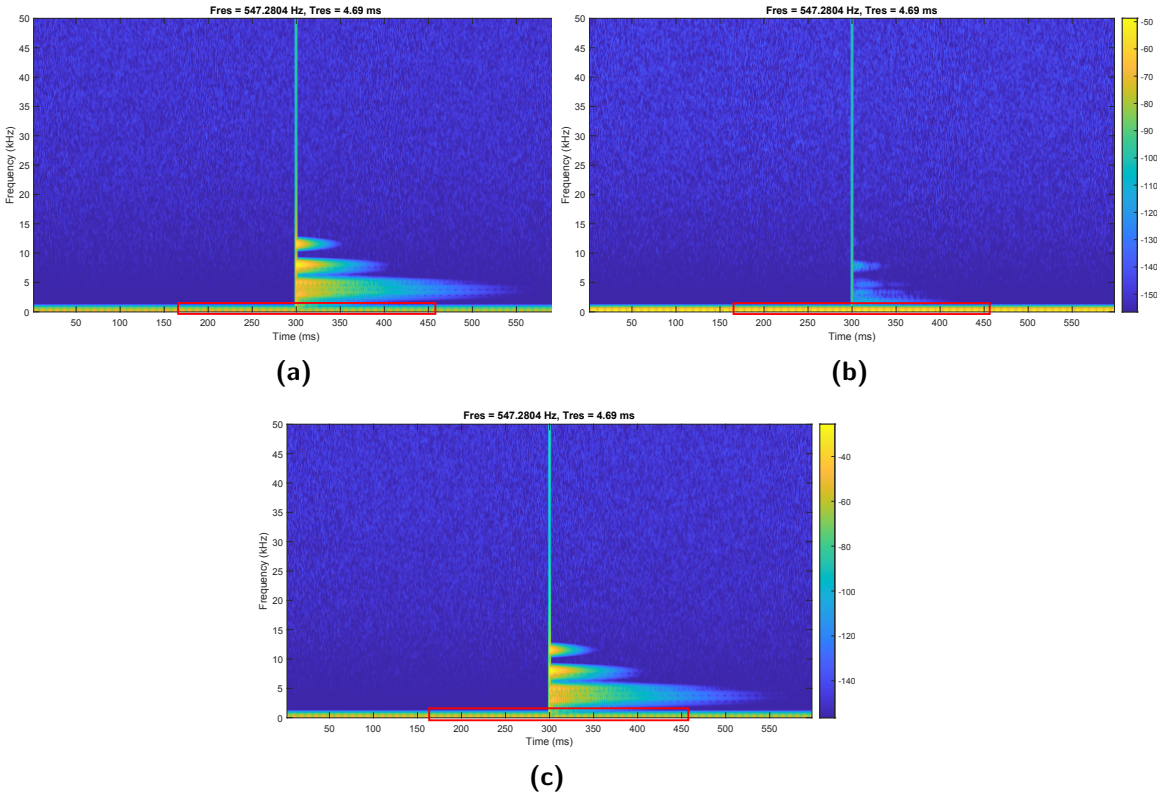


Figure 4.21: (a) Phase R (with fault), (b) Phase S (no fault), and (c) Phase T (with fault).

Figure 4.22 shows the voltage signals on the three phases R, S and T. It can be noted that the fault occurs on all three phases. In this scenario, phases R, S, and T reduce their respective voltage magnitude which of course occurs due to the fault in all three phases.

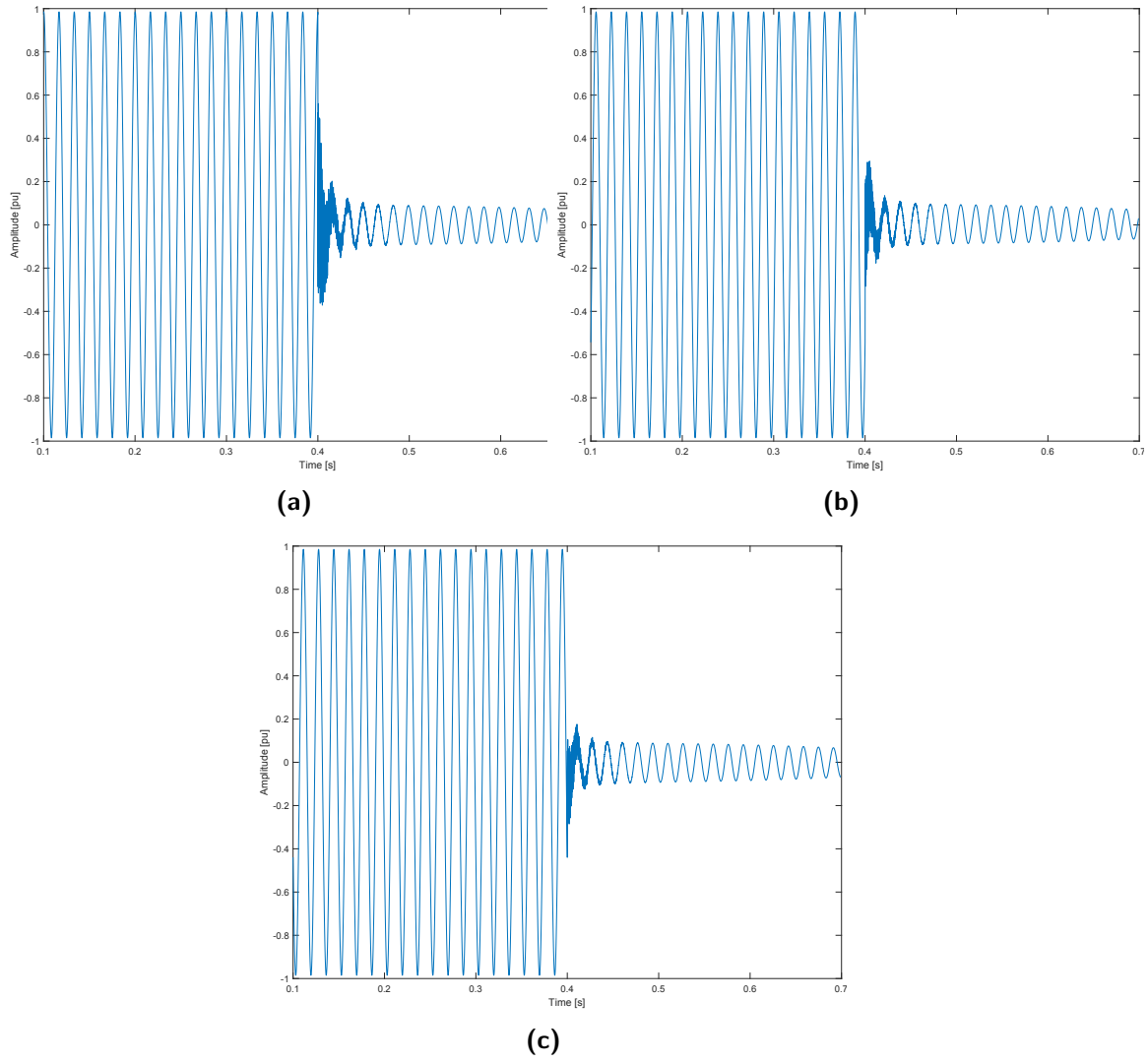


Figure 4.22: (a) Fault in phase R, (b) Fault in phase S, and (c) Fault in phase T.

