ROBUST STRATEGY FOR FAULT LOCATION IN ELECTRIC DISTRIBUTION SYSTEMS

Universidad Tecnológica de Pereira

Ever Julián Correa Tapasco

ROBUST STRATEGY FOR FAULT LOCATION IN ELECTRIC DISTRIBUTION SYSTEMS

Ever Julián Correa Tapasco

This dissertation is submitted for the degree of
Ph.D in Engineering

Pereira, May 14, 2021
UNIVERSIDAD TECNOLÓGICA DE PEREIRA
Doctorate in Engineering

ROBUST STRATEGY FOR FAULT LOCATION IN ELECTRIC DISTRIBUTION SYSTEMS

Main advisor: **Juan José Mora Flórez**, Ph.D.
Universidad Tecnológica de Pereira
Co-advisor: **Sandra Milena Pérez Londoño**, Ph.D.
Universidad Tecnológica de Pereira

Committee

Ph.D. Joaquim Meléndez Frigola, Universitat de Girona, Girona, Spain.
Ph.D. Ernesto Pérez González, Universidad Nacional de Colombia, Medellín, Colombia.
Ph.D. Andrés Ricardo Herrera Orozco, Universidad Tecnológica de Pereira, Pereira, Colombia.

Pereira, May 14, 2021
Doctorado en Ingeniería con énfasis en Ingeniería Eléctrica
Universidad Tecnológica de Pereira
Cl. 27 # 10-02, Pereira, Risaralda
TEL: (+57)(6)3137122
www.utp.edu.co

Universidad Tecnológica de Pereira

Doctoral Program in Engineering

Ph.D Thesis

ROBUST STRATEGY FOR FAULT LOCATION IN ELECTRIC DISTRIBUTION SYSTEMS

Ever Julián Correa Tapasco

Main advisor:
Juan José Mora Flórez, Ph.D.
Coadvisor:
Sandra Milena Pérez Londoño, Ph.D.

Pereira, May 14, 2021

Abstract

Power quality is a relevant aspect in power distribution systems, which considers service continuity and distortion of the current and voltage waveforms, among others; in the first, the fault location plays a fundamental role. Considering the requirement of reliable tools to improve the distribution system operation, this dissertation analyses, on the one hand, the confidence of two fault location approaches versus the waveform distortion and on the other hand, the adequate selection of its adjusting-variables when using learning-based fault locators (LBFL).

The first part of this document presents a generalised application of an impedance-based fault location method for power distribution systems. Also, it analyses the effects of the waveform distortion on its performance, by comparing the location errors in the case of distorted and not distorted measurements of voltages and currents. Different sources of waveform distortion, such as transformer saturation, the low resolution of the measuring devices, harmonics, low sampling frequency, signal noise and DC offset are analysed. Next an LBFL specially customised for distribution systems is tested in an unbalanced 34.5 [kV] distribution feeder, also considering the effects of the waveform distortion on its performance.

On the other hand, as the LBFL requires the adequate selection of its adjusting-variables, as these are strongly interrelated, this dissertation presents a strategy based on an iterative approach to determine their best values. This strategy is based on sampling and learning theories, which helps to improve LBFL performance. This second part is evaluated in two unbalanced distribution feeders, the IEEE34 test system and a local utility 34.5 [kV] system.

According to the obtained results, in the first part, the worst performance of the impedance-based fault location is obtained in the case of distortions caused by current transformer saturation. An additional circumstance of reduced performance is in such cases of the simultaneous presence of low sampling rate and high harmonic distortion. The analysed values of the resolution, signal noise, presence of DC offset and potential transformer saturation, cause a negligible reduction on the locator performance. In the case of LBFL, the lowest LBFL performance is 96.5 [%] in such cases of not distorted measurements of voltage and current. Additionally, when considering polluted inputs, the worse performance is obtained in the case of distortions caused by the low sampling frequency, current transformer saturation and high harmonic distortion.

Finally, in the last part of this dissertation, the results present a fine selection of the LBFL adjusting-variables, due to the high performance in the case of single-phase, phase-to-phase and three-phase faults location. As a result, satisfactory values for the five adjusting-variables are obtained, and the expected LBFL performance is higher than 96 [%], considering diverse operating conditions and fault scenarios.

Acknowledgments

To God and my family, for their tremendous support, especially my life partner Laura Valentina Vanegas-Ortiz and my sister Nora Luz Correa-Tapasco.

This work was supported by the National Scholarship Program Doctorates of the Colombian Ministry of Science, Technology and Innovation (Minciencias), by calling contest 617-2013. Additionally, it was partially supported by the Universidad Tecnológica de Pereira (UTP), research project 6-20-6. The contained work presented in this dissertation was developed in the ICE3 research group.

This research adventure is successfully finished thanks to the collaboration of many people who give me their unconditional help. I especially highlight the director of the Research group on power quality and stability (ICE3) and principal advisor of the work: Juan José Mora-Flórez. I also highlight the collaboration of the co-advisor Sandra Milena Pérez-Londoño and the members of the research group ICE3, especially the former colleague Walter Gil-González.

List of abbreviations

This is a short list of the main abbreviations used in this dissertation.

CT	:	Current transformers.
DC	:	Direct current.
DCO	:	Direct current offset.
DB	:	Database.
DS	:	Decimal-scaling.
EF	:	Equivalent feeder.
f	:	Features obtained from the voltage and current signals.
F	:	Fault/faulted node.
FN	:	Final/farthest node.
H	:	Harmonics.
LBFL	:	Learning-based fault locator.
LS	:	Line section.
m	:	Fault distance.
MMAD	:	Median and median absolute deviation.
MM	:	Min-max.
n	:	Type of data normalisation.
N	:	Noise.
p	:	Parameters of the classification technique.
P	:	Pre-fault.
PT	:	Potential transformers.
r	:	Size of the representative fault database.
R	:	Resolution of the measuring device.
RBF	:	Radial basis function (Gaussian kernel).
SAIDI	:	System average interruption duration index.
SAIFI	:	System average interruption frecuency index.
Sf	:	Sampling frequency.
SF	:	Sigmoid function.
SVM	:	Support vector machines.
z	:	Definition of the zone division.
ZS	:	Z-score.

Contents

Chapter 1

Introduction

This chapter presents the motivation of this dissertation focused on improving the performance in fault location using impedance-based (IBFL) and learning-based fault locators (LBFL). The research related to the signal disturbance effects on the fault locators and the generalised application strategy to the IBFL, the adjusting process of LBFL, as well as contributions and publications derived from this research are also introduced.

1.1 Motivation

Power quality is one of the main issues in power distribution systems, as this includes the service continuity, waveform distortion, voltage regulation and transient behaviour. Service continuity and waveform distortion of the input voltages and currents might affect equipment connected to the power grid; some would keep operating regardless of changes of the waveform, while others work improperly or stop functioning in the case of service discontinuity [1, 2].

In the case of waveform distortion, most of the research is oriented to determine which cases could be detected and mitigated [1]. On the other hand, service discontinuity is a problem, where the design and the fault management play an important role as solution alternatives in distribution systems [1–10].

As electric power system operators require to maintain high power quality standards, the influence of the waveform distortion in the power continuity management process have to be analysed [5]. Additionally, in the case of distribution systems, continuity is a complex issue due to the presence of faults, which are the leading causes of supply interruptions and then responsible for unacceptable indices (SAIDI and SAIFI) [10–12]. Therefore, fault location strategies are commonly used [13–30]; however, due to the stochastic nature, faults are hardly avoidable [6, 31, 32].

As a result of the exposed, significant research efforts have been conducted to develop fault location methods, aimed to help network operators to improve service continuity by reducing interruption time and cost [6, 12]. Fault location allows improving continuity indexes in three forms. First, speeding up the restoration process; second, enabling a re-connection that reduces the affected area; and third, locating areas with non-permanent faults, where maintenance is required to avoid future faults [6, 33].

1.2 Problem statement

Fault location methods can be classified into two main groups: (*i*) those that use sparse measurements obtained in several system nodes, and (*ii*) those that only use measurements in the single-end substation [10]. According to its requirements fault location methods are also classified as: impedance-based methods [33–37], strategies based on measurements taken throughout the system and classical circuit analysis [38], approaches based on travelling waves [9], learning-based methods [7,8,39–44], and approaches based on the combination of the above strategies [3,5,13,33,36,37,45–49]. It is recognised that impedance-based methods are the most mature class and the subject of most of the literature. The sparse measurement-based algorithms are also based on the fundamental frequency component of the measured current and voltage, but provided by sparse meters installed throughout the network. The third class of methods is based on analysing the high-frequency travelling waves, originated by the fault. The learning-based fault locators use the information in the fault database to determine a possible faulted zone and reduce the multiple estimation problem. Finally, the integrated methods use a combination of the previously explained approaches [6].

Additionally and according to the analysed papers, the effects of distorted voltages and currents on the fault locator performance have not received major attention [25–28, 50], and then the confidence in such cases remains unknown [5, 25, 26, 28, 51]. This is an important aspect to be strictly defined, considering that the measurements of voltages and currents used to define the databases are distorted in most cases. On the other hand, the assumption of ideal sinusoidal waveform as inputs of the fault locators is frequently considered [4,5,22], as presented in [6], more than 45 approaches are analysed. At this point, the first part of the problem addressed in this dissertation arises and is summarised in the following research question: *How is the fault locator performance affected in the case of distorted inputs?.*

Moreover, among the previously presented fault location approaches, there is no evidence of strategies that consider the strong interrelation of all of the adjusting-variables, especially in the case of the learning-based fault location methods. From [13,47,52–54], it is noted that only two adjusting-variables are considered and selected in the locator setting strategy. As a natural consequence, the second part of the problem addressed in this dissertation is defined by the following research question: *How to propose a suitable strategy aimed to select the adjusting-variables of an LBFL?*

Finally, this dissertation contributes to the fault management in electric distribution systems by proposing strategies to improve the fault location strategies.

1.3 State of the art

This section presents a summary of the literature focused on fault location in distribution systems, especially those research that is relevant to solve the proposed research questions. Also, related work to the improvement of fault location methods is shown. In each part, state of the art is compared to the ideas proposed in this dissertation.

A broad classification of fault location methods is found in [6], where the methods are classified into: impedance-based methods [14,15,37], methods based on sparse measurements [18], methods based on travelling waves [19], learning-based methods [21, 22, 55–57] and integrated methods [3, 5, 23, 24, 47]. However, [10] considers the classification problem regarding the number of measurements taken in the system. Finally, in [6] also considers an adjustable fault location method and their adjustment strategies for performance improvement. The development of this dissertation opens the door to the study of this last part of the here proposed research .

1.3.1 Summary of research related to the validation of location techniques

Research is focused on two types of location methods: impedance-based and learning-based. Its evolution is reviewed considering the most influential aspects of the development of the dissertation. Most of the location methods began in 1980 with a classical analysis of electrical circuits [14, 58]. Then in 1998, a detailed method and a simulation tool were implemented as presented in [16, 17, 30, 59, 60]. The above approach had problems with the special size and configuration characteristics of the test system and did not present a general application methodology. Between 1998 and 2009, advances in fault simulation tools with research were made, such as these presented in [17, 28, 29, 32, 61–63]. One of these research in 2009 made it possible to deepen the parallel simulation to reduce the cost in computational time to obtain fault databases [19, 64, 65]. At that time, these did not consider the adjustment of the variables, as a very appropriate option to reduce the computational cost. In 2012, a methodology for the automatic variation of parameters of the distribution system was presented in [14, 66, 67]. In 2013, a sensitivity analysis tool for learning-based methods was presented, which allowed varying some modelling parameters [20, 68]. Finally, in 2016, one of the latest improvements to the tool was presented in which the behaviour of location methods is studied in the face of different operating states that may occur in a typical distribution system. This research determines the modelling parameters that most affect its performance [36, 69, 70]. Until 2017, the effects of distorted voltages and currents on the fault locator performance have not received major attention and also evidenced the absence of research involving the optimal variables adjustment of a locator [1, 2, 5, 25–27].

1.3.2 Summary of research related to generalised fault location strategies

The scope of this dissertation is oriented to impedance-based and learning-based fault location methods.

First, the impedance-based methods operate satisfactorily in radial systems. However, the main problem is the multiple estimations of the location [6, 13, 34, 58, 71]. Additionally, the accuracy of the estimate is highly dependent on a fair distribution system model [6].

Second, the learning-based methods and within these, the most used in recent years are the SVM, which is funded in the statistical and optimisation theories, due to characteristics such as the absence of local minimum, the high capacity for generalisation and the control and use of kernel functions [39, 66, 72, 73]. Additionally, several research papers as these presented in [43, 47, 74–78] describe the SVM as one of the most successful classification learning-based approaches, specially in the case of non-linear separable classes where the RBF kernel is frequently selected due to the low number of required parameters [43, 77, 79–81].

The SVM is applied with great success in the fault location problem in distribution systems, as presented in [82–84]. These research are used to consider only some variables of the problem, such as the type of normalisation, the kernel parameters, and the decomposition/reconstruction scheme. In [85], an improved strategy was proposed in which variations in the load model of the distribution system were additionally involved [15, 86]. In [53, 87] a strategy improvement is proposed in which the zone definition issue is addressed through a process of explosion and implosion. This last study reviews, an iterative strategy that is easy to implement for a distribution utility but with the disadvantage of a high computational cost. Before presenting the need for a robust strategy, it is essential to highlight that the strategies explained above take each variable independently without considering the variable interrelation. Thus, the value they find for each variable is skewed to a particular solution; it could be a local solution and not a general one. As a consequence, other combinations that can lead to a global solution are not considered. Therefore, it is essential to review the strategy and improve it in such a way that optimal results can be obtained with a minimal computational cost compared to the previous research presented in 2016. The proposed strategy must be robust due to the consideration of the following variables: database size, feature set, system zone definition, kernel parameters, and the training and validation process [3, 43, 44, 74, 75, 77]. The variables are closely related to each other. Therefore, the problem becomes more complex, and the proposed strategy must allow an adequate model to

be obtained by contemplating that relationship throughout the iterative process.

1.3.3 Summary of research related to database pre-processing

Database pre-processing is a topic of great interest when working with large amounts of information [88]. Furthermore, the use of supervised machine learning requires relevant and significant information since the computational cost is directly proportional to the size of the database. In this way, an adequate database is essential in a robust strategy because its quality impacts the final performance of the algorithm and the representation of real conditions [87]. Therefore, it is essential to explore strategies that allow the efficient reduction of the database to obtain the minimum database that allows low computational cost with adequate performance. Until now, due to the characteristics of the fault location problem, the data sampling strategy based on Latin hypercube has been used [67, 89].

The sampling methods consist of estimating population parameters from the study of the model. These methods include generation and exploration of each parameter in a given space for the input variable (i), and their main objective is to obtain a vector of outputs of the model $y(i)$. Latin hypercube sampling does not depend on the mathematical model and guarantees the excellent performance of the sensitivity techniques used in these studies [61, 67].

The Latin hypercube sampling technique is widely used in the specialised research [28,60,67–69,85,89–91]. The technique helps to reduce the number of data needed to represent a problem, thus making machine learning algorithms require less effort to achieve satisfactory results [60, 87].

1.4 Objectives

1.4.1 General aim

To develop and validate a high-performance fault locator for distribution systems.

1.4.2 Specific objectives

(a) To determine the minimum waveform distortion requirements at the input signals of current and voltage, in the case of an impedance-based and a learning-based fault locator.

(b) To propose a learning-based fault locator for distribution systems using an adjustment strategy aimed to define the best instances of the interrelated variables and then improve the performance.

(c) To validate the effect of proposed adjusting strategy of in performance of a learning-based fault locator.

1.5 Scope and main contributions

1.5.1 Scope

This dissertation focuses on two aspects concerning fault location. The first is the effect of distortions at the input signals on the fault locator performance. The second is the development and validation of a robust fault location strategy that involves adjusting different variables to improve locator performance. This research is delimited as follows.

- The maximum tolerable distortions of the input signal to the locators are determined for two fault location methods: one impedance-based and one learning-based.

- In the case of the impedance-based fault locator, the IEEE34 nodes test system is used. This system has been used for several studies in distribution systems, as presented in [50, 92].
- In the case of the learning-based method, the 34.5 [kV] test system with 75 nodes, is used. This system has been used for previous fault location studies in electric power distribution systems, then comparisons can be done, with [42, 45, 53, 54, 69, 85, 87] .
- The robust adjustment strategy is proposed for the learning-based method where several adjusting variables have to be selected.
- The robust strategy is evaluated in the IEEE34 nodes test system and the 34.5 [kV] test system.
- The case proposed in this dissertation considers the most challenging situation, where the measurements are only at the main substation.
- The core of the learning-based fault locator is the SVM. Other techniques are not considered due to the advantages demonstrated in previous studies; however, the proposed strategy can be extended to these other techniques.
- Unbalanced load and impedance distribution systems are considered.
- AC radial distribution feeders are considered.

1.5.2 Contributions

The main contributions as the following present:

- A generalised application methodology as the here analysed, helps to use most of the impedance-based fault location methods in real distribution systems. Its fundamental concepts are the definition of the equivalent feeders and the actualisation of voltages and currents at each line section.
- In such cases where distorted measurements of voltage and current are considered as inputs, it is demonstrated that the CT saturation is the most significant distortion source in impedance-based fault location methods.
- The result analysis prove a significant effect of the sampling frequency and harmonics distortion on the impedance-based fault locator performance.
- According to the evaluation of the LBFL performance, in such cases where distorted measurements of voltage and current are considered as inputs, it is established that the sampling frequency is the most influential distortion source.
- The result analysis demonstrates a significant effect of the CT saturation and harmonics distortion on the LBFL performance.
- The proposed analysis in section 3.1 and 3.2 helps to determine the minimum requirements of the fault location methods, regarding the main sources of signal distortion in power distribution systems.
- The robust strategy defines the adjusting-variables of an LBFL. These variables are firmly related, and the strategy considers an iterative approach based on learning and sampling theory to determine their best values.

- The development of the robust strategy creates a sequence of logical steps that are used successfully for the implementation of a LBFL.
- The best values for five adjusting-variables are selected, and the expected LBFL performance is higher than 96 [%] for all fault types.
- The system restoration time can be considerably reduced using the proposed fault location strategies, then the duration and frequency rates of interruptions are reduced.
- The proposed approaches, improve the fault locator performance and then increases the power continuity in the existing distribution systems. This simple strategy enhances the SAIFI and SAIDI indexes.
- The here presented fault location alternatives are the starting point to define the actions to deal with the service continuity, as these related to the distribution system reconfiguration and corrective maintenance tasks.
- The obtained performance is high, and the cost is low, then the proposed strategies constitute an affordable solution for most of the distribution utilities.

1.6 List of publications

Next, the principal results derived from this dissertation are presented.

1.6.1 Journal papers derived from this dissertation

(a) **E. Correa-Tapasco**, and J. Mora-Flórez, and S. Pérez-Londoño, "Robustness of a generalised impedance based fault locator considering distorted measurements", in Electric Power Systems Research. Vol. 154, pp. 234-244, **2018**. https://www.sciencedirect.com/science/article/abs/pii/S0378779617303620. **SJR 2018 Q1**.

(b) **E. Correa-Tapasco**, and J. Mora-Flórez, and S. Pérez-Londoño, "Performance analysis of a learning structured fault locator for distribution systems in the case of polluted inputs", in Electric Power Systems Research. Vol. 166, pp. 1-8, **2019**. https://www.sciencedirect.com/science/article/abs/pii/S0378779618303092. **SJR 2019 Q1**.

(c) **E. Correa-Tapasco**, and J. Mora-Florez, and S. Pérez-Londoño, "Comparison of measurement features used as inputs in a learning based fault location method for power distribution systems".in "Revista UIS Ingenierías". Vol. 18, no. 1, pp. 73-80, **2019**. https://revistas.uis.edu.co/index.php/revistauisingenierias/article/view/8310. **B1**.

(d) **E. Correa-Tapasco**, and J. Mora-Flórez, and S. Pérez-Londoño, "A general strategy to define the adjusting-variables of a learning-based fault locator". **R1 - Submitted for evaluation in a Q1 Journal**.

1.6.2 Currently working paper

E. Correa-Tapasco, and I. Olmos-Pineda, and D. Pinto-Avendaño, and J. Mora-Flórez, *Performance improvement of a fault locator based on learning using database reduction through an unsupervised machine*

learning strategy. This is the result of collaborative work between the research group ICE3 of the Technological University of Pereira (UTP), and the research group of the faculty of computer science, located in the Benemérita Autonomous University of Puebla (BUAP), Mexico.

1.7 Outline

This dissertation is organised as follows.

Chapter 2 reviews the main theoretical concepts related to this dissertation. The basic concepts of impedance-based methods and learning based fault location methods are presented.

Chapter 3 presents the analysis of the performance of two structured fault locators in the case of signal distortion, one impedance-based and one learning-based. The behaviour of each locator is explained independently, and the result analysis is shown.

Chapter 4 presents the design of the robust strategy for adjusting the variables of a learning-based fault locator. The strategy is described considering the proposed stages. The test is carried out in two different distribution systems.

Chapter 5 presents the conclusions and summarises the future work.

Appendix A is devoted to reporting the most significant results obtained in the additional tests performed with the robust strategy presented in chapter 4.

Chapter 2

Basic theoretical aspects

This chapter presents the basic aspects used in this dissertation. Detailed information is out of the scope of this dissertation but it will be obtained at the provided references.

There are a wide variety of methods currently used to locate faults in the transmission network (e.g., [93]). However, fault location in distribution networks faces new problems compared to the same task on transmission lines. Transmission lines are mostly equipped with dedicated protections, metering devices, and fault locators.

In contrast, distribution networks usually have laterals and load taps along their lines, which complicates the fault location procedure. Some of the fault location problems and challenges in distribution networks are: a) Geographical dispersion of distribution networks over a wide area; b) Existence of non-homogeneous lines; c) Presence of laterals, load taps and, sometimes, single-phase and two-phase loads; d) Limited measurements, typically only available in substations e) Dynamic topology of distribution networks; f) The effect of fault resistance which is usually not negligible; g) Location of multiple faults in distribution networks due to the presence of several branches [6].

Considering the above-mentioned problems and limitations,a variety of fault location methods have been proposed specifically for fault location in the distribution systems. Based on the required inputs and their core idea, distribution network fault location methods can be classified into impedance-based methods [33–37] and learning-based methods [7, 8, 39–44].

The former find a distance to the fault and the latter a fault zone. Joining the characteristics of these two methods can remove most of the above-mentioned problems and limitations.

2.1 Impedance-based fault location methods

The impedance-based fault location methods are the most mature class and the subject of most of the literature. This methods consider measurements of voltage and current at a single end and the model of the power distribution system. Unlike the transmission lines, distribution feeders usually have many intermediate loads,while voltage and current measurements are available only at the substation. Therefore, the impedance-based methods start the fault location process from the first line section, and iteratively solve the equations which describe fault steady state condition for all line section, one by one, to make an estimation of the distance to the fault. The main steps of the fault location method used are described below [6].

2.1.1 Fault type estimation

The fault type is required to estimate the distance to the faulted node. To determine the fault type, a strategy based on the phase and the zero-sequence over-current is used [59]. This strategy proposes a comparison between the measured phase currents and a defined threshold current, to determine the faulted phases. Additionally, a comparison between the estimated zero sequence current and the ground threshold current is performed, to determine the presence of ground faults.

2.1.2 Fault distance estimation and fault resistance estimation

The power distribution systems in Figs. 2.1 and 2.2 represent the un-faulted and faulted feeders, respectively. The following variables are defined: $V^{\beta}_{(x)}$ identifies the phase voltages at node (x); $I^{\beta}_{(x,y)}$ corresponds to the phase currents flowing from node (x) to node (y); $I^{\beta}_{(Y_x)}$ is the phase currents in the load admittance $Y_{(x)}$; $Z_{L(x,y)}$ identifies the phase impedance of the line section between nodes (x) and (y); $Z_{c(dy)}$ corresponds to the phase equivalent impedance downwards node (y); variable β is used to define the pre-fault (P) or fault (F) steady states and m is the per unit fault distance from node x to the faulted node F. The $Z_{c(dy)}$ in Fig. 2.2 is obtained from Fig. 2.1 using Eq. (2.1).

$$Z_{c(dy)} = \frac{V^{P}_{(x)}}{I^{P}_{(x,y)}} - Z_{L(x,y)} \tag{2.1}$$

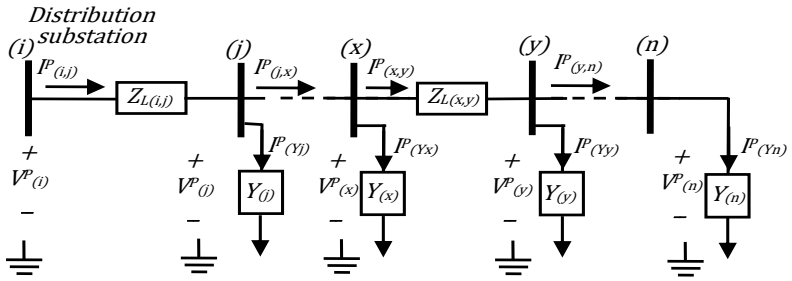

Figure 2.1: Equivalent model of an unfaulted power distribution feeder.

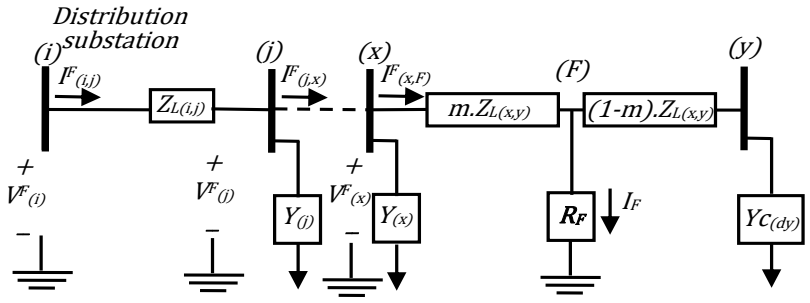

Figure 2.2: Equivalent model of a faulted power distribution feeder.

The estimated impedance from the node x (Z_{est}), in the faulted power system in Fig. 2.2 is given by Eq. (2.2). It contains three unknowns (m, R_F and I_F).

$$Z_{est} = \frac{V^F_{(x)}}{I^F_{(x,F)}} = mZ_{L(x,y)} + R_F \frac{I_F}{I^F_{(x,F)}} \tag{2.2}$$

Additionally, k_s is defined as the ratio between the fault current I_F and the current at the fault steady state $I^F_{(x,F)}$ estimated at the node x, as is given by Eq. (2.3).

$$k_s = \frac{I_F}{I^F_{(x,F)}} \tag{2.3}$$

One of the differences of the proposed method with the basic reactance approach [45,94], is the consideration of the angle of the fault impedance, given by the angle of k_s. To consider the variations of the power distribution system during fault and pre-fault conditions, the superimposed component concept is used, which is represented by the equivalent circuit shown in Fig. 2.3 [35,94].

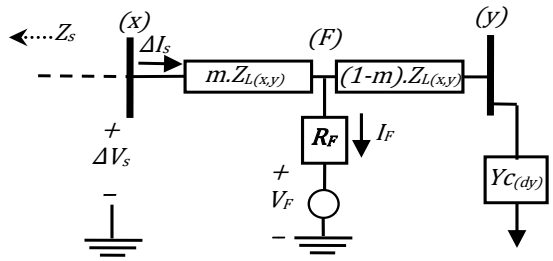

Figure 2.3: Equivalent superimposed circuit for the method.

