UNDERGROUND DISTRIBUTION CABLE INCIPIENT FAULT DIAGNOSIS SYSTEM

A Dissertation

by

MIR RASOUL JAAFARI MOUSAVI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2005

Major Subject: Electrical Engineering

UNDERGROUND DISTRIBUTION CABLE INCIPIENT FAULT DIAGNOSIS SYSTEM

A Dissertation

by

MIR RASOUL JAAFARI MOUSAVI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Karen L. Butler-Purry
Committee Members,	B. Don Russell, Jr.
	Andrew K. Chan
	G. Donald Allen
Head of Department,	Chanan Singh

December 2005

Major Subject: Electrical Engineering

ABSTRACT

Underground Distribution Cable Incipient Fault Diagnosis System. (December 2005)
Mir Rasoul Jaafari Mousavi, B.S., Isfahan University of Technology;
M.S., Sharif University of Technology
Chair of Advisory Committee: Dr. Karen L. Butler-Purry

This dissertation presents a methodology for an efficient, non-destructive, and online incipient fault diagnosis system (IFDS) to detect underground cable incipient faults before they become catastrophic. The system provides vital information to help the operator with the decision-making process regarding the condition assessment of the underground cable. It incorporates advanced digital signal processing and pattern recognition methods to classify recorded data into designated classes. Additionally, the IFDS utilizes novel detection methodologies to detect when the cable is near failure.

The classification functionality is achieved through employing an ensemble of rule-based and supervised classifiers. The Support Vector Machines, designed and used as a supervised classifier, was found to perform superior. In addition to the normalized energy features computed from wavelet packet analysis, two new features, namely Horizontal Severity Index, and Vertical Severity Index are defined and used in the classification problem.

The detection functionality of the IFDS is achieved through incorporating a temporal severity measure and a detection method. The novel severity measure is based on the temporal analysis of arrival times of incipient abnormalities, which gives rise to a numeric index called the Global Severity Index (GSI). This index portrays the progressive degradation path of underground cable as catastrophic failure time approaches. The detection approach utilizes the numerical modeling capabilities of SOM as well as statistical change detection techniques. The natural logarithm of the chronologically ordered minimum modeling errors, computed from exposing feature vectors to a trained SOM, is used as the detection index. Three modified change detection algorithms, namely Cumulative Sum, Exponentially Weighted Moving Averages, and Generalized Likelihood Ratio, are introduced and applied to this application. These algorithms determine the change point or near failure time of cable from the instantaneous values of the detection index.

Performance studies using field recorded data were conducted at three warning levels to assess the capability of the IFDS in predicting the faults that actually occurred in the monitored

underground cable. The IFDS presents a high classification rate and satisfactory detection capability at each warning level. Specifically, it demonstrates that at least one detection technique successfully provides an early warning that a fault is imminent.

DEDICATION

To my late parents
Who did everything they could to support me.

AND

To my family

Thank you for your love and support and
for everything you have given me.

ACKNOWLEDGMENTS

First and foremost, I thank God for giving me the strength and confidence to succeed. Thanks to my advisor, Dr. Karen Butler-Purry, for her patience, guidance and support. I thank her for encouraging me at difficult times and always believing in me. Thanks for all the work she provided for my assistantship throughout my Ph.D. study. I have learned from her not only about academics, but also about life in the US. I greatly acknowledge TXU Electric Delivery and Electric Power Research Institute (EPRI) for the support of this work.

I thank Dr. B. Don Russell, Dr. Andrew K. Chan, and Dr. G. Donald Allen for investing their time in serving on my committee. I thank Dr. Ricardo Gutierrez-Osuna for introducing me to the fascinating world of pattern analysis. His continued encouragement and technical assistance are very much appreciated. I would also like to thank the department faculty and staff for making my time at Texas A&M University a great experience.

Thanks to my friends and colleagues at the Power System Automation Laboratory for their friendship. I wish all of them a successful career and joyful life.

Last but not least, countless thanks to my family for their love, encouragement, and patience. I owe the completion of this work and the degree to them.

Thank you all.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1 PREFACE

The power industry has been developing into a challenging and competitive environment due to the ongoing restructuring and deregulation. This structural change has required the electric utilities to reduce operating costs and optimize usage and maintenance of electrical assets without scarifying the quality and reliability of the power delivered to the customers. Underground distribution systems are valuable assets of electric utilities, which supply power to the end customers at low voltages. Many of the system components, particularly underground cables, fail over time in part due to the deterioration of the insulating materials used in their structure. Studies reveal that cable failure rates in power systems continue to worsen as the cable ages [1].

In the past, scheduled inspections and regular maintenances were performed to assess the condition of the system and reduce the failure rates. However, the need for improvement has made it necessary to shift from scheduled maintenance to condition-based maintenance. This shift requires developing new tools and methods to prioritize and perform predictive fault diagnosis and condition assessment of underground distribution systems including power cables. In addition to degrading system reliability, cable failures cost substantial amounts of money for the utilities as replacing or repairing a cable is a very costly process. Early detection of cable faults would undoubtedly be a great benefit to the utilities enabling them to avoid catastrophic failures, unscheduled outages, and thus loss of revenues.

Manufactured with extensive use of insulation material, underground cables constitute the heart of any distribution system [1][2]. Insulation materials used in the structure of underground cable are subject to aging due to the contribution of many stress factors such as heat, moisture, and a number of electrical and mechanical stresses.

This dissertation follows the style of *IEEE Transactions on Power Systems*.

Underground cables are susceptible to deterioration due to the aggregate stress factors that adversely affect the electrical properties of the insulating materials. As such, aging and degradation of the insulating material is the most common phenomenon leading to cable failures. When the insulation undergoes stress factors, a gradual deterioration is initiated in part by voids developed in the insulation. These voids develop into channels in a tree fashion and propagate through the insulation. The gradual deterioration initiates faults called incipient faults that are precursors of a breakdown. Due to persistent aging factors, incipient faults progress and cause the cable insulation to either continue undergoing further degradation or breakdown swiftly. In either case, a catastrophic failure is inevitable. Diagnosis and detection of progressive incipient faults before developing into catastrophic failures avoid unscheduled outages and substantial loss of revenues for the electric utilities. This has motivated this research work with the goal of developing an efficient, online, and nondestructive incipient fault diagnosis system henceforth referred to as the *IFDS*. A destructive method may further degrade the cable insulation that has not already failed and an offline method operates while a section of the cable is de-energized. Thus, a nondestructive and online method is preferred over a diagnosis technique that is active and offline.

Currently, accurate permanent fault detection techniques and relatively accurate fault location methods are available for overhead distribution systems. However, fault detection and location technology for underground distribution systems is still developing. In particular, incipient fault detection technology is a relatively immature technology in need of further developments. Novel techniques need to be developed to look at specific incipient-induced activities over time, extract informative features from the measurements, and perform effective data classification and fault detection. This dissertation presents a methodology for such a detection system which incorporates advanced digital signal processing and pattern recognition methods and tailored change detection algorithms.

This work is a continuation of a research project for online fault diagnosis of underground cables that has been an ongoing project for Power System Automation Laboratory (PSAL) at Texas A&M University, College Station, since 1998. The history of the research work can be divided into two main eras. The focus of the first phase was to develop a study plan for incipient behavior and its characterization in single-phase underground cables [3]. This stage was completed in two parts. The first part of the research involved controlled testing of various aged cables, both field-aged and artificially-induced. These experiments were aptly named controlled

experiments (CE). The second part of the research involved short-term monitoring of various single-phase distribution cables in commercial as well as residential areas. During these experiments the cable was monitored for a short period ranging from a few minutes to a few hours; these were named short-term monitoring experiments (ST). The research study conducted during the first phase was focused on identifying various frequency components in the incipient activities recorded during both the controlled tests and short-term monitoring of online cables. The second phase of the research focused on the development of a long-term monitoring system and preliminary study of incipient behavior in time domain as well as in time-scale domain [4][5]. In this dissertation, a methodology was presented for an efficient, non-destructive, and online incipient fault diagnosis system (IFDS) that is able to detect underground cable incipient faults before they become catastrophic. The system provides vital information to assist the operator with the decision-making process regarding the condition assessment of the underground cable. The classification functionality of the IFDS provides class labels for input data and the detection functionality determines the near failure time at three warning levels.

The major contributions in this dissertation are in five areas. First, a study was completed to characterize underground cable incipient faults through the analysis of various data recorded during the project. This characterization was the basis for the formulation of the problem and development of appropriate methodologies to overcome the specific challenges involved in the diagnosis of incipient faults in distribution systems. One of the major challenges was due to the small magnitude of the variations in the distribution level signals which required new and sensitive signal processing techniques.

Secondly, an efficient data classification methodology was developed using the concept of ensemble learning where a combination of a rule-based and supervised classifiers were used to realize the classification functionality of the IFDS. The supervised classifiers investigated in this work include Support Vector Machines (SVM), Self Organizing Maps (SOM), K Nearest Neighbors (KNN), and Discriminant functions (DIS). Through a number of studies, it was found that the SVM performs superior and as a result, it is integrated in the IFDS to operate as the supervised classifier. In addition to the normalized energy features computed from the wavelet analysis, two new features in time domain were introduced that were applied to the high frequency current signals. These two features were termed Horizontal Severity Index (HSI), and Vertical Severity Index (VSI).

Thirdly, a novel incipient fault detection method based on the temporal analysis of arrival times of incipient abnormalities was developed and called the Global Severity Index (GSI). This qualitative measure provides the severity path of ongoing incipient faults over time and is based on the Laplace Test Statistic. Additionally, a quantitative approach based on the numerical modeling of data with SOM was developed. This method uses the natural logarithm of the minimum modeling errors in chronological order as the detection index.

Fourthly, three modified change detection algorithms were tailored to the incipient fault detection problem to predict the near failure time of cables using the quantitative detection index values. The three algorithms include Cumulative Sum (CUSUM), Exponentially Weighted Moving Averages (EWMA), and Generalized Likelihood Ratio (GLR).

Lastly, the incipient fault diagnosis system was integrated with the previously defined functionalities and tested with the field recorded data from a utility monitoring site. The tests included performance studies to assess the capability of the IFDS in predicting the incipient faults that actually occurred in underground cable during the monitoring period of three years.

1.2 ORGANIZATION

This dissertation consists of seven chapters. Chapter I provides introduction and organization of the dissertation. Chapter II is partially devoted to the literature review of the underground cable failure mechanisms, fault detection techniques, and commercially available products. Furthermore, the problem of online incipient fault classification and detection is formulated in this chapter. Overview of the field recorded data used for the research is provided in Chapter III. Also, preprocessing tasks including resampling, DC removal, and denoising operations are discussed in Chapter III. Incipient fault classification methods involving rule-based and supervised classifiers are presented in Chapter IV and the performance analysis results are discussed. Chapter V is devoted to the discussion of incipient fault detection method. In Chapter VI, an overview of the IFDS is given and the detection results are presented using field data. Finally, Chapter VII will present conclusions and future work.

CHAPTER II

LITERATURE REVIEW AND PROBLEM FORMULATION

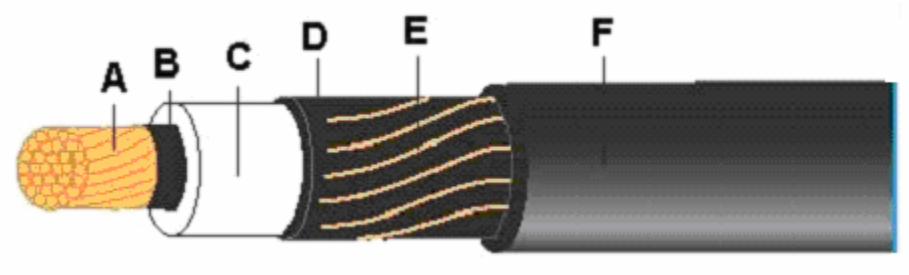
2.1 INTRODUCTION

Study of cable failures and development of accurate fault detection and location methods have been interesting yet challenging research topics in the past and present. Fault detection entails determination of the presence of a fault, while fault location includes the determination of the physical location of the fault. Accurate permanent fault detection techniques and relatively accurate fault location methods have been developed for overhead distribution systems. However, fault detection and location technology for underground distribution systems is still in developing stages. In particular, incipient faults which constitute a subcategory of cable faults have been the focus of recent investigations. These faults encompass the insulation aging process from inception to completion before leading to a catastrophic failure. From a macroscopic perspective, incipient faults refer to the abnormalities associated with any type of deterioration phenomena manifested in the cable electrical signals. The detection of these faults which occur due to the presence of a failing component is a relatively immature technology in need of further developments. Historically, the development of incipient fault detection and location systems as an ongoing area of research dates back to the early 1990s. This chapter reviews and evaluates the existing incipient fault detection and location techniques for underground cables and paves the road to formulate the incipient fault detection problem as applied in this research work. A review of cable structure and the failure mechanisms are prerequisites for the subsequent discussions that follow.

2.2 ANATOMY OF UNDERGROUND DISTRIBUTION CABLES

The core component of any underground system is the cable that supplies power from the source to the load. The longevity and reliability along with desired safety and aesthetics issues of underground cables have made underground distribution systems an unprecedented substitute for overhead distribution lines. Underground cables have been designed for various applications and voltage levels and extensive improvements in design process have been achieved. Today

pressurized cables are available up to 765 KV and even 1100 KV through the gradual advancements in materials and manufacturing processes [7]. For primary distribution systems, cables are typically designed with the following major components, conductor, conductor shield, insulation, insulation shield, concentric neutral, and jacket. These components are illustrated in Fig. 2.1.



A: Conductor B: Conductor Shield C: Insulation
D: Insulation Shield E: Concentric Neutral F: Jacket

Fig. 2.1: Anatomy of a typical single phase underground cable

The conductor can be either aluminum or copper in solid or stranded form. The selection of a conductor type depends on ampacity, voltage, physical properties, flexibility, shape, and economics [2], however it is recommended to use solid or stranded-filled conductors for reliability [1]. Conductor shields and insulation shields synergistically provide a uniform cylindrical surface next to the cable insulation to establish the most uniform possible distribution of electrical stress. Research performed on cable failures has shown that existence and development of voids or protrusions near the conductor shield-insulation interface played an important role in the failure process [1]. This region experiences extremely high electrical stresses and these irregularities help boost a non-uniform electrical field, stressing the cable insulation and eventually causing it to fail. The extruded conductor shield is a layer of semi-conducting material, used to prevent excessive electrical stress in voids between the conductor

and the insulation [8]. Insulation can be of a variety of materials such as EPR[1], XLPE[2], paper, and TRPE[3] compounds, whose thickness is a function of cable voltage rating such that the higher the voltage rating, the thicker the insulation. The extruded insulation shield also consists of a semi-conductive layer similar to the conductor shield. The function of the insulation shield is to confine the electric field within the cable, symmetrically distribute electrical stress, reduce the hazard of shock, limit radio interference, and protect cable induced potential when connected to overhead lines [8]. The shield may be a metallic tape or a non-metallic tape, drain wires, or concentric neutral wires. The outer shield is normally connected to ground.

Concentric neutral conductors serve as the metallic component of the insulation shield and as a conductor for the neutral return current [2]. Due to some mechanical and electrical considerations, concentric neutral conductors are built from copper even if the central cable conductor is aluminum.