Figure 4.23 shows the response of the Haar level 1 wavelet transform applied to a three-phase fault. For all phases, R, S and T, the amplitude of the wavelet response varies between 0.08 [pu] and 0.2 [pu].

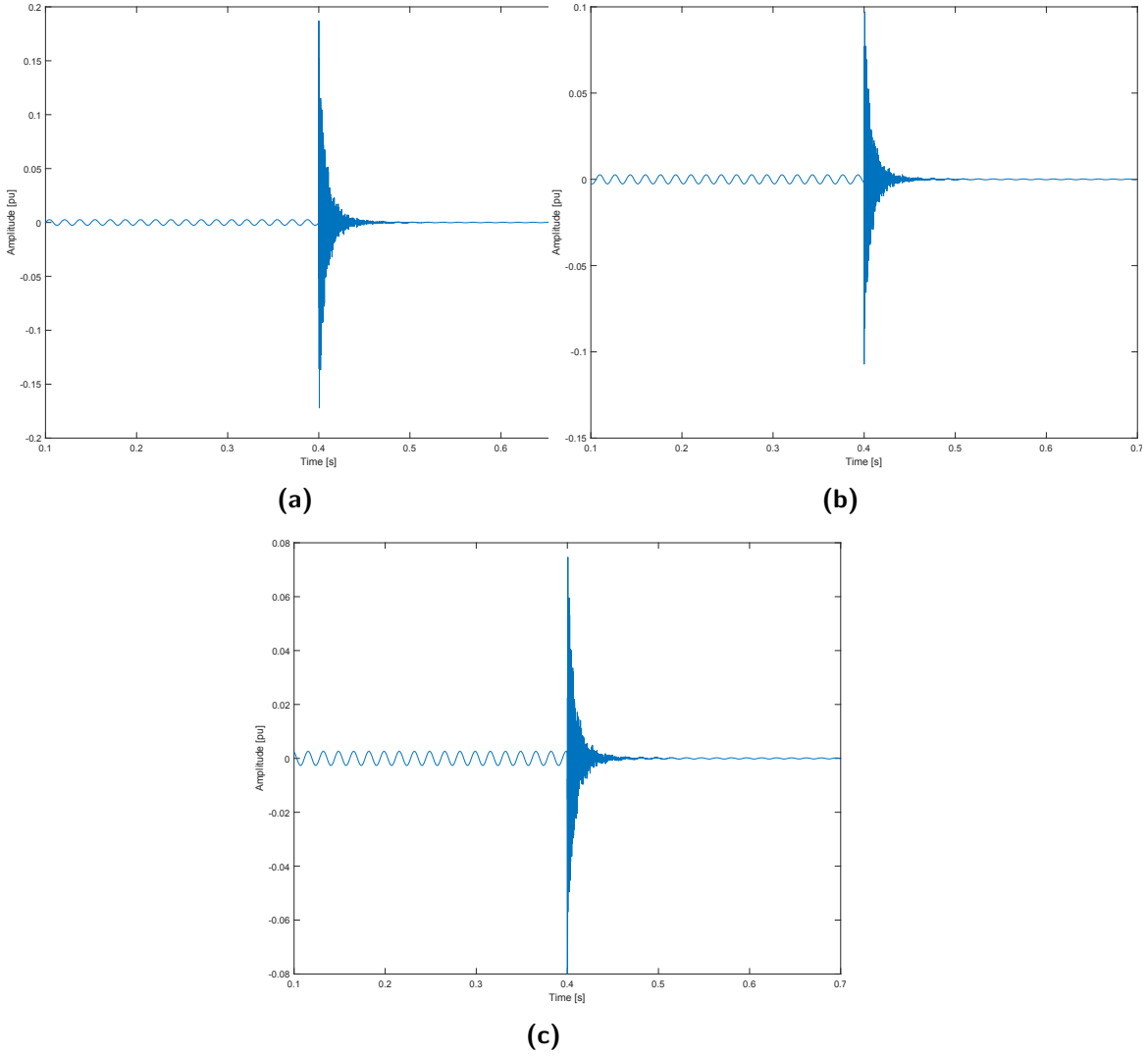


Figure 4.23: (a) Wavelet transform in phase R, (b) Wavelet transform in phase S, and (c) Wavelet transform in phase T.

Figure 4.24 shows the representation of frequency vs. time in a three-phase fault diagram. When the voltage or current signals are in steady state, only the fundamental frequency exists, as highlighted in the red boxes. When a fault occurs as in Figure 4.24 (a), (b), and (c), several additional frequencies are created in addition to the fundamental, each with different amplitude and with different duration in time, and it is observed that the fundamental frequency disappears. Therefore, it is possible to discriminate between a fault and a disturbance simply by analyzing the frequency response of the signal.