The method is based on the estimation of the fault impedance as seen from both ends of the faulted line section, it uses the pre-fault (sub-index P) and fault steady states (sub-index F) of voltages and currents. From Fig. 2.3, the impedance (Z_s) upstream node x, is estimated using Eq. (2.4).

$$Z_s = -\frac{\Delta V_s}{\Delta I_s} \tag{2.4}$$

Where:

$$\Delta V_s = V^F_{(x)} - V^P_{(x)} \; and \; \Delta I_s = I^F_{(x,F)} - I^P_{(x,F)} = |\Delta I_s| \angle \lambda_s \tag{2.5}$$

From the circuital analysis of the superimposed power system in Fig. 2.3, two voltage equations are obtained as is presented in Eq. (2.6).

$$\begin{gathered} \Delta I_s(Z_s + mZ_{L(x,y)}) + I_F R_F = V_F \\ [Z_{c(dy)} + (1-m)Z_{L(x,y)}] \cdot (\Delta I_s - I_F) - I_F R_F = V_F \end{gathered} \tag{2.6}$$

Solving the set of equations in (2.6), the fault current I_F is obtained using Eq. (2.7).

$$I_F = \Delta I_s \cdot \left(\frac{Z_{c(dy)} + Z_s + Z_{L(x,y)}}{Z_{c(dy)+(1-m)Z_{L(x,y)}}} \right) = \left| \Delta I_s \cdot \left(\frac{Z_{c(dy)} + Z_s + Z_{L(x,y)}}{Z_{c(dy)+(1-m)Z_{L(x,y)}} \right) \right| \angle \beta_s \tag{2.7}$$

Additionally, from Eqs. (2.2) and (2.7), the Eq. (2.8) is obtained.

$$\begin{aligned} & m^2 - m\frac{\left(V^F_{(x)}Z_{L(x,y)} + Z_{L(x,y)}I^F_{(x,F)}Z_c + (Z_{L(x,y)})^2 I^F_{(x,y)}\right)}{(Z_{L(x,y)})^2 I^F_{(x,F)}} + V^F_{(x)}\frac{\left(Z_{c(dy)} + Z_{L(x,y)}\right)}{(Z_{L(x,y)})^2 I^F_{(x,F)}} \\ & = R_F \Delta I_s \frac{\left(Z_{c(dy)} + Z_s + Z_{L(x,y)}\right)}{\left(Z_{L(x,y)}\right)^2 I^F_{(x,F)}} \end{aligned} \tag{2.8}$$

The Eq. (2.8) is a complex quadratic equation with two unknowns m and R_F. Solving this equation by using its real and imaginary parts, the fault distance is obtained as a per-unit value of the analysed line section.

Table 2.1 shows the values to be considered in (2.8) for pre-fault and fault steady states of voltage and current, in the case of each fault type.

Fault type	***Pre-fault voltage*** $V^P_{(x)(\alpha)}$	***Pre-fault current*** $\mathrm{I}^P_{(x,y)(\alpha)}$	***Fault voltage*** $\mathrm{V}^F_{(x)(\alpha)}$	***Fault current*** $\mathrm{I}^F_{(x,F)(\alpha)}$	***Superimposed current*** ΔI_s	***Superimposed voltage*** ΔV_s
A-G	$V^P_{(x)(A)}$	$\mathrm{I}^P_{(x,y)(A)} + kI^P_{(x,y)(0)}$	$\mathrm{V}^F_{(x)(A)}$	$\mathrm{I}^F_{(x,y)(A)} + kI^F_{(x,y)(0)}$	$3\Delta I^1_{s(A)}$	$3\Delta V^1_{s(A)}$
B-G	$\mathrm{V}^P_{(x)(B)}$	$\mathrm{I}^P_{(x,y)(B)} + kI^P_{(x,y)(0)}$	$\mathrm{V}^F_{(x)(B)}$	$\mathrm{I}^F_{(x,y)(B)} + kI^F_{(x,y)(0)}$	$3\Delta I^1_{s(B)}$	$3\Delta V^1_{s(B)}$
C-G	$\mathrm{V}^P_{(x)(C)}$	$\mathrm{I}^P_{(x,y)(C)} + kI^P_{(x,y)(0)}$	$\mathrm{V}^F_{(x)(C)}$	$\mathrm{I}^F_{(x,y)(C)} + kI^F_{(x,y)(0)}$	$3\Delta I^1_{s(C)}$	$3\Delta V^1_{s(C)}$
BC (BC-G)	$\mathrm{V}^P_{(x)(B)} - V^P_{(x)(C)}$	$\mathrm{I}^P_{(x,y)(B)} - I^P_{(x,y)(C)}$	$\mathrm{V}^F_{(x)(B)} - V^F_{(x)(C)}$	$\mathrm{I}^F_{(x,y)(B)} - I^F_{(x,y)(C)}$	$\Delta I^1_{(x,y)(B)} - \Delta I^1_{(x,y)(C)}$	$\Delta V^1_{(x)(B)} - \Delta V^1_{(x)(C)}$
AB (AB-G)	$\mathrm{V}^P_{(x)(A)} - V^P_{(x)(B)}$	$\mathrm{I}^P_{(x,y)(A)} - I^P_{(x,y)(B)}$	$\mathrm{V}^F_{(x)(A)} - V^F_{(x)(B)}$	$\mathrm{I}^F_{(x,y)(A)} - I^F_{(x,y)(B)}$	$\Delta I^1_{(x,y)(A)} - \Delta I^1_{(x,y)(B)}$	$\Delta V^1_{(x)(A)} - \Delta V^1_{(x)(B)}$
CA (CA-G)	$\mathrm{V}^P_{(x)(C)} - V^P_{(x)(A)}$	$\mathrm{I}^P_{(x,y)(C)} - I^P_{(x,y)(A)}$	$\mathrm{V}^F_{(x)(C)} - V^F_{(x)(A)}$	$\mathrm{I}^F_{(x,y)(C)} - I^F_{(x,y)(A)}$	$\Delta I^1_{(x,y)(C)} - \Delta I^1_{(x,y)(A)}$	$\Delta V^1_{(x)(C)} - \Delta V^1_{(x)(A)}$
ABC (ABC-G)	$\mathrm{V}^P_{(x)(A)} - V^P_{(x)(B)}$	$\mathrm{I}^P_{(x,y)(A)} - I^P_{(x,y)(B)}$	$\mathrm{V}^F_{(x)(A)} - V^F_{(x)(B)}$	$\mathrm{I}^F_{(x,y)(A)} - I^F_{(x,y)(B)}$	$\Delta I^1_{(x,y)(A)} - \Delta I^1_{(x,y)(B)}$	$\Delta V^1_{(x)(A)} - \Delta V^1_{(x)(B)}$

Table 2.1: Voltages and currents for different fault types [94].

Where variables in Table 2.1 are defined as:

α : Index used to identify the considered phase (A, B or C).

$\Delta I^1_{(x,y)(\alpha)} = I^{F1}_{(x,y)(\alpha)} - I^{P1}_{(x,y)(\alpha)}$: Superimposed current of positive sequence in phase α (A, B or C).

$\Delta V^1_{(x)(\alpha)} = V^{F1}_{(x)(\alpha)} - V^{P1}_{(x)(\alpha)}$: Superimposed voltage of positive sequence in phase α (A, B or C).

Finally, the positive and zero sequence line impedances are used to calculate the current derivation factor caused by the unbalance of single-phase fault, as is given in Eq. (2.9).

$$k = \frac{Z^0_{L(x,y)} - Z^1_{L(x,y)}}{Z^1_{L(x,y)}} \tag{2.9}$$

2.2 Learning-based fault location methods

The LBFL, as the here analysed, employ databases composed by the fault location and the voltage and current phasors, obtained during pre-fault (P) and fault steady stages (F), and are aimed to establish a relation between this information. These learning-based relation, which in this case is called support vectors, is next used to obtain the fault location in the case of analysing new faults. These LBFL methods determine the faulted zone and contribute on the elimination of multiple estimation of the fault location, which is a common problem in fault location approaches [95].

LBFL employ the big-data concepts applied in modelling engineering problems [7, 8, 39, 40, 42]. In the exposed approach, the fault location question is considered as a classification problem, where each fault (x : input) is related to a faulted zone of the analysed distribution system (y : output or class), as is depicted in Fig. 2.4.

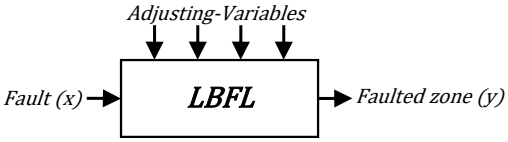

Figure 2.4: General structure of LBFL.

The LBFL considers several adjusting-variables which have to be adequately selected to guarantee the excellent performance of the fault locator. These adjusting-variables will be considered mainly in chapter 4 of this dissertation and are defined as: the features obtained from the voltage and current signals (f), the type of data normalisation (n), the parameters of the classification technique (p), the definition of the zone division (z) and the size of the representative fault database (r). The basics of each one of these adjusting-variables are following described, for use mainly in the robust strategy of chapter 4.

2.2.1 Features obtained from the voltage and current signals at the fault database (f)

In some studies such as [55, 71, 96], the features obtained from the voltage and current signals at the fault database are known as descriptors. The learning-based approaches require of databases which in this case are fault registers obtained from the relays. These registers contain some cycles, during pre-fault and fault conditions, of three-phase current and voltage signals. Additionally, measured signals normally contain thousands of samples, then, several features (f) are defined to characterise each fault and to reduce the computation time. The features which characterise the fault database, obtained using Fast Fourier Transform, are the magnitudes (Mag) and angles (Ang) of three currents and three voltages phasors, from pre-fault ($\mathbb{V}_{pre-fault,j}$ and $\mathbb{I}_{pre-fault,j}$), and fault steady states ($\mathbb{V}_{fault,j}$ and $\mathbb{I}_{fault,j}$), for each j fault register. In Eq.(2.10) the eighth sets of three-phase features obtained for voltage are presented; these defined for current

are identically obtained using the current signals and corresponds to features from $f_{(j,9)}$ to $f_{(j,16)}$. These features are selected due to the results reported in [42, 54].

$$\begin{aligned} f_{(j,1)} = Mag\left(\mathbb{V}_{pre-fault,j}\right); \quad f_{(j,2)} = Mag\left(\mathbb{V}_{fault,j}\right) \\ f_{(j,3)} = Ang\left(\mathbb{V}_{pre-fault,j}\right); \quad f_{(j,4)} = Ang\left(\mathbb{V}_{fault,j}\right) \\ f_{(j,5)} = f_{(j,1)} - f_{(j,2)}; \quad f_{(j,6)} = f_{(j,3)} - f_{(j,4)} \\ f_{(j,7)} = Mag\left(\mathbb{V}_{pre-fault,j} - \mathbb{V}_{fault,j}\right); \quad f_{(j,8)} = Ang\left(\mathbb{V}_{pre-fault,j} - \mathbb{V}_{fault,j}\right) \end{aligned} \tag{2.10}$$

A total of 48 features are proposed (16 for each phase), then a database of $J \times 48$ values is obtained, where J is the total number of faults.

2.2.2 Type of data normalisation (n)

The learning-based approaches usually require a normalisation-process to deal with features (f) on a different scale; otherwise, it may lead to a dilution in effectiveness of an equally important feature (on lower scale), because of another one with values on larger scale. In the case of multiple features as the here presented, fault current variations which are normally very high compared to voltage variations are used. Among other things, avoid normalisation may lead to poor data models while performing data mining operations [52, 53]. Some approaches such as Min-Max (MM) in Eq. (2.11), Z-score (ZS) in Eq. (2.12), Decimal-Scaling (DS) in Eq. (2.13), Median and median absolute deviation (MMAD) in Eq.(2.14) and Sigmoid function (SF) in Eq.(2.15) normalisation strategies (n), are normally used [40, 49, 52, 53].

$$f'_{(j,k)} = \frac{f_{(j,k)} - min\left[f_{(j,k)}\ \forall j\right]}{max\left[f_{(j,k)} \forall j\right] - min\left[f_{(j,k)}\ \forall j\right]} \tag{2.11}$$

$$f'_{(j,k)} = \frac{f_{kj} - mean\left[f_{(j,k)}\ \forall j\right]}{std\left[f_{(j,k)}\ \forall j\right]} \tag{2.12}$$

$$f'_{(j,k)} = \frac{f_{(j,k)}}{10^S}, \quad S \in \left(\left|f_{(j,k)}\right|\right) < 1 \tag{2.13}$$

$$f'_{(j,k)} = \frac{f_{(j,k)} - mean\left[f_{(j,k)}\ \forall j\right]}{mean\left(\left|f_{(j,k)} - mean\left[f_{(j,k)}\ \forall j\right]\right|\right)} \tag{2.14}$$

$$f'_{(j,k)} = \frac{1}{1 + e^{-T_{(j,k)}}}, \quad T_{(j,k)} = \frac{f_{(j,k)} - mean\left[f_{(j,k)}\ \forall j\right]}{std\left[f_{(j,k)}\ \forall j\right]} \tag{2.15}$$

From the previous equations, feature $f'_{(j,k)}$ corresponds to the normalised feature at the three-column sub-matrix, corresponding to the feature k. However, for simplicity, the normalised feature is also presented as f along with this dissertation.

2.2.3 Parameters of the classification technique (p)

As specified at the beginning part of this chapter, fault location is defined as a classification problem, aimed to establish a relation between the inputs or features (f) and the classes or zones (z). Then, the core of the proposed LBFL is the Support Vector Machines (SVM), used as a classification technique. The SVM is based on quadratic programming, several constraints and kernel transformations; this is a learning algorithm successfully applied in many classification and regression problems in engineering applications [40, 42, 43, 47].

The optimisation problem in Eq. (2.16) has an error penalisation constant (C), f_j represents each fault consisting of features in a N dimensional space, y_j represents a target value or class (zone) [42]. Relaxation variables (ξ_j) are used to define the "soft margin", while weight (w) and bias (b) are the function controls [66].

$$\begin{aligned} &\min_{w,b} \frac{1}{2}(w \cdot w) + C\sum_{j=1}^{n} \xi_j \\ &subject\ \ to:\ y_j(w \cdot f_j + b) \geq 1 - \xi_j\ \ and\ \ \xi_j \geqslant 0,\ \forall_j \end{aligned} \tag{2.16}$$

Additionally, to the high SVM performance in classification tasks, another advantage is related to the reduced number of parameters to adjust (p). In the case of using the Gaussian RBF kernel in Eq. (2.17), only two parameters are required, the first one is related to the kernel transformation parameters (σ) and the other is the penalisation constant of the optimisation problem (C) [42,66]. In the here exposed approach, the strategy proposed in [66] is used as a complement to determine the SVM parameters.

The SVM parameters obtained by [66], use a redefinition of the optimisation problem, which includes weight (w), bias (b), penalisation constant (C) and the kernel parameter (σ). The new optimisation problem maximises the angles defined by each support vector and a reference point defined at the same n-dimensional representation space. The optimisation algorithm uses the gradient descent method.

$$F(f_j, y) = exp\left(-\frac{\|f_j - y\|^2}{2\sigma^2}\right) \tag{2.17}$$

2.2.4 Definition of the zone division (z)

As a classification tool, the LBFL relates each fault with a faulted zone, which is a specific section of the analysed distribution system. In this task and in the case of only defining one zone at the distribution system, the LBFL performance is always perfect. However, as the number of zones increases, the performance decreases, establishing a compromise between the LBFL performance and the number of zones. On the other hand, the division in zones of the distribution system also considers its topology, the distance between consecutive poles, the line type (overhead or underground), the conductor characteristics, the load nodes, location of protective devices and recommendations of the maintenance staff [12].

As a consequence of the previous exposed, several zone divisions (z) are suggested. One of them is normally proposed by the utility operator, and the others are defined as variations of the former, which are tested to determine the possible better performance in fault location.

2.2.5 Size of the representative fault database (r)

The fault database contains several faults records (J), obtained under different operating conditions which consider variations of the load size, the main substation voltage, the system frequency and the line section length. Typically, the databases used in data-mining contains hundreds of data, which are used to train and test the learning-based approaches. In this dissertation, one representative database is defined as the smaller database, which is a portrayal of the information available at the former.

To obtain the representative database, a sampling strategy and a criterion of the number of data at the representative database, are considered. In the approach proposed in chapter 4, the Latin hyper-cube sampling strategy is proposed due to the adequacy to this experiment design and considering that this ensures that the fault database is sampled evenly, but still with the same probability trend [6, 90]. Additionally, the criterion of the number of data considers a minimum number of fault registers per node at each representative

database. Then, several representative databases (r) are proposed in the methodology presented in chapter 4, by increasing the minimum number of faults per node included in such databases.

2.3 Support vector machines as the core of the LBFL

According to several references as [22,43,47,51,55,72,74–76,97,98], one of the best performance learning-based algorithms to data analysis is the known as SVM, which is based on quadratic programming, several constrains and kernel transformations. This algorithm has been successfully used in many classification and regression problems, therefore the use of SVM as classification technique to assist fault location is here considered [39,51,71]. The main characteristic of the SVM are next described.

2.3.1 SVM in case of linear cases

The SVM basis comes from the statistical learning theory and these are defined initially as a binary classification technique. SVM are also derived from the optimisation of the separation threshold defined by the perceptron in neural networks, the use of kernel methods and the generalisation theories [99]. Classification using SVM requires of training and testing data sets. In the training set, each data (fault) consists of f_j features in a N dimensional space and y_j represents a target value called class or zone, usually denoted as 1 or -1, which in this case corresponds to two faulted zones (z) [42], as in Eq. (2.18) [71,98].

$$f_j \in R^N \; and \; y_j \in \{+1, -1\} \tag{2.18}$$

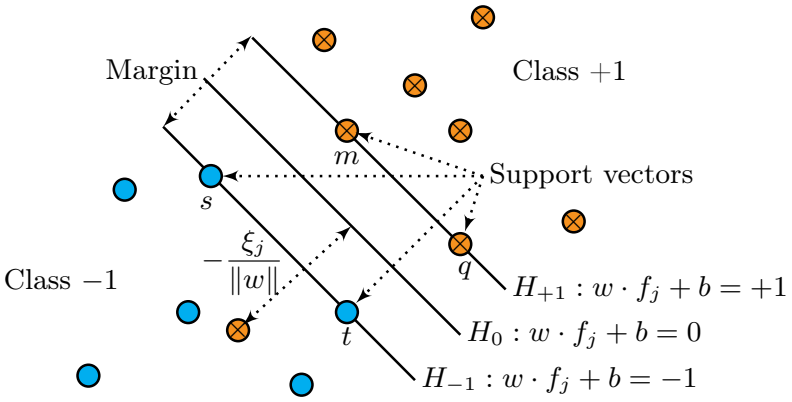

Figure 2.5: Separating hyperplanes in a SVM.

The classifier helps to develop a model, which can predict the class label of unknown data (faulted zone), considering a confidence indicator of performance. Although the SVM algorithm is defined as a binary classification, this is used to classify faults in more than two classes using decomposition and reconstruction methods [43,76,99,100]. The SVM basic algorithm determines an optimal separation hyperplane H given by $y_j = w \cdot f_j + b = 0$, which has the maximum margin to the training nearest pattern, forcing the generalisation of the learning machine as shown in Fig. 2.5 [97, 99]. Weight (w) and bias (b) control the function and those data points that the margin pushes up against are called "*support vectors*" (m,q,s and t). To find the

optimal separation hyperplane, the optimisation problem presented in Eq. (2.19) is solved, considering that margin is inversely proportional to $(w \cdot w)^{\frac{1}{2}}$ [66, 71, 76, 100].

$$\min_{w,b} \frac{1}{2}(w \cdot w), \ subject\ to: y_j(w \cdot f_j + b) \geq 1, \forall_j \tag{2.19}$$

2.3.2 SVM considering soft margin

The basic SVM formulation considers the nonexistence of mixed classes (zones). To deal with this, the SVM are reformulated by considering relaxation variables (ξ) to define the "$soft\ margin$". Thus, the optimisation problem presented in Eq. (2.19) is now given as Eq. (2.16), where C is the error penalisation constant [76,100].

2.3.3 SVM in non-linear cases

To consider non-linear separable classes (zones), it is possible to transform the input into a new higher dimension space, where these are linearly separable. The transformation function $F(\cdot)$ is defined as the inner product of the input data in the original classification space. Linear SVMs are extended to non-linear cases by using an appropriate kernel function [76, 99, 101]. When a Radial Basis Function (RBF) is chosen as kernel function, two parameters (constant C and kernel parameter σ) are required. In this dissertation, the Gaussian RBF kernel presented in Eq. (2.17) is used [43, 77, 81, 99].

2.4 Sources of input signal distortion to the locators

According to IEEE Standard 1159 of 2019 [102], there are three electromagnetic phenomena that affect power quality: the RMS variations, transient disturbances and waveform distortions [35]. The latter phenomena can be classified into two groups: those produced by the measuring devices and those originated by the electromagnetic environment [1, 2, 16, 25–27].

The considered waveform distortions are originated by low sampling frequency (Sf), saturation of measuring transformers (CTs and PTs), resolution of the measuring devise (R), harmonics (H), noise (N) and the presence of DC offset component (DCO) [1, 26, 28, 35, 102–109].

For the proposed analysis of chapter 3, the waveform distortion caused by measuring devices, considers three sources: sampling frequency (Sf), saturation of current and potential transformers (CTs and PTs) and resolution (R) [1, 5, 26–30, 103–105, 110]. The sampling frequency affects the reconstruction of the waveform by the measuring device [28, 103]; saturation occurs when currents exceeding the dynamic operating range of the CT cause magnetisation of the core to be independent of the current and thus produces distortion and reduction in the secondary current [26, 27, 110], and finally, low resolution of measuring devices might conceal relevant information regarding the waveform [104, 105].

Likewise, three additional distortion sources caused by the electromagnetic environment are considered: harmonics (H), noise (N) and DC offset (DCO). The harmonics are caused by the increasing presence of non-linear loads and these produce changes in waveform of both current and voltage. Electrical noise corresponds to undesirable electrical signals, which distort the current and voltage measurements. Finally, DC offset cause not symmetric current and voltage signals [1, 102, 106–109, 111].

Regarding the waveform distortion caused by the equipment, it is important to remark the importance of the cost associated to measuring devices, since better and more accurate measurements are expected after an adequate investment on equipment. Therefore, the utility managing staff is always interested to determine the minimum requirements, to achieve balance between the adequate performance of fault locators and the reasonable cost of the equipment.

2.5 Summary of the chapter

This chapter presented the basic theoretical aspects of the dissertation, mainly based on the use of SVM in fault locators using data mining. Without ruling out the fact that the robust strategy is generalised and can be applied with another classification technique. The basic concepts of impedance-based methods and learning-based fault location methods were also presented. Finally, the most common distortion sources of the input signal to the locators were described.

Chapter 3

Performance analysis of a structured fault locator in the case of waveform distortion

The first part of this chapter presents a generalised application of an impedance-based fault location method for power distribution systems. Also, it analyses the effects of the waveform distortion on its performance, by comparing the location errors in the case of distorted and not distorted measurements of voltages and currents. Different sources of waveform distortion, such as transformer saturation, the low resolution of the measuring devices, harmonics, low sampling frequency, signal noise and DC offset are analysed. Finally, an LBFL specially customised for distribution systems is tested in an unbalanced 34.5 [kV] distribution feeder, also considering the effects of the waveform distortion on its performance.

3.1 Structured impedance-based fault locator considering distorted measurements

The proposed analysis methodology is divided into two sections: the first one is oriented to define a structured strategy based on impedance to determine the fault location in power distribution systems, and the second section is devoted to design the process to determine the influence of the considered distortion sources. The main aspects of the proposed methodology are next described and presented in Fig. 3.1.

3.1.1 Implementation of the generalised strategy for fault location method

The fault location method presented in section 2.1, is used to determine the fault location considering that the faulted line section is identified, e.g. between the faulted nodes (x) and (y) [94]. As in a real distribution system, the faulted line section has to be determined and an equivalent of the feeder have to be proposed, this step is devoted to propose a generalised fault locator application strategy.

The inputs of the fault location method are the measurements of voltages and currents during pre-fault and fault steady states, the fault type and also the equivalent feeders of the power distribution system.

The fault location process starts by the analysis of the first line section ($LS = 1$) of the first selected equivalent feeder. At this line section, the estimation of the fault distance m is performed as is indicated in

Eq. (2.8), considering the fault type. In the case of estimate a value of m between 0 and 1, it is assumed that the actual is the faulted line section, and the distance to the fault in the line section corresponds to the product of m and the length of the analysed line section. Otherwise, voltages and currents have to be calculated at the beginning of the next line section, considering the voltage drop along the actual analysed line section and the currents flowing to tapped loads. To obtain the estimated fault distance, the length of the previous analysed and non-faulted section lined is added to the previous obtained distance according to Eq. (3.1).

$$Estimated\ fault\ distance = m \times Faulted\ line\ section\ length + \sum_{LS=1}^{LS=Last\ unfaulted\ LS} Length_{Ls} \tag{3.1}$$

Next, the main stages of the general strategy are described.

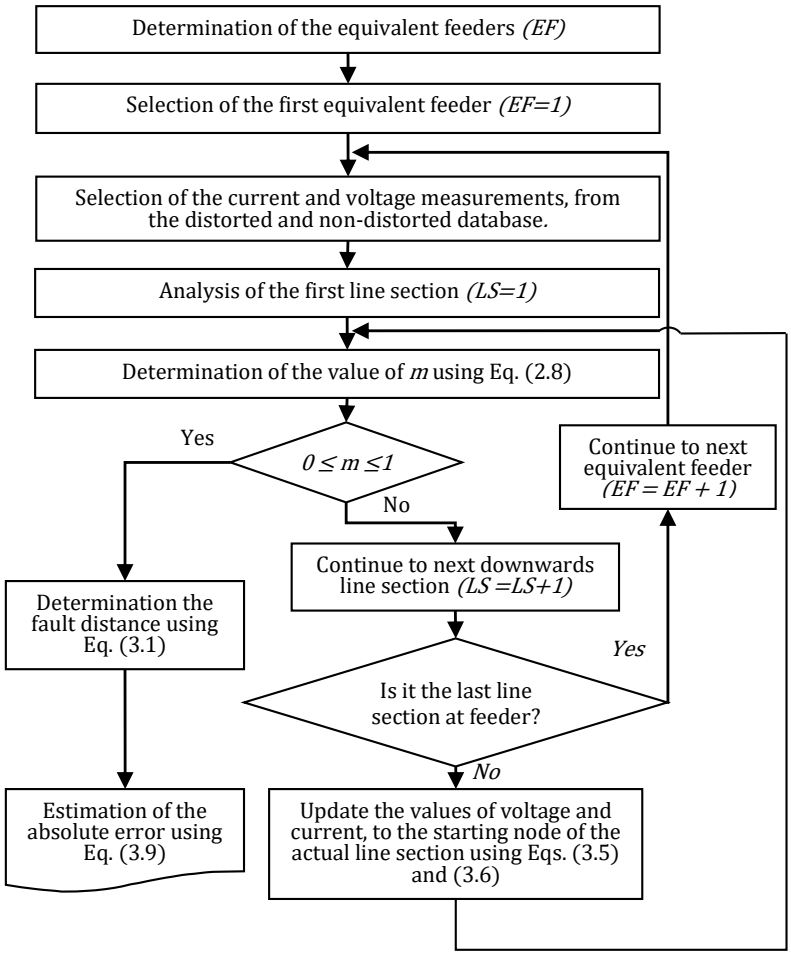

Figure 3.1: Proposed methodology of generalised fault locator.

a. Determination of the equivalent feeders (EF)

Power distribution systems have radial topology and contain several feeders, laterals, tapped loads, different conductor gauges and serial impedances, as shown in Fig. 3.2. From any distribution system it is possible to obtain several equivalent feeders. One equivalent feeder is defined as a simplification of the original power system, which considers one single path from the main distribution substation to one final node (FN). The number of equivalent feeders is equal to the number of final nodes, and then in the case of the distribution system in Fig. 3.2, there are five (5) equivalent feeders (FN1 to FN5). An equivalent feeder, from the main power substation to the final node 1 ($FN1$) is depicted in Fig. 3.3.