The cable jacket is the outermost layer of the cable. The purpose of the jacket is to provide mechanical, thermal, chemical, and environmental protection. It can be made of polyethylene, polyvinyl chloride, nylon, as well as other plastics. Certain cables use a sheath or armor instead of a jacket, which provide a much better protection to the cable than a jacket [2].

The first widely accepted concentric neutral cables were unjacketed. The bare concentric neutral (BCN) cables were directly buried exposing the concentric neutral conductors to the surrounding soil and consequently provided very effective ground. This design was desired from a personnel safety point of view in case of a dig-in. Due to the presence of a low resistance path through neutral conductors, adequate fault current could be conducted to operate protective devices. The low resistance between the neutral and earth would also reduce the touch potential at the dig-in site, significantly [1].

Despite the numerous advantages of BCN cables, major durability problems hindered their wide installment in underground systems. Soon engineers found that cable moisture and/or concentric neutral corrosion played a major role in increasing the failure rate of unjacketed underground cables. Due to the lack of a protective jacket, BNC cables were subject to corrosion. Once corroded, the only neutral current path was through ground rods which were a totally unsatisfactory condition from the safety and reliability stand point. Therefore, jacketed concentric cables (JCN) achieved wide acceptance with a special attention to system grounding.

[1] Ethylene Propylene Rubber
[2] Cross-linked Polyethylene
[3] Tree-Retardant Polyethylene

It is worth mentioning that while U.S utilities installed BCN cables, European and Japanese utilities installed only jacketed cables and as a result these utilities have experienced much higher reliability than in the United States. Today, the U.S utilities mainly use jacketed cables [1].

2.3 AGING MECHANISMS IN UNDERGROUND CABLES

Deterioration of insulation is an inevitable phenomenon in underground cables leading to insulation failures. The aging is caused by single or synergistic action of several aging factors that include thermal, electrical, mechanical and environmental [9][10]. Persisting aging factors eventually fail the cable insulation through a number of mechanisms summarized in Table 2-1. Activation of aging mechanisms either change the bulk properties of the insulating materials referred to as intrinsic aging or cause degradation known as extrinsic aging. The degradation is the result of the presence of contaminants, defects, voids, and protrusions in the insulation material and their interaction with different aging mechanisms [10][11].

Under normal conditions, electrical stresses are the predominant aging factors that may fail cables through partial discharge and treeing mechanisms aggravated by the presence of water. In organic extruded dielectric and in particular in cross-linked polyethylene (XLPE) cables, the majority of cable failures are related to the treeing activity. Treeing refers to any kind of damages in the insulation medium in which the deterioration path resembles the form of a tree. This pre-breakdown phenomenon takes place in the form of either electrical trees or water tress under DC, AC and impulse voltages [12]. The primary cause of treeing in dry dielectrics is partial discharges under high electric stresses and moisture at lower electric stresses. On the other hand, not all degradation phenomena are associated with electrical stresses. Cable might fail –under abnormal conditions- through thermally aged insulation breakdown [13]. Moisture increases dielectric losses so localized heat generation is produced and thermally degrades the paper insulation. The following sections will briefly discuss electrical aging mechanisms i.e. partial discharges, electrical trees, and water trees as the most commonly sought failure phenomena.

Table 2-1:
Summary of aging mechanisms in cables [10]

Aging Factor		Aging Mechanisms
Thermal	High temperature Temperature cycling	Chemical reaction Thermal expansion Diffusion Insulation melting Anneal locked-in mechanical stresses
	Low temperature	Cracking Thermal contraction
Electrical	Voltage, AC, DC, Impulse	Partial discharges Electrical trees Water trees Charge injection Intrinsic breakdown Dielectric losses and capacitance
	Current	Overheating
Mechanical	Cyclic bending, vibration, fatigue, tensile, compressive, shear stresses	Yielding of materials Cracking Rupture
Environmental	Water, humidity Contamination Liquids, gases	Electrical tracking Water treeing Corrosion Dielectric losses and capacitance
	Radiation	Chemical reaction rate increase

2.3.1 Partial Discharges

A partial discharge (PD) is a localized gas discharge in a gas-filled void or on a dielectric surface of a solid or liquid insulation system without bridging the system electrodes [14]. PD can result from the discharge in cavities developed inside the insulation, voids between the semiconductor and dielectric, tracking discharge along an interface, or discharge from electrical or water tree growth[4] [15]. When the electric field intensity within a cavity or a crack reaches a threshold value, the gas contained in the defect ionizes, producing free electrons, which by multiple collisions initiate an avalanche. If the size of the void in the direction of the electric field is large enough, the avalanche may eventually initiate a breakdown, or discharge, across the void [16].

[4] As the discharge does not bridge the entire insulation, it is called *partial discharge*.

To initiate a PD, the cavity size must reach the critical limit for development of a discharge. For XLPE insulation, this critical size is 0.03mm for a spherical cavity filled with air at atmospheric pressure [17]. PD inception voltage is a function of cavity size, cavity location, cavity shape, insulation thickness and type, and conductor size. The larger the cavity size, the lower the inception voltage. PDs develop into electrical trees when self-sustaining PDs occur at the system operating voltage. Prolong PD activity deteriorates the wall of the cavities physically and chemically [17] that may in turn lead to the initiation of treeing. In [14], Paoletti and Golubev discuss the detailed theory of partial discharge and how it manifests itself in cables.

2.3.2 Electrical Trees

The presence of high and divergent electric stresses is the primary contributing factor to initiate and propagate electrical trees (ET) [12]. An ET may consist of many discharge paths including tree trunk and branches originating from the tree trunk. The tree structure in microscopic dimensions is highly visible in solid dielectrics and may appear in different forms. Fig. 2.2 shows typical patterns of electrical trees in polyethylene (PE).

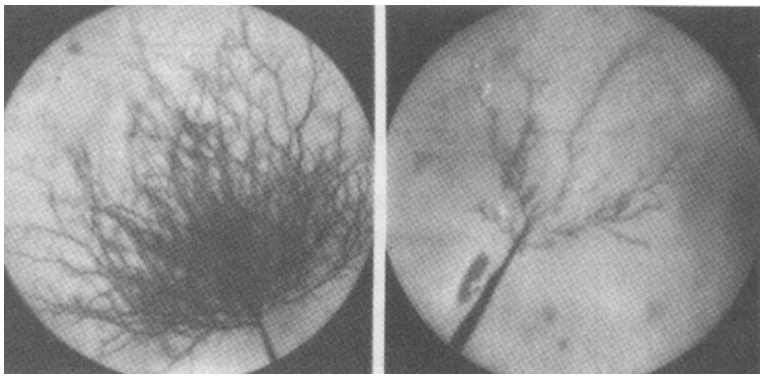


Fig. 2.2: Typical patterns of electrical trees [12]

ETs can also be initiated due to prolonged PD activities, protrusions at the semicon-dielectric interface, contamination or from a water tree conversion. The required electric field to initiate ET is 150KV/mm [17]. Once initiated, ET will normally propagate through the insulation

as a series of sporadic bursts, and when the branches of the tree bridge the length of the dielectric, a breakdown happens. Thus, ET concludes the degradation process of the insulation by leading to its failure.

2.3.3 Water Trees

Water trees are caused in the presence of moisture by the ionic contaminants especially at the semiconductor-insulation interface of cable. Unlike electrical trees, water trees typically commence at lower electric stress values and propagate more slowly through the insulation [12]. Water trees do not cause detectable partial discharges before converting to the electrical trees. Under normal operating conditions, the conversion process is caused by prolonged PD activity in cavities that are created in the water tree channel [13]. Large water trees can convert at normal operating voltages and small water trees convert due to lightning impulses. In some instances, external chemical may get trapped in the cable components and this may initiate localized PD called electrochemical treeing. Water trees fail the cable when they convert to electrical trees. Once a water tree is converted to an electrical tree, the time to failure is normally short because the initiated electrical tree propagates rapidly through the aged insulation [13]. Fig. 2.3 shows a pictorial presentation of water trees initiated in an insulating material.

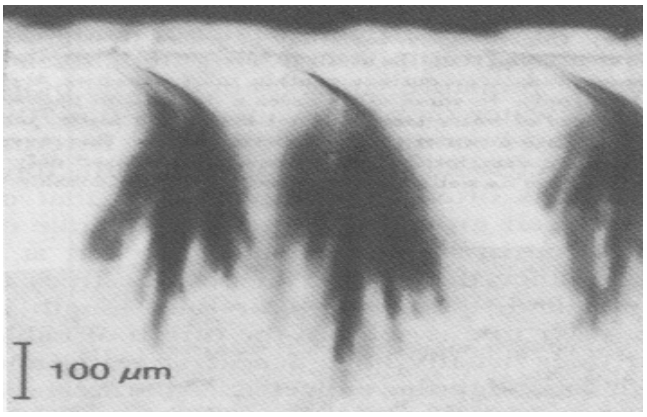


Fig. 2.3: Water tree in insulation [12]

2.4 UNDERGROUND CABLE FAULT DETECTION AND LOCATION METHODS

To date, various research studies have been conducted to develop methods for fault identification and detection in underground systems [14]-[45] and some commercial detection systems are also available for diagnostic testing [46]-[49]. The present methods, although conceptually different, can be categorized in terms of the mutually exclusive *active/passive* and *offline/online* terms. The term active describes detection schemes that require an external electric source to energize the system and generate the diagnosis signals. The opposite holds true for passive methods in which there is no external injection to the cable system. Active methods are often destructive which implies that they may further degrade cable insulation that has not already failed. Thus, the portion of the system involved in the fault must be replaced before restoring power. Offline methods consist of detection techniques that operate while a section of the cable is de-energized. Offline methods can require local outages and potential system contingencies. If a method is not an offline method, it is considered an online method in which there is no service interruption during the application of the method. Passive and online methods are preferred over active and offline diagnosis techniques because they are not destructive and can be applied without service interruption.

Existing methods target two main categories of insulation categories. While some of the methods are used to provide an overall assessment of the insulation, there are other methods that perform an incremental condition assessment of the underground cable. Diagnosis methods involving dissipation factor, harmonic analysis, DC leakage current, and return voltage measurements are some examples belonging to the first group and various methods based on partial discharge measurements constitute examples of the latter group of existing methods [18]. From a field application point of view, the existing methods can be categorized into the following classes [34]: i) Methods based on the partial discharges, ii) Methods based on the time and frequency domain reflectometry, iii) Methods utilizing the dielectric ohmic and polarization losses, and iv) Methods based on the acoustic and pressure wave techniques. The following sections discuss these methods with particular attention to the advantages and disadvantages of each method.

2.4.1 Techniques Based on Detection of Partial Discharges

In a cable system, various factors can contribute to the partial discharge, such as voids, shield protrusions, contaminants, advanced stages of water trees, electrical tree growth, etc. Partial discharge is a precursor to premature degradation of dielectric materials in a power cable [33] that reveals itself in a number of ways. Charge displacement in the void, high frequency radiation emitted by excited particles, ultrasonic sound, heat from particle impact and chemical reactions are some of the phenomena that can accompany a partial discharge [15]. The produced high frequency electromagnetic signals travel along the cable and cover a broad frequency range. In solid insulation, this frequency range typically varies from a few hundred kHz to a few hundred MHz [35] depending on the location of the PD with respect to the testing point[5]. The partial discharge detection method is based on the measurement of these pulses by high frequency inductive or capacitive sensors. The general PD testing method proceeds in three steps as shown schematically in Fig. 2.4.

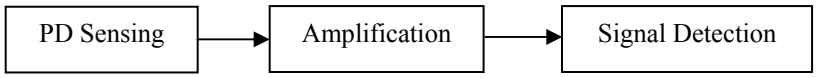


Fig. 2.4: PD testing process

In the first step, PD signals are measured by special sensors such as capacitive couplers, inductive couplers, or antennas. Since the original PD pulses do not possess sufficient magnitude, an amplification procedure is usually followed to enhance the resolution of the measured pulses. Any treatment for possible noise interference also takes place in this step. Finally, the amplified signals are analyzed in time or frequency domain to detect the PD locations.

PD measurements can be performed either offline or online [15]. Offline methods require that the cable be disconnected from the system and elevated voltages be applied to generate the diagnosis signals. De-energized cable system during the offline testing conveys no PD signal to

[5] From a Fourier analysis stand point, a pulse of ‘zero’ width should have a very wide spectrum.

be picked up by the sensors; thus, to activate the PD sites again during testing, an elevated voltage is required to re-initiate the PD. Applying this voltage does not necessarily initiate a discharge because in addition to sufficient excitation voltage, a free electron is required to initiate a discharge. Thus, a minimum test time in minutes is needed for each stress level to become effective.

The utilized sensor in offline methods involves a high frequency capacitive coupling connected at one end of the cable parallel to the conductor. These capacitors act as a filter, block the 60 HZ component and allow the very high frequency pulses associated with PD to be measured. As an important point, these capacitors must be free of PD since they are directly connected to the high voltage side and undergo the same test voltage [15]. The discharge signals are measured across external impedance which is in series with the capacitor. A resonant circuit is sometimes used to amplify the discharge pulses in time domain for a better detection capability. The PD detection circuit used in most detectors is shows in Fig. 2.5 . Due to the need for a return path through the cable shield, this method can not be applied to unshielded cables.

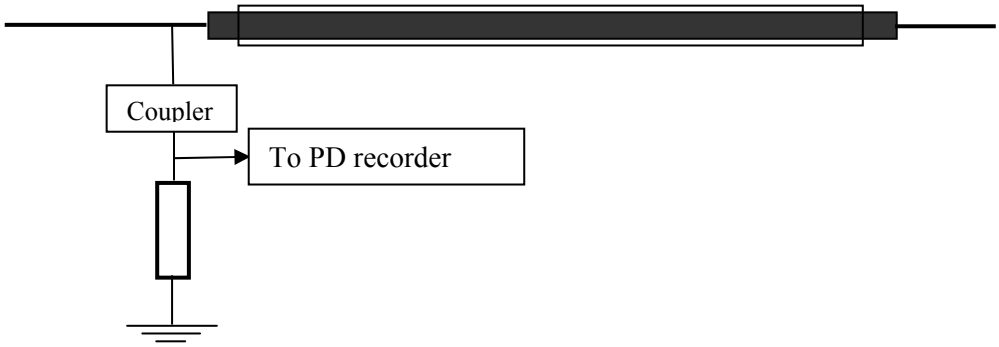


Fig. 2.5: PD detection setup

Online measurements are conducted without power interruption and while the cable is in service. There is no need for the heavy and expensive test voltage supply and the coupling capacitors are often replaced with inductive couplers, current transducers, placed around cable. For these reasons and some others discussed later, online PD measurement has gained much

favor over the conventional offline methods. The following is a discussion of various available methods that make use of PD measurements.

2.4.1.1 Partial Discharge Detection and Location Using Correlation Technique

This method makes use of PD measurements from both ends of the cable to detect and locate PD sites [41]. The system setup is shown in Fig. 2.6. The time lag τ between the two PD signals is determined by maximizing the cross-correlation function. This function $R_{12}(\tau)$ is given in [34].

$$R_{12}(\tau) = \lim_{t\to\infty}\left[\frac{1}{t}\int_0^t v_1(t)v_2(t+\tau)dt\right] \qquad (2\text{-}1)$$

where, $v_1(t)$ and $v_2(t)$ are the instantaneous values of the PD signals measured at the two ends of the cable at time t.