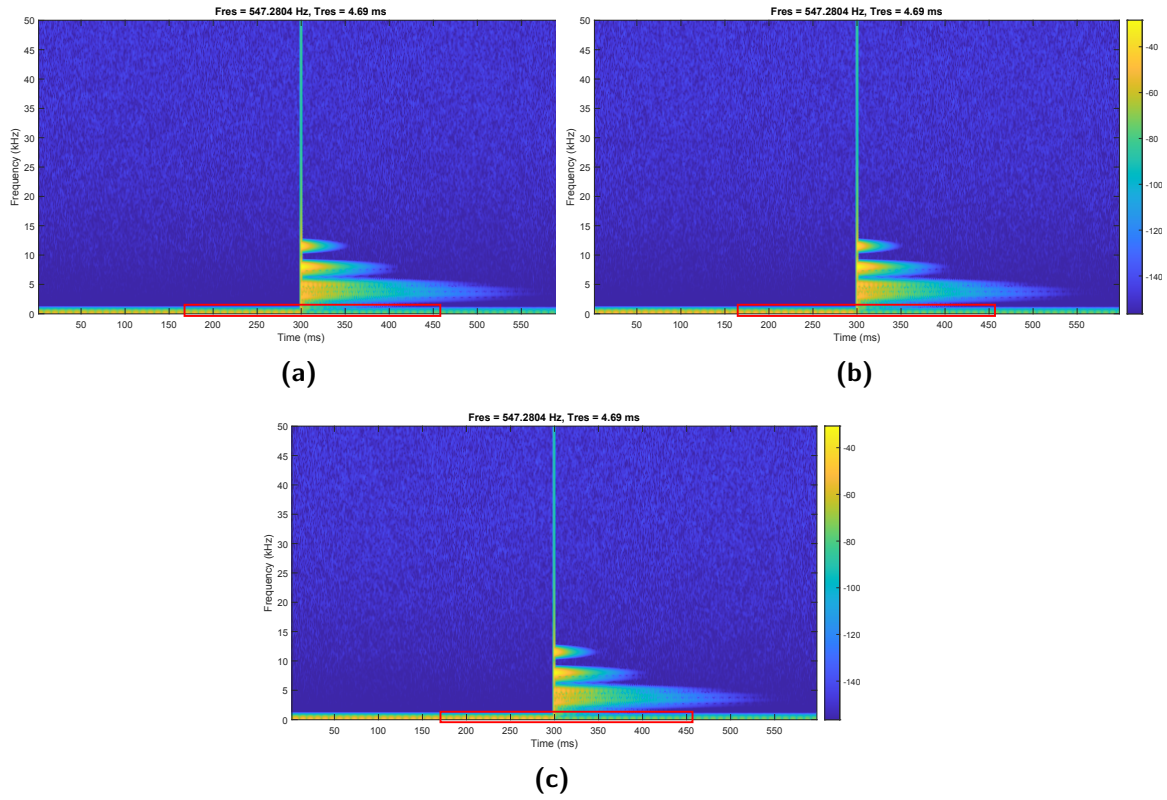


Figure 4.24: (a) Phase R (with fault), (b) Phase S (with fault), and (c) Phase T (with fault).

4.3.2 Analysis of Results

Table 4.5 summarizes the results of this work for the detection and classification of faults in transmission lines for the proposed study case. The accuracy of the algorithm developed for fault detection is 100%; however, it is important to take into account that when a single-phase fault occurs and depending on the fault resistance, disturbances may occur in the other two phases of the system. The results also indicate when there is No Fault Detected (NFD).

Table 4.5: Fault detection results achieved in this research

Fault Type	Absolute Error Phase R	Absolute Error Phase S	Absolute Error Phase T
Single-Phase 0 ohm	0	2.20E-05	2.20E-05
Single-Phase 5 ohm	NFD	NFD	4.20E-05
Single-Phase 10 ohm	NFD	3.10E-0.5	NFD
Single-Phase 20 ohm	2.10E-0.5	NFD	NFD
Two-phase 0 ohm	0	0	NFD
Two-phase 5 ohm	6.20E-05	NFD	6.20E-05
Two-phase 10 ohm	NFD	3.10E-0.5	3.10E-0.5
Two-phase 20 ohm	1.42E-0.4	1.22E-0.4	NFD
Three-Phase 0 ohm	0	0	5.00E-06
Three-Phase 5 ohm	2.120E-04	2.22E-05	2.20E-05
Three-Phase 10 ohm	1.220E-04	1.12E-05	2.12E-05
Three-Phase 20 ohm	1.64E-04	1.42E-04	2.92E-04

Table 4.6 shows the time in milliseconds taken from the algorithm to detect a fault in the test system.

Table 4.6: Fault detection times achieved in this research.

Fault Type	**Detection Time Phase R**	**Detection Time Phase S**	**Detection Time Phase T**
Single-Phase 0 ohm	0.382E-03	0.411E-03	0.3350E-03
Single-Phase 5 ohm	0.338E-03	0.356E-03	0.335E-03
Single-Phase 10 ohm	0.362E-03	0.315E-03	0.373E-03
Single-Phase 20 ohm	0.417E-03	0.296E-03	0.290E-03
Two-phase 0 ohm	0.374E-03	NFD	0.338E-03
Two-phase 5 ohm	0.366E-03	NFD	0.348E-03
Two-phase 10 ohm	NFD	0.373E-03	0.336E-03
Two-phase 20 ohm	0.387E-03	320E-03	NFD
Three-phase 0 ohm	0.345E-03	0.312E-03	0.2831E-03
Three-phase 5 ohm	0.322E-03	1.477E-03	0.3171E-03
Three-phase 10 ohm	0.467E-03	0.344E-03	0.619E-03
Three-phase 20 ohm	0.316E-03	0.278E-03	0.567E-03

Table 4.7 shows the time in seconds that the algorithm takes to classify a fault according to its type. This fault classification algorithm has an accuracy of 100%.

Table 4.7: Fault classification times achieved in this research.

Fault Type	Fault Classification time	Phases with electrical fault (proposed scenario)	Fault classification results
Single-Phase 0 ohm	0.98387	R	R
Single-Phase 5 ohm	1.64782	S	S
Single-Phase 10 ohm	0.86177	T	T
Single-Phase 20 ohm	1.08435	R	R
Two-phase 0 ohm	0.90893	R,T	R,T
Two-phase 5 ohm	0.89210	R,T	R,T
Two-phase 10 ohm	0.93385	S,T	S,T
Two-phase 20 ohm	0.93911	R,S	R,S
Three-phase 0 ohm	0.97855	R,S,T	R,S,T
Three-phase 5 ohm	0.93298	R,S,T	R,S,T
Three-phase 10 ohm	0.90478	R,S,T	R,S,T
Three-phase 20 ohm	0.92153	R,S,T	R,S,T

As it can be seen in Table 4.7, previous works were able to achieve high accuracy for fault location or fault classification, however, it is also important to analyze how long those research works took to achieve those results.

From Table 4.7, the best results for fault detection achieved were 99.7% and 100% accuracy with detection times varying from 10ms to 20ms. In this research with the developed algorithm, a 100% accuracy was achieved for fault detection in all the analyzed scenarios and the best detection time was 0.278 ms.

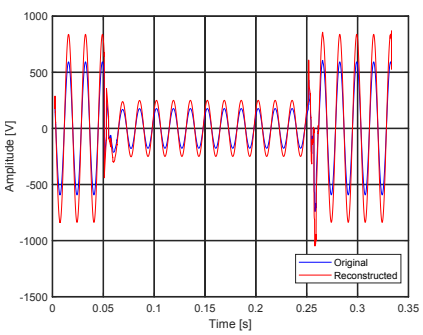
Following the results in Table 4.7, the best results for fault classification achieved were 97%, 99% and 100% accuracy with classification times ranging from 6ms to 10ms. This investigation achieved a 100% accuracy for fault classification in all the scenarios analyzed and the best detection time was 0.86177 seconds.

4.4 Signal Compression

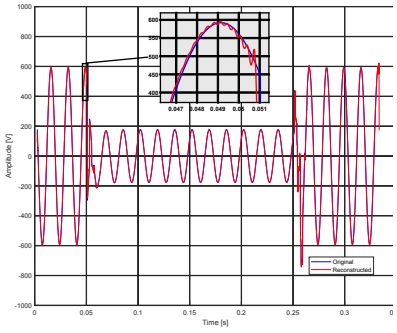
The amplitude of the reconstructed signals varies compared to the original one, as shown in Figure 4.25. The first strategy proposed in this work is the normalization of the reconstructed signal (by normalizing the results it will be easier to interpret). To achieve this objective, an algorithm was developed to search for the zero crossings in both the original and the reconstructed signal number of samples per cycle. It is important to note that this process must be performed for each phase of the electrical signals since they are 120 degrees out of phase with each other.