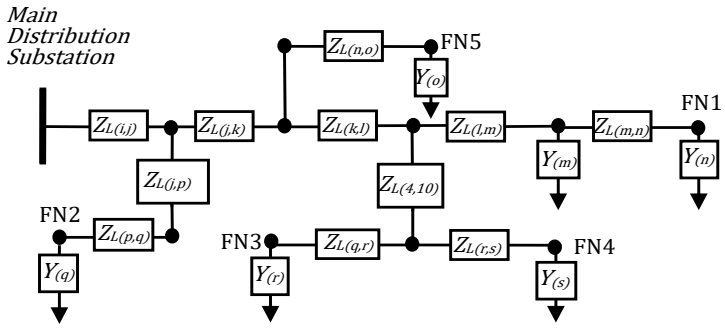

Figure 3.2: Real topology of power distribution system.

The $Y equivalent\ up\ to\ node\ FNx$ in Fig. 3.3. is defined as the inverse of the Thevenin equivalent impedance, seen from the connection node to the farthest node (FNx), at the considered lateral. In general, the equivalent admittance up to node at $(n-1)$ of a generalised power distribution feeder, as the presented in Fig. 3.4, is obtained based on the equivalent admittance up to node (n), using the matrix Eq. (3.2).

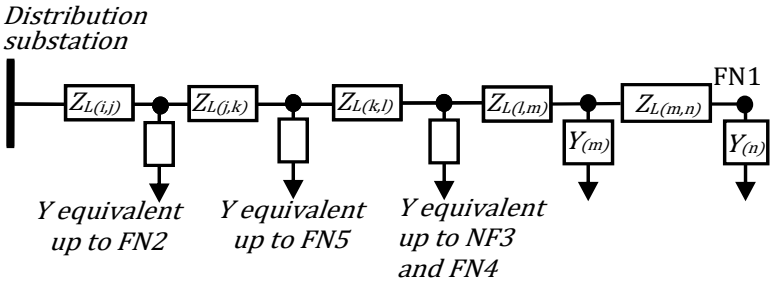

Figure 3.3: Equivalent feeder of power distribution system.

$$Y^{eq}_{(n-1)} = Y_{(n-1)} + \left[Z_{L(n-1)} + \left[Y_{(n)} \right]^{-1} \right]^{-1} \tag{3.2}$$

For each lateral, the loads are represented by their admittance $Y_{(i)}$ while the lines are represented by the serial impedance $Z_{L(i)}$. For any node i, the equivalent admittance is determined by using the matrix Eq. (3.3).

$$Y_{(i)}^{eq} = \begin{cases} Y_{(n)} & if,\ i = n \\ Y_{(i)} + \left[Z_{L(i)} + \left[Y_{(i+1)}^{eq} \right]^{-1} \right]^{-1} & if, 1 \leq i \leq n-1 \end{cases} \tag{3.3}$$

Finally, it is supposed that node n is the farthest node of the lateral and, node 1 is the node at the equivalent feeder where the lateral begins, then the $Y_{(1)}^{eq}$ represents the lumped lateral at the equivalent feeder.

Most of the distribution systems have short lines, where the capacitive effect is negligible, however the fault locator here proposed allows to consider the capacitive effect in medium, long and underground lines. Considering the shunt admittance of the line section in each node of the system as $Y_{LC(i)}/2$ and the loads represented by their admittance $Y_{Load(i)}$, then, the equivalent load admittance at each node $Y_{(i)}$ is obtained using Eq. (3.4). The total equivalent admittance of each lateral is obtained using the matrix Eq. (3.3).

$$Y_{(i)} = Y_{Load(i)} + Y_{LC(i)}/2 \tag{3.4}$$

Other alternatives to consider the capacitive effect are shown in [58] and [112].

The strategy proposed in [17] consider the equivalent systems through the calculation of series and parallel impedances, avoiding the use of the mutual impedances, which is not useful in large coupled distribution systems. The application of a strategy as the proposed in [95], helps to consider the mutual coupling. In the case of the here proposed approach, all of the line and load models consider a 3×3 matrix to include the self and the mutual impedances.

b. Determination of the occurrence and type of fault

To determine the occurrence and type of a fault, two threshold currents are defined following the strategy proposed in [59]. The former is associated with phase faults and is defined as 1.5 times the rated phase current. The latter is associated with ground faults and is defined as 2.0 times the maximum zero sequence current at the rated condition. These threshold currents are defined considering the maximum load magnitude and unbalance, following the same principles of the overcurrent protection setting [113].

The presence of a fault is determined in the case of have phase or zero sequence currents higher than the predefined current threshold. In such case, by comparing these currents, the fault type is identified [59].

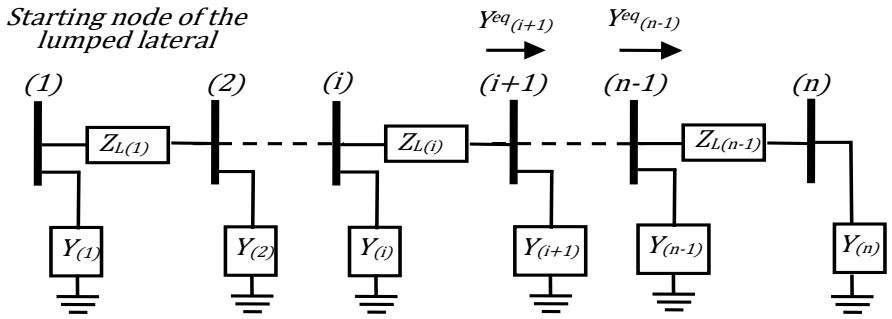

Figure 3.4: Reduction of a lateral by considering the equivalent admittance.

c. Update of currents and voltages at each line section

Normally, the voltage and current measurements are only available at the distribution substation. The fault location method defined in section 2.1 requires of the voltage and currents measurements at the beginning of each line section. Therefore, to calculate the voltages and currents in pre-fault and fault steady states in the remaining nodes of the equivalent radial, an iterative update is performed section by section of the equivalent faulted or unfaulted feeder, using Eqs. (3.5) and (3.6).

$$V^{\beta}_{(j)} = V^{\beta}_{(i)} - I^{\beta}_{(i,j)} Z_{L(i,j)} \tag{3.5}$$

$$I^{\beta}_{(j,x)} = I^{\beta}_{(i,j)} - I^{\beta}_{(Yj)} \tag{3.6}$$

Additionally, to consider the load dynamics, a complex current factor is defined as is presented in Eq. (3.7). Then the current matrix at the load $I^{\beta}_{Y_j}$ is defined as Eq. (3.8). There are other compensation techniques and one of them is in [70].

$$\delta_{(\alpha)} = \frac{I^{P}_{(i,j)}}{I^{rated}_{(i,j)}} \; \forall \; \alpha = Phases \; A, B, C \tag{3.7}$$

$$\left[I^{\beta}_{(Yj)}\right] = \delta_{(\alpha)} \left[I^{\beta}_{(Yj)(\alpha)}\right] \tag{3.8}$$

3.1.2 Distortion sources on the fault location performance

Voltage and current measurements are obtained through the Discrete Fourier Transform (DFT). It is used to find the fundamental magnitude and phase angle of the pre-fault and post-fault signal. The time windows used are fixed of one cycle length. Values for the pre-fault are obtained one cycle previous to the fault cycle. The post-fault values are obtained by sampling the fifth cycle after the fault cycle. There is no adjustment for frequency and angle considering the sampled steady-state signal.

In this step, the voltage and current measurements used to test the fault locator are divided into two groups. The first one is obtained from the non-distorted measurements, which is used as a reference, and the latter considers these distortions defined in section 2.4. The distorted measurements help to determine the influence on the locator performance by comparing the location results to these obtained in the case of non-distorted measurements.

The distorted and non-distorted measurement database contains faults at different values of fault resistance, several nodes of the analysed distribution system and all types of shunt faults. In the case of distorted measurements, sampling frequency, saturation of CTs and PTs, resolution of the measuring device, presence of harmonics, noise and DC offset currents are considered. The magnitudes of these distortions used in test are in compliance with the maximum values allowed by standards and factory settings of the measuring equipment [102, 104–107, 109]. Table 3.1 summarises the distortion source values.

At the specific case of current transformer saturation (CTs), the proposed instances are defined from Fig. 3.5, where "Icc" is the peak value of the maximum short circuit current at the distribution substation of the analysed power system. The distorted measurements are obtained by confine the original signal into the saturation boundaries. At this research, the proposed instances for CT saturation are given as a function of ΔI, which is defined as the difference between the peak values of the maximum short circuit and the rated currents. The selected saturation boundaries are presented in Table 3.1.

In the case of potential transformers (PTs), the saturation threshold is defined on the basis of the rated voltage $Vrated$.

The resolution of the measuring device (R) in Table 3.1 is given by the capability of store information in units of micro, mili and integers of the measured variable [105]. Distortion (H) is obtained by adding the same total harmonic distortion in voltage and current signals, according to Table 3.1 and considering the third, fifth and seventh harmonics [106]. Noise perturbation (N) is included by adding a $200 - kHz$ signal in percentage of the fundamental magnitude of the voltage and current signals [107]. Finally, the DC offset (DCO) is a positive constant added to the voltage and current measurements [111].

Distortion source	**Instances / Analysed scenarios**							**Units / Description**
	1	**2**	**3**	**4**	**5**	**6**	**7**	
Sf	512	256	128	64	32	16	8	Samples/cycle
CTs	No Sat.	(5/6) ΔI	(4/6) ΔI	(3/6) ΔI	-	-	-	ΔI
PTs	No Sat.	1.1 Vrated	1.05 Vrated	-	-	-	-	Vrated
R	100 (Default)	6 (micro)	3 (milli)	0 (integer)	-	-	-	Decimal units
H	0 [%]	1 [%]	3 [%]	5 [%]	-	-	-	THDi = THDv
N	0 [%]	10 [%]	20 [%]	-	-	-	-	[%] of V or I (f=200kHz)
DCO	0 [%]	2.5 [%]	5 [%]	-	-	-	-	[%] of V or I

Table 3.1: Instances of the considered scenarios and magnitudes of each distortion source.

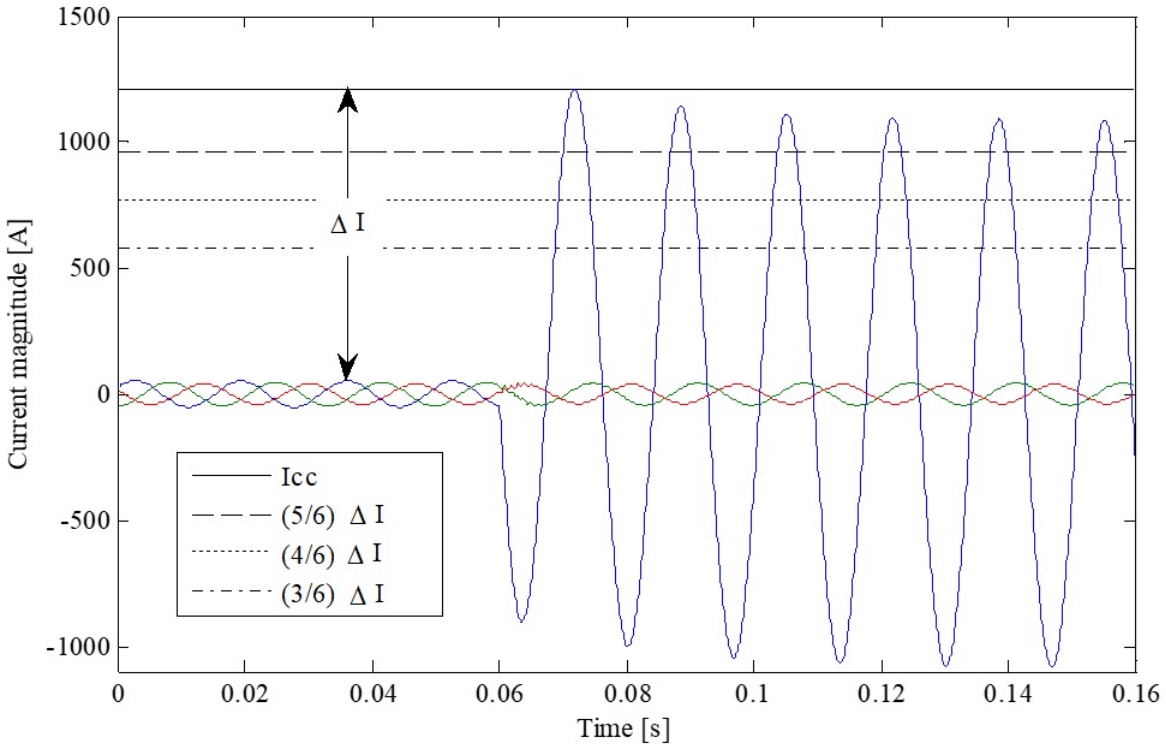

Figure 3.5: Definition of the instances for saturation of the current transformers (CTs).

The number of possible cases of distorted measurements is obtained as the product of the number of instances considered at each distortion source, which according to Table 3.1 corresponds to 12096 ($7 \times 4 \times 3 \times 4 \times 4 \times 3 \times 3$) cases. At the non-distorted measurements, the instance "1" in all of the distortion sources of Table 3.1 is considered.

Finally and having determined the cases of distorted and non-distorted measurements of voltage and current, the fault location process is performed for each case. The objective is to determine these distortion cases which cause minimum or negligible affectation on the performance of the fault locator. The accepted affectation on the fault locator performance is given by an affordable patrolling distance defined by the utility maintenance team, then the affectation analysis is performed by the comparison of the maximum absolute fault location errors and the defined acceptable patrolling thresholds.

3.1.3 Description of the proposed test

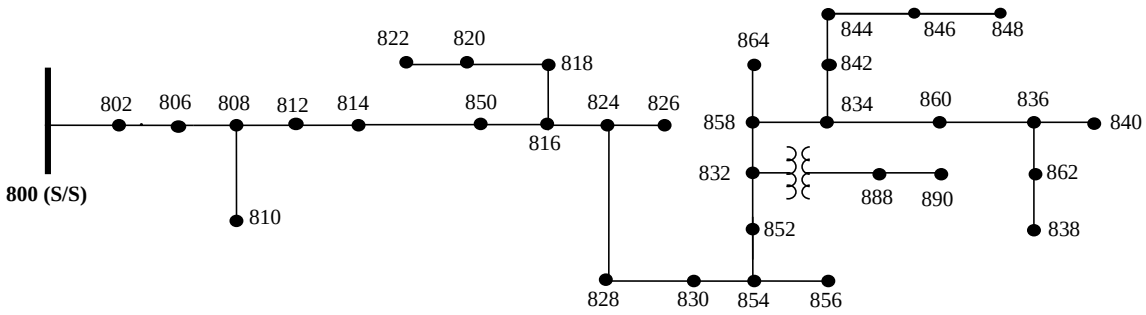

Figure 3.6: IEEE 34-bus test feeder.

The following sections is devoted to report the most significant results obtained in the tests performed. Tests are performed on the IEEE 34-bus test feeder presented in Fig. 3.6 [114].

Initially, test considering non-distorted measurements of voltage and current are performed to identify the greatest absolute error in fault location. This error is used as a validation reference to determine the robustness of the proposed fault location strategy considering distorted measurements.

The fault database contains measurements obtained by considering three types of faults (single phase, phase to phase and three phase faults), three different values of fault resistance (0.0002 Ω, 20 Ω, and 40 Ω.), and four nodes located along the main feeder of the distribution power system (808, 812, 828 and 834) [30].

The fault location absolute error (A_e) is obtained by a comparison of the real fault distance and the estimated fault distance, as is proposed in Eq. (3.9) [20].

$$A_e = \left(\frac{Real\ fault\ distance - Estimated\ fault\ distance}{Total\ equivalent\ feeder\ length} \right) \times 100\ [\%] \tag{3.9}$$

3.1.4 Preliminary tests results

In Table 3.2 a comparison of the performance of the proposed fault locator is presented, considered the previous reported results given in [115]. According to the results, the proposed approach has best performance in most of the cases here analysed.

Fault Distance [km]	Fault type	Estimated error [%]							
		Rf = 0.0002 [Ω]				Rf = 10 [Ω]			
		Proposed Method	You [115]	Salim [95]	Filomena [112]	Proposed Method	You [115]	Salim [95]	Filomena [112]
24.8328	a-g	0.0154	0.1209	0.1100	0.1209	0.3127	0.0278	0.0759	0.0057
	a-b-g	0.0423	0.0906	0.1116	0.0906	0.0725	0.0243	0.1181	0.0180
	a-b-c-g	0.0171	0.1512	0.1653	0.1498	0.0182	0.0843	0.1749	0.0802
36.6522	a-g	0.0092	0.2483	0.1905	0.2483	0.2106	0.2106	0.0393	0.2150
	a-b-g	0.0754	0.2963	0.3894	0.2963	0.0312	0.2428	0.4950	0.2420
	a-b-c-g	0.0039	0.4044	0.4718	0.3978	0.0231	0.3940	0.6186	0.3847
56.8026	a-g	0.0258	0.8707	1.2382	0.8707	0.0458	0.9921	1.6802	0.9926
	a-b-g	0.1223	1.8426	2.1337	1.5831	0.1389	2.1244	2.6025	1.6827
	a-b-c-g	0.0312	1.4412	1.9053	1.4270	0.0531	1.6358	2.4745	1.6178

Table 3.2: Comparison of the proposed method with some currently referenced methods.

According to Table 3.3, the performance of the proposed approach is adequate to the expected application. The maximum reported error is lower than 1.5 [%] in all of the analysed load dynamic cases.

Scenarios	Fault type	Estimated maximum error [%]
Low load [40-80]	a-g	0.9235
	a-b-g	0.7165
	a-b-c-g	0.3441
Average load [80-120]	a-g	0.2606
	a-b-g	0.2530
	a-b-c-g	0.4156
High load [120-150]	a-g	1.4701
	a-b-g	1.1212
	a-b-c-g	0.5229

Table 3.3: Maximum fault location error, considering the load dynamic in three scenarios: low, average and high load.

3.1.5 Fault locator performance results considering non-distorted measurements

The obtained performance of the impedance-based fault locator considering the non-distorted measurements, for each case of shunt faults, is presented in Figs. 3.7, 3.8 and 3.9. Each figure shows 12 estimations of the absolute error calculated using Eq. (3.9), considering three values of fault resistance and four nodes along the main feeder.

The highest absolute error in fault location is 1.2 [%], which corresponds to a single-phase fault with $Rf = 40$ [Ω] (see Fig. 3.7). This value is taken as validation reference error for the proposed methodology, and then these distorted measurements of current and voltages that cause fault location absolute errors near to 1.2 [%] are considered as not severely affected by the analysed distortion.

The maximum absolute error considering non-distorted measurements and phase to phase faults, corresponds to 0.3 [%] in the case of $Rf = 40$ [Ω] (see Fig. 3.8). On the other hand, Fig. 3.9 shows the maximum absolute error considering non-distorted measurements and three phase faults, which corresponds to 0.09 [%] in the same case of $Rf = 40$ [Ω].

An error of 1.2 [%] as validation reference implies that maintenance staff should patrol around 694 meters, forwards or backwards along the main feeder, from the fault location given by the generalised fault locator. In this section, an evaluation of the effects of signal distortion was performed to identify such cases, which cause a minor effect on fault location. To define these minor effects in terms of the fault location absolute error, a threshold of 2.5 [%] is proposed, i.e. a maximum backward and forward patrolling distance of approximately 1.5 $[km]$ is the admissible uncertainty distance along the equivalent feeder.

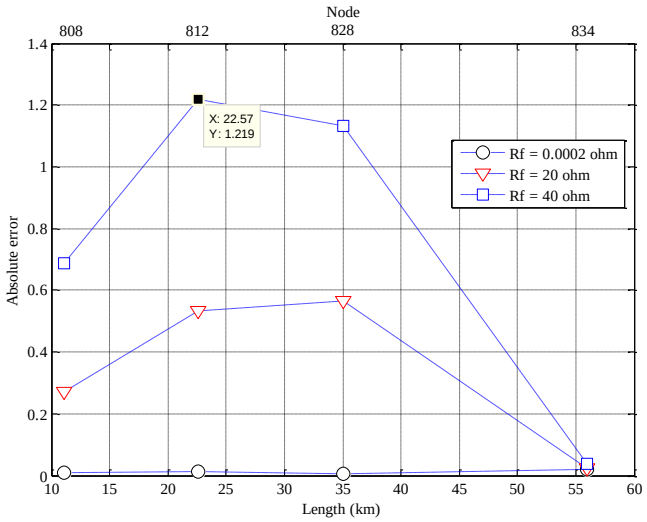

Figure 3.7: Fault location absolute error considering non-distorted measurements to single phase fault (A).

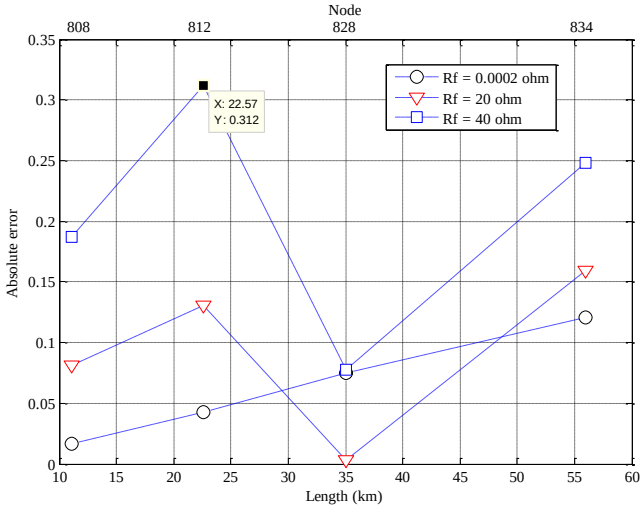

Figure 3.8: Fault location absolute error considering non-distorted measurements to phase to phase fault (AB).

3.1.6 Fault locator performance results considering distorted measurements

From the results, the discrimination of those distortion cases with errors greater than 2.5 [%], which is the defined acceptable threshold, is proposed. It is necessary to analyse the maximum absolute error of each one of the 12096 cases of distorted measurements of voltage and current. The absolute error in fault location considering distorted measurements is shown in Figs. 3.10, 3.11 and 3.12, for the three fault types. At these figures, two data clusters representing those cases whose absolute errors exceed the 2.5 [%] threshold are observed; one of them approximately within the range of 6 [%] to 14 [%] and the other roughly 3.5 [%]. The identified distortion sources correspond of saturation of current transformers equivalent to $(3/6) \cdot \Delta I$ and $(4/6) \cdot \Delta I$. The total cases of distortion in this situation are 6048. For other circumstances of CT saturation

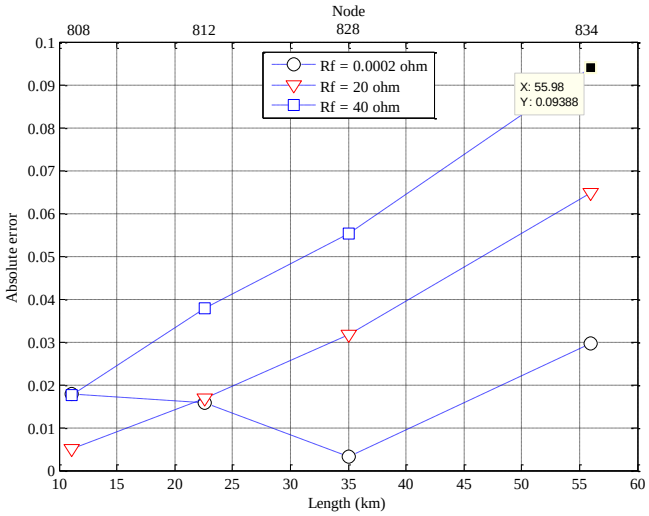

Figure 3.9: Fault location absolute error considering non-distorted measurements to three phase fault (ABC).

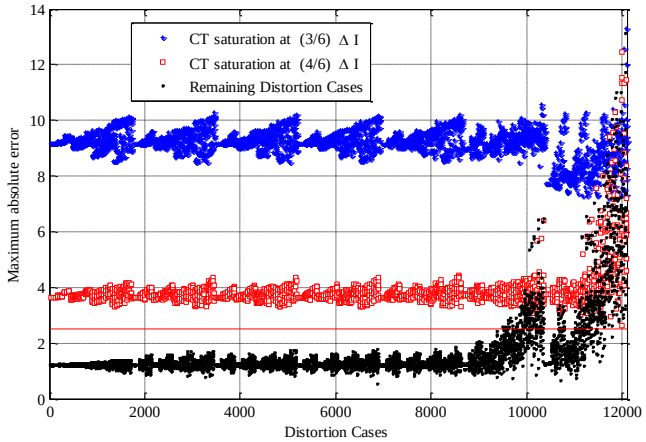

Figure 3.10: Fault location absolute error considering distorted measurements to single phase fault (A).

over $(5/6) \cdot \Delta I$, there is no clear evidence of its negative effect in the fault locator performance.

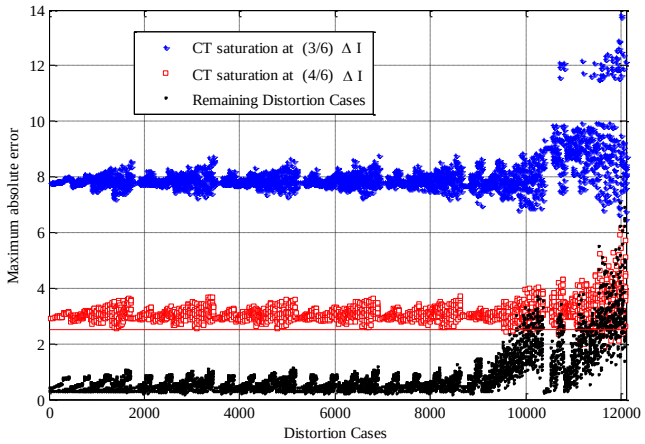

Figure 3.11: Fault location absolute error considering distorted measurements to phase to phase fault (AB).

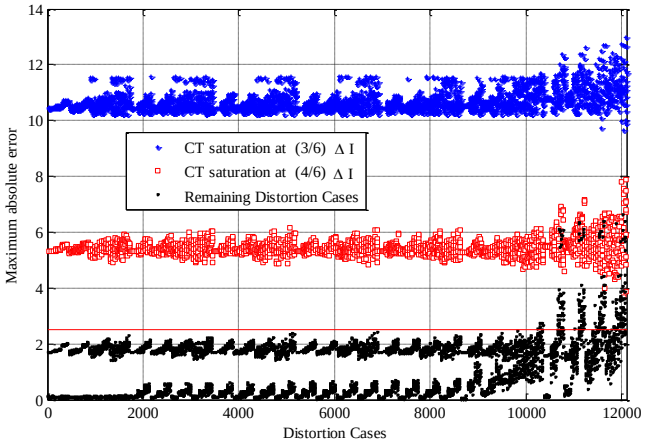

Figure 3.12: Fault location absolute error considering distorted measurements to three phase fault (ABC).

The remaining cases where the performance was considerably better than these that consider high CT saturation are next analysed. From those remaining distortion cases, a single set of cases that exceeds the threshold of 2.5 [%] is found, as is shown enclosed in a dotted rectangle in Figs. 3.13, 3.14 and 3.15. After analysing these cases, sampling frequency (Sf) and harmonics (H) were identified as the distortion sources that severely affects the fault locator performance, and these consists of cases which have in common the sampling frequency of 8 samples/cycle and a total harmonic distortion of 5 [%] in current and voltage. As a consequence, in such cases of harmonic distortion equal or higher 5 [%] and simultaneously a low sampling frequency device, considerable affects the locator performance, especially in phase to ground and phase to phase faults, as is presented in Figs. 3.13 and 3.14.