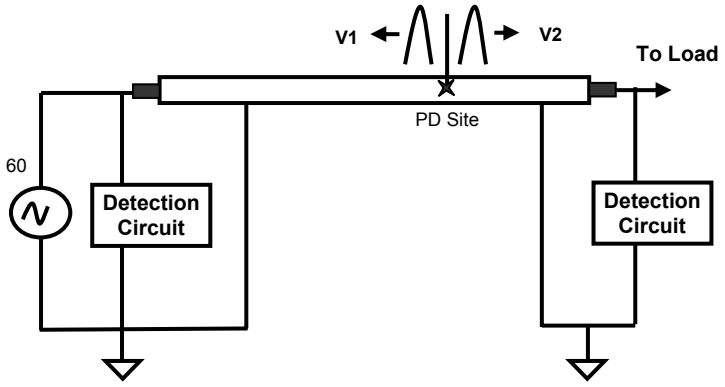


Fig. 2.6: Circuit setup for PD detection and location using correlation technique [34]

This cross correlation function allows the determination of the time lag τ in a noisy environment since the PD signal is unlikely to correlate with noise signals from other independent sources, however the traveling signals suffers from distortions as the result of the attenuation and dispersion when they travel towards the two ends of the cable. Thus, the PD signals detected at both the ends of the cable may be quite dissimilar. For this reason, the

correlation function is used in conjunction with threshold logic. The advantage of this technique is that it can be performed online and there is no need for any high voltage test sources. Nevertheless, it requires that PD signals be measured at each end of the cable which may not be possible in the field [34].

2.4.1.2 Partial Discharge Detection and Location Using Reflectometry Technique

The reflectometry-based technique uses the arrival time of the original PD signal and its reflection from the open end of the cable to detect the PD site while the service is interrupted (offline). Fig. 2.7 shows the system setup [34][42][43]. The location of PD is determined by (2-2).

$$\frac{X}{L}=\frac{2t_X}{2t_L} \tag{2-2}$$

where, $2t_x$ is the time lag between the arrival of the original PD pulse traveling directly from the PD site, and that of the other PD pulse reflected from the open end of the cable [34]. The time lag $2t_L$ is the time lag measured between the arrival of the direct PD pulse and its reflection from the open end as illustrated in Fig. 2.8.

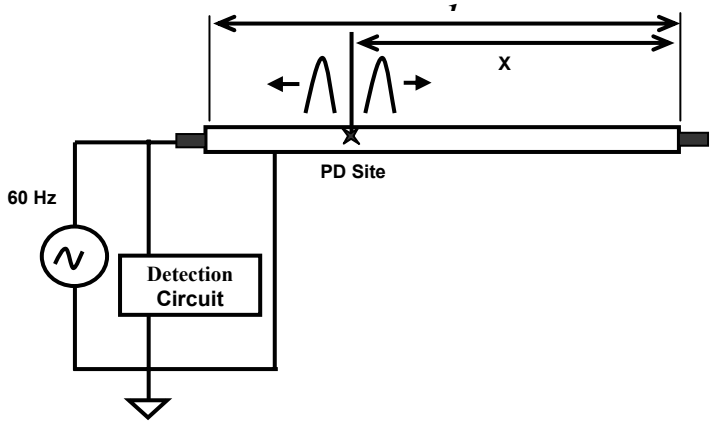


Fig. 2.7: PD detection using reflectometry techniques [34]

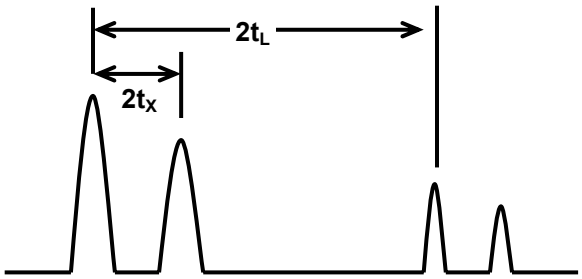


Fig. 2.8: PD signal detection time [34]

As it is implied, this method requires only one PD detection circuit at one end of the cable and the other end of the cable is left open during the measurements. Although various methods are available for PD extraction and reconstruction, this method has a major disadvantage in its application in the field, as the service in the cable must be interrupted to conduct the test.

2.4.1.3 Partial Discharge Detection and Location Using Modulated X-Rays

In this method, which is not widely known, the detection and location of PD are accomplished using modulated X-rays directed through established PD sites in a dielectric medium [34][44]. These X-rays reduce the discharge level, but increase the number of discharge pulses per cycle. Hence, if a cable with PD is scanned with chopped X-rays, partial discharges can be expected in X-ray chopping frequency. By tuning the detection circuit at the X-ray modulation frequency, the signal-to-noise ratio is improved and as a result a higher sensitivity is achieved. Although this method can potentially be used in the field, a number of safety issues concerning the radiations applied in the field hinder widespread and safe utilization of this technique, and thus it seems to be impractical.

2.4.1.4 Challenges Involved in PD-based Techniques

PD signals are normally subject to fast transients and non-periodic components which only appear in the high frequency spectrum. During PD activity, the location of discharge remains the same but the magnitude and number of the pulses can vary considerably with time, voltage, temperature, load and humidity, making the detection a difficult task [18]. Moreover, attenuation

along the cable length, particularly at the higher frequencies and background noise in the field can compound the methodology.

From application viewpoint, PD detection is often hampered by the huge volume of data due to the high sampling frequency [24] and the problem of noise interference worsens the data usage. On the other hand, not all the degradation phenomena are associated with PD. Cable might fail through thermally aged insulation breakdown [13]. Moisture increases dielectric losses so localized heat generation is produced and thermally degrades the paper insulation. Under such circumstances, PD may only be initiated at advance stages of such degradation. Existing online detection methods cover some of the shortcomings associated with the offline techniques, but in contrast to offline PD diagnosis, the sensitivity of online PD detection and the results interpretation are strongly influenced by the local external disturbances and operation conditions of the cable section [25] and expertise of the data interpreter. Additionally, PD measurement might not indicate the time to failure. If electrical trees develop first, the time to failure is normally short because the initiated electrical tree propagates swiftly through the aged insulation [13].

In order for a PD to progress and cause a failure, self-sustained discharges must be present at the system operating voltage i.e. the discharge extinction voltage should not exceed the line voltage [13]. Presence of self-sustained discharges is in turn a function of defect type. If the PD inception and extinction voltage are both less than the system voltage, there is a good chance of self-sustained discharges. Another type of defects might have an inception voltage greater than the system voltage but their extinction voltage is less than the line voltage. For these defects, discharges can initiate due to an abnormal system condition or disturbances. Once initiated, these discharges are self-sustained and visible to detection methods. Lastly, some defects have the inception and extinction voltage greater than the line voltage. Partial discharges due to these defects can only be ignited by transient over-voltages and once the over stress condition is removed, the partial discharge disappears, consequently.

The type of PD that exists in the cable is another limiting factor. The detectability of partial discharges in cavities highly depends upon the size of the cavity. Roughly speaking, as the size of the cavity increases, the magnitude of the PD increases and as a result, the detection becomes easier. Both offline and online methods can detect these discharges but there is a minimum size for a detectable cavity. This limit also depends on the cable length. Using online methods, the cavity size can be as small as 0.1 mm if the cable length being tested is less than 200 meters [15].

Any cavities smaller than 0.5mm for shorter cables and 0.7 mm for longer cables are not visible to offline methods whereas they might fail the cable and thus there might not be enough time for early fault detection [15]. Partial discharges forming electrical trees are detectable and can be picked up by both online and offline methods. Recall that electrical trees constitute the final phase of the degradation process. As the tree length increases, the PD magnitude increases and makes it easier to be sensed by the couplers. If the PD possesses sufficient magnitude, the sensitivity of the detection method need not be high. Tracking discharges due to cable splices are also visible to both methods. PD magnitude increases as the track length extends. Finally, partial discharges that occur between the cable neutral and semiconductor are often severe enough to be detected by both the offline and online methods [15].

2.4.2 Techniques Based on Time-Domain Reflectometry

The premise utilized in Time-Domain Reflectometry (TDR) techniques is that the degradation of the insulation translates into the variations of the cable surge impedance defined by the per-unit inductance and capacitance of the cable. Because of this impedance variation at the deteriorated spots of the cable, a square wave signal injected to the cable will be partially reflected from the location of degradation. This reflected signal and the signal reflected at the open end of the cable are used to calculate the time lag and in turn, to determine the location of the incipient fault [45][34]. A similar reflectometry approach assumes that an incipient fault site behaves as a region of dielectric nonlinearity. Therefore, when two signals of different frequencies are fed into the cable under test, new signals having frequencies equal to the sum and the difference of the original signals are generated at the fault site. The propagated and reflected (from the open end of the cable) signals of these new signals can be used to determine the location of the fault site.

The drawbacks associated with method stems from the inherent characteristic of the power cable to attenuate high frequency components. The high frequency components are the only signals reflected from the incipient fault site, but they are highly attenuated and distorted while traveling in the cable. Thus, conventional TDR techniques will not be effective in detection of incipient faults in long cables. Furthermore, TDR techniques require heavy and expensive high voltage test equipment and can be applied when the cable is not in service i.e. they are offline methods [45].

2.4.3 Techniques Based on Measurement of Dielectric Losses

Dielectric losses including ohmic and non-ohmic losses are important measures of insulation quality. For a perfect insulation these losses are negligible; but as the cable ages and trees extend throughout the insulation medium, the associated dielectric losses increase over time. The increase in ohmic losses gives rise to the increase in the DC component of the ac charging current, whereas the increase in non-ohmic (polarization) losses results in the increase in the dielectric dissipation factor i.e. $tan\delta$. This is the premise used in the methods that are based on dielectric losses.

These methods measure the DC component of the ac charging current and the dielectric dissipation factor to assess the degree of deterioration in the cable insulating material [36][37]. It is been argued that measurement of the DC component in the ac charging current combined with the $tan\delta$ readings, provide an accurate way to detect incipient fault sites in the cable.

Unlike most of the other techniques discussed so far, this technique is able to indicate the existence of water trees while the cable is in service and doesn't require heavy expensive test equipments. Additionally, it can be used for long cables. Nonetheless, it possesses a number of drawbacks. The detection of incipient faults using this technique becomes possible only after a large number of trees have developed but even then, the DC current involved are in nano-amperes and thus it prove to noise interference [34]. Further, this method can not provide a clue as to the location of the fault.

2.4.4 Techniques Based on Acoustic and Pressure Wave

As mentioned earlier, PD is not a silent phenomenon and usually accompanied by charge displacement, radiation, ultrasonic sound, heat, and chemical reactions. Acoustic methods detect the sound energy emanating from the PD sites. Also, the acousto-optical techniques can be used for detection purposes in which the light intensity in a fiber optic waveguide is modulated by the sound waves generated by the partial discharges. As implied, the main power of these techniques lies in the detectability of the acoustic waves and thus the water trees are not detectable unless the degeneration is bad enough to emanate sound. Furthermore, the need for implementing a large number of waveguides makes this method very expensive for long lengths of cable.

The pressure wave techniques are based on the detection of dielectric space charges at an incipient fault site. For this technique, a short rise-time pressure wave is fed into the cable using

any one of the several techniques [38][39]. The pressure wave travels along the dielectric with the speed of sound and encounters dielectric space charges at the incipient fault sites. As the result of this encounter, the space charge is displaced and the dielectric permittivity is altered because of the variation of the local concentration of charges thus generating a potential between the cable conductors [39]. Monitoring this transient potential field along the cable provides the means to approximate the space charge distributions along the cable and thereby locate the incipient fault sites. This technique although seems potentially promising and appealing to researchers, has not been tested in the field and needs further investigations. It requires pressure wave generators and needs monitoring the transient potential filed all along the cable at short intervals.

2.5 COMMERCIALLY AVAILABLE PRODUCTS

Traditionally, the HIPOT[6] test at elevated DC voltage, very low frequency, and power frequency has been applied to power cables to diagnose the damaged regions. This is a destructive test that may identify incipient failures, but may not completely reveal whether a cable will function properly during high ambient temperatures or under high load currents. In recent years, a number of non-destructive commercial products have been developed to evaluate and assess the insulation condition of the cable system in both offline and online fashion. This section will provide a review of the techniques utilized in these products. A general assessment of these techniques is also provided.

2.5.1 DTE Energy Technologies

DTE Energy has developed a passive condition assessment technique called Cable|Wise that makes use of the RF emissions from cables, splices, and terminations to provide an assessment of the remaining life of the cable and its accessories [46]. The emissions are recorded for 15 minutes while the cable system is in service. The recorded data is analyzed by an expert to identify the equipment that emits the RF signal and estimate the probability of failure or remaining life of each component. Fig. 2.9 shows the typical experimental setup for Cable|Wise data acquisition. The principles of the PD measurements were discussed in section 2.4.1 and the discussion of PD challenges in section 2.4.1.4 applies to the approach developed by DTE.

[6] High Potential Withstand (HIPOT)

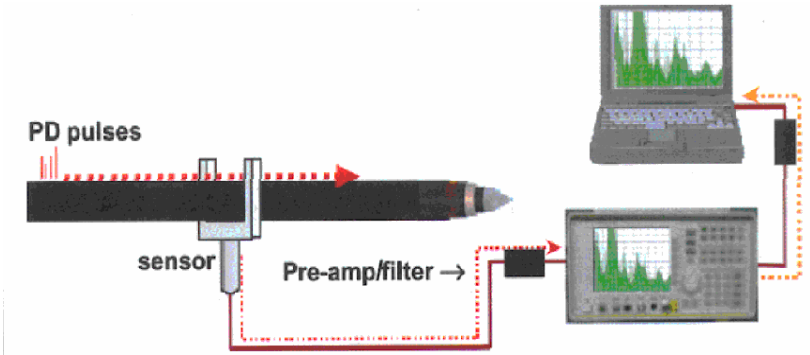


Fig. 2.9: Experimental setup used by DTE energy technologies for data acquisition [46]

This technique depends on detecting low energy, high frequency (300 MHz) emissions while the cable is in service under normal operating conditions. The detection of these emissions is not a simple task as they are well within the energy levels of background noise and other electromagnetic signals. Cable|Wise utilizes a combination of custom built high frequency sensors, amplifiers, and noise rejection methods to accomplish the detection task [46]. A report given to the user provides the probability of failure of each cable section, cable accessory, transformer, and switchgear connected to the circuit. Refs. [13][15][17][20][21][22][35] provide some detail information about the solution methodologies, however the interpretation of the results is conducted by an expert in the lab and there is no automatic, on-site decision making system. This product does not take the past behavior of the cable into account and is not a condition monitoring system.

2.5.2 IMCORP

Imcorp, a company developed by the work of Mashikian and his colleagues [16][31][32][33], operates under an exclusive license from the University of Connecticut. It provides an array of cable diagnostic services for medium and high voltage shielded power cables. The latest technology developed in Imcorp allows tests to be carried out on the cable systems to locate defects and assess the overall condition of the system. The testing methodology is based on the pulse reflectometry and time-of-arrival of partial discharges, discussed in 2.4.1.2,

that proceeds in four steps, namely topology identification, sensitivity calibration, HV diagnostic test, and data analysis [47].