From algorithm 7, two values are obtained for the maximum steady magnitude value and the position for both the original and the compressed signal. Finally, the reconstructed signal is multiplied by the result of the division between the maximum magnitude of the original signal and the maximum magnitude of the reconstructed signal. Algorithm 7 performs this process from step 3 to step 5.

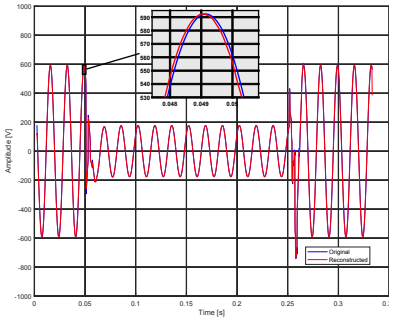
The next problem to be solved is the ripple that occurs in the reconstructed signal as shown in Figure 4.25 when enlarging at the points of maximum magnitude of each electrical phase. In order to solve this problem, it is proposed to apply the moving average filter described by E.q. 3.15.



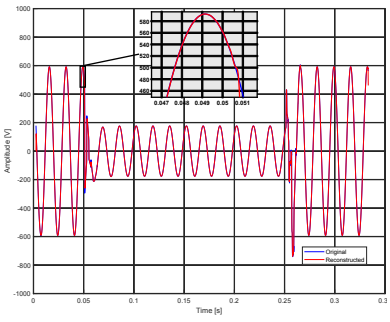
(a) Wavelet Scaling effect



(b) Wavelet ripple effect



(c) Wavelet Shifting effect



(d) Ripple, Shifting and scaling correction

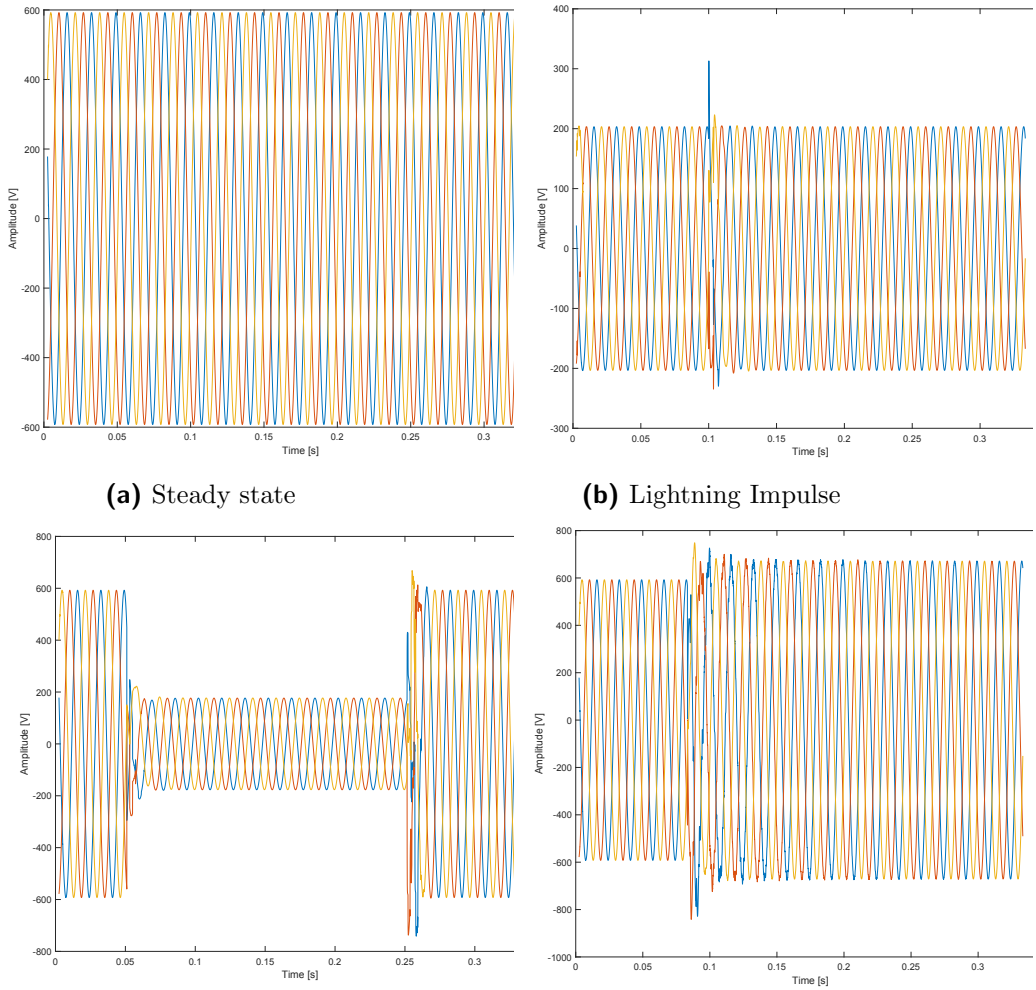
Figure 4.25: Wavelet Faults

Finally, due to the characteristics of the wavelet transform, specifically, change of scale and translation with respect to the original signal, a slight displacement

between the original signal and the reconstructed signal can be seen when it is magnified, as shown Figure 4.26. To correct the presented displacement between the signals, the approach proposed in this study is to use the positions of the maximum points calculated in one cycle of the signal in the original and the reconstructed steady state. The difference between the positions of these two points allows the reconstructed signal to shift until they remain in phase, as shown in Figure 4.26.

To analyze variables related to power quality, this research proposes a sampling frequency of 200 kHz, which allows the measurement of very fast signals especially during transient conditions.. Transients are produced by atmospheric discharges, and they are presented as pulses with a duration in the order of microseconds and typically can last from 50 ns to 1ms. In addition, using the same sampling frequency, three more signals were analyzed: steady-state voltage signals, a three-phase fault and a surge.

Each analyzed signal generates a 66168×4 matrix, each column carries specific information with respect to time, voltage R, voltage S and voltage T respectively. In addition, the generated files have a size of 1,889,931 bytes. Finally, the computer equipment used in the development of the experiment has the following characteristics: Processor (Intel (R) Xeon (R) E-2176M CPU @ 2.70GHz) and 64 GB of RAM. Figure 4.26 shows the representation of each signal as a function of time.