Additionally, the individual presence of distortion sources associated to low sampling frequency (8 samples

/ cycle) and high harmonic distortion (5 [%]), causes performance problems, as in the case of phase to ground fault location, as is shown in Fig. 3.13.

In Figs. 3.16, 3.17 and 3.18, the gradual effect of the Sf is highlighted using several colours, where the last black coloured data correspond to the 8 samples per cycle sampling frequency.

3.1.7 Analysis of the obtained results

The low Sf and the CTs saturation, viewed as distortion sources, are strongly related with the equipment quality, which forces utilities in the acquisition of appropriate equipment. Higher sampling frequencies led to lower absolute errors even considering the higher values of the analysed harmonic distortion (H) and the remaining sources of distortion presented in Table 3.1.

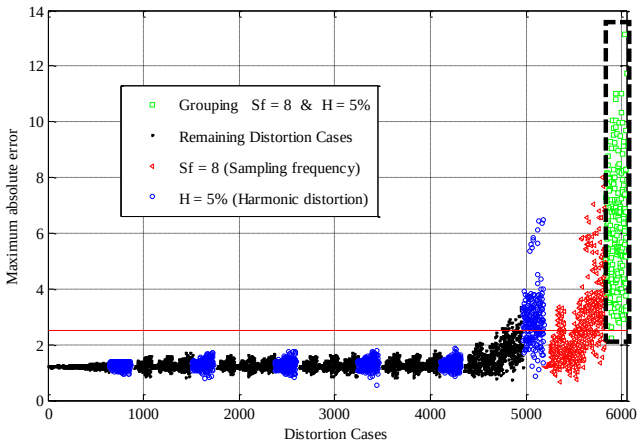

Figure 3.13: Fault location absolute error considering some distorted measurements to single phase fault (A).

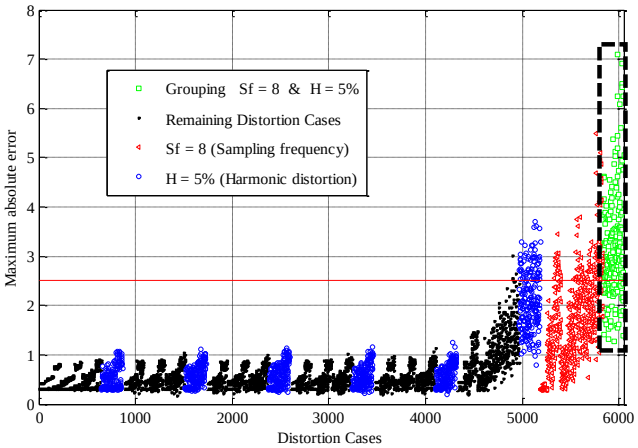

Figure 3.14: Fault location absolute error considering some distorted measurements to phase to phase fault (AB).

At the analysis process, the following results are worthy of mention: first, low performance was ascribed to three types of waveform distortions, low Sf (specifically 8 samples per cycles), saturation of CTs at values lower than $(5/6) \cdot \Delta I$ and total harmonic distortion for values higher than 5 [%]. Second, after ruling out the aforementioned distortion sources, the admissible values of constraints obtained in this study is given by a minimum Sf of 16 samples per cycle, a maximum CT saturation of $(5/6) \cdot \Delta I$ and a maximum total distortion of 3 [%] for the voltage and current measurements.

Finally, the considered values of the remaining distortion sources as saturation of PTs, presence of noise (N), resolution of the measuring device (R) and the presence of the DC offset component do not affect the

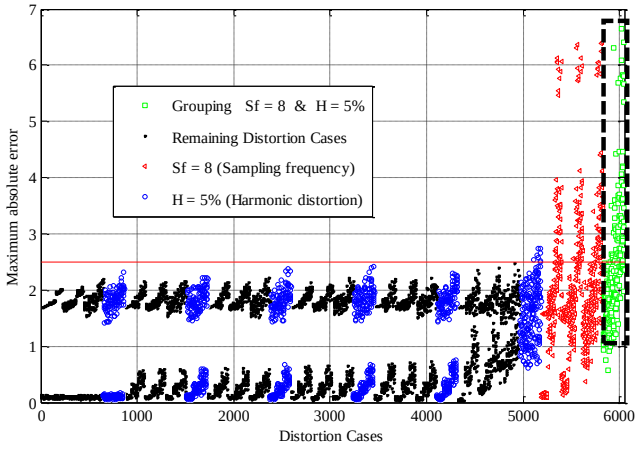

Figure 3.15: Fault location absolute error considering some distorted measurements to three phase fault (ABC).

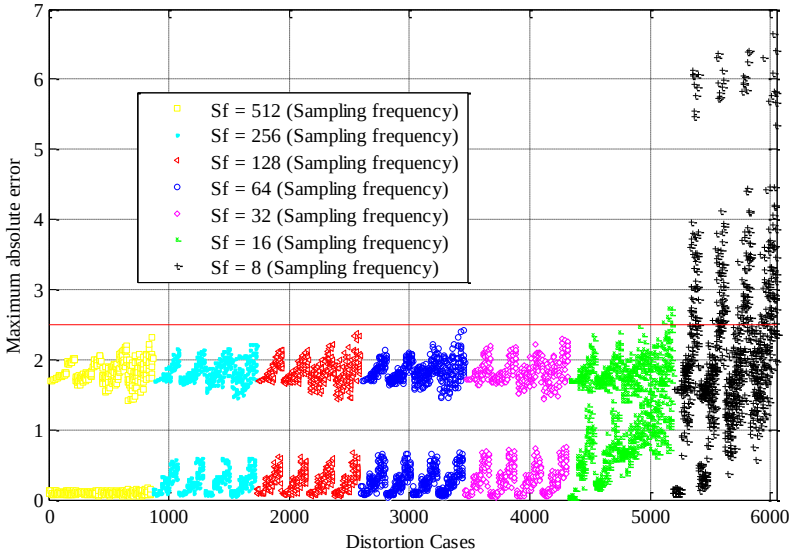

Figure 3.16: Fault location absolute error considering different sampling frequency to single phase fault (A).

fault locator performance, as it was demonstrated by the tests.

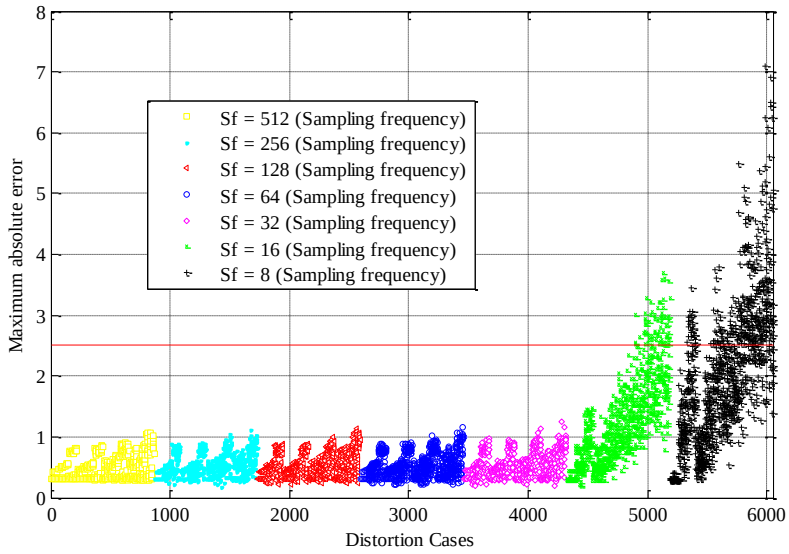

Figure 3.17: Fault location absolute error considering different sampling frequency to phase to phase fault (AB).

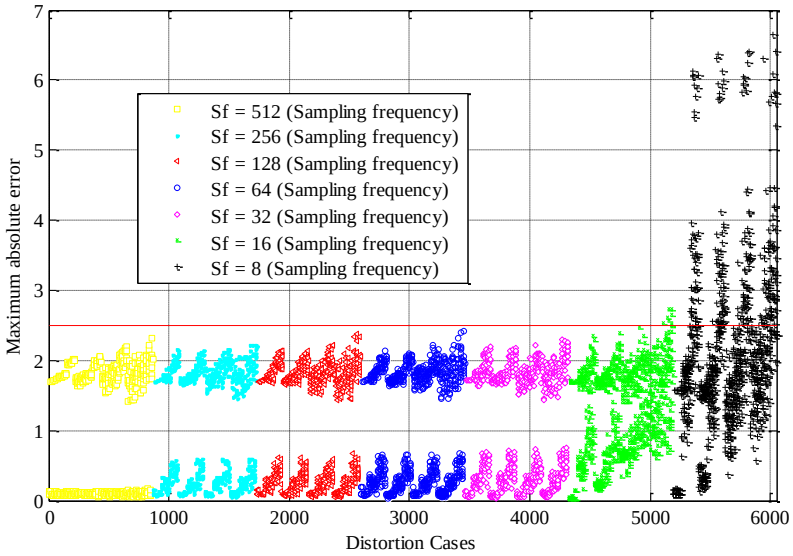

Figure 3.18: Fault location absolute error considering different sampling frequency to three phase fault (ABC).

3.2 Structured learning-based fault locator considering distorted measurements

The approach here presented is a LBFL, specially customised for distribution systems, whose core are the SVM. The LBFL uses the SVM as a learning tool to determine the fault location as the result of a classification task. The basic aspects of the LBFL are shown in section 2.2. The LBFL considers several adjusting-variables which have to be adequately selected to guarantee the excellent performance of the fault locator. However, in the case of this section where the effect of the distorted signals is sought, all variables are left fixed and an exhaustive search for the best values of them is not carried out to analyze the performance of the LBFL.

This section of the chapter seeks to make a comparison of an undisturbed scenario versus several disturbance scenarios. Therefore, the problem is relaxed by setting the variables to values that show right results, for example using the zone definition (z) with the minimum of zones in the system and the best parameters are selected using simple cross-validation. Also taking as a reference previous studies that show correct results.

The LBFL is implemented in four separated modules; three of them are aimed to locate faults considering each fault type (single phase, phase to phase and three phase faults) and one module is used in fault type identification.

The validation methodology proposed to analyse the performance of the LBFL is divided in three stages, as is depicted in Fig. 3.19. The former is oriented to determine the descriptor database obtained from the measurements of current and voltage signals obtained during fault situation (features); the second stage is oriented to adjust the LBFL by the iterative application of training and testing processes. Finally, the third stage is devoted to the performance comparison of the LBFL by considering inputs composed by distorted and non-distorted measurements of voltage and current.

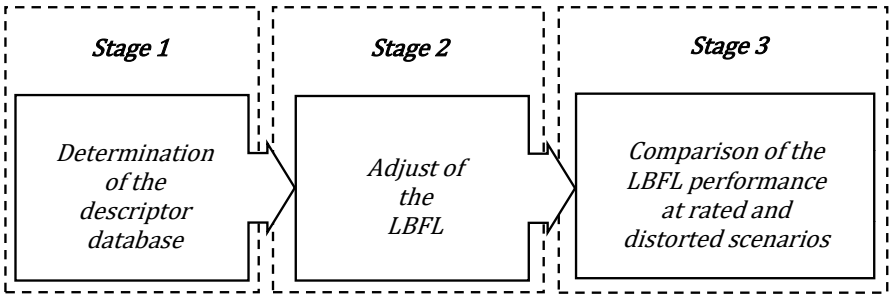

Figure 3.19: Proposed fault locator validation methodology.

3.2.1 Stage 1. Determination of the descriptor database

Three steps compose this stage, as is shown in Fig. 3.20. The former is the zone definition at the analysed distribution system; the second is the conformation of the fault database composed by the measurements of current and voltage during faults and considering several operating conditions of the distribution system. The latter is oriented to obtain the descriptor database from the fault data-set. A descriptor is here defined as a linear combination of the current and voltage measurements during pre-fault and fault condition, that is, a features combination as shown in section 2.2.1, to consider the variation in such signals, due the presence of a fault. The descriptors are used as input of the LBFL [71].

a) Step 1. Division in zones of the distribution system

As the LBFL uses the SVM as its core, a classification problem is here modelled to predict which class (zone) has to be assigned to an input (set of features that represents a fault). Then, to define the fault location as a classification problem, a class corresponds to a part or zone of the distribution system.

The division in zones of the distribution system is performed considering the following criteria: topology of the distribution system, distance between consecutive structures, line type (overhead or underground), conductor characteristics and load type [51, 55, 97, 98]. The adequate knowledge of the distribution system allows additionally, on the analysis of critical zones and the reconfiguration strategies in case of permanent faults. The zone definition on the distribution system also considers subgroups of nodes where there are critical loads, areas of easy access facilities, protective devices, presence of customers, distance between structures and the recommendations of the maintenance staff.

b) Step 2. Conformation of the fault database

The fault database helps to represent the behaviour of the distribution system under transient and steady state conditions of pre-fault and fault. This is composed by voltage and current measurements at the main feeder substation, considering different distribution system operating conditions. Different operating conditions are used at this research to adjust and test the LBFL. Each operating condition is given by variations of the load size, network frequency, main substation equivalent and line section length [29, 71, 111].

c) Step 3. Determination of the descriptor database

As the fault database is composed by time domain sampled voltage and current signals, measured at the main distribution substation, then the descriptors are here proposed to represent the main characteristics of those measurements, reducing the size of this database. The proposed descriptors are obtained as the variation of the three phase voltages and currents, from the pre-fault to the fault steady state. This new database contains 12 descriptors, which corresponds to the difference of magnitudes and angles of three currents and

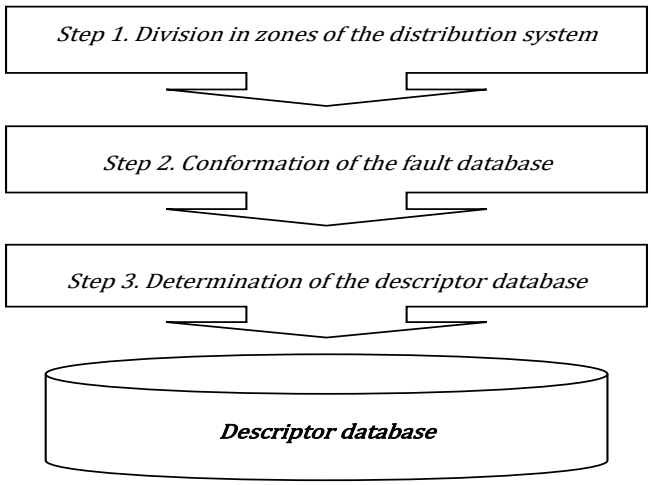

Figure 3.20: Definition of steps at the stage 1.

three voltages from the pre-fault and fault steady states. This descriptor database was successfully used in previous problems as is described in [71].

3.2.2 Stage 2. Adjust of the LBFL

The conditioning of the LBFL is performed using two phases: training and testing. In training phase, the inputs (descriptors) and the outputs (zones) are presented to the SVM to determine the support vectors. These are used in the testing phase as a learning map, which helps to determine the output (zone), in the case of a single input (descriptor file), which corresponds to a fault. This testing phase is performed using a descriptor database not previously used in training phase, to determine the confidence or performance of the adjusted LBFL.

At this stage, three different databases are used. The first one considers faults at different operating conditions of the distribution feeder. The second database considers faults only at the rated operating condition of the distribution feeder. The third database is obtained by considering faults at the rated operating condition, but the obtained measurements of voltage and current are affected by the power quality disturbances presented in section 2.4.

From the previous defined conditions, the fault database and the corresponding descriptor databases are obtained. These descriptors are used in stage 2, as depicted in Fig. 3.21.

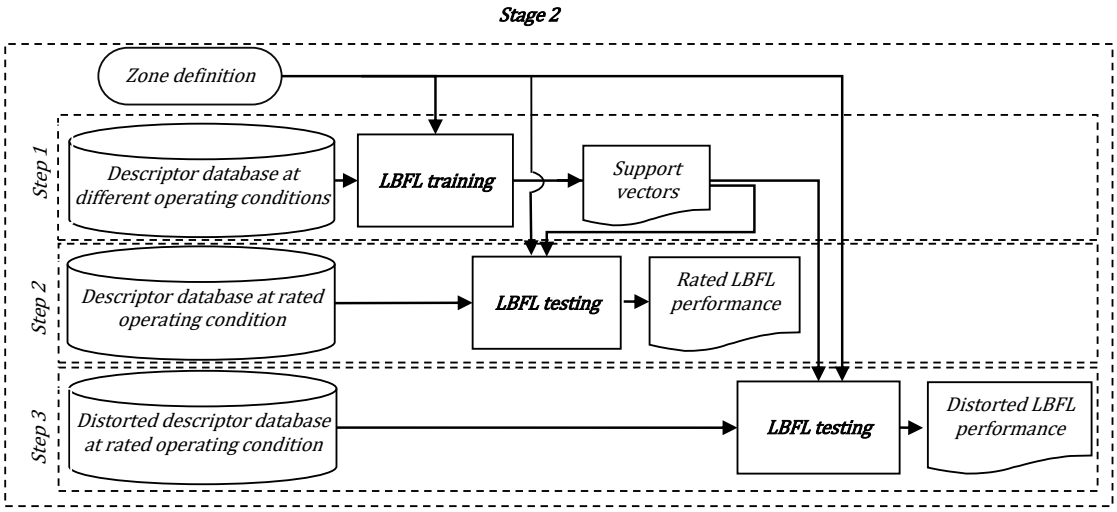

Figure 3.21: Functional structure of stage 2.

a) Step 1. Training the LBFL with different operating conditions of the distribution system

The LBFL is trained using the descriptor database obtained from faults considering different operating conditions and the respective faulted zone (output). As previously mentioned, the operating conditions are defined by variations on the load size from 0.4 to 1.1 of the rated load, network frequency from 0.95 to 1.05 of the rated frequency, main substation equivalent from 0.9 to 1.1 of its value and line section length from 0.9 to 1.1 of its value, to consider normal uncertainties [55, 111]. As a result of this step, the support vectors, which are used to determine the faulted zone, are obtained.

b) Step 2. Testing the LBFL with the rated operating condition of the distribution system

At this step, the LBFL previously trained in step 1, is tested using the descriptor database obtained from faults, at the rated condition of the distribution system (input). In this testing stage, only the inputs are presented to the LBFL, then the faulted zone (output) is obtained for each set of descriptors. The obtained faulted zone is next compared to the real faulted zone to determine the performance of the LBFL at the rated operating condition. The performance (Per) in fault location is defined, according to Eq. (3.10) [71].

$$Per = \frac{Number\ of\ faults\ where\ the\ zone\ is\ well\ identified}{Total\ number\ of\ analysed\ faults} \times 100\ [\%] \tag{3.10}$$

c) Step 3. Testing the LBFL with distorted measurements of voltage and current faults at the rated operating condition

Similarly to the previous step, the LBFL is tested using the descriptor database obtained from faults of the distribution system considered at the rated condition, but the measurements of current and voltage are distorted.

The considered distortions are: the sampling frequency (Sf), considered by using different sampling frequencies of the measurements of voltage and current, obtained for each fault. Saturation of current transformers (CTs) and potential transformers (PTs) is obtained by confine the measurements of voltages and currents into the saturation boundaries. Resolution (R) is considered by round out the measurements of voltages and currents following to the proposed value. Finally, harmonics (H), noise (N) and D offset (DCO) are defined by adding these signals to the voltage and currents measurements [35, 106–109]. The magnitudes of the distortion sources here used, follow the maximum values allowed by standards and the default factory settings of the measuring equipment, as was presented in Table 3.1 [1, 102, 109].

The combination of the number of instances at each distortion source defines the total number of distorted scenarios, which corresponds to 12095. Remembering that at the non-distorted measurements, the instance "1" in all of the distortion sources of Table 3.1 is considered. The 12096 scenarios are completed, which for comparison purposes are the same as in section 3.1.2. The performance of the LBFL, in the case of each distorted scenario, is estimated using Eq. (3.10).

3.2.3 Stage 3. Comparison of the LBFL performance at rated and distorted scenarios

The LBFL performance comparison is performed considering distorted and not distorted scenarios.

The objective of the comparison is to determine the distorted scenarios, which cause affectation on the performance of the fault locator. The accepted affectation on the LBFL performance is given by a comparison of the performance obtained at the rated non-distorted case and the performance of each one of the proposed distorted scenarios. The accepted difference in performance is defined in section 3.2.8.

3.2.4 Proposed tests and discussion

The LBFL performance, considering the disturbance scenarios previously proposed and tests on a 34.5 $[kV]-$ 75 bus distribution feeder, implemented in ATP and presented in Fig. 3.22, is here analysed. The configuration of the overhead line corresponds to the presented in Fig. 3.22, the conductor is ACSR 350 [kCM] in a PI model for each line section. The average length of each line section between two consecutive nodes is 250 [m], the average soil resistivity is 100 $[\Omega/m]$, the frequency is 60 $[Hz]$ and the skin effect is considered. The loads are presented in Table 3.4.

Nodes are classified as defined in section 3.2.1, into five (5) zones, as is presented in Fig. 3.22.

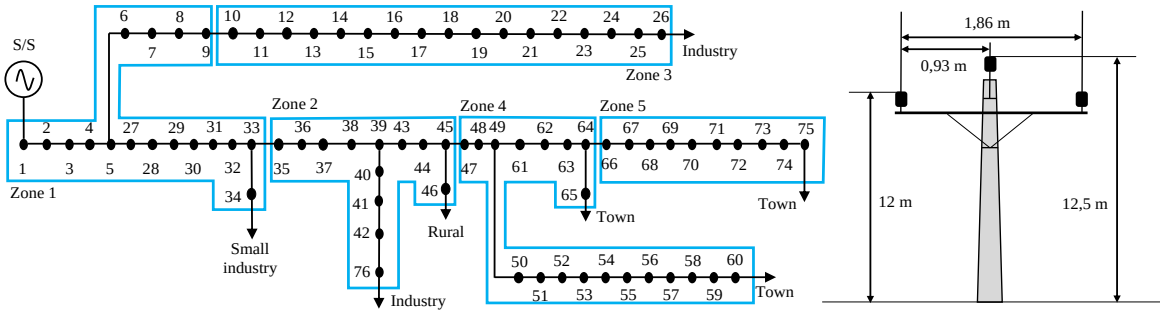

Figure 3.22: System 34.5 kV -75-bus test feeder implemented in ATP.

Node	Rated apparent power [kVA]	Power factor
26	270	0.80
34	410	0.80
46	275	0.91
60	2640	0.92
65	595	0.91
75	1220	0.91
76	312	0.89

Table 3.4: Load size.

3.2.5 Description of the fault databases

As presented in section 3.2.1, three different fault databases are used.

The first fault database is obtained by considering 200 different operating conditions of the distribution system, as is described in the step 1 of stage 2, presented in section 3.2.1. Considering five different fault resistances (0 [Ω], 10 [Ω], 20 [Ω], 30 [Ω], and 40 [Ω]) and 75 nodes, each operating condition has 1125 faults, 375 corresponding to each fault type (single phase, phase to phase and three phase faults), the total of faults at this database is 225000 [30]. The obtained descriptor database is used in training the LBFL.

The second fault database is obtained at the rated operating condition of the analysed distribution system. It contains measurements of voltage and currents of 1125 faults, considering three types of faults (single phase, phase to phase and three phase faults).

Finally, the third fault database is obtained at the rated operating condition of the analysed distribution system, but the measurements of voltage and current are distorted considering 12095 distortion scenarios, as presented in Table 3.1. According to the explained, this database contains information of 13606875 faults.

As summary, Table 3.5 shows the three different fault databases considered to obtain the used training and testing databases. The number of faults presented in such table are these considered in the case of single phase faults. In the case of the other two types of faults, similar information is obtained.

Description of the used fault database	Inputs and outputs for the adjust of the LBFL
First database: Composed by 200 different operating conditions, each one defined by 375 faults. The total of faults at this database is 375×200 = 75000.	LBFL input: Difference of magnitudes and angles of three currents and three voltages, from the pre-fault to the fault steady state. A total of 12 descriptors characterise each fault. LBFL output: Zone where the fault is located.
Second database: Contains faults only at the rated operating condition of the distribution feeder. The total of faults at this database is 5×75 = 375.	
Third database: Faults at the rated operating condition, but the obtained measurements of voltage and current are affected by the power quality disturbances. The total number of distorted scenarios is 12 095. The total of faults at this distorted descriptor database is 375×12095 = 4536625.	

Table 3.5: Summary of inputs and outputs considering the fault database for single-phase faults.

3.2.6 LBFL performance in case of non-distorted measurements

Using the descriptor database at the rated operating condition, as the input to the trained LBFL in this testing process, the obtained performance is 97.6 [%] in the case of single-phase faults. According to Eq. (3.10), this result means that 366 faults are classified in the correct faulted zone, from 375 analysed faults. The obtained performance in the case of phase-to-phase faults and in case of three phase faults are 96.5 [%] and 99.2 [%], respectively.

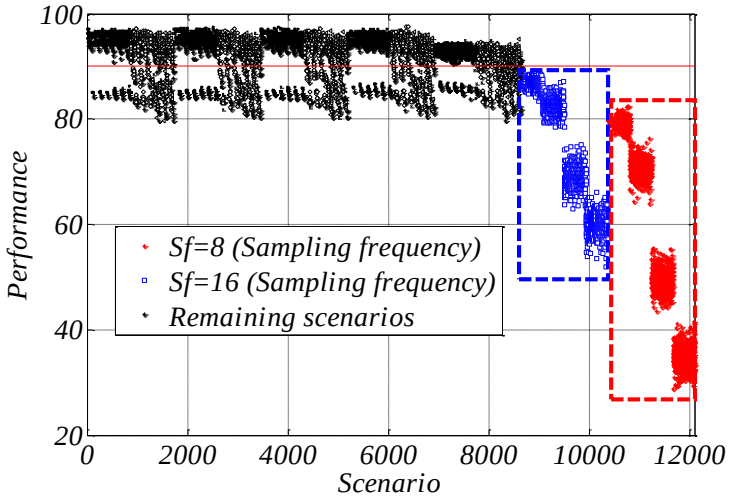

Figure 3.23: LBFL Performance for distorted measurements to single phase fault (A).

3.2.7 LBFL performance in case of distorted measurements

The LBFL performances in the case of use the distorted descriptor database at the rated operating condition as input, and considering the three types of shunt faults, are presented in Figs. 3.23, 3.24 and 3.25. Considering the reference performance as the lowest from these obtained using non distorted measurements, then these cases of performance lower to 96.5 [%] are considered as severely affected by distortions.

The vertical axis of Figs. 3.23 to 3.28 represents the performance that is calculated using the Eq. (3.10), which is related to the number faults where the faulted zone is correctly identified, pondered by the total number of faults analysed in the fault location process. The horizontal axis represents the faults considered

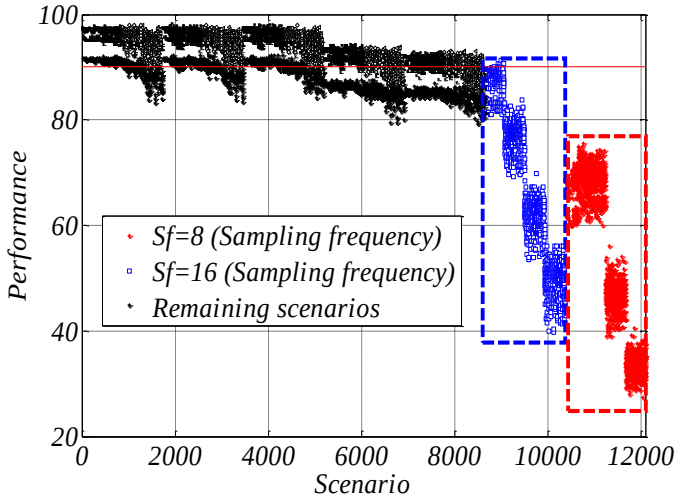

Figure 3.24: LBFL Performance for distorted measurements to phase to phase fault (AB).