The first stage involves a low-voltage TDR mapping to locate cable joints and neutral discontinuities, and to determine the cable length. This test uses a low-voltage pulse generator to inject pulses of magnitude between four to eight volts and width of 20 to 650 nanoseconds. The original pulse and its reflected version is recorded and digitally processed to locate joints and large cable anomalies. The sensitivity calibration is performed to calibrate the diagnostic system and to determine the smallest pulse, capable of making a round trip on the cable. In the third stage, a continuous 60 Hz voltage signal is applied to the cable for a few seconds and the relevant PD data is measured and recorded. After removing noise components, the processed data is further analyzed to prepare a summarized report on the location of joints, and defects and the type of defects throughout the cable length [47]. Although, this method uses a 60 Hz signal, it can be destructive. Furthermore, it is not operated as an online condition monitoring system that is able to detect an imminent fault.

2.5.3 PowerTech

PowerTech offers a number of services for condition assessment, laboratory evaluation and failure analysis of cables that may be performed in the laboratory, on cable samples, or on-site, using one or more electrical tests, but none of these include an online monitoring system. Lab assessment is made to determine whether cables of similar type and age should remain in service and it may include dissection, water tree counts, and various small-sample chemical tests. The cable is checked for voids, protrusions, contaminants, water trees or other damage. Such off-site and offline laboratory tests can provide the knowledge of cable condition and cause of failure, but only tell the story on the sample examined and thus, on-site condition assessment is needed to evaluate the insulation of the cables still in the ground. On-site condition assessment is carried out on de-energized cable systems using a number of tests, including partial discharge, dissipation factor, voltage recovery and leakage current measurement. However, no single test provides an accurate account of insulation condition [48].

The technology used for XLPE cables at PowerTech is a low voltage DC test method called Leakage current (I) pico Ampere test (LIpATEST). During this offline test, a negative, step-DC voltage is applied to a cable and the resulting leakage current is measured with a sensitive, highly filtered, DC pico-ammeter for one minute at each voltage step. The voltage level used in

this test is less than half that of the recommended levels for aged cables using traditional DC hipot testing. Cables with large and extensive water trees show significant departures in the linearity of leakage current vs. test voltage as shown in Fig. 2.10 [48].

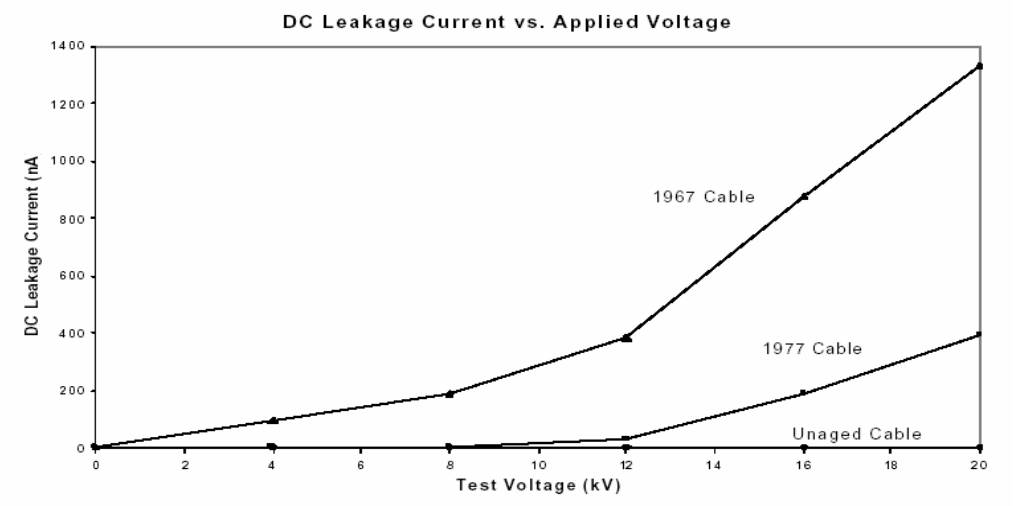


Fig. 2.10: Variation of DC leakage current versus applied voltage in the LIpATEST

2.5.4 EA Technology

EA Technology –a company based in UK- has developed condition assessment instruments that are handheld portable detectors [49]. They can locate underground cable faults or partial discharge activity from within high voltage insulation material. PDM03 is a partial discharge monitoring device that is non-intrusive and designed to monitor discharges over a period of time and indicate variations caused by changes in the operating environment, such as voltage surges, temperature and humidity. The data gathered is stored onto disk, which can then be analyzed to indicate levels of discharge and severity over a period of up to two months. This product, as it appears, is not an online condition monitoring system and tests need to be repeated at least every two months.

PDL1 is another device that locates partial discharge activity from within high voltage insulation material by measuring transient earth voltage signals (TEV) generated by small discharges. The magnitude of the TEV signal is measured using the instrument in single probe mode. Using the instrument in two-probe mode indicates the source of the discharge in relation to the probes [49].

2.6 OVERALL ASSESSMENT OF ONLINE AND OFFLINE CABLE FAULT DETECTION METHODS

In this section, an overall assessment of underground cable fault detection methods including both online and offline techniques is presented. As shown, there are a number of different techniques for failure detection in underground cables; nevertheless, partial discharge measurement constitutes the core component of most commonly used methods. As a result, this assessment often targets techniques based on partial discharges. Whether performed online or offline, the underground cable fault diagnosis methods are subject to a number of drawbacks and technical boundaries due to the following factors.

The high frequency attenuation in solid cables is one of the main constraints in the application of RF emitted signals. The frequency-dependant attenuation characteristic of the cable is such that the attenuation factor is directly proportional to the signal's squared frequency i.e. the higher the frequency, the greater the degree of attenuation. Additionally, the presence of sheath and its conductivity increases the attenuation [21]. As the result, the high frequency pulses attenuate and shift to lower frequencies as they travel along the cable [22]. This imposes a limit on the PD pulse measurements as to the length of the cable that can be tested. Generally, the sensitivity of the measurement decreases significantly as the length of the cable increases. Online methods can have an advantage over offline techniques in this respect since online measurements can be done at several short distance intervals provided that the cable is physically accessible.

Some detection techniques involve measurements of high frequency pulses that possess small energy. Noise components also convey the same characteristics and as a result noise interference is an inevitable problem in such measurements. Noise contamination reduces the sensitivity of the detection method especially in a very noisy environment such as highly populated areas. Testing the cables at shorter intervals might alleviate noise interference; however it may not always be possible to do so. Another approach is to measure the signals

within a frequency range where little interference is expected. The downside is that pulses such as partial discharge signals may contain frequencies that overlap the noise frequencies. If so, these frequencies will remain undetected unless other measurements are made possible at different locations. After all, noise interference is one of the main challenges in field measurements that needs to be overcome in a subtle way.

When offline testing is conducted on a feeder with branched circuits, multiple reflections from these circuits make it very difficult to distinguish relevant reflections and therefore the detection becomes unfeasible. This is also a challenge encountered in traditional reflectometry methods. Physical disconnection of multiple circuits can potentially alleviate the detection complexity if it is economically justified. Unlike offline methods, online detection methods can overcome this problem by testing each branch separately and independently from each other [15].

It is possible that multiple discharges can occur simultaneously at different cavities distributed along the cable. Similar to the multiple circuits case, in the offline testing, multiple reflections superimpose and appear as an average value [15]. As a result, the location accuracy is decreased or lost. For a cable with a few discharge locations, the TDR method provides satisfactory results; however as the number of discharge sites increases, the level of accuracy in TDR approach is dramatically lowered.

Cable characteristics such as attenuation and propagation speed differ from one type of cable to another type. Thus, methods based on TDR seem to be difficult to generalize easily as they rely on varying cable characteristics. For example, the propagation speed is not a constant value and it changes from one cable to another. Pertaining to the cable structure, some offline detection techniques might require a return current through the cable shield, which might not be available for some types of cables. Therefore, these methods are not applicable to unshielded cables or shielded cables with shield discontinuities. On the contrary, online methods can overcome this problem if the cable section can be tested independently and if return current is not required.

Although some of the challenges involved in condition assessment of cable were addressed in existing techniques, the demand for an online, non-destructive, and efficient condition monitoring system has not yet been met. None of the techniques discussed in this section feature such a system. In particular, there is no method that looks at the trend and specific behavior of incipient faults over time and incorporates the past behavior of cable in the decision making

process. Also, the interpretation and analysis of the results is made by an expert who must be available. There is no automatic, computer based pattern classifier that makes the decisions online and without the need for a continued human service. Realizing the need for the development of an online, nondestructive, and efficient diagnosis system, this dissertation presents a methodology for such incipient fault diagnosis system (IFDS) to detect underground cable incipient faults before they become catastrophic. The system provides vital information to help the operator with the decision-making process regarding the condition assessment of the underground cable. It incorporates advanced digital signal processing and pattern recognition methods and utilizes novel detection methodologies to classify recorded data into designated classes and detect when the cable is near failure

2.7 PROBLEM FORMULATION

2.7.1 Introduction

Underground cable diagnostic testing and condition assessment has been an emerging technology. To address the shortcomings associated with the existing techniques, novel methods need to be developed. The focus of the work reported in this dissertation is to develop an efficient, online, and passive system that uses measured voltage and current values over a period of time to diagnosis cable incipient faults. Previous efforts made in this project have led to the development of a monitoring system to collect data from an underground distribution system [3][4]. Further, a preliminary data analysis was conducted to characterize underground cable incipient behavior and evaluate some of the ideas that could be used in the design process [5][6]. In continuation of the previous work, this research work presents a methodology for the diagnosis system that utilizes advanced signal analysis and statistical pattern recognition methods and modified change detection techniques. The following sections formulate the important aspects of the system commencing with a formal definition of incipient faults.

2.7.2 Definition of Incipient Faults

Underground cable incipient faults are the primary causes of catastrophic failures in the distribution systems. These faults develop in the extruded cables from gradual deterioration of the solid insulation due to the persisting stress factors. The initial incipient activity is caused by the electrical stresses applied to the voids or protrusions near the conductor shield insulation

interference. This region undergoes an extremely high electrical stresses and such irregularities serve as stress amplifiers when they produce a non-uniform electrical field. Once initiated, the gradual damage propagates locally through the insulation in the form of a tree and the incipient process develops. The aging in the insulation can progress due to the contribution of electrical stresses in the form of partial discharges i.e. electrical trees or from the presence of moisture in the form of water trees. Electrical trees are swift whereas the propagation time of the water trees is expressed in years [34]. Water trees fail the cable when they convert to electrical trees as a result of heat generation or under other stress factors. Once this happens, the time to failure is normally short because the initiated electrical tree propagates rapidly through the already weakened dielectric. The only window for detection is during the conversion process [13]. Electrochemical trees are also likely to develop which are believed to be due to the presence of chemicals in the region [1]. Regardless of the type of aging mechanism, the term *incipient fault* encompasses the insulation treeing process from inception to completion before leading to a catastrophic failure. From a macroscopic perspective, incipient faults refer to the abnormalities associated with any type of deterioration phenomena manifested in the cable electrical signals.

2.7.3 Characterization of Incipient Faults

The very important step to detect underground cable incipient faults is to characterize the behavior of the cable undergoing incipient faults. As shown in Fig. 2.11, this characterization can be typically realized by the analysis of a large database of incipient faults, which could be obtained through computer simulations or field experiments. Generating simulation data is less expensive, time consuming, and labor demanding than that of the field data; however, accurate models need to be available for system components as well as incipient faults to establish an analytical model of the distribution system that can be used to obtain data through computer simulations.

As discussed, incipient faults develop and progress through various stages which are not easy to model. They convey a random, complex, and nonlinear behavior such that the deterministic models developed for short circuit studies lack sufficient accuracy. The modeling task becomes more difficult and challenging when one realizes that it involves other power equipment such as distribution transformers, overhead lines, and switch gears besides the underground cable. To date, there is no complete model for underground cable system to use in

the computer simulations of incipient faults. As such, the computer modeling of incipient faults is not practical and an alternative solution needs to be adopted.

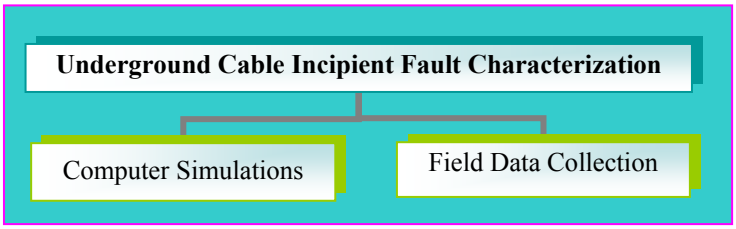


Fig. 2.11: Approaches to characterize underground cable incipient behavior

Collecting data through monitoring sites is the most appropriate way of gathering information about the behavior of cables being exposed to incipient faults at various degrees. As said earlier, many stress factors contribute to the development and progress of incipient faults. The way these factors interact and affect the insulation structure influences the behavior of the resulting incipient fault and therefore one can not assume a perfect realization is achieved if either one of the stress factors is ignored or weighted less than others. In an actual monitoring site, every factor that can somehow affect the cable response is considered including those which are not even known to be influential in the aging process. Thus, creating a database of incipient faults from field recorded data assures that accurate information is collected to characterize cable incipient behavior.

In this research work, three sets of field data were collected: data from controlled experiments conducted at the Texas A&M Downed Conductor Testing Facility on field aged and intentionally damaged cables (CE Data Set); short-term on-line monitoring conducted at various service sites of TXU utility (ST Data Set), and long-term, on-line monitoring conducted at one site of the utility (LT Data Set). The controlled experiments were conducted at the Downed Conductor Testing Facility at Texas A&M University Riverside campus over several noncontiguous months [3]. The purpose of the CE Data set was to study the incipient fault characteristics on typical cable samples that were either field aged or intentionally damaged. This study would reveal specific signatures of faults without the interference of other system

phenomena. The short-term monitoring of underground cables was conducted at various commercial and residential TXU sites over several noncontiguous months. The purpose of this data collection was to generate data that might allow the observation of general characteristics over several sites. The LT Data set includes data collected over a long period of time and the purpose was to observe trends of characteristics [5][6]. Also, the data were used to study the behavior of the underground system before and after faults that occurred in the cable lateral.

2.7.4 Typical Characteristics of Incipient Faults

Deterioration of insulating materials typically conveys itself as sporadic arcing, which is believed to increase in severity as the equipment nears failure. Most often, the insulating material undergoes a gradual aging process before a catastrophic fault occurs. During this period, the electrical properties of the insulation alter adversely and incipient behavior commences.

Incipient behavior is portrayed as a spike or series of spikes (burst) in measured current waveforms of a distribution system. This is a direct result of an ongoing aging and deterioration process in the insulation medium used in cables and other power equipment. The abnormalities introduced by incipient behavior are indicators of system health. As the system ages, this atypical behavior tends to exhibit itself conspicuously. Persisting incipient abnormalities can eventually lead to catastrophic failure and system unscheduled outages. Hence, they are precursors to an imminent fault. This was observed during field experiments conducted on a distribution transformer [50]. Fig. 2.12 shows advanced stage of incipient behavior that preceded failure of the transformer.