(a) Steady state

(b) Lightning Impulse

(c) 3 Φ fault

(d) Swell

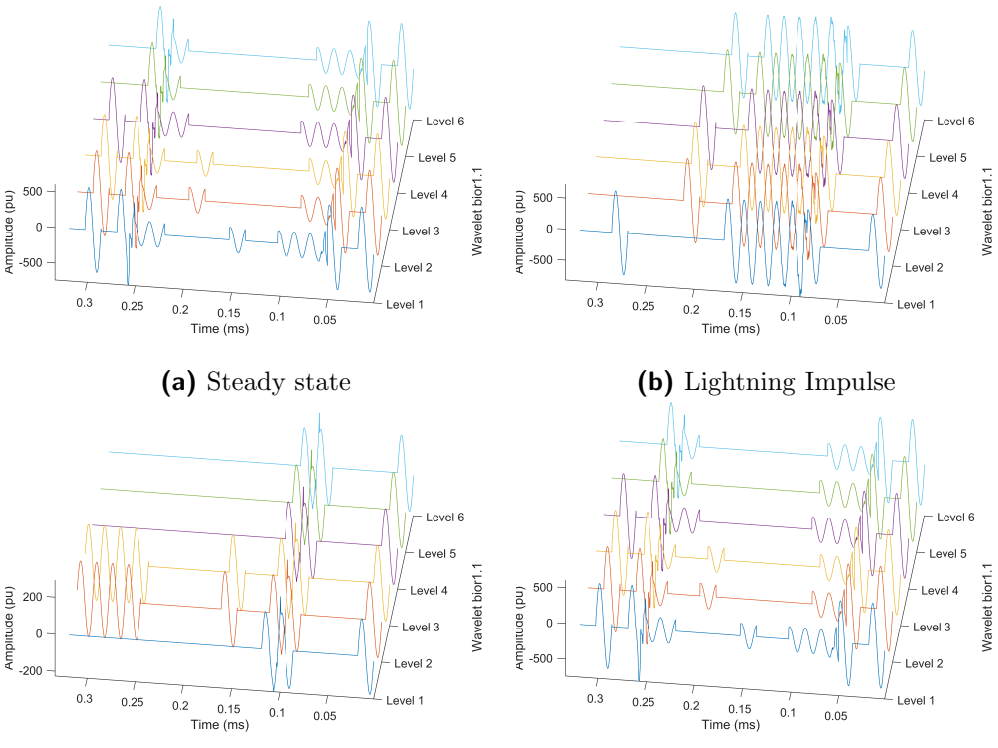
Figure 4.26: Original signals

Figure 4.27 (a) shows the compressed signal in steady state using wavelet bior1.1, were it can be seen that in steady state, it is necessary to have the information of the first cycle and depending on the wavelet level, the algorithm selects the necessary signals for the subsequent reconstruction.

Figure 4.27 (b) shows several signals that are necessary to reconstruct a perturbation caused by the atmospheric discharge phenomenon. It can be seen that the number of signals per cycle increases compared to a single signal in steady state; it is also shown that there are cycles in which the signals are zero because they are represented by another signal while maintaining a high level of RTE.

Figure 4.27 (c) represents the signal of a three-phase fault; it is observed that the fault generates a phase shift when compared to the steady state signal. This makes necessary a large number of signals represented by cycles to reconstruct the original signal.

Figure 4.27 (d) shows the result of the compression algorithm and the proposed window for disturbances such as sag, swell or harmonics. This solution is proposed considering that after a certain time the system tends to stabilize and future signals can be represented by specific signals in certain cycles.



(a) Steady state

(b) Lightning Impulse

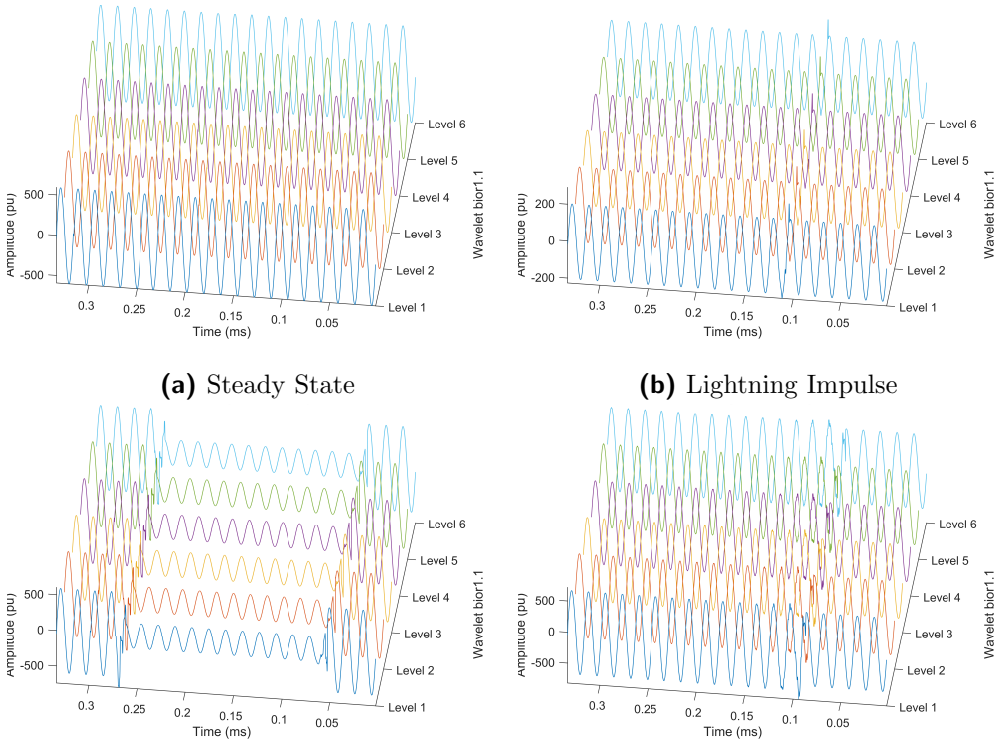
(c) 3 Φ fault

(d) Swell

Figure 4.27: Different wavelet levels for reconstruction.

Finally, in Figures 4.27 (a), (b), (c) and (d) it can be observed that as the wavelet level increases, a lower number of sample cycles is needed because some characteristics of the small amplitude signals are lost, allowing only a few signals (if necessary) to represent the complete set.

Figure 4.28 shows the reconstruction of each of the analyzed signals represented in the 6 levels of the wavelets. It can be noted that there is a similarity between the different levels of each of the signals, thus corroborating the results shown in Tables 4.8 and 4.9.



(a) Steady State

(b) Lightning Impulse

(c) 3 Φ fault

(d) Swell

Figure 4.28: Signals reconstruction.

Table 4.8 shows the compression performed by the wavelet at each level. From the data shown in the Table 4.8, it is evident that the compression percentage for each signal is the same in all levels; moreover, this percentage does not depend on the characteristics of the signal to be analyzed. The size of each signal is presented in bytes and it can be observed that the compression level depends not only on the wavelet level but also on the characteristics of each of the signals to be analyzed.

Table 4.8: Comparison of compression results between different wavelet levels.