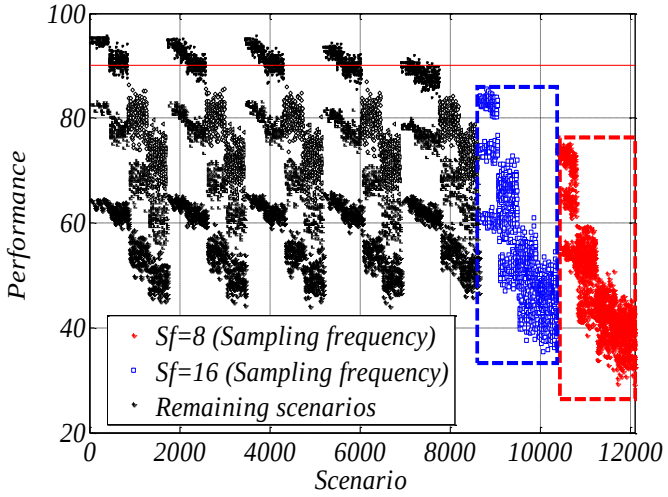

Figure 3.25: LBFL Performance for distorted measurements to three phase fault (ABC).

in each one of the 12095 distorted scenarios of the third fault database summarised in Table 3.5. A distorted scenario corresponds to a fault database that is affected by the distortion sources presented in Table 3.1, as is described in Table 3.5 for the third fault database.

From Figs. 3.23 and 3.24 it is observed that the performance is very low, for such scenarios that consider low sampling frequency. Low performance indicates that the error on the fault location is very large which means that the obtained faulted zone is different of the real zone where the fault is. In this case, a performance lower than 90 [%], indicates that when considering low sampling frequency, there are more than 38 faults classified in an incorrect faulted zone, from 375 analysed faults in each distorted scenario. According to the strategy here proposed and the expectation of the utility engineers, to efficiently locate and attend the distribution system contingency, a fault location error higher than 10 [%] is considered as unacceptable.

Taking into account that error is usually increased in the boundaries of the defined zones and considering an acceptable error margin of 10 [%], a threshold of 90 [%] is proposed in this paper as reference performance to identify such cases where the distortion sources cause a minor effect on fault location. The Fig. 3.25 shows, enclosed in a dashed rectangle, two data clusters representing those scenarios whose performance is lower than the 90 [%] threshold; one of them approximately within the range from 30 [%] to 80 [%] and the other roughly of 35 [%] to 85 [%]. The distortion sources which cause this performance corresponds to a specific case of low sampling frequency equivalent to 8 (red) and 16 (blue) samples per cycle. The total number of distortion scenarios in this situation is 3456.

The performances for single phase and phase to phase faults are very similar (see Figs. 3.23 and 3.24), while at the case of three phase fault the performance is severely affected. The Fig. 3.25 also shows that there are other distortion sources which affects the performance, which drops below 80 [%], making it very sensitive to the signal distortions.

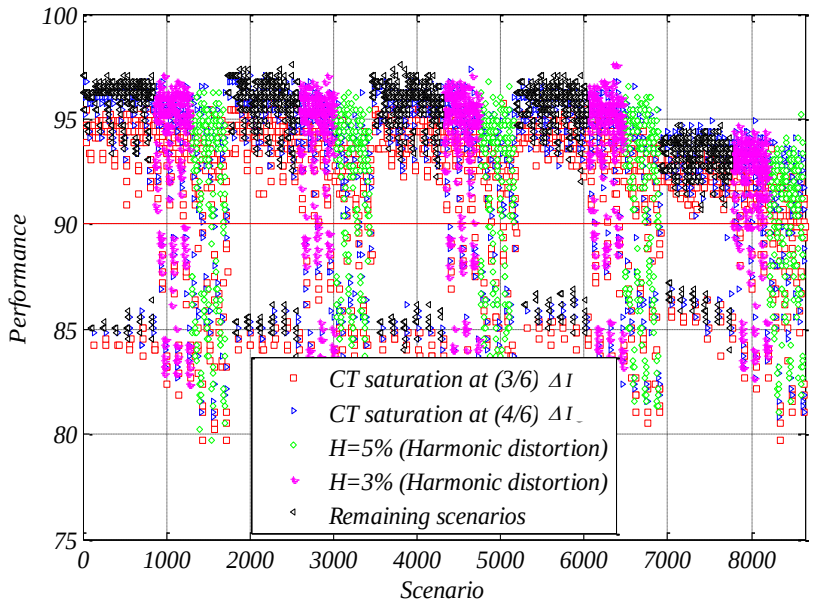

Figure 3.26: Performance of 8640 clustered fault scenarios to single phase fault (A).

The remaining fault scenarios where the performance was considerably better than these that consider low sampling frequency ($Sf = 8$ and $Sf = 16$) are presented in Figs. 3.26, 3.27, and 3.28. From those remaining

8640 fault scenarios, four different scenarios (represented by red squares, blue triangles, green diamonds and

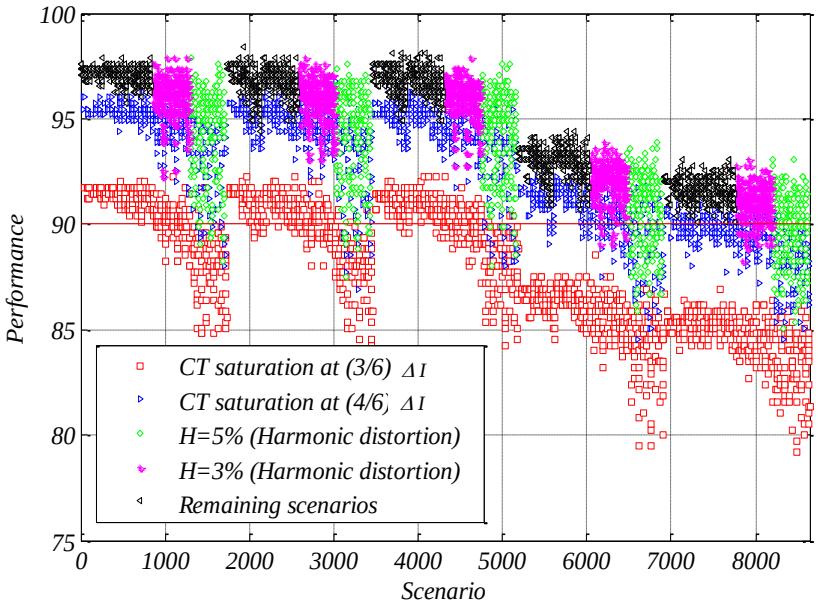

Figure 3.27: Performance of 8640 clustered fault scenarios to phase to phase fault (AB).

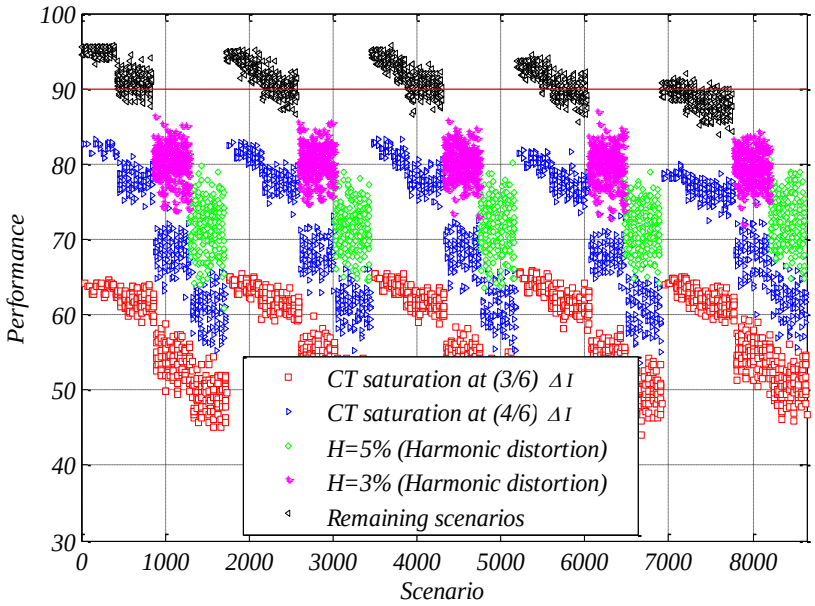

Figure 3.28: Performance of 8640 clustered fault scenarios to three phase fault (ABC).

pink asterisks), that cause poor LBFL performance are shown in Fig. 3.28. This figure shows that in case of three-phase fault, the performance is severe affected by distortions, presenting values lower than 75 [%].

After analysing these fault scenarios, saturation of current transformers (CT) equivalent to $(3/6)\cdot\Delta I$ and $(4/6)\cdot\Delta I$ and total harmonic distortion (H) equivalent to 5 [%] and 3 [%] in current and voltage signals, were identified as the distortion sources that cause severe reduction on the performance in case of the three phase faults.

Consequently, such fault scenarios of high saturation of current transformers equal or lower $(4/6)\cdot\Delta I$ and those with a total harmonic distortion equal or higher 3 [%], severely affects the locator performance, especially in the case of three-phase faults.

3.2.8 General comments

Considering distortion sources which causes low performance, some results are worthy of mention. Initially, low performance was described to three types of waveform distortions, a) low sampling frequency (8 and 16 samples per cycles), b) saturation of CTs at values lower than $(5/6)\cdot\Delta I$ and c) total harmonic distortion for values higher than 3 [%]. Next, after ruling out the aforementioned distortion sources, the admissible values consider a minimum sampling frequency of the recording device of 32 samples per cycle, a maximum CT saturation of $(5/6)\cdot\Delta I$ and maximum total harmonic distortion of 2 [%] for both, voltage and current signals. Finally, the remaining distortion scenarios which consider saturation of PTs, presence of noise (N), resolution of the measuring device (R) and the presence of the DC offset component (DCO), cause negligible effect on the LBFL performance, in such cases when the distortion magnitude meet the standards presented in Table 3.1.

Finally, the low sampling frequency and the CTs saturation are related with the equipment quality, and as these seriously affect the LBFL performance, which force utilities in the acquisition of appropriate equipment. Higher sampling frequencies led to high performance even considering higher values of the remaining sources of distortion.

3.3 Summary of the chapter

This chapter showed the analysis of the performance of two structured fault locators in the case of signal distortion, one impedance-based and one learning-based. The behavior of each locator was explained independently and the final results of the analysis were shown.

https://drive.google.com/drive/u/3/folders/1ZFYfy$_U$$VN3bqe15QW1D6hrLq3GqTgxeL$

Chapter 4

Robust strategy to adjusting-variables of a learning-based fault locator

The fault location problem in distribution systems is here considered by using the LBFL which requires the adequate selection of its adjusting-variables. As these variables are strongly interrelated, the robust (high performance) strategy proposed in this chapter uses an iterative approach to determine their adequate values, based on sampling and learning theories, which helps to improve the LBFL performance. According to the analysed tests, the results present a successful selection of the adjusting-variables, due to the high performance in the case of locating single-phase, phase-to-phase and three-phase faults, on IEEE34-bus test feeder, where a database of 140000 fault registers is used. As a result, satisfactory values for the five adjusting-variables are obtained, and the expected LBFL performance is higher than 96 [%] for all fault types.

4.1 Robust strategy proposed

The robust strategy here presented is based on a LBFL, specially customised for distribution systems, whose core are the SVM. The basic aspects of the SVM are shown in section 2.3. Four stages compose the proposed strategy shown in Fig. 4.1, which is oriented to define the five adjusting-variables of the LBFL and presented in the section 2.2. Considering that these adjusting-variables are interrelated, an iterative strategy is proposed as a contribution to determine their best configuration. The first stage determines the best features (f_{best}) and the data normalisation approach (n_{best}). The best zone division (z_{best}) and the best parameters of the SVM (p_{best}) are defined in the second stage. The algorithm in the third stage relates the first and second stages to define the best values of the previous four adjusting-variables through an iterative strategy. Finally, in the fourth stage, the best representative database (r_{best}) and the expected LBFL performance (Per_{exp}) are determined. All of the previous stages are following described.

4.1.1 Stage 1. Selection of features and normalisation

This stage is divided into five steps, to define the best normalisation (n_{best}) and the best set of features (f_{best}), as presented in Fig. 4.2. The definition of these two adjusting-variables is an iterative process of LBFL training and testing, using a small representative database, to reduce the computation time. Since the computational cost increases exponentially with respect to the increase in the fault records in the database.

In step 1 of stage 1, several sets of features are obtained from the 48 features (f) defined in section 2.2.1. In step 2, one of the normalisation strategies (n) presented in section 2.2.2 is applied. The obtained

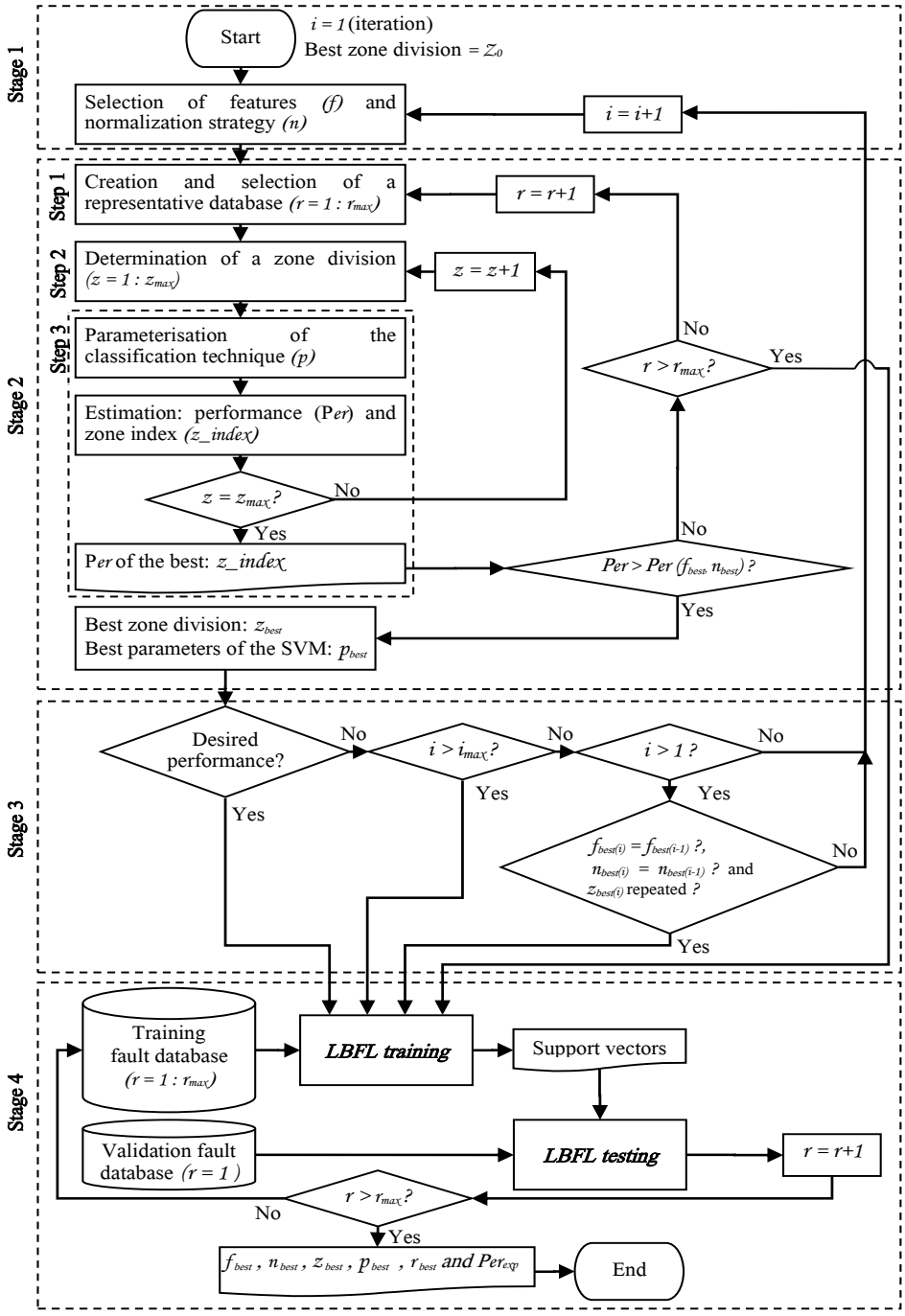

Figure 4.1: General strategy to define the adjusting-variables of a LBFL.

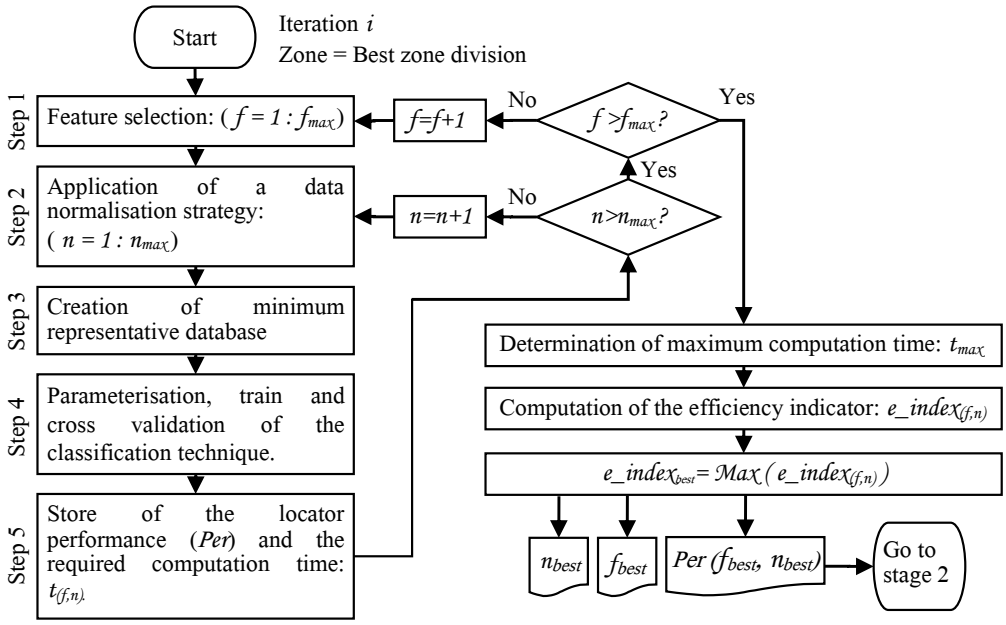

Figure 4.2: Stage 1. Selection of features and normalisation strategy.

combination of f and n adjusting variables is used to create a representative database in step 3. In general, for each combination of f and n, a representative database is obtained considering two conditions: i) sampling the complete database through the Latin Hypercube strategy [6,90], and ii) having at least 25 [%] representative faults per node of the analysed distribution system and that value must not be less than three representative faults per node.

Next and using a cross-validation strategy of at least four groups, adequate parameters of the SVM are obtained in step 4, like these with the best performance (Per) in fault location, as defined in Eq. (4.1).

$$Per = \frac{Number\ of\ faults\ where\ the\ zone\ is\ well\ identified}{Total\ number\ of\ analysed\ faults} \times 100\ [\%] \tag{4.1}$$

Finally, in step 5, the LBFL performance and the computation time of the cross-validation process are stored and compared at each iteration.

Steps 1 to 5 are repeated to evaluate all of the pairs of f and n. Next, the maximum computation time (t_{max}) for all of all of the pairs of f and n is determined; finally the efficiency indicator of the LBFL ($e\_index_{(f,n)}$), defined in Eq. (4.2), is estimated for each pair of f and n.

$$e\_index_{(f,n)} = \left[\alpha \times Per_{(f,n)} + (1 - \alpha) \times \left(\frac{t_{max} - t_{(f,n)}}{t_{max}} \right) \right] \times 100\ [\%] \tag{4.2}$$

In Eq. (4.2), constant α is a value between 0 and 1, which is used to define the degree of importance of the performance (Per) against the execution time ($t_{(f,n)}$). The maximum efficiency indicator e_index_{best} defines the best combination of features (f_{best}) and normalisation (n_{best}), which are the outputs of stage 1. The execution time is directly related to the computational cost, the more time the more computational cost.

Meta-heuristic algorithms have the concept of elitism when one thing is given priority over another [116, 117]. In this case, priority is not given only to performance and that behaviour makes the strategy

not elitist against performance. In this way, opportunities are given to those combinations that have low performance and at the same time, have a good computational cost. This process of the strategy allows it to become robust, as it explores alternatives other than the local optimal versus performance, and this favours the exploration of the solution space, in which the global optimal can be found. It should also be considered that in the final stage, the elitist is applied, to select the best combination of variables throughout the whole process of executing the strategy.

In each iteration of this stage, the best solutions are stored when considering performance. The best performance solution may be different from the best combination of features (f_{best}) and normalisation (n_{best}). This is due to the best indicator: e_index_{best}.

All of the previous processes in stage 1, requires of the adjusting-variable related to the zone division (z). Then, at the first iteration ($i = 1$) of the proposed methodology presented in Fig. 4.1, the best zone division (z_0) is the one defined by the utility operator. Lather, this zone division is modified by the best obtained in the next stages.

4.1.2 Stage 2. Selection of best zone division and the best parameters of the SVM

The three main steps which define this stage are next described.

The first step is oriented to create several sets of representative databases, which contain f_{best} and n_{best} defined in the previous stage, by sampling the database using Latin hypercube sampling and considering several values for the minimum number of faults per node. Next and using the set of representative databases, the smallest database not previously used is selected.

In the following step, a set of different zone divisions (z) is determined by considering the zone length and field restrictions to locate the fault. These zone divisions met the requirements imposed by the utility operator, related to the size and the geographical definition of the zone. These new zones provide an alternative to improve LBFL performance.

The third step is aimed to find a new set of SVM parameters (p), which varies mainly depending on the representative database and the zone definition. As a complement to the methodology presented in [66], the parameterisation strategy makes a comparison of the parameters obtained at the moment by [66] and those stored as parameters of the best solutions and then selects the best. Next, the performance defined in Eq. (4.1) and the zone index in Eq. (4.3) are estimated. Finally, this step is carried out until the whole set of possible zone definitions has been evaluated ($z = z_{max}$). Then best zone division selected as this that has higher z_index.

$$z\_index = \left[\beta \times Per + (1 - \beta) \times \frac{N_z}{N_n} \right] \times 100 \ [\%] \tag{4.3}$$

In Eq. (4.3), N_z is the number of zones, and N_n is the number of nodes in the distribution system, β is used to prioritise either the performance (Per) or the ratio between number of zones and nodes ($\frac{N_z}{N_n}$). This z_index is proposed due to the number of defined zones is inversely proportional to the performance; therefore, β is selected for the utility operator given priority between N_z and (Per). This z_index is then directly related to the fault management process.

The preceding steps of stage 2 are performed until one of the following conditions is fulfilled. The first one is met when the performance of the best zone division is better than the one obtained in stage 1. In this case, the best zone division is updated. The second condition is satisfied when none of the analysed zone definitions improves LBFL performance.

4.1.3 Stage 3. Iterative strategy

The iterative strategy is composed of four independent comparisons, driven by the results obtained at stage 2, and aimed to determine the inputs of stage 4.

The first comparison considers the LBFL performance; in the case of performance higher than the desired by the utility operator, the process continues in stage 4; otherwise, a new iteration $(i+1)$ is executed, and the process continues in stage 1.

The second comparison is related to the number of iterations; in the case of exceeding the maximum number, the process continues in stage 4; otherwise, the third comparison is performed.

The third comparison considers the case wherein two consecutive iterations, a new set (f_{best}, n_{best}) is found, and also the zone definition is repeated in previous iterations, then the process continues in stage 4; otherwise, a new iteration is proposed $(i+1)$.

The best solutions are also known as incumbent solutions are stored in each iteration. The third comparison is validated with data from the best solutions. Remember that the best solutions are stored independently from the results of each iteration. Therefore there will be two types of results per iteration. The third comparison uses the first type of result that is one of the best solutions.

Finally, the last comparison considers the number of representative datasets (r); in the case of exceeding the maximum, the process continues in stage 4; otherwise continues in stage 2.

The outputs of stage 3 are the best set of features (f_{best}), the best normalisation strategy (n_{best}), the best SVM parameters (p_{best}) and the best zone definition (z_{best}).

At last, in such cases where the process ends before to reach the maximum iterations and the desired performance, it is possible to determine that the fault database analysed do not possibilities the utility operator desires.

4.1.4 Stage 4. Training and testing of the LBFL

From the best solutions obtained by stage 3, the required adjusting variables to perform the LBFL training are defined. The training database contains the set of features (f_{best}), the type of normalisation (n_{best}), the parameterisation (p_{best}) and the zone definition (z_{best}) of the analysed distribution system, which presents the best performance during the development of the overall strategy on previous stages.

Using the obtained adjusting-variables, stage 4 determines the best representative database (r_{best}) used to train the LBFL. In this stage, an iterative training-testing is proposed to select the minimum size of the representative fault database that generates the best performance. Each representative fault database used in training process considers a different minimum number of faults per node. The testing of the LBFL is performed using a fault database which is different from the one used in training.

Once the training representative databases are obtained, a continuous training-testing process is proposed using such databases. At each iteration, the training database is increased from the smallest to the biggest one, and the trained LBFL is tested using the previously mentioned testing database. The best representative fault database (r_{best}), is obtained when the slope in the LBFL performance curve, obtained by the successive training-testing procedure, is near to zero. The process ends when all the representative databases are used in LBFL training $(r > r_{max})$, as it is depicted in stage 4 of Fig. 4.1.

The performance when the five adjusting-variables are selected (f_{best}, n_{best}, p_{best}, z_{best} and r_{best}), is a measure of the LBFL confidence for field applications (Per_{exp}).

Finally, the proposed strategy is enough general to be used as a reference in the case of other fault location approaches, which require the selection of several interrelated adjusting-variables.

4.2 Proposed test and result analysis

This section is devoted to report and analyse the most significant results obtained in the tests. The proposed strategy is used to adjust four different LBFL; three of them are aimed to locate faults considering each fault type (single-phase, phase to phase and three-phase faults) and one of them is used for fault type identification.

4.2.1 Test system

The 24.9 $[kV]$ IEEE34-bus test feeder in Fig. 3.6 is used to test the fault location approach [50, 114]. The IEEE34-bus test feeder is implemented in ATP [29, 35].

4.2.2 Range definition for the adjusting-variables

On the presented tests, the evaluation of the proposal considers the following ranges for the five adjusting - variables: four combinations of features (f), two types of normalisation (n),a SVM parameterisation strategy (p), 10 different zone definitions (z) and a representative database (r) with 5000 records per node.

Despite the constrains here defined for the presented tests, the proposed methodology is enough general to consider other testing ranges, which involves more instances for each variable.

a) Feature selection (f)

Four combinations of features (f) are presented in Table 4.1. These combinations are selected from the features defined at section 2.2.1 in Eq. (2.10).