Incipient faults unlike short circuit faults typically do not draw sufficient currents from the line to trigger the protective devices, therefore there is no protection mechanism implemented in the distribution system to detect them. When the cable undergoes an incipient activity, the level of variations in the monitored signals is far below the sensitivity of most conventional fault diagnosis devices.

Incipient faults may convey intermittent, asymmetric and sporadic spikes, which are random in magnitude and could involve sporadic bursts as well. Unlike, short circuit faults that have a deterministic behavior, incipient faults exhibit complex, nonlinear and dynamic characteristics and may not last for a definite period. They may persist in the system from as little as several days to several years. The nature of these faults does not allow for predicting a

linear or deterministic behavior since there are many factors that influence the aging process in a nonlinear fashion.

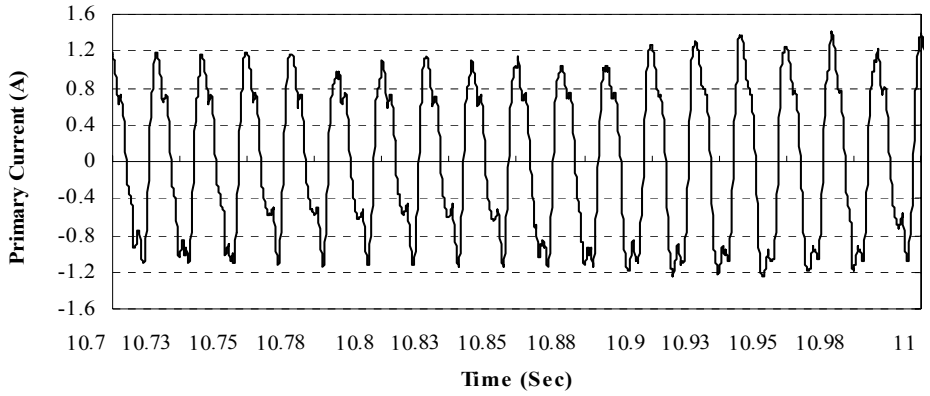


Fig. 2.12: Primary current of a distribution transformer during advanced incipient behavior preceding its failure

The abnormalities associated with incipient faults usually manifest themselves in the high frequency spectrum and possess small values, thus noise interference is a potential problem that needs to be handled appropriately. An example of incipient-like behavior conveyed in the high frequency current signal is shown in Fig. 2.13. This signal resulted from filtering of all aspects of the signal except 2-7.5 KHz. The dominant 60 Hz component was removed so that the low energy high frequency components can become visible. Spikes represent a very fast transient at the inception time of the abnormality.

In addition to the challenges associated merely with the incipient faults, there are other characteristics resulting from the interference of distribution system events with the cable incipient behavior. Comparing with transmission power systems, normal events in typical distribution systems occur more frequently and exhibit more dynamic behavior. Load changes and various switching operations take place very often and many diverse load characteristics can be recognized in such systems. The resulting challenge is that the behavior of some abnormalities in distribution systems due to normal switching operations might resemble the

behavior attributed to incipient behavior. For example, the level of variations in the current signal due to a load change in such a system is sometimes low. Thus, a load change could mimic an incipient abnormality if insufficient features and characteristics are considered.

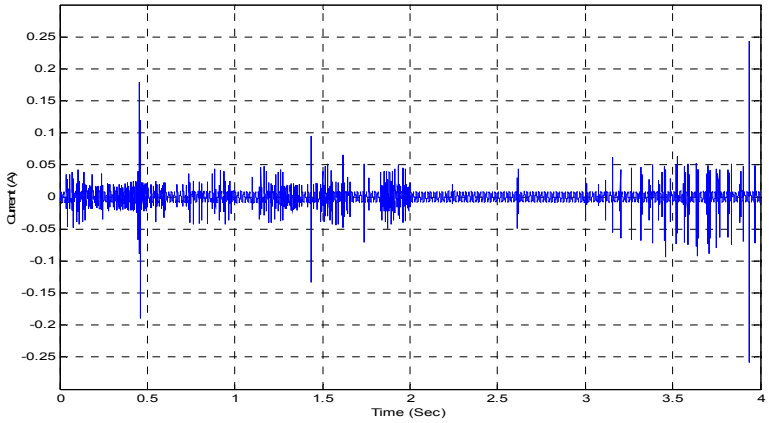


Fig. 2.13: High frequency current signal during an incipient behavior

Recall that incipient faults are also known to introduce small variations in the signals at least at the early stage of their development. The consequences of these common characteristics are two fold. Firstly, normal events and incipient faults can not be distinguished by the magnitude change in time domain only. Secondly, new and specific requirements are imposed on the diagnosis methods in that they must be sensitive and selective. Thus, the diagnosis system must look at the specific activities over time and monitor the system on its gradual path from healthy to sick and analyze the measurements to recognize and detect such faults.

2.7.5 Recognition of Incipient and Non-incipient Patterns in Data

The data obtained through various experiments contain vital information about the system behavior that needs to be extracted using an efficient pattern recognition methodology. This task becomes challenging when one realizes that not all the collected data contain information about the incipient behavior of the cable. It is perceivable that a portion of the collected data contains

normal system behavior or non-incipient events. The non-incipient data can successfully be recognized by incorporating a priori knowledge about the characteristics of such events. Moreover, the gradual nature of incipient faults requires that a large amount of data be collected to provide a sufficient timeframe to monitor and track the changes. Thus, the pattern recognition method is not as straightforward as it appears and must overcome the curse of the large database and irrelevant data samples. Following the forgoing discussion, a complex yet efficient pattern recognition system was developed to perform efficient data processing and pattern classification. The various aspects of the developed pattern analysis system is illustrated in Fig. 2.14 and discussed hereafter.

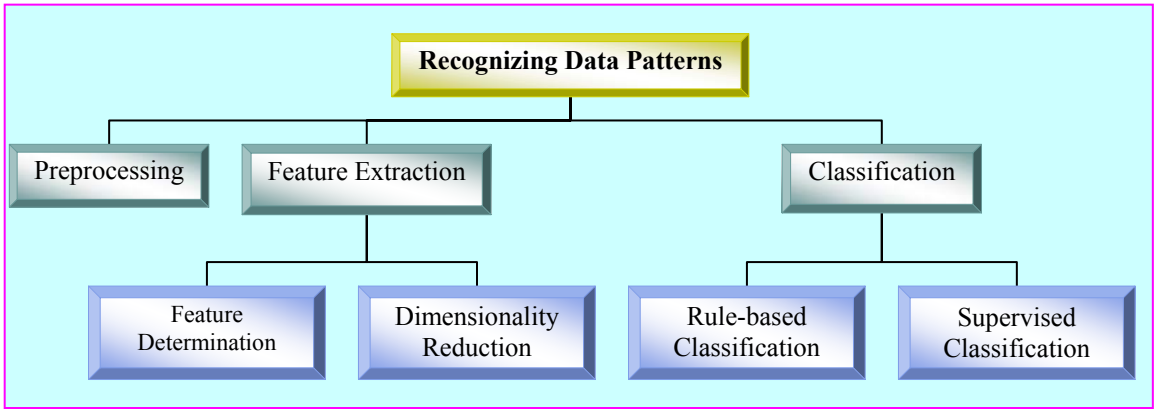


Fig. 2.14: Pattern recognition aspects of the IFDS

2.7.5.1 Preprocessing of Field Recorded Data

When collecting field data, the requirements of the data acquisition system are met but not those of the processing tasks necessarily. The nature of the data most often requires preprocessing schemes to be implemented before advancing to the analysis phase. Preprocessing includes various data manipulation operations that prepare the raw data for the subsequent analyses without scarifying the accuracy and information content of the signals. The significance of the preprocessing steps could be as important as improving the efficiency or as vital as

avoiding wrong assignments. As shown in Fig. 2.15, the preprocessing tasks resulting from three inevitable facts associated with field recorded data include resampling, DC removal, and denoising.

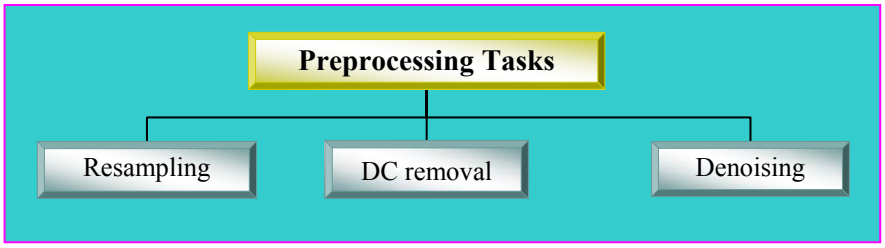


Fig. 2.15: Preprocessing tasks implemented for the field recorded data

During the recording of data in the field, due to some practical reasons, the actual sampling rate might be higher than what is actually needed for a particular signal. Thus, resampling the digital signal becomes a necessary preprocessing step to eliminate inevitable redundancies and achieve an effective sampling rate. This task is usually accomplished by down-sampling of the signal.

Another preprocessing task stems from the fact that the presence of DC components on the recorded signals to be called *measurement DC offset* is an inevitable feature of real world applications. This measurement DC offset needs to be estimated and suppressed before advancing to any data analysis operation. The DC removal procedure can be applied in real time. As an alternative, off-line DC processing can be adopted through which the spurious DC values are estimated and eliminated from the digital signal.

Finally, the curse of noise contamination requires another preprocessing task i.e. denoising. The physical signals being recorded or transmitted in the field are often contaminated with noise from the environment or surrounding electrical equipment. As such, one of the problems that primarily arises in the analysis of the data is the noisy content of the signals and its removal without sacrificing signal features. Noisy data adversely influences the signal processing results and as a result, denoising is a necessary step in the field data preprocessing. This task aims to remove the noise and keep the important signal features as mush as possible [51]. A wavelet-

based denoising procedure was adopted to efficiently remove the noise components. This method uses the wavelet decomposition theory to concentrate the signal energy to a small number of large coefficients. The coefficients, which fall into a rejection band, are dropped, and those that are not located within the band are reserved or shrunk.

2.7.5.2 Feature Extraction

The goal of feature extraction in general is to obtain a set of sufficient and necessary features that discriminate classes with a high degree of accuracy. This important procedure is a key step for the success of any classifier. The feature extraction process involves two steps: feature determination and dimensionality reduction. First, a number of raw features are obtained. Second, the raw features are projected onto a lower dimensional space by various means such as multivariate statistical techniques in order to reduce the dimensionality while preserving a good classification rate [52][53].

In this work, three sets of features in time domain, frequency domain, and wavelet domain were investigated. The time domain features included statistical information, RMS, and spike features. Statistical features include maximum, minimum, range, and mean, variance, skewness, and kurtosis of the signal. RMS features encompass the number of step changes and percentage change calculated using RMS signal. Spike features on the other hand entail maximum spike magnitude, spike frequency, duration and inception angle. Fourier features are the calculated harmonics of the signal using Fourier analysis. Wavelet features encompass level-dependent energy and histogram features of the detail decompositions. It has been shown that wavelet analysis performs very well when working with essentially non-stationary power signals [54][55][56]. Since, the incipient fault high frequency data are non-stationary and transient in nature, utilizing powerful analysis techniques such as wavelets would be of a great benefit. Every feature represents a portion of a signal's characteristics and the combination of all features would model the desired signal exclusively. In the feature determination phase, the presence of potential redundancies is not a subject matter at this point.

The objective of dimensionality reduction is to keep the dimensionality of the pattern recognition problem (i.e. the number of features) as small as possible while preserving good classification accuracy. Dimensionality reduction can be accomplished by means of feature selection or feature extraction as shown in Fig. 2.16. The term feature selection refers to

techniques that select the best subset of the input features set. Methods that create new features based on transformations and combinations of the original feature set are called feature extraction methods. The choice between feature selection and extraction depends on the application domain.

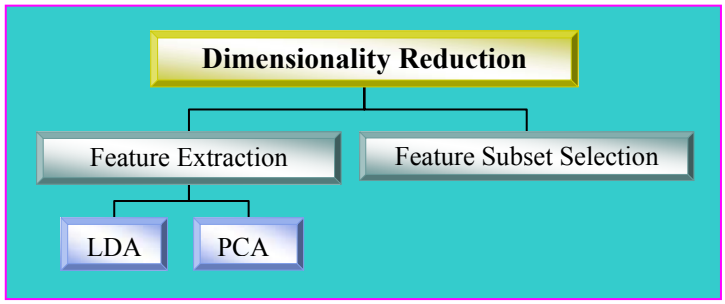


Fig. 2.16: Dimensionality reduction approaches

As shown in Fig. 2.16, two feature extraction methods were investigated in this work. They are namely Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The objective of PCA is to perform dimensionality reduction while preserving as much as the randomness (variance) in the high-dimensional space as possible [52][57]. PCA performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance. The main limitation of PCA is that as an unsupervised method, it does not consider class separability information. There is no guarantee that the direction of maximum variance will contain good features for discrimination.

The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible [52][58]. In LDA, interclass separation is measured by Fisher criterion. The solution proposed by Fisher is to maximize the function that represents the difference between the means of the classes (between-class scatter) normalized by a measure of the within-class scatter. LDA will fail when the discriminatory information is not in the mean of the data but rather in the variance.

Feature extraction methods based on LDA and PCA perform some type of mathematical transformation on the feature vector and therefore the resulting features no longer preserve their physical units and consequently extracting meaningful rules from the classifier results becomes cumbersome. In such situations, the feature subset selection (FSS) methods are employed. Feature subset selection requires a search strategy to select candidate subsets and an objective function to evaluate these candidates. There are a large number of search strategies among which Sequential Forward Selection (SFS) is chosen for further analysis [52].

2.7.5.3 Classification of Patterns

Generally speaking, in the field of pattern analysis, a pattern consists of a pair of given variables (χ, ω) where χ is a vector of observations namely feature vector and ω is the concept behind the observation i.e. the label. Classification involves pattern assignments as to which class a particular sample belongs. Once a feature selection or extraction procedure finds a proper representation, a classifier can then be designed using a number of possible approaches. The question whether one classifier is better than another classifier can barely be answered in practice unless we try to answer the question "Does this classifier solve our problem or not?".

The trivial method for the analysis of incipient data patterns would consider a single classifier expected to learn a number of complex decision-making functions and respond to an unknown input appropriately. Considering the dimension of the problem, sophistication involved in the data structure of incipient faults, and expected accuracy, the success of this method is questionable. Instead, one can break down the entire feature space into separate regions and design individual experts which are responsible for responding to a portion of the input space.

The idea of using multiple experts is referred to as ensemble learning in pattern analysis where instead of one expert to make the decision, an ensemble of experts are employed each of which is responsible for responding to a specific pattern and the outcome is the combination of the individual responses [59]. This technique is useful in a number of ways to improve the accuracy and efficiency of the overall method. The compromise is that one must be able to build components that are more accurate than chance and, more importantly, that are independent from each other. This is achievable when one can effectively utilize a priori knowledge to filter out the observed events that exhibit known behavior. Understanding the distribution system specific

characteristics and the many known normal and abnormal events associated with the types of distribution loads are essential in this design process.