Wavelet Compression in bytes						
Signal	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
All signals	972,522	483,306	250,021	118,528	61,602	30,670
Compression %	48.5419	74.4273	86.7790	93.7284	96.740	98.377
Windowing in bytes						
Steady State	41.547	21,504	30,624	16,035	8,793	3,723
Compression %	97.8017	98.8622	98.379	99.151	99.534	99.803
Lightning	117,321	59,706	39,954	30,579	20,559	4,959
Compression %	93.7923	96.8408	97.886	98.382	98.912	99.737
Fault 3Φ	269,043	135,867	87,885	40,089	20,757	13,329
Compression %	85.7644	92.8110	95.349	97.878	98.901	99.294
Swell	306,348	154,674	78,447	39,762	20,625	10,929
Compression %	83.7905	91.8159	95.849	97.896	98.908	99.421

Table 4.9 provides a summary of the results obtained from the analyzed signals: steady state, atmospheric discharge, triphasic faults and waves. The RTE, NMSE and COR results are shown for each wavelet level (these four parameters are found in the literature review). Furthermore, this research considers important to analyze two additional parameters that have not been considered in previous works. These parameters are: **Compression Time**, which is the time elapsed from the time of loading the file (with the original data) until a new file with the compressed signal is generated; and **Recovery Time** which is the time from the time the file with the compressed signal is loaded until a new file with all the reconstructed signal information is generated.

Table 4.9: Compression results: RTE, NMSE, COR compression time and recovery time.

Bior 1.1	RTE(%)	NMSE	Cross Correlation	Compression Time	Recovery Time
Steady State Signal					
Level 1	99.9143	0.019565	0.990646	0.259778	0.290082
Level 2	99.9059	0.001650	0.999645	0.092780	0.443042
Level 3	99.9357	0.001768	0.999437	0.082618	0.134150
Level 4	99.9315	0.001972	0.999356	0.048862	0.049667
Level 5	99.9486	0.002044	0.998720	0.039826	0.086259
Level 6	99.9479	0.001902	0.998788	0.036617	0.101389
Lightning Signal					
Level 1	99.8847	0.011983	0.994585	0.283787	0.445981
Level 2	99.8620	0.001972	0.999704	0.130686	0.482466
Level 3	99.8861	0.002771	0.999183	0.042138	0.089791
Level 4	99.8814	0.000380	0.999216	0.039053	0.114254
Level 5	99.9988	0.000098	0.999944	0.039361	0.231550
Level 6	99.9415	0.000434	0.999925	0.048162	0.115147
Fault 3Φ Signal					
Level 1	99.77542	0.013614	0.992070	0.244907	0.246346
Level 2	99.90412	0.006783	0.996128	0.120452	0.445679
Level 3	99.93682	0.008608	0.995380	0.039305	0.110532
Level 4	99.95180	0.007468	0.996024	0.054989	0.091779
Level 5	99.84289	0.003184	0.997622	0.074945	0.097257
Level 6	99.84292	0.002965	0.997731	0.052374	0.113188
Swell Signal					
Level 1	99.9383	0.004697	0.997960	0.521757	0.794249
Level 2	99.9103	0.003014	0.998941	0.089559	0.409472
Level 3	99.9314	0.004772	0.997957	0.050329	0.126874
Level 4	99.9210	0.003857	0.998466	0.116466	0.065097
Level 5	99.9595	0.000132	0.999731	0.077322	0.157097
Level 6	99.9825	0.001105	0.999534	0.036522	0.144762

Chapter 5

Conclusions and Further Research

The results obtained by the proposed algorithms, which are reported in Chapter 4 lead to several considerations.

The heuristic algorithm developed finds a quasi-optimal solution and solves the problem known as np-complete setcover. The result is the sizing of the electrical distribution network by georeferentially locating the transformers and the route of the electrical feeders in medium and low voltage, optimizing the cost of implementation of the electrical distribution network. Finally, it was verified that the network design provides electric service coverage to 100% of the buildings, power losses do not exceed 1.8% of the total energy flow and voltage profiles are acceptable with maximum voltage drops of 3%.

The results obtained show that it is possible to generate an optimal dictionary from the steady state signals. The dictionary contains the most relevant information of the different types of faults and depending on the optimization methods signals with 40% of missing data can be recovered as performed by Orthogonal Matching Pursuit with run times of about 2 seconds. If faster run times are required, the solution is Matching Pursuit with run times of 0.2 seconds, but it requires 80% of samples.

The results obtained show that the algorithm developed for electrical fault diagnosis using time domain signals is classified within the ultra high speed models with detection times of less than 1ms and 100% reliability. The results of the experiments show that the Haar wavelet is the most susceptible to abrupt changes in the signals as a consequence of electrical faults. The algorithm developed for fault classification requires the establishment of thresholds depending on the type of fault. The sampling frequencies must be higher than 100 kHz, with lower sampling frequencies, it has been shown that the error in detection, classification and localization increases and therefore decreases the reliability.

The algorithms developed for data compression show that the quality indexes of the reconstructed signals such as RTE, NMSE and X-COR have been improved with respect to the analyzed literature since the change in amplitude, ripple and displacement of the signal with respect to the original has been taken into account. Depending on the type of fault, data compression can reach ratios of 99.803% by applying the cycle-based signal comparison system.

Finally, this PhD dissertation has generated new models focused on the new smart grids, providing a solution to engineering problems in planning, deployment, fault diagnosis and data compression.

The creation of artificial intelligence-based fault classification methods, using the dictionaries of each type of fault for training, validation and testing, thus improving the classification based on thresholds proposed in this thesis, is planned as future work.

Bibliography

[Abdullah, 2018] Abdullah, A. (2018). Ultrafast transmission line fault detection using a dwt-based ann. *IEEE Transactions on Industry Applications*, 54(2):1182–1193.

[Abedi et al., 2019] Abedi, M. H., Hosseini, H., and Jalilvand, A. (2019). Subtransmission substation expansion planning considering load center uncertainties of size and location. *International Journal of Electrical Power and Energy Systems*, 109(January):413–422.

[Aggarwal et al., 1999] Aggarwal, R. K., Xuan, Q. Y., Dunn, R. W., Johns, A. T., and Bennett, A. (1999). A novel fault classification technique for double-circuit lines based on a combined unsupervised/supervised neural network. *IEEE Transactions on Power Delivery*, 14(4):1250–1256.

[Aghaei et al., 2014] Aghaei, J., Muttaqi, K. M., Azizivahed, A., and Gitizadeh, M. (2014). Distribution expansion planning considering reliability and security of energy using modified PSO (Particle Swarm Optimization) algorithm. *Energy*, 65:398–411.

[Ahmadi et al., 2019] Ahmadi, S. A., Vahidinasab, V., Ghazizadeh, M. S., Mehran, K., Giaouris, D., and Taylor, P. (2019). Co-optimising distribution network adequacy and security by simultaneous utilisation of network reconfiguration and distributed energy resources. *IET Generation, Transmission and Distribution*, 13(20):4747–4755.

[Ahmadigorji et al., 2017] Ahmadigorji, M., Amjady, N., and Dehghan, S. (2017). A novel two-stage evolutionary optimization method for multiyear expansion planning of distribution systems in presence of distributed generation. *Applied Soft Computing Journal*, 52:1098–1115.

[Aleem et al., 2014] Aleem, S. A., Shahid, N., and Naqvi, I. H. (2014). Methodologies in power systems fault detection and diagnosis. *Energy Systems*, 6(1):85–108.