Table 4.1: Combinations of features used in LBFL tests

Combination	**Features (f)**
1	$[f_{(j,5)},\ f_{(j,6)},\ f_{(j,13)},\ f_{(j,14)}]$
2	$[f_{(j,7)},\ f_{(j,8)},\ f_{(j,15)},\ f_{(j,16)}]$
3	$[f_{(j,1)},\ f_{(j,2)},\ f_{(j,3)},\ f_{(j,4)},\ f_{(j,9)},\ f_{(j,10)},\ f_{(j,11)},\ f_{(j,12)}]$
4	$[f_{(j,2)},\ f_{(j,4)},\ f_{(j,10)},\ f_{(j,12)}]$

b) Type of normalisation (n)

From the five normalisation strategies presented in section 2.2.2, $MinMax$ (MM) and $Zscore$ (ZS) are the most commonly used according to the adequate obtained results in previous approaches; therefore these

are taken into account in this generalised strategy. As previously exposed, it is possible to include other normalisation alternatives [42].

c) SVM parameterisation strategy (p)

In stage 1, the SVM parameters (p) are selected according to the methodology presented in [66], where a radial basis function kernel is used to consider non linearly separable classes. As a complement to the methodology presented in [66], the parameterisation strategy makes a comparison of the parameters obtained by [66] and those stored as parameters of the best solutions and then selects the best.

d) Determination of the zone division (z)

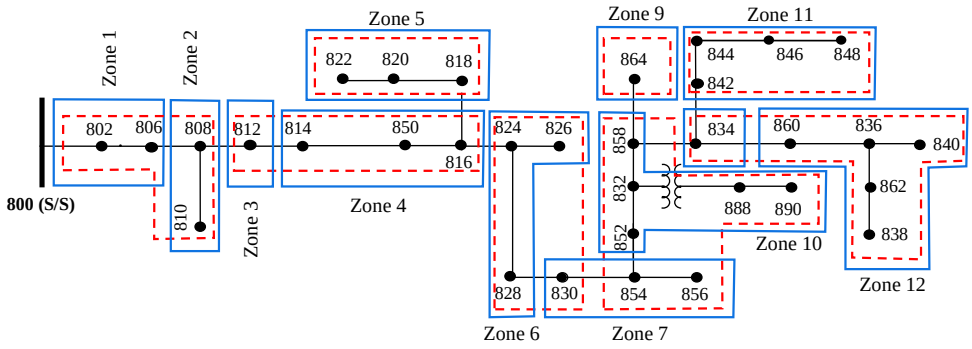

Figure 4.3: IEEE 34-bus test feeder used in LBFL tests.

The nodes are grouped as in Fig. 4.3, where two zone definitions are shown: i) the proposed by the utility operator (dashed lines) and, ii) the best zone definition (continuous line) obtained using the proposed strategy for single-phase faults. The initial definition is given by the distribution utility operator of the proposed test system (8 zones) [50,114], and this zone division is presented using dashed lines in Fig. 4.3, in the case of single-phase faults. At the presented tests, 10 different zone definitions are proposed, all of them eliminate the multiple estimations [71].

e) Representative database (r)

The used database contains 140000 faults for each fault-type. These are obtained from several operating conditions of the distribution system, which include random variations between 40 [%] to 110 [%] of rated load, ± 3 [%] of network frequency,± 10 [%] of the main substation rated voltage and ± 10 [%] of the line section length. Fault resistances from 0 to 80 [Ω] are also considered.

The Latin Hypercube sampling strategy is used to reduce the database to representative database [6,90]. To select a set of representative databases, an analysis of the fault registers per node is considered, as proposed in section 2.2.5. In the IEEE34-bus test feeder system, the complete database for a fault type is 140000 fault records, of which 5000 records are per node. The system has records on 28 nodes. By taking 35 records per node from the database, a complete database of 980 fault records is obtained. The 25 [%] of those records per node is 8.75, which is approximately nine (9). Then the smallest representative database ($r = 1$) defined by Latin Hypercube sampling strategy is the one containing at least nine (9) faults in each node. Additional representative databases are defined by using at least 15 ($r = 2$), 20 ($r = 3$), 25 ($r = 4$) and 50 ($r = 5$) faults

in each test system node. In this way, an incremental representative database of 810, 974, 1147, 1307 and 2329 fault registers (J) is obtained in the case of $f = 1$ and $n = 1$. These representative databases are used in the second stage for this combination of f and n.

On the other hand, the minimum signal requirements consider a sampling frequency of 32 samples per cycle, CT saturation lower than 17 [%] of the difference between the maximum peak current and the rated current, and a total harmonic distortion lower than 2 [%] for voltage and current signals, as suggested in section 3.2.

4.2.3 Results and analysis - stage 1

At this stage, as the number of faults J at the representative database r, depends on the combination of the initially analysed adjusting-variables f and n, then and aimed to establish an adequate comparison the validation time is considered as the average per fault, as is presented the fourth column in Table 4.2. Using this validation time and the obtained performance Per, the $e\_index_{(f,n)}$ in column sixth of Table 4.2 is obtained to establish an adequate relation between the performance and the time required in the off-line LBFL setting.

As previously defined, the first stage zone definition has (8) zones as defined by the utility operator. In this initial stage, considering the number of intermediate results, only these for the single-phase fault in the first iteration are shown in Table 4.2. The results for other fault types are similar to these presented for single-phase fault and presented at the end of this subsection.

All of the above allows finding the combination that has the lowest computational cost when evaluating a single fault. By determining the combination with the least computational time, that combination is allowed to be evaluated in the next stage and thus be able to obtain results from the system with the least computational time. It is also important to highlight that in the second stage of the strategy an increase in the database is considered, and due to this, it is very probable that the computational cost will increase considerably in the second stage when using an inappropriate combination.

The combination with the lowest computational cost may be different from the combination with the best performance. Therefore, those two results are stored later in stage three. The result of this first stage in its first iteration is seen in Table 4.2. The result of the combinations with the lowest computational cost is stored in Table 4.4 and in Table 4.5 is listed the combinations with the best performance, incumbents or best solutions in Table 4.5.

Table 4.2: Results of the first stage for the IEEE34 system at the first iteration, in the case of single-phase faults.

f	n	J	**Validation time / fault** $[ms]$	***Per*** *[%]*	$e\_index_{(f,n)}$ *[%]*
1	1	810	5.3	78.3	64.8
1	2	810	3.6	71.2	64.7
2	1	600	5.8	75.8	61.1
2	**2**	**600**	**2.4**	**67.8**	**66.2**
3	1	581	3.6	62.5	57.8
3	2	581	4.9	43.9	38.4
4	1	581	5.9	67.0	53.6
4	2	581	4.6	60.3	52.6

In table 4.2, the eight rows correspond to the combinations between (n) and (f). From the $e\_index_{(f,n)}$, the best combination is the fourth with a value of 66.2 [%]. From the exposed, a value of $\alpha = 0.8$ is selected,

Table 4.3: Results at the second stage for the smallest representative database ($r = 1$) and all zone definitions for the single-phase fault locator (IEEE34 system).

Zone definition (z)	**Number of zones**	p		**Validation time / fault [ms]**	Per [%]	z_index [%]
		C	σ			
1	8	1024	0.80155	2.4	67.8	60.2
2	9	16384	0.38849	4.7	68.7	61.7
3	**13**	**64**	**0.76375**	**4.6**	**69.6**	**65.3**
4	12	2048	0.76375	2.4	60.5	57.3
5	14	1024	0.76117	7.6	65.3	62.6
6	14	16384	0.38849	5.8	63.9	61.5
7	16	16384	0.38849	6.1	61.3	60.9
8	15	16384	0.38849	5.2	57.8	57.3
9	16	512	0.75392	5.4	63.1	62.3
10	17	64	0.75338	6.2	64.5	64.2

i.e., 80 [%] of priority is given to performance and 20 [%] to the validation time of each fault record. In this way, the combination is obtained with a validation time per fault of 2.4 [ms] and a performance of 67.8 [%]. In this case, the fourth combination is the one with the lowest validation time per fault; however, it is important to remember that the selected combination is not always the one with the shortest time. In Appendix A, the same strategy applied to another system is shown, in that system, the combination selected in the first stage is not the one with the shortest time, as shown in Table A.2. As a consequence, the outputs of stage 1 are $f_{best} = 2$ and $n_{best} = 2$ in the case of single-phase faults.

In the case of phase-to-phase and three-phase faults the obtained results at the $r = 1$ in stage 1 are $f_{best} = 2$, $n_{best} = 1$ and $f_{best} = 2$, $n_{best} = 1$, respectively.

The validation time in this first stage is considered very important because, for the second stage, it is necessary to increase the database and validate with other types of zone definition. To efficiently perform this procedure, variables with lower computational cost must be considered. The validation time depends on the number of faults; therefore, the time used in the $e\_index_{(f,n)}$ is required to validate a fault.

The total time used to execute the first stage is 6 [h] for single-phase fault, 1 [h] for phase-to-phase fault and 1.5 [h] for three-phase faults.

The high computational cost in the first stage for single-phase faults has nothing to do with the number of fault records J. It is a cost associated with the complexity of the core of the LBFL proposed for the system under study that is the SVM, used as a classification tool. The computational cost depends on the system under review, the results of the 34 [kV] -75 bus test system in appendix chapter A.3 for the three fault types understudy was less than 2 [h].

4.2.4 Results and analysis - stage 2

In the first stage, the adjusting-variables $f = 2$ and $n = 2$ are defined in the case of the single-phase fault locator. In the second stage, the 10 different zone definitions are considered; the utility operator determines the first one. The remaining zone definitions are analysed in this stage, as is presented in Table 4.3, in the case of using the smallest representative database ($r = 1$ or $J = 600$).

From Table 4.3 and considering the zone index (z_index), it is found that the best combination has a value of 65.3 [%], which corresponds to a 13 zones (zone definition number 3, $z = 3$), a performance (Per) of 69.6 [%] and a validation time per fault of 4.6 [ms]. It is denoted from this validation time, the adequate selection obtained from the e_index in the previous stage, considering this approach as an iterative process where the

time is an important aspect. Additionally, a value of $\beta = 0.8$ is used to prioritise either the performance (Per) and the ratio between the number of zones and nodes ($\frac{N_z}{N_n}$), because as previously explained, performance is usually higher in the case of a low number of zones and an adequate balance is required.

The total time used to execute the second stage is 7.1 $[h]$ for single-phase fault, 1.6 $[h]$ for phase-to-phase fault and 3.2 $[h]$ for three-phase faults.

According to the results for the smallest representative database ($r = 1$ or $J = 600$) at second stage, the best adjusting variables are $f_{best} = 2$, $n_{best} = 2$ and $z_{best} = 3$ (13 zones defined in the power system) for the single-phase fault locator.

Finally, the following sets of adjusting-variables $f_{best} = 2$, $n_{best} = 1$, $z_{best} = 4$ (12 zones defined in the power system) and $f_{best} = 2$, $n_{best} = 1$, $z_{best} = 4$ are obtained for the smallest representative database ($r = 1$), in the case of phase-to-phase and three-phase fault locator, respectively.

4.2.5 Results and analysis - stage 3

Stage 3 relates stages 1 and 2, then the proposed strategy continues with the remaining iterations, repeating the first stage with the new zone definition obtained in the second stage, as is depicted in Fig. 4.1.

There are two types of results in this third stage, those used with the lowest computational cost in the strategy, with them it can evaluate situations in which it could find local optimal and, on the other hand, the best solutions also known as incumbent solutions, which define the optimal global.

Table 4.4: Results of the third stage in the IEEE34 system for all fault locators. First stage combination results with the lowest computational cost.

Fault type	i	f_{best}	n_{best}	z_{best}	**Number of zones**	p_{best}		Per [%]
						C	σ	
Single-phase	1	2	2	1	8	1024	0.80155	67.8
	2	1	2	3	13	64	0.76375	69.6
	3	1	1	4	12	1024	0.73296	88.7
	4	1	2	5	14	512	2.89689	74.2
	5	**1**	**1**	**4**	**12**	**16384**	**0.38849**	**90.4**
Phase to phase	1	2	1	1	8	32	0.44670	66.4
	2	2	1	4	12	8	0.43522	68.6
	3	2	1	5	14	16	0.42648	63.8
	4	2	2	4	12	256	0.66716	63.8
	5	2	1	5	14	8192	0.60458	70.8
	6	2	2	4	12	32	0.71247	70.0
	7	2	2	2	9	128	0.67134	77.7
	8	**4**	**1**	**3**	**13**	**16384**	**0.38849**	**79.2**
	9	2	1	4	12	16384	0.38849	67.8
Three-phase	1	2	1	1	8	128	0.91324	81.0
	2	1	1	4	12	4096	0.23652	79.1
	3	4	2	7	16	4096	1.87393	89.6
	4	**4**	**1**	**4**	**12**	**4096**	**0.59863**	**93.9**
	5	2	1	2	9	262144	0.61253	74.7
	6	2	1	4	12	4096	0.27185	77.6

The strategy internal results are giving in Table 4.4, also known as first stage combination results with the lowest computational cost. On the other hand, the best solutions or the combinations with the best performance or incumbents are showing in Table 4.5. The first iterations for all fault types are presented.

In the case of the single-phase fault locator, as best results the following adjusting-variables are obtained: $f_{best} = 1$, $n_{best} = 1$ and $z_{best} = 4$ (12 zones defined in the power system) and p_{best} is defined by $C = 16384$ and $\sigma = 0.38849$.

These adjusting variables are selected, due to the maximum performance with that zone definition (90.4 [%]), obtained at iteration $i = 5$. In the case of phase-to-phase and three-phase fault locators, the results are also presented in Table 4.5. In all fault types, the third stop comparison shown in section 4.1.3 was fulfilled, as is shown in high quality solutions from Table 4.5.

Additionally, and as it is observed from Table 4.5, the LBFL performance is higher than in the previous stages. This is due to the increasing in the representative database size; then as higher is the representative database, usually higher performances are reached, as a more significant number of faults are used in LBFL training.

Table 4.5: Best solutions of the third stage for the IEEE34 system of all fault locators. Combinations with the best performance (incumbents).

Fault type	i	f_{best}	n_{best}	z_{best}	**Number of zones**	p_{best}		Per [%]
						C	σ	
Single-phase	1	1	1	1	8	2048	0.76375	78.3
	2	1	1	3	13	64	0.76375	83.9
	3	1	1	4	12	4096	0.72190	89.4
	4	1	1	5	14	1024	0.80155	88.6
	5	**1**	**1**	**4**	**12**	**16384**	**0.38849**	**90.4**
Phase to phase	1	4	1	1	8	16384	0,66716	78,8
	2	4	1	4	12	1024	0.42648	80.3
	3	4	1	5	14	16384	0.38849	79.3
	4	1	1	4	12	128	0.66339	75.9
	5	4	1	5	14	256	0.66716	71.6
	6	4	1	4	12	8192	0.60458	78.2
	7	2	2	2	9	4	2.16942	77.7
	8	4	1	3	13	16384	0.38849	79.2
	9	**4**	**1**	**4**	**12**	**16384**	**0.38849**	**84.2**
Three-phase	1	4	1	1	8	131072	0.80094	88.6
	2	4	1	4	12	8192	1.88546	92.7
	3	4	2	7	16	4096	1.87393	89.6
	4	4	1	4	12	4096	0.59863	93.9
	5	4	1	2	9	8192	1.85937	91.9
	6	**4**	**1**	**4**	**12**	**524288**	**2.26378**	**96.4**

The total iterations for single-phase fault, phase-to-phase fault and three-phase fault were 5, 9 and 6 respectively. The total time used to execute the third stage is $65,5$ $[h]$ for single-phase fault, 23.4 $[h]$ for phase-to-phase fault and $28,2$ $[h]$ for three-phase faults. The previous time is the minimum and is calculated by multiplying the number of iterations with the sum of the times of the first two stages. The time can increase considerably in the second stage because the computational cost is not linear with the increments of the database.

4.2.6 Results and analysis - stage 4

From previous stages, the adjusting-variables f_{best}, n_{best} ,z_{best} and p_{best} are defined, then this last stage defines the best representative database r_{best} and the expected LBFL performance (Per_{exp}) are determined. Training and testing of LBFL are sequentially performed. Initially, the smallest representative database is

used in training, and each iteration consists of using a new representative database from the smallest to the biggest one. The testing database is always the same.

The training process is performed using different-size representative databases. In the case of single-phase faults, the representative database varies from 48 to 36382 faults; for phase-to-phase faults, this varies from 50 to 38921 faults, and in the case of three-phase faults this varies from 50 to 35292 faults. The testing database

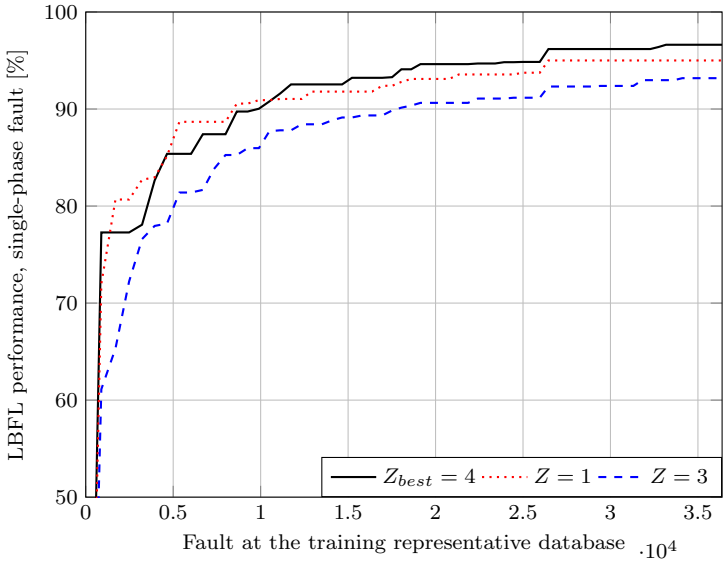

Figure 4.4: Single-phase fault performance for training using several representative databases.

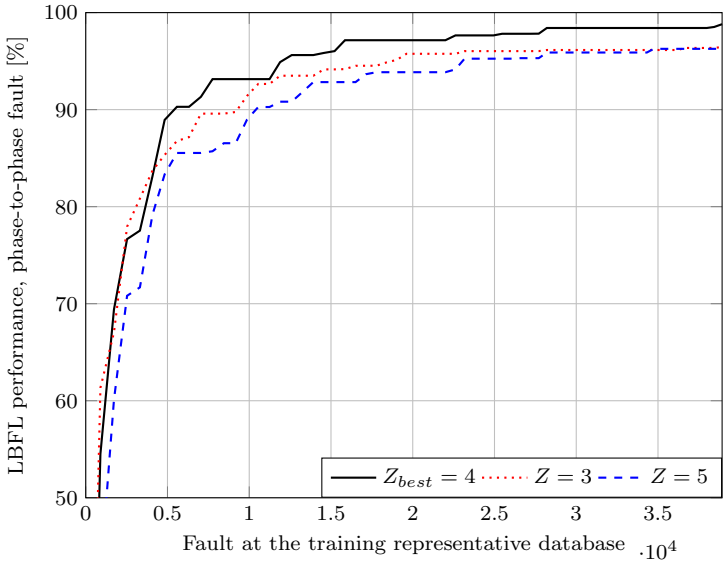

Figure 4.5: Phase-to-phase fault performance for training using several representative databases.

used in the three types of faults contains 1400 faults not previously used in the training stage. Fig. 4.4 shows the obtained LBFL performance for three different sets of adjusting-variables, in case of single-phase faults. In all cases, the adjusting-variables related to features, normalisation and SVM parameters are these obtained in the previous stages (f_{best}, n_{best} and p_{best}), while the zone definitions are: the z_{best} (continuous line), the zone defined by the utility operator or a zone better than the one defined by the operator (dashed line), and one additional zone definition with excellent performance in the intermediate stages (dotted line). According to the incremental training, the size of the minimum representative database (r_{best}) for the single-phase fault type is 26450 faults. Additionally, a measure of the LBFL confidence is the expected performance (Per_{exp}), which in the case of single-phase faults is over 96.18 [%]. In the case of phase-phase faults, r_{best} = 28200 faults and Per_{exp} = 98.41 [%], while for three-phase faults, r_{best} = 10980 faults and Per_{exp} = 98.37 [%], as shown Figs. 4.5 and 4.6. The performance of the fault type identification LBFL module is 99.13 [%].

Fig. 4.7 shows the obtained LBFL performance for three different sets of adjusting-variables, in case of single-phase faults. In all cases, the adjusting-variables related to zone definition, normalisation and SVM parameters are these obtained in the previous process (z_{best}, n_{best} and p_{best}), while the features are: the f_{best} (continuous line) and two features with promising performances but below the best (dashed and dotted line). According to the incremental training in Figs. 4.7, 4.8 and 4.9, the selection of f_{best} for all faults types studied with the strategy was successful.

The Fig. 4.10 first considers two parameters close to those of the incumbent p_{best} = (C : 16384 ; σ : 0.38849), in case of single-phase faults. On the other hand, it consider the parameters incumbent (best) p_0 = (C : 32768 ; σ : 0.4766) on the 34.5 [kV] - 75 bus feeder system shows in Appendix A and labeled as: $p = 0$. It is observed that the parameters close to p_{best} show a similar performance, while the parameters of another system show a lower performance. Therefore, the strategy shows parameters that lead to the best performance. And it shows that the parameters are unique to each system since, regardless of whether they are the best, it is important to obtain parameters with information from the system under study. Finally, According to the incremental training in Figs. 4.10, 4.11 and 4.12 the selection of p_{best} for all faults types studied with the strategy was successful.

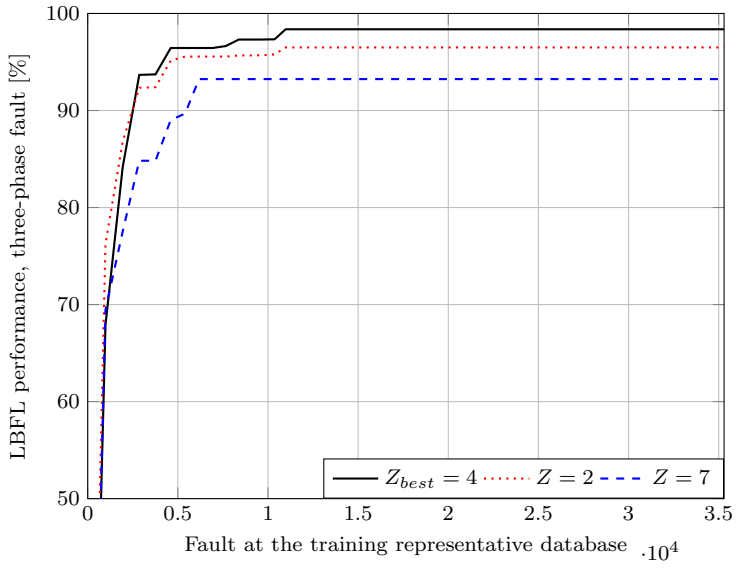

Figure 4.6: Three-phase fault performance for training using several representative databases.

As presented in the test, the adequate size of the training database (r_{best}) is defined by the one that contains enough information to identify the fault zone adequately.

Finally, the time required to determine all of the adjusting-variables and to train the complete LBFL (single-phase, phase-to-phase, three-phase faulty zone locators and fault type identifier), in an off-line setting process, is 117.1 $[h]$. Having adjusted the LBFL, the on-line locating average time is 5.2 $[ms]$, according to the

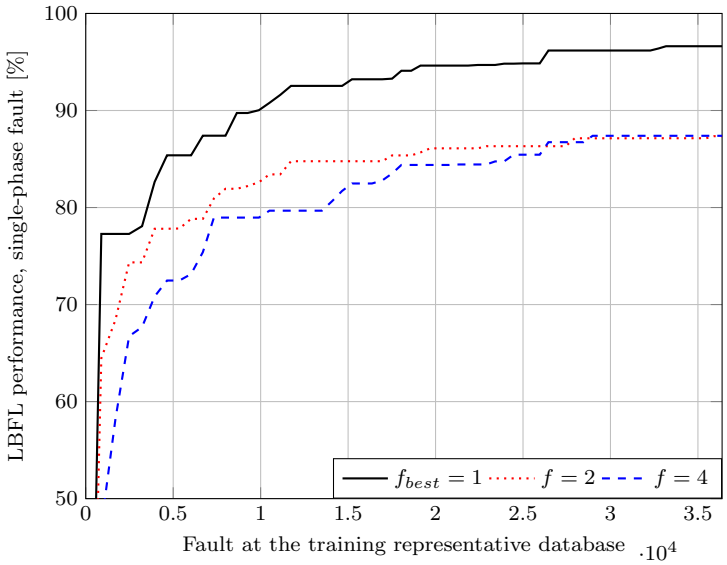

Figure 4.7: Single-phase fault performance for training using several representative databases.

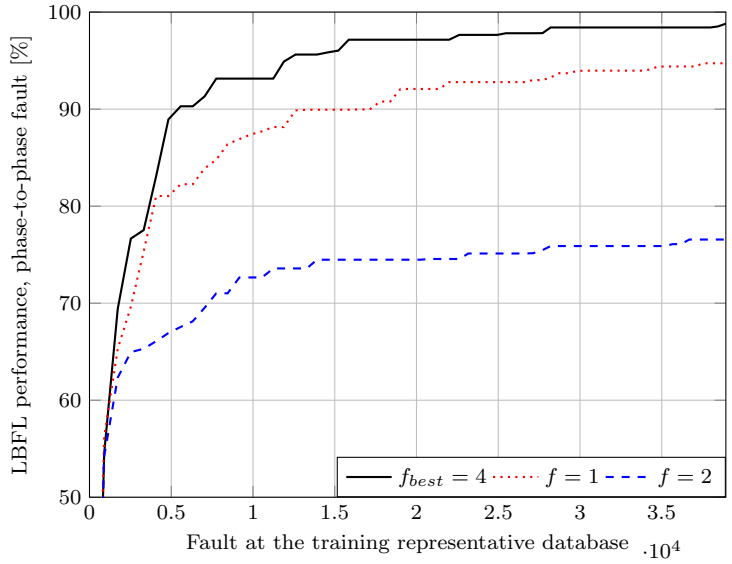

Figure 4.8: Phase-to-phase fault performance for training using several representative databases.

performed tests. This locating time is short enough, considering that faulty zone location is used to handle the distribution system reconfiguration and to guide the maintenance staff actions to correct the actual and to avoid future faults. The estimated computational effort is based on tests performed in a common $i5$ Core, 1.6 $[GHz]$, 8 $[GB]$ RAM PC.

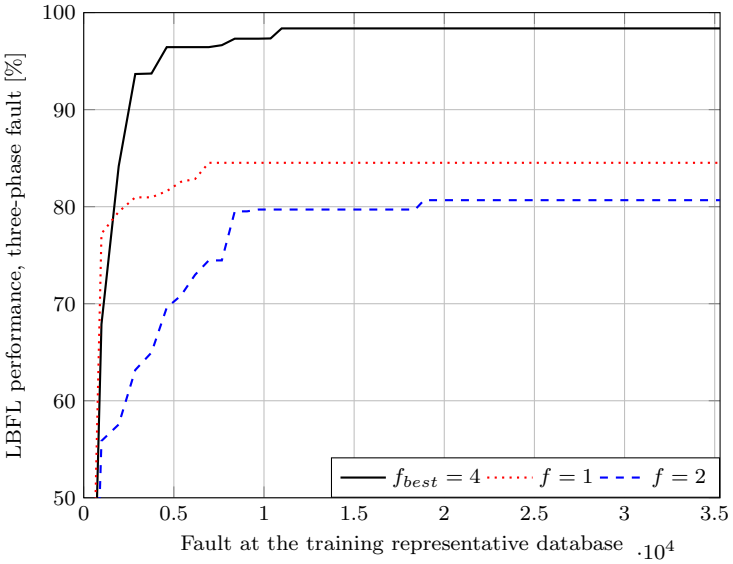

Figure 4.9: Three-phase fault performance for training using several representative databases.

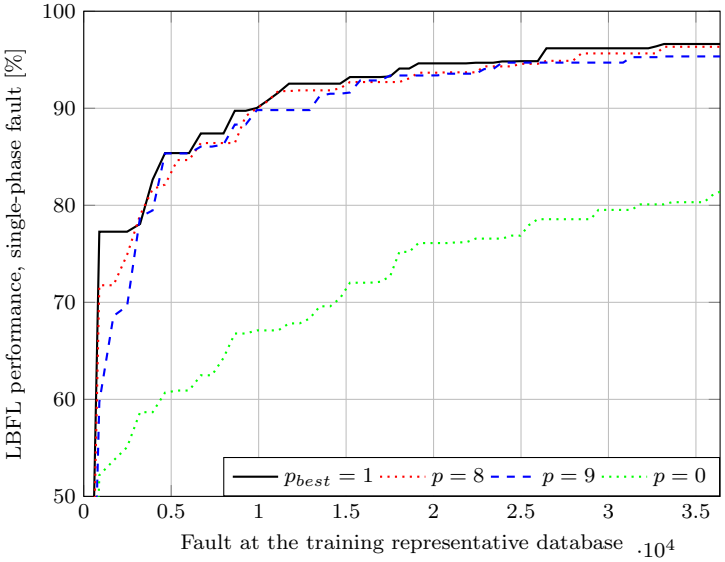

Figure 4.10: Single-phase fault performance for training using several representative databases.