As mentioned, in a typical distribution system, it is expected that switching events occur frequently. This implies that when the underground cable is monitored for a period of time, there are many events that are not associated with incipient faults. Unlike fields such as power quality, in the realm of incipient faults diagnosis, these events are not of much interest. To recognize the patterns of non-incipient data, one might consider trivial approaches such as visual inspection and manual filtering of the recorded data but when encountering a tremendous number of data records, this solution is a tedious, inefficient, potentially inaccurate, and very time-consuming process. As such, it is imperative that new, automatic, and efficient methods be developed to recognize the non-incipient power system events and avoid an unnecessary increase in the complexity of the final diagnosis system.

As shown in Fig. 2.14, the non-incipient patterns are recognized by utilizing a particular first-order classifier known as the rule-based classifier created from machine learning methods. Rule-based classifiers are attractive solutions when classes of data can be characterized by general relationships among entities [53]. These classifiers can be deduced automatically from the data and expressed as a set of crisp or fuzzy *if-then* logical statements. The specific kind of rules implemented for non-incipient data are called "1-rules" that classify an input capture based upon a single feature i.e. they are 1-level decision trees [60]. The biggest advantage of rule-based classifiers is that the reasons for their decisions are readily verified. In other words, the results generated by these classifiers are comprehensible and the interpretation of the results is straightforward [61]. Crisp logical rules are the core components of the rule-based classifier, however, establishment of these rules for the data recorded from a distribution system require prior knowledge about the system events and their characteristics. Despite the relative simplicity, the rule-based classifiers have been shown to perform well on most commonly used datasets [60]. In this work, a rule-based classifier was designed to classify DC classes and various other load related disturbances conveyed by the shape of the RMS current signal. Further discussion about this classifier will be provided in Chapter IV.

Rule-based classifiers are optimal choices when the number of rules is small and the accuracy of the classification is adequately high. The if-then expressed information used by the rule-based classifier can be extracted for the non-incipient data but the patterns manifested in the incipient data are not expressible by linear, single level, and crisp relationships. In fact, as

described earlier, the complexity involved in these patterns is the actual challenge of the classification task. Furthermore, rule-based classifiers using crisp rules identify one class as the result and thus can only provide a black-and-white picture, whereas incipient data vary according to some gradation defined by the severity. Therefore, a set of other complex classifiers known as supervised classifiers were utilized for the purpose of recognizing the patterns in the incipient data.

Classification in the supervised mode involves a two-step process: prototype selection and classifier design. In prototype selection, a subset of given data is selected to train the classifier. Once the prototype data becomes available, the classifier can be trained according to its specific training procedure to recognize the complex relationships among the data and provide a colored picture as opposed to the black-and-white picture developed by the typical rule-based classifiers. The supervised classifiers deployed in the system are Support Vector Machines (SVM), Self-Organizing Map (SOM), K-Nearest Neighbors (KNN), and Discriminant classifiers. These classifiers and their results will be further discussed in Chapter IV .

2.7.6 Detection of Incipient Faults

Degradation and development of incipient faults are gradual in nature and do not occur suddenly. The incipient activity might persist in the cable for a few days, months, or even years before it actually develops into a cable fault characterized by a short circuit fault. Realistically, the incipient activity of the underground cable may experience a gradual incremental increase over the course of cable operation before a catastrophic failure takes place. The mission of the IFDS is to monitor the condition of the system as long as the monitored parameters are below a specified normal threshold. Once the threshold is passed, the system must detect this situation before catastrophic failure occurs so that corrective and preventive actions can be taken so that an outage does not occur. Therefore, the incipient fault detection problem translates into a change detection problem where the goal is to keep the process somewhere below a threshold or “in-control”. Unlike single change point detection methods, in this problem there is no target value for the change parameter and the monitoring continues as long as the underground cable condition is believed to be in green-zone and a catastrophic failure is not imminent.

As presented in Fig. 2.17, the developed detection method makes use of a numerical modeling approach supplemented by a severity measure. The severity measure is based on the temporal analysis of arrival times of incipient abnormalities through a new index called the

Global Severity Index (GSI). This qualitative measure provides the severity path of ongoing incipient faults over time and is computed from the Laplace Test Statistic [62]. The second approach is based on the numerical modeling of data with SOM. As a quantitative detection approach, it uses the natural logarithm of the minimum modeling errors in chronological order as the detection index. To solve the change detection problem, three algorithms are investigated which include Cumulative Sum (CUSUM), Exponentially Weighted Moving Averages (EWMA), and Generalized Likelihood Ratio (GLR). The detection methods and algorithms will be presented in Chapter V and the detection results will be provided in Chapter VI.

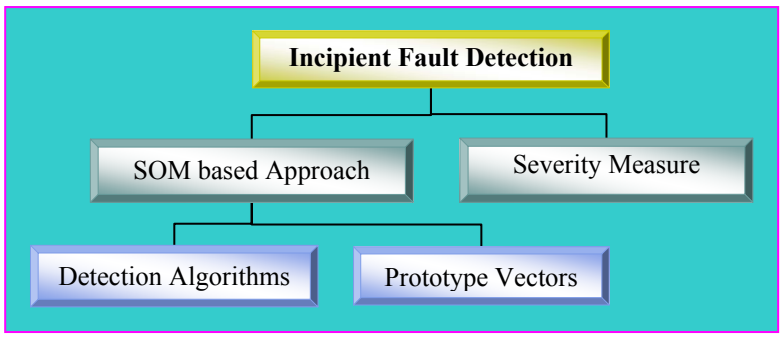


Fig. 2.17: Units of the incipient fault detection method

2.8 THE OPERATIONAL ASPECTS OF THE IFDS

The operational aspects of the system are presented in Fig. 2.18. Characterization, preprocessing, feature extraction, classification, detection, and postprocessing are major attributes of the system. Characterization provides vital information about the behavior and nature of the incipient faults. Section 2.7.3 elaborated on this. Preprocessing encompasses a set of operations that prepares the raw data for subsequent processing tasks. DC removal, resampling, and denoising are the preprocessing tasks that were considered. Chapter III provides more information on this. Feature extraction is an important step prior to the classification through which a number of time domain and wavelet domain features are extracted to feed to the classifiers. This crucial step is discussed in Section 4.3. Classification is an important

functionality of the system that is accomplished by the combination of rule-based and supervised classifiers. Chapter IV discusses this aspect of the system in detail. As an important functionality of the system, the detection is accomplished through a numerical modeling approach. Chapter V elaborates on this important aspect of the system. Finally, postprocessing is used to interpret the results and provide user-friendly outputs. The system outputs are discussed in Chapter VI.

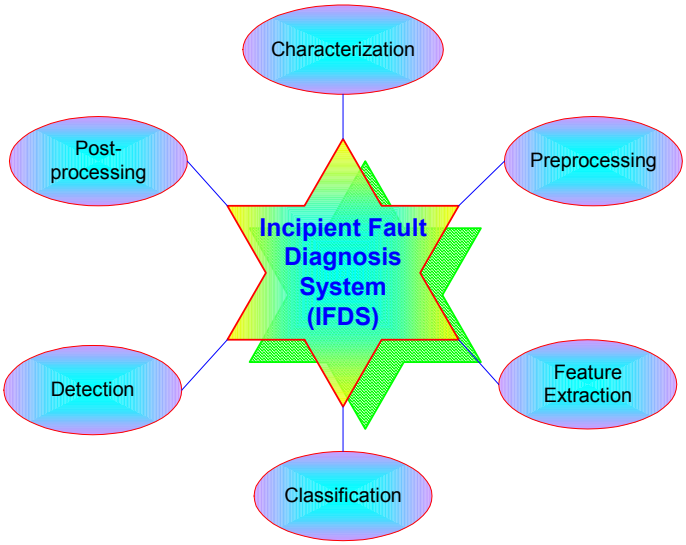


Fig. 2.18: Constituting concepts of the IFDS

2.9 SUMMARY

In this chapter, a literature review on the underground cable incipient fault detection and location methods were provided and each method was concisely discussed and assessed. Also, a number of commercially-available products were introduced and reviewed. The results of this review showed that each fault detection method targets a particular phenomenon preceding cable failure to identify the existing fault. The advantages and disadvantages associated with existing methods were reviewed and the need for the development of a passive, online, and cost effective condition monitoring system for underground cables was justified through discussion of the

incipient fault characteristics and gradual deterioration of the cable insulation. The diagnosis requires a system that is efficient; works online i.e. no power interruption is necessary, and features a passive method i.e. no external signal injection is required.

In the formulation of the problem, various aspects of the IFDS were introduced and discussed. Field data were used to study and characterize incipient faults. Dealing with distribution level underground system incipient faults intrinsically imposes challenges due to the unique characteristics associated with such systems and faults. These characteristics limit the use of the methods that have already been used in other applications. Accordingly, novel methods were developed to meet the specific needs of the various data analysis tasks. The proposed system considers the forgoing features and overcomes the complexity involved by utilizing a priori knowledge, advanced signal processing, pattern analysis and change detection techniques.

In the next chapter, the data acquisition setup as well as the monitored signals will be introduced and then the required preprocessing tasks will be presented.

CHAPTER III

FIELD RECORDED INCIPIENT DATA AND PREPROCESSING METHODS

3.1 INTRODUCTION

This chapter reviews the long term monitoring system and discusses the recorded data from an underground distribution lateral. Using the monitoring system, three basic electrical signals, namely voltage, phase current, and neutral current are observed. The current signals are further filtered to generate four outputs, with different magnitude resolution and frequency spectrum. The voltage signal along with the filtered phase current signals are collectively called *data* in this work. Understanding the format and structure of the recorded signals comprising the data is necessary to establish the required preprocessing tasks performed in the first stage of any pattern analysis problem involving real world data. An overview of recorded data and a complete discussion of the preprocessing tasks implemented in the IFDS constitute the focus of this chapter. The three preprocessing tasks include resampling, DC removal and denoising. The later i.e. noise removal is the most critical preprocessing action that requires a greater attention as it directly affects the overall performance of the IFDS.

3.2 MONITORING SYSTEM AND RECORDED DATA

An underground distribution cable laterally installed in a residential area was chosen to collect on-line data [4]. This site was selected as the most appropriate location to capture possible incipient abnormalities. Fig. 3.1 shows the data collection site, including the distribution transformer and the underground cable. The underground distribution cable lateral is fed from a standard 7200 V distribution feeder and supplies power to the 7200V/120V/240V, 100 KVA, 60 Hz distribution transformers.

Data collection was performed using an on-line monitoring system installed at the site. The monitoring system, whose block diagram is shown in Fig. 3.2, comprises three basic components: signal transducers, analog signal conditioning unit, and a computer-based data

acquisition system. The signal transducers transform the voltage signals to levels acceptable by the signal-conditioning unit. They also transform the current signals into equivalent voltage signals of acceptable range. The transformed signals are then fed into the analog signal-conditioning unit whose functions are to act as an isolation unit, and to filter the signals into various frequency ranges. Signals from the analog signal-conditioning unit are finally fed to the digital data-acquisition system embedded in the computer.

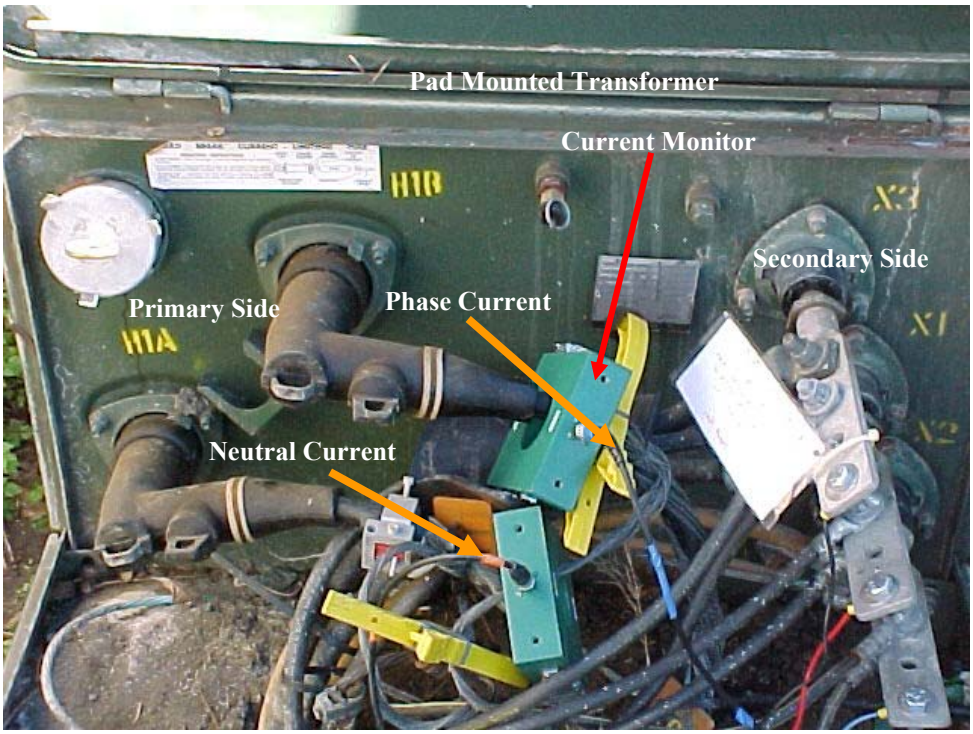


Fig. 3.1: Monitoring site including transformer and cable

Using the monitoring system, three basic electrical signals, namely voltage, phase current , and neutral current are observed. The system records the signals for one-second duration every 15 minutes. Moreover, various statistical and frequency parameters of these signals, namely

average, maximum, minimum, standard deviation, and magnitude of the harmonics are calculated and recorded.

In the signal-conditioning unit, the phase and neutral current signals generate four outputs, as shown in Table 3-1, to increase the magnitude resolution during data-acquisition. The notch in three of the output signals is to remove the dominant fundamental frequency (60Hz), thereby improving the magnitude resolution in the given frequency range. Note that although the neutral and notch low frequency signals are recorded at the site, the IFDS primarily relies on the voltage, phase current, and notch high frequency current signals to perform its operation.

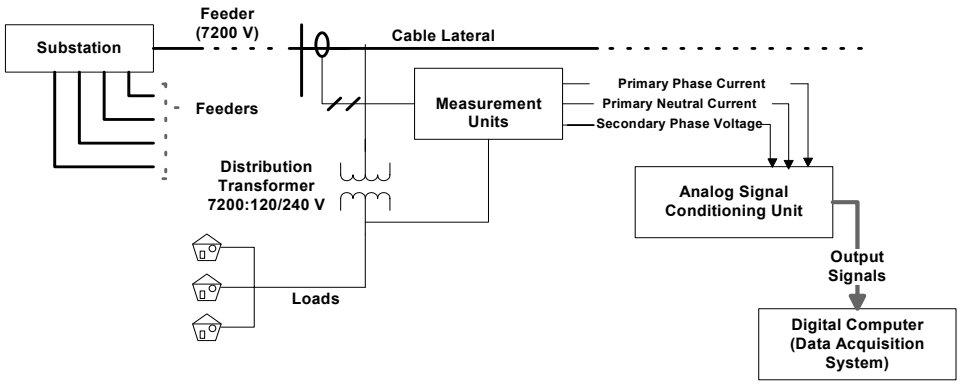


Fig. 3.2 : Block diagram of the monitoring system

Table 3-1:
Categories of current signals

Category	Frequency Range of Output Signals	Sampling Rate (samples/second)
Low Frequency Signal	0 - 780 Hz	1920
Notch Low Frequency Signal	0 -780 Hz, Notch at 60 Hz	15360
Notch High Frequency Signal Scale 1 and 10	2 - 7.5 KHz, Notch at 60 Hz	15360

3.3 DATA PREPROCESSING

3.3.1 Introduction

When collecting field data, the requirements of the data acquisition system are met but not those of the processing tasks necessarily. The nature of the data most often requires preprocessing schemes to be implemented before advancing to the analysis phase. Preprocessing includes various data manipulation operations that prepare the data for the subsequent analyses without scarifying the accuracy and information content of the signals. The significance of the preprocessing steps could be as important as improving the efficiency or as vital as avoiding wrong assignments. The preprocessing tasks included in the IFDS are resampling, DC removal, and denoising.