[American Society of Civil Engineers, 2020] American Society of Civil Engineers (2020). Failure to Act: The Economic Impact of Current Investment Trends in Electricity Infrastructure. *Judicial Review in EU Law*, 1(1):46.

[Arasteh et al., 2016] Arasteh, H., Sepasian, M. S., Vahidinasab, V., and Siano, P. (2016). SoS-based multiobjective distribution system expansion planning. *Electric Power Systems Research*, 141:392–406.

[Arasteh et al., 2019] Arasteh, H., Vahidinasab, V., Sepasian, M. S., and Aghaei, J. (2019). Stochastic System of Systems Architecture for Adaptive Expansion of Smart Distribution Grids. *IEEE Transactions on Industrial Informatics*, 15(1):377–389.

[Arauz, 2019] Arauz, J. (2019). Smart Cities and the Dire Need for a Course Correction. *2018 IEEE International Smart Cities Conference, ISC2 2018*, 1(1):1–6.

[Baimel et al., 2016] Baimel, D., Tapuchi, S., and Baimel, N. (2016). Smart grid communication technologies- overview, research challenges and opportunities. *Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2016*, 1(1):116–120.

[Banol Arias et al., 2018] Banol Arias, N., Tabares, A., Franco, J. F., Lavorato, M., and Romero, R. (2018). Robust Joint Expansion Planning of Electrical Distribution Systems and EV Charging Stations. *IEEE Transactions on Sustainable Energy*, 9(2):884–894.

[Basavaraj, 2016] Basavaraj, S. (2016). Approach for Power Quality Monitoring and Data Compression [I. *lnternational Conference on Power and Energy Systems:Towards Sustainable Energy (PESTSE)*, 1(1):4–8.

[Candes and Tao, 2006] Candes, E. J. and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425.

[Carta et al., 2018] Carta, D., Pegoraro, P. A., and Sulis, S. (2018). Impact of Measurement Accuracy on Fault Detection Obtained with Compressive Sensing. *9th IEEE International Workshop on Applied Measurements for Power Systems, AMPS 2018 - Proceedings*, pages 1–5.

[Cheng Hong, 2000] Cheng Hong, S. E. (2000). A B-Spline Wavelet Based Fault Classification Scheme for High Speed Protection Relaying. *Electric Machines & Power Systems*, 28(4):313–324.

[Daim et al., 2018] Daim, T. U., Oliver, T., and Phaal, R. (2018). Technology roadmapping. *Technology Roadmapping*, pages 1–783.

[Das and Reddy, 2005] Das, B. and Reddy, J. V. (2005). Fuzzy-logic-based fault classification scheme for digital distance protection. *IEEE Transactions on Power Delivery*, 20(2 I):609–616.

[de Quevedo et al., 2019] de Quevedo, P. M., Muñoz-Delgado, G., and Contreras, J. (2019). Impact of electric vehicles on the expansion planning of distribution

systems considering renewable energy, storage, and charging stations. *IEEE Transactions on Smart Grid*, 10(1):794–804.

[Donoho, 2006] Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal l 1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829.

[El-Zonkoly, 2013] El-Zonkoly, A. M. (2013). Multistage expansion planning for distribution networks including unit commitment. *IET Generation, Transmission and Distribution*, 7(7):766–778.

[Fei et al., 2019] Fei, W., Ji, G., Sharma, D., and Jiang, J. N. (2019). A New Traveling Wave Representation for Propagation of Energy Transients in Power Lines from a Quantum Perspective. *2018 North American Power Symposium, NAPS 2018*, pages 1–6.

[Gayathri, 2015] Gayathri (2015). Double Circuit EHV Transmission Lines Fault Location with RBF Based Support Vector Machine and Reconstructed Input Scaled Conjugate Gradient Based Neural Network. *International Journal of Computational Intelligence Systems*, 8(1):95–105.

[Gholizadeh-Roshanagh and Zare, 2019] Gholizadeh-Roshanagh, R. and Zare, K. (2019). Electric power distribution system expansion planning considering cost elasticity of demand. *IET Generation, Transmission and Distribution*, 13(22):5229–5238.

[Gholizadeh-Roshanagh et al., 2020] Gholizadeh-Roshanagh, R., Zare, K., and Marzband, M. (2020). An A-Posteriori Multi-Objective Optimization Method for MILP-Based Distribution Expansion Planning. *IEEE Access*, 8:60279–60292.

[Gupta and Tripathy, 2015] Gupta, O. H. and Tripathy, M. (2015). An innovative pilot relaying scheme for shunt-compensated line. *IEEE Transactions on Power Delivery*, 30(3):1439–1448.

[Hung et al., 2014] Hung, D. Q., Mithulananthan, N., and Bansal, R. C. (2014). An optimal investment planning framework for multiple distributed generation units in industrial distribution systems. *Applied Energy*, 124:62–72.

[IEEE Std 1159-2019, 2019] IEEE Std 1159-2019 (2019). *IEEE Recommended Practice for Monitoring Electric Power Quality*, volume 2019. IEEE, New York.

[Isermann and Ballé, 1997] Isermann, R. and Ballé, P. (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5):709–719.

[Jeddi et al., 2019] Jeddi, B., Vahidinasab, V., Ramezanpour, P., Aghaei, J., Shafie-khah, M., and Catalão, J. P. (2019). Robust optimization framework for

dynamic distributed energy resources planning in distribution networks. *International Journal of Electrical Power and Energy Systems*, 110(February):419–433.

[Jia et al., 2018] Jia, Y., Li, T., Zhang, D., and Wang, Z. (2018). Design and analysis of a hardware-efficient compressed sensing architecture for data compression in power quality data acquisition. *Proceedings - 2018 10th International Conference on Advanced Computational Intelligence, ICACI 2018*, 1(1):300–307.

[Jooshaki et al., 2019] Jooshaki, M., Abbaspour, A., Fotuhi-Firuzabad, M., Moeini-Aghtaie, M., and Lehtonen, M. (2019). MILP Model of Electricity Distribution System Expansion Planning Considering Incentive Reliability Regulations. *IEEE Transactions on Power Systems*, 34(6):4300–4316.

[Kaewmamuang et al., 2019] Kaewmamuang, K., Siritaratiwat, A., Surawanitkun, C., Khunkitti, P., and Chatthaworn, R. (2019). A novel method for solving multi-stage distribution substation expansion planning. *Energy Procedia*, 156:371–383.

[Li et al., 2018] Li, H., Ma, Y., and Fu, Y. (2018). An improve d RIP-base d performance guarantee for sparse signal recovery via simultaneous orthogonal matching pursuit. *Signal Processing*, 144:29–35.

[Maldonado, 2018] Maldonado, M. G. R. (2018). Wireless sensor network for smart home services using optimal communications. *International Conference on Information Systems and Computer Science, INCISCOS 2017*, 2017-Novem:27–32.

[Mamiş et al., 2013] Mamiş, M. S., Arkan, M., and Keleş, C. (2013). Transmission lines fault location using transient signal spectrum. *International Journal of Electrical Power and Energy Systems*, 53(1):714–718.