4.3 General comments

The robust strategy to adjusting-variables of a learning-based fault locator has some advantages for its implementation compared to traditional general strategies, which are:

1. The use of the strategy reduces computational time.

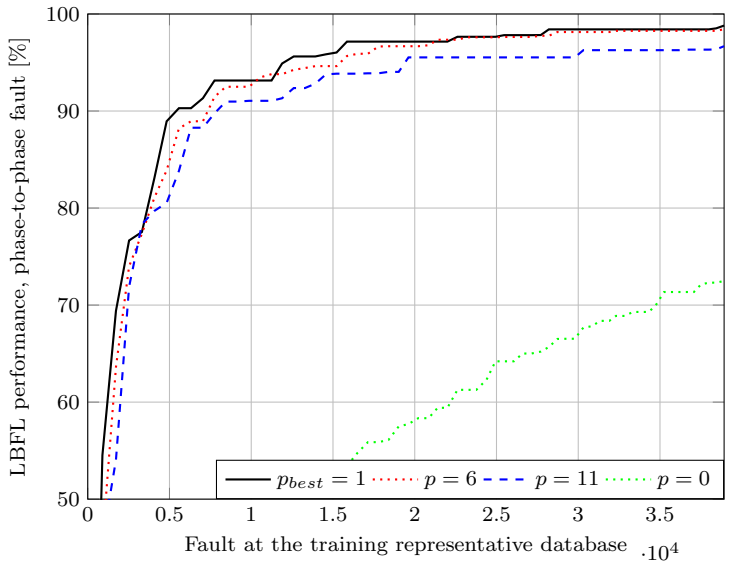

Figure 4.11: Phase-to-phase fault performance for training using several representative databases.

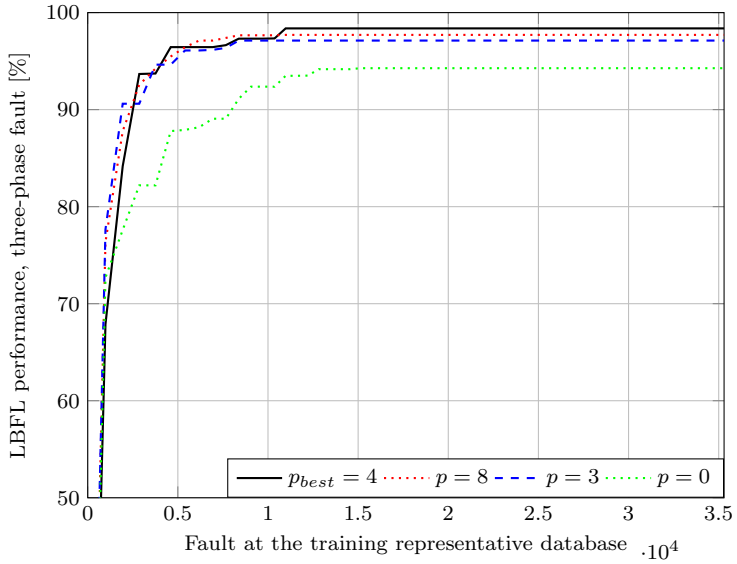

Figure 4.12: Three-phase fault performance for training using several representative databases.

2. The results of the first stage are preserved by increasing the iterations, therefore, a good selection strategy is evidenced.
3. The adjustment of the variables is a procedure that is traditionally done in a decoupled way and when analysing them together, a global solution is found, not a particular one.
4. The strategy is general enough to be able to include other variables in addition to those shown.
5. The confidence of a fault locator is increased by improving its performance with an adequate adjustment of its variables.
6. The implementation of location strategies that use measures taken in a single terminal (S / E) is encouraged, which makes their implementation more economical.
7. In the last stage it is verified that the best zone definition allows the best performance in fault location. However, zone definition alone is not enough to perform well.

4.4 Summary of the chapter

This chapter presented the design of the robust strategy for adjusting the variables of a learning-based fault locator. The strategy was described with each of its stages. The test was carried out in distribution systems. It was found that performance is significantly improved and that the solutions found to highlight a promising answer for future applications.

Chapter 5

Conclusions

This chapter summarises the main conclusions of this dissertation and provides recommendations for future work.

5.1 Main results.

Below, the conclusions about the main problems addressed in this dissertation are presented.

5.1.1 General conclusions

Power quality is a topic of great interest to the operators and users of the distribution utilities. The current importance is associated with the fact that the participation of private capital in the electricity sector that requires establishing remuneration guidelines and minimum quality indices. Among the several important aspects of quality include waveform, continuity of service and customer service.

Supply continuity indices are directly associated with the fault location. In this way, the system restoration time can be significantly lowered, reducing the duration and frequency rates of interruptions.

This dissertation is oriented on improving the performance of fault locators considering two fundamental aspects: to identify the distortion sources of the input signal that most affect locators and adjusting the locator variables through the robust strategy in order to reduce the error of fault zone location.

Previous studies do not consider the waveform distortions, and as a result of this research work, the sources of distortion that most affect the performance of the locators based on measurements taken in the substation are found. On this side, the performance of two different methods against distortions of the input signal is analysed. Satisfactory and significant results for possible applications in the field were obtained.

On the other hand, the development of the robust strategy creates a sequence of logical steps that are used successfully for the implementation of a fault zone locator.

5.1.2 Conclusions associated with the structured impedance-based fault locator considering waveform distortion

Section 3.1 presents a generalised application methodology which helps to use most of the impedance-based fault location methods in real distribution systems. Its fundamental concepts are the definition of the equivalent feeders and the actualisation of voltages and currents at each line section.

Additionally, according to the evaluation of the fault locator performance, in such cases where distorted measurements of voltage and current are considered as inputs, it is noticed that the CT saturation is the most influential distortion source. This is due because the impedance-based methods rely on pre-fault and post-fault current and voltage magnitudes, then not trustable recordings of magnitudes lead to unreliable estimations. Additionally, the result analysis shows a significant effect of the sampling frequency and harmonics distortion on the fault locator performance.

According to the performed research, the accepted specifications of the equipment must consider the following aspects: a) minimum CTs saturation level at (5/6) of the difference between the peak fault current and the rated current of the analysed power system, b) minimum Sf of 16 samples by cycle, and finally, c) maximum total harmonic distortion of 3 %. Furthermore, the values in Table 3.1 for noise, resolution, PT saturation and DC offset, have a reduced effect in the fault location performance.

Finally, the proposed analysis in section 3.1 helps to determine the requirements of these impedance-based fault location methods, regarding the main sources of signal distortion in power distribution systems. By accomplishing these requirements, the performance of the fault locator is assured in real distribution systems.

5.1.3 Conclusions associated with the structured LBFL considering distorted measurements

A structured methodology to apply a learning-based fault locator in real distribution systems is presented in section 3.2. Additionally, according to the evaluation of the LBFL performance, in such cases where distorted measurements of voltage and current are considered as inputs, it is noticed that the sampling frequency is the most influential distortion source. This is due that learning-based methods rely on phasors of pre-fault and post-fault current and voltage signal, then not trustable recordings of phasors lead to unreliable estimations. Additionally, the result analysis shows a significant effect of the CT saturation and harmonics distortion on the LBFL performance. In the case of three-phase faults, these distortions cause a strong impact on performance distortions, because the information contained in all of the phases is severely affected.

Finally, the proposed analysis in section 3.2 helps to determine the minimum requirements of these learning-based fault location methods, regarding the main sources of signal distortion in power distribution systems.

5.1.4 Conclusions associated with the robust strategy to define the adjustment of the variables of an LBFL

A robust strategy to define the adjusting-variables of an LBFL is presented in chapter 4. As these variables are firmly related, then the robust strategy considers an iterative approach based on learning and sampling theory, to determine their best values. According to the tests and the performance obtained, the results present an adequate LBFL adjusting in the case of single-phase, phase-phase and three-phase faults on two unbalanced distribution feeders, the IEEE34 system and the 34.5 [kV] system. A database of 140000 fault registers obtained using different fault resistances, in the distribution system under different operating conditions, is used to adjust and to evaluate the proposed strategy on the IEEE34 system. As a result, the best values for five adjusting-variables are selected, and the expected LBFL performance is higher than 96 [%] for all fault types.

The proposed approach in chapter 4 is useful to improve the fault locator performance and then to increase the power continuity in the actual distribution systems. This simple strategy enhances the SAIFI and SAIDI indexes, by the easy and fast identification of the faulty zone, in the case of permanent and transient faults. In this way, the development of a robust method based on stages makes a more straightforward implementation of the fault locator based in data mining techniques.

The average time required to determine one faulty zone location, which is performed on-line, is less than 10 $[ms]$. This convenient fault location alternative is the starting point to define the actions to deal with the service continuity, as these related to the distribution system reconfiguration and corrective maintenance tasks.

The case proposed in this dissertation considers the most challenging situation, where the measurements are only at the main substation. As presented in the testing section, the obtained performance is high, and the cost is low, then the proposed strategy constitutes an affordable solution for most of the distribution utilities.

Finally, the proposed strategy is oriented to determine the adjusting-variables; it is not restricted to some specific inputs. However, it is tested using only measurements at the main substation, obtaining satisfactory performance results, useful to improve the power continuity in distribution systems.

5.2 Future research

Next, some possible research topics related to fault location in electric distribution systems are suggested.

- Include other adjusting-variables in the proposed strategy presented in chapter 4, as normalisation methods, kernel type and SVM parameters.
- Perform a comparative analysis of the robust strategy proposed in chapter 4, using other classification techniques.
- Analyse the feature selection, as an excessive number of non-significant features not only increases the computational effort but also degrades the classification accuracy.
- Consider the optimal location of the measuring devices in the case of multiple sources along the distribution feeder.
- Perform experimental work considering cooperative research with distribution utilities. Utility measurements have to be used to adjust the fault locators.

Appendix A

34.5 $[kV]$ - 75 bus test system with the robust strategy proposed

This chapter is devoted to report the most significant results obtained in the additional tests performed. The proposed robust strategy presented in chapter 4, is used to adjust four different LBFL; three of them are aimed to locate faults considering each fault type (single-phase, phase to phase and three-phase faults) and one of them is used for fault type identification. According to the analysed tests, the results present a successful selection of the adjusting-variables, due to the high performance in the case of locating single-phase, phase-to-phase and three-phase faults, in a 75 node distribution system, where a database of 75000 fault registers is used. As a result, satisfactory values for the five adjusting-variables are obtained, and the expected LBFL performance is higher than 95 [%] for all fault types.

A.1 Test system

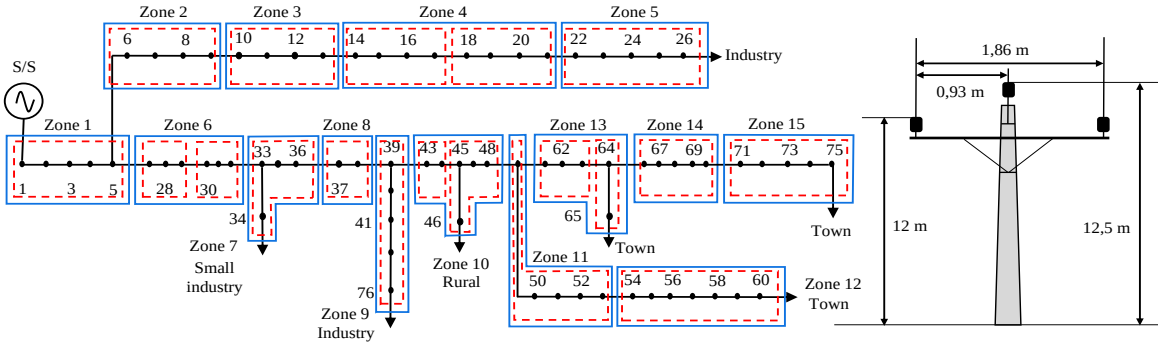

Figure A.1: 34.5 $[kV]$ 75-bus test feeder used in LBFL tests.

Tests system consist of a 34.5 $[kV]$ - 75 bus feeder from a local distribution utility, 60 $[Hz]$, implemented in ATP [29, 35, 92]. The average length of each line section between two consecutive nodes is 250 $[m]$, the average soil resistivity is 100 $[\Omega/m]$, and the skin effect is considered. The overhead line configuration is

Table A.1: Load size.

Node	Rated apparent power [kVA]	Power factor
26	270	0,80
34	410	0,80
46	275	0,91
60	2640	0,92
65	595	0,91
75	1220	0,91
76	312	0,89

presented in Fig. A.1, the conductor is ACSR 350 $[kCM]$ in a PI model for each line section. The rated loads are presented in Table A.1 and the used measurements are obtained at the distribution substation (node 1).

The used database contains 75000 faults for each fault-type. These are obtained from several operating conditions of the distribution system, which include random variations between 40 [%] to 110 [%] of rated load, ± 3 [%] of network frequency,± 10 [%] of the main substation rated voltage and ± 10 [%] of the line section length. Fault resistances from 0 to 80 [Ω] are also considered.

The minimum signal requirements consider a sampling frequency of 32 samples per cycle, CT saturation lower than 17 [%] of the difference between the maximum peak current and the rated current, and a total harmonic distortion lower than 2 [%] for voltage and current signals, as suggested in section 3.2.

The nodes are grouped as in Fig. A.1, where two zone definitions are shown: i) the proposed by the utility operator (dashed lines) and, ii) the best zone definition (continuous line) obtained using the proposed strategy for single-phase faults.

A.2 Range definition for the adjusting-variables

On the presented tests, the evaluation of the proposal considers the following ranges for the five adjusting-variables: four combinations of features (f), two types of normalisation (n), 13 different zone definitions (z), and a database of 75000 faults using fault resistances from 0 to 80 Ω for each fault-type.

Despite the constrains here defined for the presented tests, the proposed methodology is enough general to consider other testing ranges.

a) Feature selection (f)

Four combinations of features (f) are presented in Table 4.1. These combinations are selected from the features defined at section 2.2.1 in Eq. (2.10).

b) Type of normalisation (n)

From the five normalisation strategies presented in section 2.2.2, $MinMax$ (MM) and $Zscore$ (ZS) are the most commonly used according to the adequate obtained results in previous approaches; therefore these are taken into account in the generalised strategy of section 4.1. Remember that it is possible to include other normalisation alternatives [42].

c) SVM parameterisation strategy (p)

In stage 1 of the strategy, the SVM parameters (p) are selected according to the methodology presented in [66], where a radial basis function kernel is used to consider non linearly separable classes. As a complement to the

methodology presented in [66], the parameterisation strategy makes a comparison of the parameters obtained at the moment by [66] and those stored as parameters of the best solutions and then selects the best.

d) Determination of the zone division (z)

The initial definition is given by the distribution utility operator of the proposed test system (19 zones), and this zone division is presented using dashed lines in Fig. A.1, in the case of single-phase faults. At the presented tests, 13 different zone definitions are proposed, all of them eliminate the multiple estimations, and each zone size is smaller than three kilometres length.

e) Representative database (r)

The Latin Hypercube sampling strategy is used to reduce the representative database [6, 90]. To select a set of representative databases, an analysis of the fault registers per node is considered, as proposed in section 2.2.5. In the 34.5 $[kV]$ - 75 bus feeder system the complete database for a fault type is 75000 fault records, of which 1000 records are per node. The system has records on 75 nodes. Taking 12 records per node from the database, a complete database of 900 fault records is obtained. The 25 [%] of those records per node is three (3). Then the smallest representative database ($r = 1$) defined by Latin Hypercube sampling strategy is the one containing at least three (3) faults in each node. Additional representative databases are defined by using at least six ($r = 2$), nine ($r = 3$), 15 ($r = 4$) and 50 ($r = 5$) faults in each test system node. In this way, an incremental database of 881, 1267, 1549, 2491 and 6176 fault registers (J) is obtained in the case of $f = 1$ and $n = 1$. These databases are used in the second stage for this combination of f and n.

A.3 Results and analysis - Stage 1

Due to the number of intermediate results, only these for the single-phase fault in the first iteration are shown. The results for other faults types are similar. Additionally, the first stage zone definition has 19 zones as defined by the utility operator.

Table A.2: Results of the first stage for the 34.5 $[kV]$ system at the first iteration, in the case of single-phase faults.

f	n	J	**Validation time / fault** $[ms]$	***Per* [%]**	$e\_index_{(f,n)}$ **[%]**
1	1	881	6.8	74.3	63.4
1	**2**	**783**	**3.4**	**65.3**	**64.1**
2	1	660	8.5	53.2	42.6
2	2	698	3.2	42.2	46.3
3	1	959	6.9	48.8	42.8
3	2	775	3.7	36.5	40.4
4	1	775	6.5	66.8	58.2
4	2	866	3.8	37.7	41.2

In Table A.2, the eight rows correspond to the combinations between (n) and (f). From the $e\_index_{(f,n)}$, the best combination is the second with a value of 64.1 [%]. From the exposed, a value of $\alpha = 0.8$ is selected, that is, 80 [%] of priority is given to performance and 20 [%] to the validation time of each fault record. In this way, the combination is obtained with a validation time per fault of 3.4 $[ms]$ and a performance of 65.3 [%]. Although the fourth combination has the lowest validation time (3.2 $[ms]$), and the first one has the

highest performance (74.3 [%]), the second combination has the best $e\_index_{(f,n)}$. As a consequence, the outputs of stage 1 are $f_{best} = 1$ and $n_{best} = 2$ in the case of single-phase faults.

In the case of phase-to-phase and three-phase faults the obtained results at the $r = 1$ in stage 1 are $f_{best} = 1$, $n_{best} = 1$ and $f_{best} = 4$, $n_{best} = 2$, respectively.

The validation time in this first stage is considered very important because, for the second stage, it is necessary to increase the database and validate with other types of zone definition. To efficiently perform this procedure, variables with lower computational cost must be considered. The validation time depends on the number of faults; therefore, the time used in the $e\_index_{(f,n)}$ is required to validate a fault.

The total time used to execute the first stage is 1.8 $[h]$ for single-phase fault, 1.4 $[h]$ for phase-to-phase fault and 1.7 $[h]$ for three-phase faults.

A.3.1 Results and analysis - Stage 2

In the first stage, the adjusting-variables $f = 1$ and $n = 2$ are defined in the case of the single-phase fault locator. In the second stage, the 13 different zone definitions are considered; the utility operator defines the first one. The remaining zone definitions are analysed in this stage, as is presented in Table A.3, in the case of using the smallest representative database ($r = 1$ or $J = 783$).

Table A.3: Results at the second stage for the smallest representative database ($r = 1$) and all zone definitions for the single-phase fault locator (34.5 $[kV]$).

Zone definition (z)	**Number of zones**	p		**Validation time / fault [ms]**	Per [%]	z_index [%]
		C	σ			
1	19	16384	0.38849	3.4	65.3	57.3
2	14	5000000	0.47660	2.4	70.2	59.9
3	**15**	**16384**	**0.38849**	**2.5**	**78.7**	**67.0**
4	16	65536	0.34124	2.6	69.0	59.5
5	17	5000000	19.00000	2.7	69.8	60.4
6	18	16384	0.38849	2.8	71.5	62.0
7	19	128	0.91324	2.9	65.0	57.1
8	20	4096	0.23652	3.0	62.8	55.6
9	21	32768	0.47660	3.1	65.3	57.8
10	20	5000000	0.4766	3.0	59.8	53.2
11	18	4096	0.27185	2.8	65.5	57.2
12	17	16384	0.38849	2.8	68.6	59.4
13	15	32768	0.47660	2.7	70.7	60.6

From Table A.3 and considering the zone index (z_index), it is found that the best combination has a value of 67.0 [%], which corresponds to a 15 zones (zone definition number 3, $z = 3$), a performance (Per) of 78.7 [%] and a validation time per fault of 2.5 $[ms]$. It is denoted from this validation time, the adequate selection obtained from the e_index in the previous stage, considering this approach as an iterative process where the time is an important aspect. Additionally, a value of $\beta = 0.8$ is used to prioritise either the performance (Per) and the ratio between the number of zones and nodes ($\frac{N_z}{N_n}$), because as previously explained, performance is usually higher in the case of a low number of zones and an adequate balance is required.

The total time used to execute the second stage is 2.1 $[h]$ for single-phase fault, 1.8 $[h]$ for phase-to-phase fault and 1.9 $[h]$ for three-phase faults.

According to the results for the smallest representative database ($r = 1$ or $J = 783$) at second stage, the best adjusting variables are $f_{best} = 1$, $n_{best} = 2$ and $z_{best} = 3$ (15 zones defined in the power system) for the single-phase fault locator.

Finally, the following sets of adjusting-variables $f_{best} = 1$, $n_{best} = 1$, $z_{best} = 2$ (14 zones defined in the power system) and $f_{best} = 4$, $n_{best} = 2$, $z_{best} = 2$ are obtained for the smallest representative database ($r = 1$), in the case of phase-to-phase and three-phase fault locator, respectively.

A.3.2 Results and analysis - Stage 3

Stage 3 relates stages 1 and 2, then the proposed strategy continues with the remaining iterations, repeating the first stage with the new zone definition obtained in the second stage, as is depicted in Fig. 4.1.

In table A.4, all of the iterations executed by the proposed strategy for the three fault types are presented. As results for the single-phase fault locator, the following adjusting-variables are obtained: $f_{best} = 1$, $n_{best} = 1$ and $z_{best} = 13$ and p_{best} is defined by $C = 32768$ and $\sigma = 0.4766$. These adjusting variables are selected, due the maximum performance at this zone definition (89.5 [%]), obtained at iteration $i = 3$ and confirmed in iteration $i = 5$.

In the case of phase-to-phase and three-phase fault locators, the results are also presented in Table A.4. In all fault types, the third stop comparison shown in section 4.1.3 was fulfilled, as is shown in high quality solutions from Table A.4. For the first type of fault, the strategy ends in iteration 5 ($i = 5$), however, two more iterations ($i = 6$ and $i = 7$) are left to observe the behavior. From observation it is found that the strategy repeats $z_{best} = 3$ and $z_{best} = 2$. That finding can lead to confusion as to which is the best zone definition. If performance (Per) is taken as a selection criterion, it is found that $z_{best} = 13$, however in stage 4, $z_{best} = 3$ and $z_{best} = 1$ (it is found that $z_{best} = 2$ has a worse performance with the p_{best} compared to $z_{best} = 1$) are evaluated in the same Fig. A.2 to show that $z_{best} = 13$ is really the best.

Additionally, and as it is observed from Table A.4, the LBFL performance is higher than in the previous stages. This is due to the increasing in the representative database size; then as higher is the representative database, usually higher performances are reached, as a more significant number of faults are used in LBFL training.

The total iterations for single-phase fault, phase-to-phase fault and three-phase fault were 5, 5 and 7 respectively. The total time used to execute the third stage is $19,5$ [h] for single-phase fault, 16 [h] for phase-to-phase fault and $25,2$ [h] for three-phase fault. The previous time is the minimum and is calculated by multiplying the number of iterations with the sum of the times of the first two stages. Time may increase in the second stage because the computational cost is directly proportional and not linear to the increments of the database.

A.3.3 Results and analysis - Stage 4

From previous stages, the adjusting-variables f_{best}, n_{best} ,z_{best} and p_{best} are defined, then this last stage defines the best representative database r_{best} and the expected LBFL performance (Per_{exp}) are determined. Training and testing of LBFL are sequentially performed. Initially, the smallest representative database is used in training, and each iteration consists of using a new representative database from the smallest to the biggest one. The testing database is always the same.

The training process is performed using different-size representative databases. In the case of single-phase faults, the representative database varies from 50 to 38483 faults; for phase-to-phase faults, this varies from 49 to 28783 faults, and in the case of three-phase faults this varies from 55 to 39137 faults. The testing database used in the three types of faults contains 750 faults not previously used in the training stage.

Table A.4: Best solutions of the third stage for the 34.5 $[kV]$ system of all fault locators. Combinations with the best performance (incumbents).

Fault type	i	f_{best}	n_{best}	z_{best}	**Number of zones**	p_{best}		Per [%]
						C	σ	
Single-phase	1	1	1	1	19	16384	0.3885	74.3
	2	1	1	3	15	16384	0.3885	87.5
	3	1	1	13	15	32768	0.4766	89.5
	4	1	1	2	14	16384	0.3885	84.5
	5	**1**	**1**	**13**	**15**	**32768**	**0.4766**	**89.5**
	6	1	1	3	15	16384	0.3885	87.5
	7	1	1	2	14	16384	0.3885	84.5
Phase to phase	1	1	1	1	19	65536	0.3412	57.7
	2	1	1	2	14	16384	0.3885	79.2
	3	1	2	13	15	32768	0.4766	68.1
	4	1	1	3	15	16384	0.3885	73.7
	5	**1**	**1**	**13**	**15**	**5000000**	**0.4766**	**81.9**
Three-phase	1	4	2	1	19	128	0.9132	75.7
	2	4	2	2	14	4096	0.2365	81.1
	3	4	1	13	15	5000000	19.431	84.5
	4	4	1	7	19	16384	0.3885	81,9
	5	4	1	3	15	32768	0.2235	78.7
	6	4	1	11	18	4096	0.2718	81.6
	7	**4**	**1**	**13**	**15**	**5000000**	**19.431**	**84.5**

Fig. A.2 shows the obtained LBFL performance for three different sets of adjusting-variables, in case of single-phase faults. In all cases, the adjusting-variables related to features, normalisation and SVM

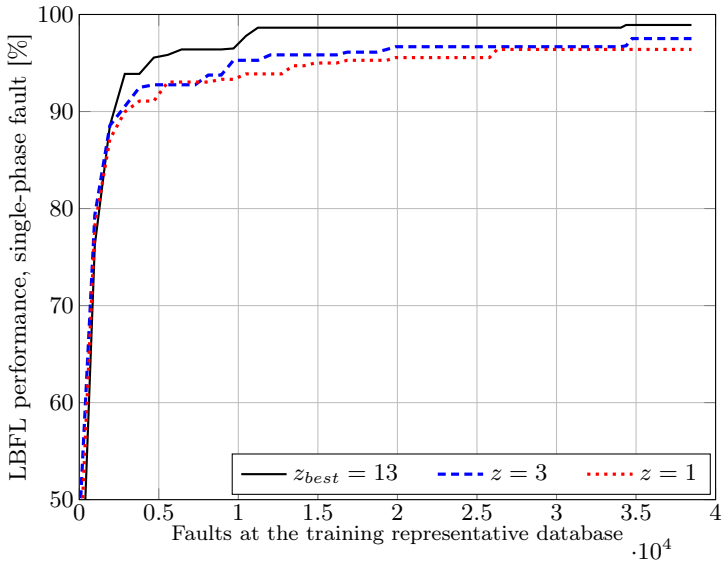

Figure A.2: Single-phase fault performance for training using several representative databases.

parameters are these obtained in the previous stages (f_{best}, n_{best} and p_{best}), while the zone definitions are: the z_{best} (continuous line), the zone defined by the utility operator (dotted line), and one additional zone definition with excellent performance in the intermediate stages (dashed line). According to the incremental training, the size of the minimum representative database (r_{best}) for the single-phase fault type is 11230

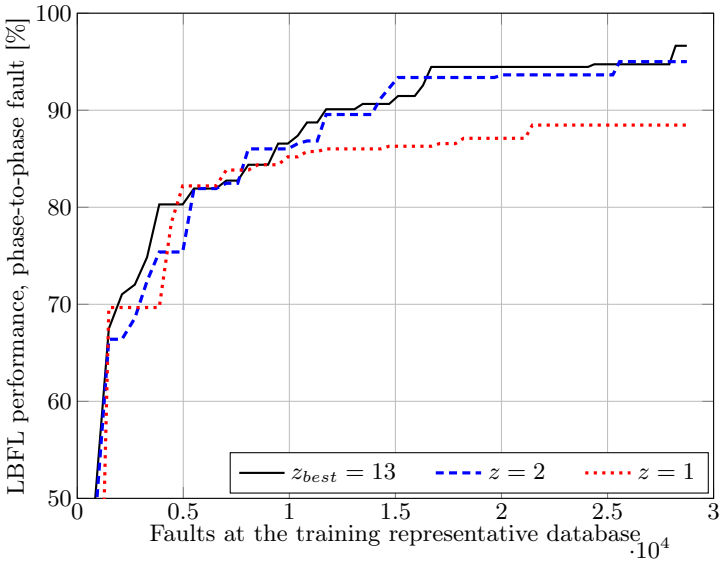

Figure A.3: Phase-to-phase fault performance for training using several representative databases.