3.3.2 Resampling

Often resampling is one of the preprocessing tasks. For practical reasons, the sampling rate might sometimes be higher than what is required by the Nyquist criterion. If a signal is already limited to lower frequencies by low pass anti-aliasing filters, naturally there might be redundancy in the signal that should be removed. Thus, the signal must be decimated by an appropriate factor to yield an effective sampling rate.

The specific motivation for the use of preprocessing for long term data is to reduce redundancy. As discussed earlier, the sampling rate for the notch low frequency signals was set at 15360 Hz for practical reasons. This sampling rate gave a frequency range of 0 – 7680 Hz. Nevertheless, these signals were limited to 0 – 780 Hz by a low pass anti-aliasing filter in the analog signal-conditioning unit. Thus, to reduce redundancy, these signals were decimated by a factor of eight yielding an effective sampling rate of 1920 samples/sec. Table 3-2 shows the modified data format of the signals that are recorded in the experiment site.

Table 3-2:
Modified data format for long-term monitoring experiments

Signal Name	Frequency Range	Sampling Rate (Samples/Sec)
Voltage Signal	0 – 780 Hz	1920
Low Frequency Current Signal	0 – 780 Hz	1920
Notch Low Frequency Current Signal	0 – 780 Hz	1920
Notch High Frequency Current Signal(x1&x10)	2 – 7.5 KHz	15360

3.3.3 DC Removal

Another important preprocessing operation encountered in real world data acquisition applications aims to remove the measurement DC components. Power system analysis is generally concerned with sinusoidal voltage and current waveforms and their variations in time or frequency domain. These waveforms can be either generated through computer models or actually recorded in the field. Many times, the analysis results using the recorded or simulated data are similar. However, there is a very important distinction between field and simulated data. Field recorded signals are natural realizations of system behavior whereas simulated data are artificial signals that are produced by a computer program under simplifying assumptions. A simulated pure sinusoidal signal by definition is symmetric and possesses no DC components; but, a recorded sinusoidal waveform from a power system is not expected to have zero mean over its period. In other words, the presence of DC components on the recorded signals referred to as *measurement DC offset* is an inevitable fact of real world applications. Accordingly, the measurement DC offset needs to be estimated and suppressed before advancing to any data analysis operation. The DC removal procedure can be applied in real time. As an alternative, the off-line DC processing can be adopted through which the spurious DC values are estimated and eliminated from the digital signal.

The measurement DC offset is usually unidirectional and possesses small values. Thus, an estimate based on a long signal observation window should provide satisfactory results. For each signal, a single estimated value is calculated and subtracted from each signal sample. Mathematically, this amounts to subtraction of a straight line from the waveform. Fig. 3.3 shows an example in which graph (a) shows the mean values of a current signal calculated in each cycle

before preprocessing and plot (b) is the mean values of the same signal after removing the measurement DC offset. As implied, the original signal is not symmetric and contains negative DC offset over one second length of the recording window. After preprocessing, it is approximately symmetric and the remaining DC values in each cycle are negligible.

It should be noted that although the DC components are often the intrinsic elements of recorded signals, various power system events might also induce non-symmetric, DC-offset components. Therefore, there should be a distinction between the measurement DC values and event-induced DC elements in the recorded signal. This discrimination can be realized by utilizing the characteristics that are associated with the two classes. Section 4.2.3 will discuss this issue in more detail.

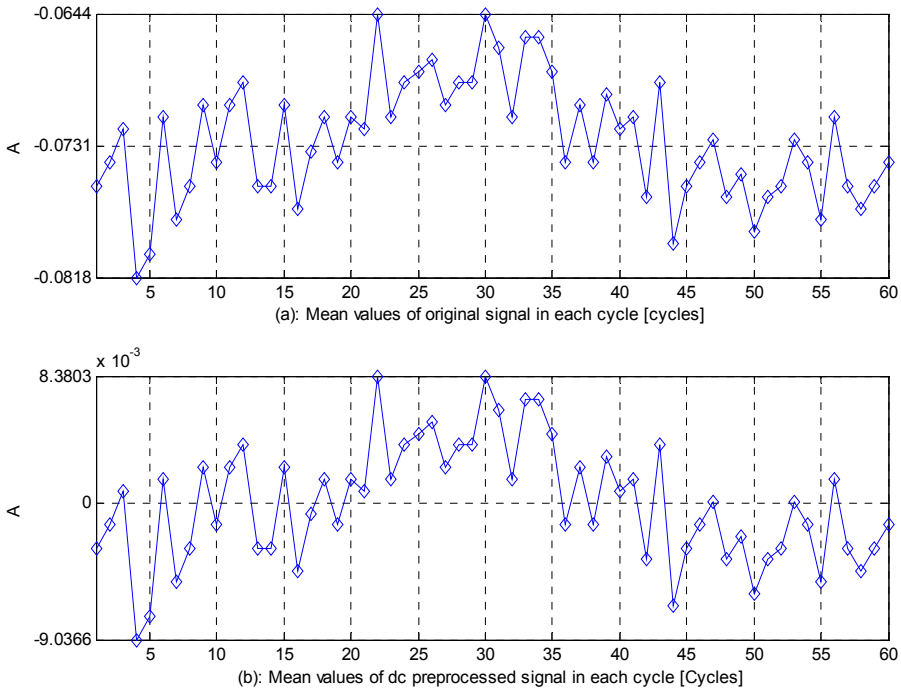


Fig. 3.3 : DC preprocessing of current signal

3.4 DENOISING

3.4.1 Introduction

An important preprocessing operation includes denoising the field data. In various power system applications, it is common to collect data such as voltage and current measurements. The sampling frequencies of these signals usually vary from a few KHz to hundreds of MHz depending on the application and type of analysis. As many power system events are generally transient phenomena diminishing in a few cycles, the data are sampled at relatively high rates to capture the useful information and meet the Nyquist criterion. At the same time, the physical signals being recorded or transmitted at such frequencies are often contaminated with noise from the environment or surrounding electrical equipment. As such, one of the problems that primarily arises in the analysis of the data is the noisy content of the signals and its removal without sacrificing signal features. Noisy data adversely influences signal processing results and as a result, denoising is a necessary step in the field data preprocessing. This task aims to remove the noise and keep the important signal features as mush as possible [51].

In this particular application, the denoising operation is not necessary for all signals. The voltage and low frequency current signal along with the notch low frequency signal were already filtered by analog low pass filters in the signal-conditioning unit at the cut-off frequency of 780 Hz. As the noise signal is characterized by its high frequency content, the denoising action is not required for these signals. On the other hand, the high frequency signals covering a frequency range of 2 to 7.5 KHz may contain noise components. Thus, denoising was performed on the notch high frequency signals.

3.4.2 Denoising Approaches

Noise removal can technically be performed by hardware or software approaches [19]. Hardware approaches use new sensors and differential circuits to discriminate noise, whereas the software denoising approaches provide an alternative solution by taking advantage of advanced signal processing techniques. A noisy signal is recognized by the high frequency components that possess low energy values compared with the signal total energy. This characteristic is the key element employed in the widely used denoising algorithms.

The denoising process can technically be viewed as performing a linear or non-linear filtering of the input signal. One of the conventional denoising methods is the Wiener filter and

its variants. These methods rely on the statistical characteristics of the input signal and produce a minimum mean square error output [63][64]. The original Wiener filter is usually used for stationary signals; however, there are other versions that consider non-stationary signals by segmenting them into short-time essentially stationary blocks similar to what is performed in the Short Time Fourier Transform. However, these methods also simultaneously remove high frequency components of the signal while denoising. In this work, a software denoising approach based on the signal wavelet decomposition is adopted. Before discussing this method, a concise introduction to wavelet signal representation is beneficial.

3.4.3 Wavelet Signal Analysis

A wavelet is a transient waveform that has a zero average value and decays quickly to zero. Due to its transient nature, the analysis of transient signals with a wavelet is naturally more descriptive than that of periodic functions such as sine and cosine bases. Wavelet transform is a relatively new signal processing method but its mathematical foundation was established in the nineteenth century by Joseph Fourier [65]. The concept of wavelets in its present form was first proposed by Jean Morlet and his team in France and the main algorithm was developed by Stephane Mallat in 1998 [66]. Wavelet transform can be performed on continuous and discrete functions respectively referred to as the continuous waveform transform (CWT) and discrete wavelet transform (DWT). The two transforms are discussed next.

3.4.3.1 Continuous Wavelet Transform

Wavelet analysis consists of decomposing a signal into shifted and scaled versions of the original wavelet often called *the Mother Wavelet* and denoted by $\psi(t)$. Mathematically, the wavelet transform of a continuous function $f(t)$ is defined by (3-1) where the continuous variables a and b are the scale and shift factor, respectively [65].

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \qquad (3\text{-}1)$$

As implied, the continuous wavelet transform of the function $f(t)$ is the sum over all time of the function multiplied by scaled, translated versions of the wavelet function. The results $CWT(a,b)$ are called the wavelet coefficients. From an intuitive point of view, the wavelet

coefficients represent the similarity of the input function to the scaled and translated version of the original wavelet. Thus, the input signal in time domain is transformed into a two-dimensional wavelet space of scale and translation (time). For every given scale a and translation b, there is a corresponding wavelet transform coefficient $CWT(a,b)$. The set of all wavelet coefficients constitute the wavelet representation of the input function $f(t)$ with respect to the original wavelet $\psi(t)$.

CWT uses a variable size windowing technique to capture the low and high frequency information over the time-scale region. This is realized by the wavelet function $\psi(.)$ in which the scale factor determines the window size. Scaling a wavelet simply implies stretching it. The wavelet scales and frequency are interrelated. The higher the scale a, the more stretched the wavelet function and thus the lower the captured frequency. On the contrary, the lower the scale a, the more compressed the wavelet and thus the higher the captured frequency [65].

In addition to the mother wavelet, the wavelet transform usually involves an additional function. A complement of the coefficients corresponding to ψ is needed to recover the original signal. The complementary coefficients are obtained by another function called the scaling function ϕ, which is very similar to the mother function. The scaling function can be interpreted as the impulse response of a low-pass filter and is not defined for all wavelets [66].

3.4.3.2 Discrete Wavelet Transform

The continuous wavelet coefficients can theoretically be calculated on any desired scale and shift values resulting in tremendous amount of data. The more efficient and just as accurate approach is however to apply dyadic scales and translations, which are based on powers of two. The Discrete Wavelet Transform (DWT) provides such an analysis [66]. The implementation algorithm first developed by Mallat in 1998 is called two-channel subband coder in the signal processing literature. This algorithm involves filtering and down sampling steps as depicted in Fig. 3.4. Down sampling is performed by keeping only one point out of every two points in each of the two emerging signals.

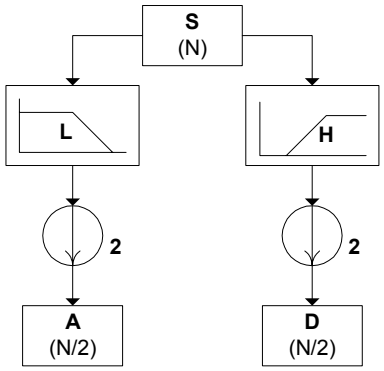


Fig. 3.4: Discrete wavelet analysis at the basic level

As shown, the input signal is passed through two complementary high pass and low pass filters and decomposed into two emerging signals unique in frequency content. These signals are called approximations and details where approximations contain the low frequency components of the signal and details encompass the high frequency elements at every scale. The resulting decomposition produces twice as many signal samples as the original signal. Therefore, down sampling follows the filtering operation to remove the redundancy without losing useful information. The decomposition procedure is repeated at each subsequent scale using the approximation signal from the previous scale. This results in the wavelet decomposition tree depicted in Fig. 3.5.

Sometimes, it is desired to recover the original signal without loss of information. Of course, there is no point decomposing a signal and immediately reconstructing it. When reconstructing is performed, usually the wavelet coefficients are modified before performing the reconstruction step. The reconstruction is in fact the synthesis stage, which is defined by the Inverse Discrete Wavelet Transform (IDWT). It involves up-sampling and filtering steps as shown in Fig. 3.6. Up-sampling a signal simply means extending the signal by inserting zeros between the samples. This figure illustrates the decomposition and reconstruction processes utilizing low and high pass filters called quadrature mirror filters [65][67].The choice of these filters not only determines whether perfect reconstruction is possible, it also determines the shape of the wavelet to perform the analysis. The high pass filter determines the wavelet function, which produces the details and the low pass quadrature filter determines the scaling function, which associated with the approximations of the wavelet decomposition.

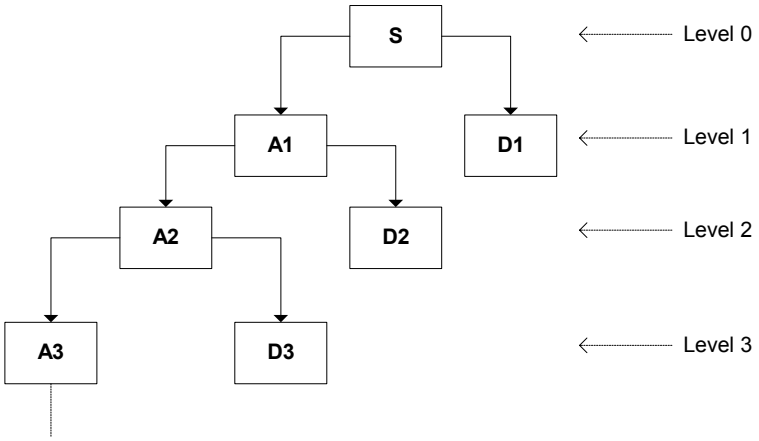


Fig. 3.5: Multi-level wavelet decomposition tree

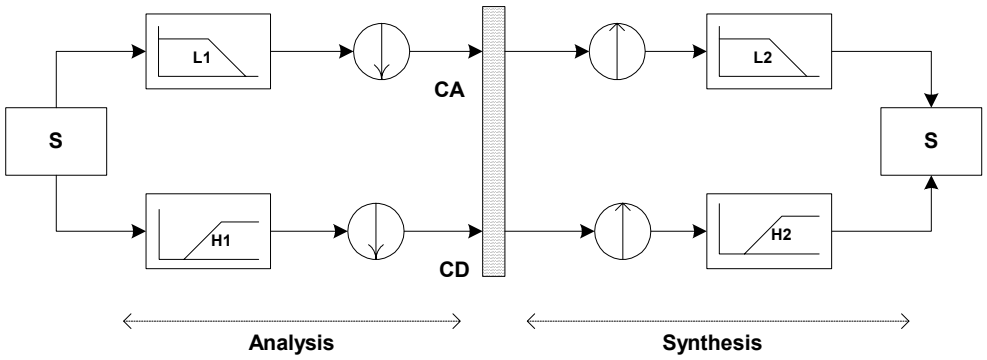


Fig. 3.6: Wavelet decomposition and reconstruction processes

3.4.4 Wavelet-based Denoising

Wavelet-based denoising is a nonlinear technique that was originally developed by Donoho et al. in 1994 [68]. This method uses the wavelet decomposition theory to concentrate the signal energy to a small number of large coefficients. The coefficients, which fall into a rejection band, are dropped, and those that are not located within the band are reserved or shrunk. As said, the wavelet transform involves functions that are well localized in both time and frequency domains

allowing the unique features of the signal to be easily identified during the transformation. Grossman showed that for white noise, the variance and amplitude of the details decrease regularly as the level increases whereas the amplitude and variance of the transformed target signal are irrelevant to the level change [69]. Furthermore, the signal is usually decomposed into a few large coefficients whereas the noise component gives rise to small coefficients only. This is the property of the wavelet transformation that helps suppress the noise part of the signal. Another property of the wavelet analysis is the linearity property. It implies that the wavelet coefficients of the linear combination of two signals are equal to the linear combination of their wavelet coefficients [65]. The same holds true for the corresponding approximation and detail components.