[Mazhari and Monsef, 2015] Mazhari, S. M. and Monsef, H. (2015). Distribution Substation Planning. *IEEE SYSTEMS JOURNAL*, 9(4):1396–1408.

[Mazon et al., 2000] Mazon, A. J., Zamora, I., Miñambres, J. F., Zorrozua, M. A., Barandiaran, J. J., and Sagastabeitia, K. (2000). New approach to fault location in two-terminal transmission lines using artificial neural networks. *Electric Power Systems Research*, 56(3):261–266.

[Mehrtash et al., 2020] Mehrtash, M., Kargarian, A., and Conejo, A. J. (2020). Graph-Based Second-Order Cone Programming Model for Resilient Feeder Routing Using GIS Data. *IEEE Transactions on Power Delivery*, 35(4):1999–2010.

[Munoz-Delgado et al., 2016] Munoz-Delgado, G., Contreras, J., and Arroyo, J. M. (2016). Multistage generation and network expansion planning in distribution systems considering uncertainty and reliability. *IEEE Transactions on Power Systems*, 31(5):3715–3728.

[Narayanan et al., 2018] Narayanan, S., Sahoo, S. K., and Makur, A. (2018). Greedy pursuits based gradual weighting strategy for weighted l1-Minimization. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2018-April:4664–4668.

[Niknam et al., 2012] Niknam, T., Kavousifard, A., and Aghaei, J. (2012). Scenario-based multiobjective distribution feeder reconfiguration considering wind power using adaptive modified particle swarm optimisation. *IET Renewable Power Generation*, 6(4):236.

[Raza et al., 2020] Raza, A., Benrabah, A., Alquthami, T., and Akmal, M. (2020). A review of fault diagnosing methods in power transmission systems. *Applied Sciences (Switzerland)*, 10(4).

[Reddy, 2007] Reddy (2007). A wavelet-neuro-fuzzy combined approach for digital relaying of transmission line faults. *Electric Power Components and Systems*, 35(12):1385–1407.

[Reddy, 2008] Reddy (2008). Adaptive-neuro-fuzzy inference system approach for transmission line fault classification and location incorporating effects of power swings. *IET Generation, Transmission and Distribution*, 2(2):235–244.

[Roostaee et al., 2017] Roostaee, S., Thomas, M. S., and Mehfuz, S. (2017). Experimental studies on impedance based fault location for long transmission lines. *Protection and Control of Modern Power Systems*, 2(1).

[Ruiz et al., 2020] Ruiz, M., Inga, E., and Simani, S. (2020). Scalable Electrical Distribution Networks Planning for Medium and Low Voltage Considering Capacity on Transformers and Voltage Drop. *Communications in Computer and Information Science*, 1154 CCIS:75–88.

[Ruiz and Montalvo, 2020] Ruiz, M. and Montalvo, I. (2020). Electrical faults signals restoring based on compressed sensing techniques. *Energies*, 13(8).

[Sahoo and Makur, 2015] Sahoo, S. K. and Makur, A. (2015). Signal recovery from random measurements via extended orthogonal matching pursuit. *IEEE Transactions on Signal Processing*, 63(10):2572–2581.

[Silva et al., 2006] Silva, K. M., Souza, B. A., and Brito, N. S. D. (2006). Fault detection and classification in transmission lines based on wavelet transform and ann. *IEEE Transactions on Power Delivery*, 21(4):2058–2063.

[Sipoli et al., 2014] Sipoli, D., Bosco, J., Junior, A. L., Cláudio, A., and Delbem, B. (2014). Multi-Objective Evolutionary Algorithm for single and multiple fault service restoration in large-scale distribution systems. *Electric Power Systems Research*, 110:144–153.

[Tleis, 2019] Tleis, N. (2019). *Power Systems Modelling and Fault Analysis*. Elsevier Ltd, London, second edition.

[Vahidinasab et al., 2017] Vahidinasab, V., Tabarzadi, M., Arasteh, H., Alizadeh, M. I., Mohammad Beigi, M., Sheikhzadeh, H. R., Mehran, K., and Sepasian, M. S. (2017). Overview of Electric Energy Distribution Networks Expansion Planning. *IEEE Access*, 8:34750–34769.

[Valsan and Swarup, 2009] Valsan, S. P. and Swarup, K. S. (2009). High-speed fault classification in power lines: Theory and fpga-based implementation. *IEEE Transactions on Industrial Electronics*, 56(5):1793–1800.

[Vélez M. et al., 2014] Vélez M., V. M., Hincapíe I., R. A., and Gallego R., R. A. (2014). Low voltage distribution system planning using diversified demand curves. *International Journal of Electrical Power and Energy Systems*, 61:691–700.

[Verma et al., 2020] Verma, M. K., Mukherjee, V., Yadav, V. K., and Ghosh, S. (2020). Constraints for effective distribution network expansion planning: an ample review. *International Journal of Systems Assurance Engineering and Management*, 11(3):531–546.

[Wang and Keerthipala, 1998] Wang, H. and Keerthipala, W. W. L. (1998). Fuzzy-neuro approach to fault classification for transmission line protection. *IEEE Transactions on Power Delivery*, 13(4):1093–1104.

[Wang et al., 2016] Wang, J., Kwon, S., Li, P., and Shim, B. (2016). Recovery of sparse signals via generalized orthogonal matching pursuit: A new analysis. *IEEE Transactions on Signal Processing*, 64(4):1076–1089.

[Xie et al., 2018] Xie, S., Hu, Z., Zhou, D., Li, Y., Kong, S., Lin, W., and Zheng, Y. (2018). Multi-objective active distribution networks expansion planning by scenario-based stochastic programming considering uncertain and random weight of network. *Applied Energy*, 219(March):207–225.

[Yadav and Swetapadma, 2015] Yadav, A. and Swetapadma, A. (2015). A novel transmission line relaying scheme for fault detection and classification using wavelet transform and linear discriminant analysis. *Ain Shams Engineering Journal*, 6(1):199–209.

[Yao et al., 2014] Yao, W., Zhao, J., Wen, F., Dong, Z., Xue, Y., Xu, Y., and Meng, K. (2014). A multi-objective collaborative planning strategy for integrated power distribution and electric vehicle charging systems. *IEEE Transactions on Power Systems*, 29(4):1811–1821.

[Youssef, 2004] Youssef, O. A. S. (2004). Combined fuzzy-logic wavelet-based fault classification technique for power system relaying. *IEEE Transactions on Power Delivery*, 19(2):582–589.

[Zakernezhad et al., 2021] Zakernezhad, H., Nazar, M. S., Shafie-khah, M., and Catalão, J. P. (2021). Multi-level optimization framework for resilient distribution system expansion planning with distributed energy resources. *Energy*, 214.

[Zeng et al., 2014] Zeng, B., Zhang, J., Yang, X., Wang, J., Dong, J., and Zhang, Y. (2014). Integrated planning for transition to low-carbon distribution system with renewable energy generation and demand response. *IEEE Transactions on Power Systems*, 29(3):1153–1165.