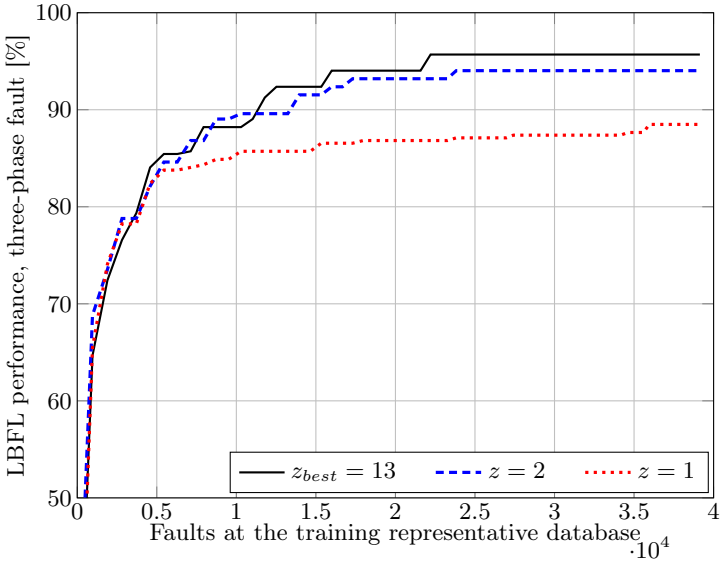

Figure A.4: Three-phase fault performance for training using several representative databases.

faults. Additionally, a measure of the LBFL confidence is the expected performance (Per_{exp}), which in the case of single-phase faults is over 98.58 [%]. In the case of phase-phase faults, $r_{best} = 27690$ faults and $Per_{exp} = 96.46$ [%], while for three-phase faults, $r_{best} = 22210$ faults and $Per_{exp} = 95.54$ [%], as shown Figs. A.3 and A.4. The performance of the fault type identification LBFL module is 99.21 [%].

As presented in the test, the adequate size of the training database (r_{best}) is defined by the one that contains enough information to identify the faulty zone adequately.

Finally, the time required to determine all of the adjusting-variables and to train the complete LBFL (single-phase, phase-to-phase, three-phase faulty zone locators and fault type identifier), in an off-line setting process, is 60.7 [h]. Having adjusted the LBFL, the on-line locating average time is 4.7 [ms], according to the performed tests. This locating time is short enough, considering that faulty zone location is used to handle the distribution system reconfiguration and to guide the maintenance staff actions to correct the actual and to avoid future faults. The estimated computational effort is based on tests performed in a common $i5$ Core, 1.6 [GHz], 8 [GB] RAM PC.

Bibliography

[1] A. Baggini . *Handbook of Power Quality*. June, 2008.

[2] W. E. Reid. Power quality issues - standards and guidelines. *Industry Applications, IEEE Transactions on*, 32(3):625 – 632, June, 1996.

[3] André Luís Da Silva Pessoa, Mário Oleskovicz, and Paulo Martins. Sensibility analysis of a fault location method based on ANN, WPT and decision tree in distribution systems. *Journal of Control, Automation and Electrical Systems*, 31, 05 2020.

[4] M. Mirzaei , Z. Kadir , E. Moazami , and H. Hizam . Review of fault location methods for distribution power system. *Australian Journal of Basic and Applied Sciences*, 3(3):2670–2676, July, 2009.

[5] A. P. Bíscaro, R. F. Pereira, M. Kezunovic, and J. S. Mantovani. Integrated fault location and power-quality analysis in electric power distribution systems. *IEEE Transactions on Power Delivery*, 31(2):428–436, April, 2016.

[6] A. Bahmanyar , S. Jamali , A. Estebsari , and E. Bompard . A comparison framework for distribution system outage and fault location methods. *Electric Power Systems Research*, 145:19– 34, April, 2017.

[7] R. Sowah, N. Dzabeng, A. Ofoli, A. Acakpovi, K. Koumadi, J. Ocrah, and D. Martin . Design of power distribution network fault data collector for fault detection, location and classification using machine learning. In *2018 IEEE 7th ICAST*, pages 1–8, August, 2018.

[8] T. S. Rao, S. T. Ram, and J. V. Subrahmanyam . Fault recognition and diagnosis based on a decision tree for power distribution systems. *Journal of Computational Mechanics, Power System and Control (JCMPS)*, 2(2), 2019.

[9] S. Shi, B. Zhu, A. Lei, and X. Dong. Fault location for radial distribution network via topology and reclosure-generating traveling waves. *IEEE Transactions on Smart Grid*, 10(6):6404–6413, November, 2019.

[10] R. Paternost , T. Ferreira , and F. Trindade . Performance of a fault location method in the context of modern electric power distribution systems. pages 1–6, September, 2019.

[11] Colombia: CREG. Comisión de regulación de energía y gas, calidad del servicio. *Resolución CREG 015 de 2018 (capítulo 5)*, January, 2018.

[12] D. Carta, P. A. Pegoraro, S. Sulis , M. Pau , F. Ponci , and A. Monti . A compressive sensing approach for fault location in distribution grid branches. pages 1–6, September, 2019.

[13] Nikolaos Sapountzoglou, Jesus Lago, Bart De Schutter, and Bertrand Raison. A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids. *Applied Energy*, 276:115299, 2020.

[14] S. Das, N. Karnik, and S. Santoso. Distribution fault-locating algorithms using current only. *IEEE Transactions on Power Delivery*, 27(3):1144–1153, July, 2012.

[15] G. Morales-España, J.J. Mora-Flórez, and H. Vargas Torres. Fault location method based on the determination of the minimum fault reactance for uncertainty loaded and unbalanced power distribution systems. pages 803 – 809, December, 2010.

[16] M. Choi, S. J. Lee, D. S. Lee, and B. G. Jin. A new fault location algorithm using direct circuit analysis for distribution systems. *Power Delivery, IEEE Transactions on*, 19(1):35 – 41, February, 2004.

[17] S. Lee, M. Choi, S. H. Kang, B. G. Jin, D. S. Lee, B. S. Ahn, N. S. Yoon, H. Y. Kim, and S. B. Wee. An intelligent and efficient fault location and diagnosis scheme for radial distribution systems. *Power Delivery, IEEE Transactions on*, 19(2):524 – 532, May, 2004.

[18] A. Bahmanyar and S. Jamali. A new fault location method for distribution networks using sparse measurements. *International Journal of Electrical Power Energy Systems*, 81:459–468, October, 2016.

[19] A. Borghetti, M. Bosetti, C. A. Nucci, P. Mario, and A. Abur. Integrated use of time-frequency wavelet decompositions for fault location in distribution networks: Theory and experimental validation. 25(4):3139 – 3146, October, 2010.

[20] A. Rafinia and J. Moshtagh. A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ann and fls. *International Journal of Electrical Power Energy Systems*, 55:261–274, February, 2014.

[21] M. Pourahmadi Nakhli and A. A. Safavi. Path characteristic frequency-based fault locating in radial distribution systems using wavelets and neural networks. *Power Delivery, IEEE Transactions on*, 26(2):772 – 781, April, 2011.

[22] L. Ye, D. You, X. G. Yin, K. Wang, and J. Wu. An improved fault-location method for distribution system using wavelets and support vector regression. *International Journal of Electrical Power Energy Systems*, 55:467–472, February, 2014.

[23] D. Gazzana, G. Ferreira, A. Bretas, A.L. Bettiol, A. Carniato, L. Passos, A.H. Ferreira, and J. E. M. Silva. An integrated technique for fault location and section identification in distribution systems. *Electric Power Systems Research*, 115:65–73, October, 2014.

[24] R. Dashti and J. Sadeh. Fault section estimation in power distribution network using impedance-based fault distance calculation and frequency spectrum analysis. *Generation, Transmission Distribution, IET*, 8:1406–1417, August, 2014.

[25] G. Weng, F. Huang, Y. Tang, J. Yan, Y. Nan, and H. He. Fault-tolerant location of transient voltage disturbance source for dg integrated smart grid. *Electric Power Systems Research*, 144:13–22, March, 2017.

[26] E. Hajipour, M. Vakilian, and M. Sanaye-Pasand. Current-transformer saturation prevention using a controlled voltage-source compensator. *IEEE Transactions on Power Delivery*, 32:1–1, June, 2016.

[27] M. Kaczmarek. Secondary current distortion of inductive current transformer in conditions of dips and interruptions of voltage in the power line. *Electric Power Systems Research*, 137:1–5, April, 2016.

[28] W.L.A. Neves, N.S.D. Brito, B.A. Souza, A.V. Fontes, K. Dantas, A.B. Fernandes, and S.S.B. Silva. Sampling rate of digital fault recorders influence on fault diagnosis. pages 406 – 411, December, 2004.

[29] J. J. Mora, J. Bedoya, and J. Meléndez. Extensive events database development using atp and matlab to fault location in power distribution systems. *IEEE PES TD: Latin America, Caracas*, January, 2006.

[30] J. Dagenhart. The 40- ground-fault phenomenon. *Industry Applications, IEEE Transactions on*, 36(1):30 – 32, February, 2000.

[31] T.A. Short. *Electric Power Distribution Handbook*. CRC Press , 2018.

[32] I.D. Serna Suárez, G.H. Carrillo Caicedo, and R. Vargas Torres. Revisión de técnicas de estado estable y transitorio para la localización de fallas en sistemas de distribución. *Revista UIS Ingenierías*, 9(1):23–38, 2010.

[33] A. Bahmanyar and A. Estebsari . A practical integrated fault location method for electrical power distribution networks. pages 1–5, June, 2018.

[34] F. Aboshady, D. Thomas, and M. Sumner. A new single end wideband impedance based fault location scheme for distribution systems. *Electric Power Systems Research*, 173:263 – 270, 2019.

[35] E. J. Correa Tapasco, J. J. Mora Flórez, and S. Perez Londoño. Robustness of a generalized impedance based fault locator considering distorted measurements. *Electric Power Systems Research*, 154:234–244, 2018.

[36] M. Daisy and R. Dashti . Single phase fault location in electrical distribution feeder using hybrid method. *Energy*, 103:356–368, May, 2016.

[37] C. Orozco, A. Bretas, R. Chouhy Leborgne , A. Herrera-Orozco, and J. Marín. Active distribution network fault location methodology: A minimum fault reactance and fibonacci search approach. *International Journal of Electrical Power Energy Systems*, 84:232–241, January, 2017.

[38] M. Majidi and M. Etezadi. A new fault location technique in smart distribution networks using synchronized/nonsynchronized measurements. *IEEE Transactions on Power Delivery*, 33(3):1358–1368, June, 2018.

[39] Y. Mohammadia , M. H. Moradi , and R. Chouhy Leborgne . A novel method for voltage-sag source location using a robust machine learning approach. *Electric Power System Research*, 145:122 – 136, 2017.

[40] A. Shahsavari, M. Farajollahi, E. Stewart, E. Cortez, and H. Mohsenian-Rad. A machine learning approach to event analysis in distribution feeders using distribution synchrophasors. In *2019 International Conference on SGSMA*, pages 1–7, May, 2019.

[41] N. Coleman, C. Schegan, and K. Miu. A study of power distribution system fault classification with machine learning techniques. In *2015 North American Power Symposium (NAPS)*, pages 1–6, October, 2015.

[42] E. Correa-Tapasco , J. J. Mora-Flórez, and S. Perez-Londono. Performance analysis of a learning structured fault locator for distribution systems in the case of polluted inputs. *Electric Power Systems Research*, 166:1 – 8, 2019.

[43] Jair Cervantes, Farid Garcia-Lamont, Lisbeth Rodríguez-Mazahua, and Asdrubal Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 2020.

[44] J. Liang, T. Jing, H. Niu, and J. Wang. Two-terminal fault location method of distribution network based on adaptive convolution neural network. *IEEE Access*, 8:54035–54043, 2020.

[45] C. Grajales-Espinal, J. J. Mora-Flórez, and S. Perez-Londono. Advanced fault location strategy for modern power distribution systems based on phase and sequence components and the minimum fault reactance concept. *Electric Power Systems Research*, 140:933 – 941, 2016.

[46] C. Orozco, J. Mora, J. Marín, J. C. Velez, M. da Silva, and S. Perez. Fault location system for active distribution networks. In *2019 IEEE PES ISGT Conference - Latin America*, pages 1–6, September, 2019.

[47] M. Shafiullah, M. Ijaz, M. Abido, and Z. Al-Hamouz. Optimized support vector machine and wavelet transform for distribution grid fault location. In *2017 11th IEEE International Conference on CPE-POWERENG*, pages 77–82, April, 2017.

[48] J. Xu, Z. Wu, X. Yu, and C. Zhu. Robust faulted line identification in power distribution networks via hybrid state estimator. *IEEE Transactions on Industrial Informatics*, 15(9):5365–5377, September, 2019.

[49] P. Barra, A. Pessoa, T. Menezes, G. Santos, D. Coury, and M. Oleskovicz. Fault location in radial distribution networks using ann and superimposed components. In *2019 IEEE PES ISGT Conference - Latin America*, pages 1–6, September, 2019.

[50] N. d. O. P. Westin, C. A. V. Guerrero, J. M. d. C. Filho, N. B. Pereira, P. M. Silveira, and T. C. d. Oliveira. Comparison between voltage sag simulation methodologies in distribution systems. In *2020 19th International Conference on Harmonics and Quality of Power (ICHQP)*, pages 1–6, 2020.

[51] D. Shakya and S. N. Singh. Svm based fault location and classification using fuzzy classifier for pq monitoring. pages 1 – 8, August, 2008.

[52] P. Sang, S. Shu, and M. Zhou. An online fault detection model and strategies based on svm-grid in clouds. *IEEE/CAA Journal of Automatica Sinica*, 5(2):445–456, March, 2018.

[53] D. Arredondo Arteaga , W. Gil González, and J. Mora Flórez. Methodology for selection of attributes and operating conditions for svm-based fault locator's. *Tecnura*, 21:15–26, March, 2017.

[54] E. J. Correa Tapasco, J. J. Mora Flórez, and S. M. Pérez Londono. Comparison of measurement features used as inputs in a learning based fault location method for power distribution systems. *Revista UIS Ingenierías*, 18:73–80, February, 2019.

[55] J. J. Mora Flórez, V. Barrera Nunez, and G. Carrillo Caicedo. Fault location in power distribution systems using a learning algorithm for multivariable data analysis. *in IEEE Transactions on Power Delivery*, 22(3):1715–1721, July, 2007.

[56] X. Magagula, Y. Hamam, J. Jordaan, and A. Yusuff . A fault classification and localization method in a power distribution network. pages 1337–1343, September, 2017.

[57] H. Livani and C. Y. Evrenosoglu . A fault classification method in power systems using dwt and svm classifier. pages 1–5, May, 2012.

[58] R. H. Salim, K.C.O. Salim, and A. Bretas. Further improvements on impedance-based fault location for power distribution systems. *Generation, Transmission Distribution, IET*, 5(4):467 – 478, May, 2011.

[59] R. Das . *Determining the locations of faults in distribution systems.* PhD thesis, January, 1998.

[60] M. D. Mckay , R. J. Beckkman , and W. Conover . A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, January, 2000.

[61] K. T. Fang , F. Li, and A. Sudjianto. Design and modeling for computer experiments. 304, January, 2006.

[62] J. Bedoya, J. Meléndez, and J. J. Mora-Flórez. Implementación de protecciones y simulación automática de eventos para localización de fallas en sistemas de distribución de energía. *Revista Ingeniería y Competitividad, Universidad del Valle*, 8, July, 2006.

[63] F. Jaimes Flórez and R. J. Villar Cruz. Caracterización de circuitos de distribución para estudios de calidad en sistemas de energía eléctrica. 2012.

[64] L. Pérez Hernandez, S. Pérez Londoño, and J. Mora Flórez. Diseño de una herramienta eficiente de simulación automática de fallas en sistemas eléctricos de potencia. *Trabajo de grado, Universidad Tecnológica de Pereira, Colombia*, 77(164):178–188, 2010.

[65] S. Kotsiantis , D. Kanellopoulos , and P. Pintelas. Data preprocessing for supervised learning. *International Journal of Computer Science*, 1(2):111–117, January, 2006.

[66] Cheng-Hsuan Li, Hsin-Hua Ho, Yu-Lung Liu, Chin-Teng Lin, Bor-Chen Kuo, and Jin-Shiuh Taur. An automatic method for selecting the parameter of the normalized kernel function to support vector machines. *Journal Inf. Sci. Eng.*, 28:1–15, 01 2012.

[67] J. Marín. Análisis del efecto de la variación de parámetros de modelo de un sistema de distribución sobre las metodologías de localización de fallas paralelas. *Tesis Maestría, Universidad Tecnológica de Pereira, Colombia*, 2013.

[68] N. Alzate. Influencia de la variación de parámetros del sistema de potencia en la localización de fallas con métodos de clasificación. *Trabajo de grado, Universidad Tecnológica de Pereira, Colombia*, August, 2013.

[69] N. Alzate. Metodología para el análisis de sensibilidad de los localizadores de fallas ante los parámetros de modelado del sistema de potencia. *Tesis Maestría, Universidad Tecnológica de Pereira, Colombia*, February, 2016.

[70] M. Ramos, A. Bretas, D. Bernardon, and L. Pfitscher. Distribution networks HIF location: A frequency domain system model and WLS parameter estimation approach. *Electric Power Systems Research*, 146:170–176, May, 2017.

[71] J. J. Mora Flórez, G. Morales España, and S. Perez Londono. Learning-based strategy for reducing the multiple estimation problem of fault zone location in radial power systems. *IET Generation, Transmission Distribution*, 3(4):346–356, April, 2009.

[72] T. Zhang and Z. Zhou. Optimal margin distribution machine. *IEEE Transactions on Knowledge Data Engineering*, 32(06):1143–1156, jun 2020.

[73] J. J. Mora Flórez. Localización de faltas en sistemas de distribución de energía eléctrica usando métodos basados en el modelo y métodos basados en el conocimiento. *Tesis Doctoral, Universidad de Girona, España*, 2006.

[74] T. Dai and Y. Dong. Introduction of svm related theory and its application research. In *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pages 230–233, 2020.

[75] V. V. Vijayachandran and U. J. Shenoy. Implementation of support vector machine based relay coordination scheme for distribution system with renewables. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, pages 1–1, 2020.

[76] Debasmita Pradhan, Biswajit Sahoo, Bijan Bihari Misra, and Sudarsan Padhy. A multiclass svm classifier with teaching learning based feature subset selection for enzyme subclass classification. *Applied Soft Computing*, 96:106664, 2020.

[77] G. Sharma, A. Panwar, I. Nasiruddin, and R. C. Bansal. Non-linear ls-svm with rbf-kernel-based approach for agc of multi-area energy systems. *IET Generation, Transmission Distribution*, 12(14):3510–3517, 2018.

[78] S. Beheshtaein, R. Cuzner, M. Savaghebi, S. Golestan, and J. M. Guerrero. Fault location in microgrids: a communication-based high-frequency impedance approach. *IET Generation, Transmission Distribution*, 13(8):1229–1237, 2019.

[79] F. Bai and R. Liu. Improved nonparallel hyperplanes support vector machines for multi-class classification. *In 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pages 1–5, 11 2018.

[80] J. Bao, J. Nie, C. Liu, B. Jiang, F. Zhu, and J. He. Improved blind spectrum sensing by covariance matrix cholesky decomposition and rbf-svm decision classification at low snrs. *IEEE Access*, PP:1–1, 07 2019.

[81] A. Zafari, R. Zurita-Milla, and E. Izquierdo-Verdiguier. A multiscale random forest kernel for land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2842–2852, 2020.

[82] G. Morales y A. Gómez. Estudio e implementación de una herramienta basada en máquinas de soporte vectorial aplicada a la localización de fallas en sistemas de distribución. *Trabajo de Grado. Universidad Industrial de Santander*, page 88, 2005.

[83] K. Gayathr and N. Kumarappan . Accurate fault location on ehv lines using both rbf based support vector machine and scalcg based neural network. *Expert Syst. Appl.*, 37:8822–8830, December, 2010.

[84] L. Shalabi and Z. Shaaban . Normalization as a preprocessing engine for data mining and the approach of preference matrix. pages 207 – 214, June, 2006.

[85] W. J. Gil González. Clasificador robusto basado en máquinas de soporte vectorial para la localización de fallas en sistemas de distribución. *Tesis de maestría, Universidad Tecnológica de Pereira, Colombia*, 2013.

[86] A. R. Herrera Orozco. Análisis de los efectos de la variación de los parámetros del modelo de línea, de carga y de fuente, en la localización de fallas en sistemas de distribución. *Trabajo de grado, Universidad Tecnológica de Pereira, Colombia*, 2013.

[87] D. J. Arredondo A. Desarrollo y validación de la metodología generalizada para localización de fallas basada en técnicas de minería de datos. *Tesis de Maestría, Universidad Tecnológica de Pereira*, August, 2016.

[88] T. Dasu and T. Johnson. *Exploratory Data Mining and Data Cleaning*, volume 101. January, 2003.

[89] M. Liefvendahl and R. Stocki. A study on algorithms for optimization of latin hypercubes. *Journal of Statistical Planning and Inference*, 136:3231–3247, September, 2006.

[90] S. Mousavirad, A. Bidgoli, and S. Rahnamayan. Tackling deceptive optimization problems using opposition-based de with center-based latin hypercube initialization. In *2019 14th ICCSE*, pages 394–400, July, 2019.

[91] Andrés Felipe Zapata T. Implementación y comparación de técnicas de localización de fallas en sistemas de distribución basadas en minería de datos. *Tesis Maestría, Universidad Tecnológica de Pereira, Colombia*, 2013.

[92] Radial IEEE 34 node test feeder, distribution system subcommittee, available from:. *IEEE.ORG*, https://site.ieee.org/pes-testfeeders/resources/, accessed 27 September 2020.

[93] Milad Gil, Ali Akbar Abdoos, and Majid Sanaye-Pasand. A precise analytical method for fault location in double-circuit transmission lines. *International Journal of Electrical Power Energy Systems*, 126:106568, 2021.

[94] D. Novosel, D. Hart, Y. Hu, and J. Myllymaki. System for locating faults and estimating fault resistance in distribution networks with tapped loads. *US Patent number 5,839,093*, November, 1998.

[95] R. Salim, M. Resener, A. Filomena, K. Oliveira, and A. Bretas. Extended fault-location formulation for power distribution systems. *Power Delivery, IEEE Transactions on*, 24(2):508–516, April, 2009.

[96] J. Gutiérrez Gallego, J. Mora Flórez, and S. Pérez Londoño. Strategy based on genetic algorithms for an optimal adjust of a support vector machine used for locating faults in power distribution systems. *Revista Facultad de Ingeniería*, (53):174–187, July, 2010.

[97] R. A. Agrawal and D. Thukaram. Identification of fault location in power distribution system with distributed generation using support vector machines. *IEEE PES Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–6, 2013.

[98] S. Shilpa Gururajapathy, H. Mokhlis, H. Illias, A. H. Abu Bakar, and L. Awalin. Fault identification in an unbalanced distribution system using support vector machine. *Journal of Electrical Systems*, 12:786–800, October, 2016.

[99] B. Schölkopf and A.J. Smola . *Learning with Kernels - Support Vector Machines, Regularization, Optimization and Beyond.*, volume 98. January, 2001.

[100] P. H. Chen, C. J. Lin, and B. Schölkopf. A tutorial on -support vector machines. *Applied Stochastic Models in Business and Industry*, 21:111 – 136, March, 2005.

[101] C. C. Chang and C. J. Lin. Libsvm: a library for support vector machines. *ACM Trans. Intell. Syst. Technol*, 2(3):1–27, January, 2011.

[102] I. Std. IEEE Std 1159-2019. IEEE recommended practice for monitoring electric power quality. *Institute of Electrical and Electronics Engineers*, June, 2019.

[103] F. Bonavolontà, M. D'Arco , G. Ianniello , A. Liccardo , R. Schiano , R. Schiano Lo Moriello, L. Ferrigno , M. Laracca , and G. Miele . On the suitability of compressive sampling for the measurement of electrical power quality. pages 126–131, May, 2013.

[104] A. Baccigalupi, D. Carnì, D. Grimaldi, and A. Liccardo. Characterization of arbitrary waveform generator by low resolution and oversampling signal acquisition. *Measurement*, 45(10):2498–2510, December, 2012.

[105] I. Std. IEEE Std C57.13-2016. IEEE standard requirements for instrument transformers - revision of IEEE Std c57.13-2008. January, 2016.

[106] I. Std. IEEE Std 519-2014. IEEE recommended practice and requirements for harmonic control in electric power systems. 2014.

[107] I. Std. IEEE Std. 368-1977. IEEE recommended practice for measurement of electrical noise and harmonic filter performance of high-voltage direct-current systems. *Institute of Electrical and Electronics Engineers*, 1977.

[108] A. Hooshyar and M. Sanaye-Pasand. Accurate measurement of fault currents contaminated with decaying dc offset and ct saturation. *IEEE Transactions on Power Delivery - IEEE TRANS POWER DELIVERY*, 27(2):773–783, April, 2012.

[109] I. Std. IEEE Std. 1057-1994. IEEE standard for digitizing waveform recorders. *Institute of Electrical and Electronics Engineers*, 1994.

[110] D. Yu, J. Cummins, Z. Wang, H. J. Yoon, and L. Kojovic. Correction of current transformer distorted secondary currents due to saturation using artificial neural networks. *Power Delivery, IEEE Transactions on*, 16:189 – 194, June, 2001.

[111] I. Std. IEEE Std 1547-2008. IEEE standard for interconnecting distributed resources with electric power systems. 2008.

[112] A. Filomena, M. Resener, R. Salim, and A. Bretas. Fault location for underground distribution feeders: An extended impedance-based formulation with capacitive current compensation. *International Journal of Electrical Power Energy Systems*, 31:489–496, October, 2009.

[113] J. L. Blackburn and T. J. Domin. Protective relaying: Principles and applications. 2014.

[114] P. Schneider, B. A. Mather, B. C. Pal, C. W. Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias S. Paudyal, T. E. McDermott, and W. Kersting. Analytic considerations and design basis for the IEEE distribution test feeders. *IEEE Transactions on Power Systems*, PP(99):1–1, 2017.

[115] D. You, L. Ye, X. G. Yin, Q. Yao, K. Wang, and J. Wu. A new fault-location method with high robustness for distribution systems. *Electronics and Electrical Engineering*, 19(6), June, 2013.

[116] R. Gallego R., E. Toro, and A. Escobar Z. *Técnicas heurísticas y metaheurísticas.* Universidad Tecnológica de Pereira. Vicerrectoría de Investigaciones , 2015.

[117] D. Escobar V., R. Gallego R., and A. Escobar Z. Hybrid solution methodology: Heuristic-metaheuristic-implicit enumeration 1-0 for the capacitated vehicle routing problem (cvrp). *International Journal Of Advanced Computer Science And Applications*, 7(3):259–268, 2016.