The general wavelet-based denoising method proceeds in three steps: decomposing the noisy signal, thresholding the wavelet coefficients, and reconstructing the signal [65] as shown in Fig. 3.7. In the first step, the signal is transformed into the wavelet domain and represented by its coefficients at different scales. This stage requires that three elements be chosen. They are namely, the type of mother wavelet, the order of the mother wavelet, and the level of the decompositions. Once these parameters are set, the wavelet decomposition can be performed. The second step is crucial in efficient noise cancellation. The key questions are i) how to perform the thresholding and ii) how to find the appropriate threshold value for each scale given the noise model. The reconstruction is defined by the Inverse Discrete Wavelet Transform which involves up-sampling and filtering steps as shown already in Fig. 3.6. At the completion of this stage, the denoised signal is represented in time domain.

The success of the noise cancellation can be assessed using a number of measures. One of the measures involves the signal-to-noise ratio (SNR) defined by (3-2). The goal of the denoising is to improve the SNR. It should be noted that in fact this measure is not a real SNR value as the numerator is not the noise free signal. In practice, the real noise signal is not known. Another measure evaluates the mean squared error (MSE) between the input signal and denoised signal and minimizes it. In this case, the denoising action is guided by a reasonable minimum MSE as defined in (3-3).

$$SNR = 10\log\left(\frac{\sum s_i^2}{\sum (s_i - f_i)^2}\right) \tag{3-2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(s_i - f_i)^2 \tag{3-3}$$

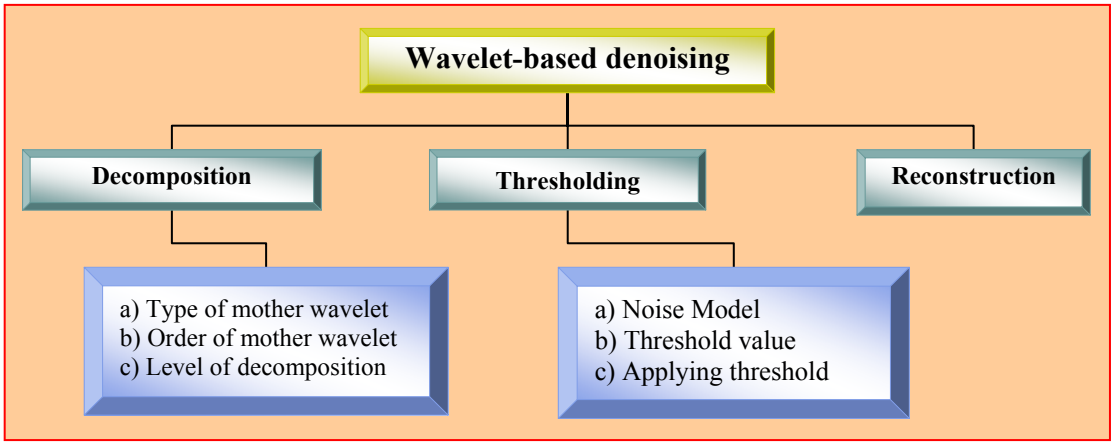


Fig. 3.7: Wavelet-base denoising procedure

3.4.4.1 Decomposition

Consider denoising the typical notch high frequency current signal shown in Fig. 3.8 along with its estimated frequency spectrum. From the time domain waveform, dominant asymmetric spikes with positive peak of 236mA and negative peak of 190mA are seen around 0.61 seconds. The estimated spectrum reveals frequency components spread over the entire pass band, among which the 2.25 and 5 kHz components are dominant.

As discussed, during the decomposition stage, the signal is transformed into the wavelet domain and represented by its coefficients at different scales. This stage requires that three elements be chosen. They are namely, the type of mother wavelet, the order of the mother wavelet, and the level of the decompositions. From previous studies, it was found that Db4 wavelet at level five provided satisfactory performance [51]. The resulting coefficients are shown in Fig. 3.9. The spikes in each of the details one through three show the contribution of that frequency band to the spike in the original waveform. Details four and five comprise much smaller energy and apparently do not contribute to the spike. Another observation is that the noise level is different in each detail, which implies that each detail should be thresholded with a different threshold value.

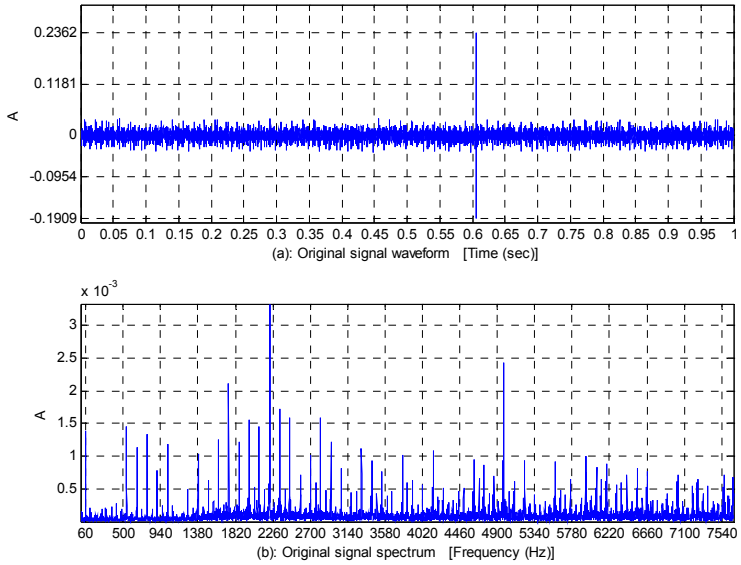


Fig. 3.8: Phase notch high frequency signal and corresponding frequency spectrum before denoising

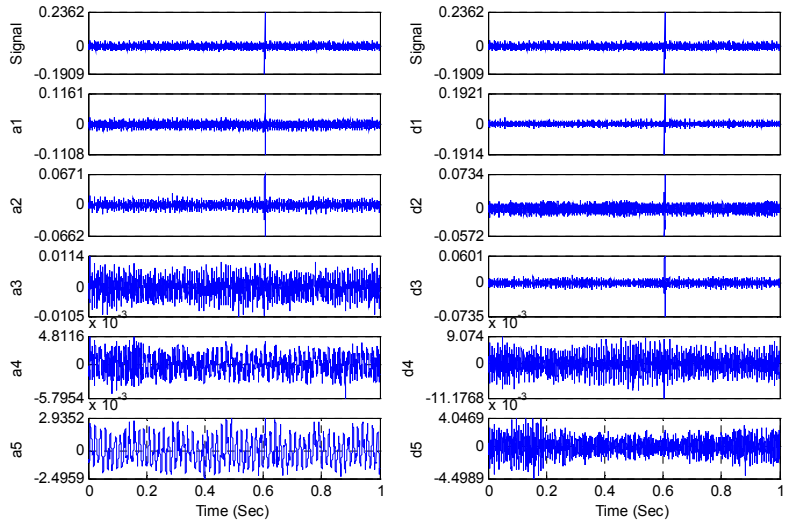


Fig. 3.9: Phase notch high frequency signal in db4 wavelet domain before denoising

3.4.4.2 Thresholding

As a crucial stage in efficient noise cancellation, this step involves adopting appropriate thresholding method and threshold selection rule with respect to the estimated noise structure. The important aspects of the thresholding stage are discussed below.

3.4.4.2.1 Thresholding Methods

The thresholding methods are technically grouped into the following categories: hard thresholding, soft thresholding, and semi-soft thresholding [65]. The hard thresholding simply sets all the signal elements whose absolute values are lower than the threshold to zero and retains the others unchanged. In the soft thresholding, however, in addition to setting to zero the elements whose absolute values are lower than the threshold, the other nonzero elements are shrunk towards zero. Semi-soft thresholding employs two threshold values (T1 < T2) to provide a compromised solution between the hard and soft thresholding. The coefficients whose values are greater than T2 are kept unchanged, those less than T1 are set to zero, and those between T1 and T2 are shrunk by a weighted average of T1 and T2 denoted by w [64][70]. These methods are mathematically expressed in (3-4), (3-5), and (3-6), respectively. $sign(s)$ returns the sign of the signal and $|\,|$ denotes the absolute value.

$$Soft\,Thresholding: \begin{cases} sign(s).(|s| - threshold) & if \quad |s| > threshold \\ 0 & if \quad |s| \le threshold \end{cases} \tag{3-4}$$

$$Hard\,Thresholding: \begin{cases} s & if \quad |s| > threshold \\ 0 & if \quad |s| \le threshold \end{cases} \tag{3-5}$$

$$SemisoftThresholding: \begin{cases} s & if \quad |s| > T_2 \\ ws & if \quad T_1 \le |s| \le T_2 \\ 0 & if \quad |s| < T_1 \end{cases} \tag{3-6}$$

Fig. 3.10 describes graphically the thresholding methods when applied to a pure sinusoidal waveform. As seen, the hard thresholding method results in discontinuities at ± 0.5 while the soft thresholding procedure is smooth, avoids fictitious oscillations, and produces visually appealing results. On the other hand, the hard thresholding approach achieves a smaller mean

square error (MSE) [71]. The semi-soft thresholding method provides more flexibility at the expense of the need for establishing two thresholds and a weight vector. In short, there is an equivalent bias-variance trade-off between noise reduction and smoothing of the signal and thus the selection of each of these methods depends on the denoising objective.

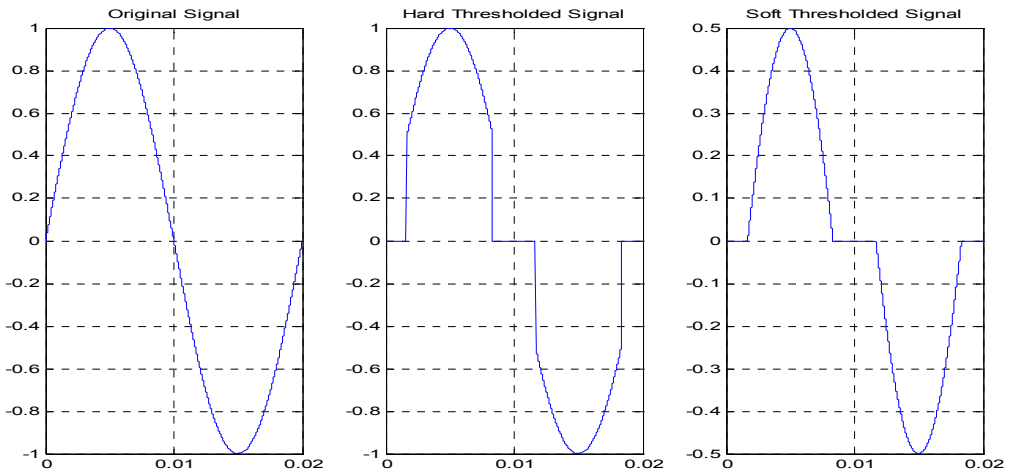


Fig. 3.10: Hard and soft thresholding of a sine signal with a threshold value of 0.5

3.4.4.2.2 Noise Models and Threshold Selection Rules

Generally, electrical signals are exposed to noise interference from many different sources. Radio and TV station broadcasts, mobile telecommunications, corona in high voltage equipment, arcing faults and discharges in adjacent circuits are among the common noise sources. Noise introduced by some of these sources has a well-known structure and frequency range, but in general, the noise structure must be estimated for the selection of the thresholds in the denoising process.

Let s be an observation (corrupt) signal, e be a pure noise signal, and f be the noise-free signal. The general noise model assumes that the observation signal is equal to the noise-free signal superimposed by the noise [65]. This mathematically implies that

$$s(n) = f(n) + \sigma\, e(n) \qquad n = 1,\ldots, N \tag{3-7}$$

where N denotes the length of the signals and σ represents the noise level. The model in its general form does not impose any constraints on the noise signal and it could theoretically model any function. In practice, however, the noise signal is considered a random variable, independently distributed as a Gaussian function. Depending upon the assumption made about the noise structure, the general noise model can be used to describe the denoising formulation. The four noise models listed in Table 3-3 are discussed next.

Table 3-3:
Noise types

Type No.	Description
I	White noise with unit noise level
II	Unscaled white noise
III	Non-white noise
IV	Non-white noise with non-stationary variance

The type I noise model is the simplest one, which assumes that the noise can be modeled by a zero mean, constant variance σ^2 Gaussian white noise. If this is the case, a fixed threshold based on the signal length or statistical properties of the wavelet coefficients can be applied to the signal decompositions at each level. In practice, at least small deviations from this assumption are tolerable by the thresholding methods [72]. The type II noise model assumes the basic model with unscaled noise. The noise level is estimated using the first detail coefficients. The type III noise model takes the non-white noise into account for which a level-dependant estimation of the noise level is required to rescale the thresholds. In other words, in this case σ is substituted by σ_{level}. Unlike the first three models in which the noise variance is assumed to remain constant, in the type IV noise model, it can vary with time. If this noise model holds true, it requires level-by-level time-dependant thresholds to be appropriately applied to the wavelet coefficients.

In practice, the noise structure is difficult to estimate due to its dynamic and stochastic nature. However, when the noise signal is observed for a relatively long window size, the law of large sample sizes can be applied. Fig. 3.11 depicts a pure noise signal waveform along with its corresponding frequency spectrum. This signal was recorded during a feeder outage in the monitored system. Since there was no current flowing through the phase conductor, the measured values were merely perceived as the pure noise signal during the de-energization period. As the signal spectrum shows, there is a colored noise with dominant high frequency components concentrated around 5 kHz. In addition, small low frequency components including 60 Hz are seen from the non-ideal notch filter. Concerning the noise origin, it is noted that the AM/FM radio and TV stations could not have contributed to the noise signal since the notch high frequency signals were filtered by the band pass filter with a high cutoff frequency of 7.5 kHz. These sources typically introduce noise at frequencies ranging from a few hundred kHz to a few hundred MHz.

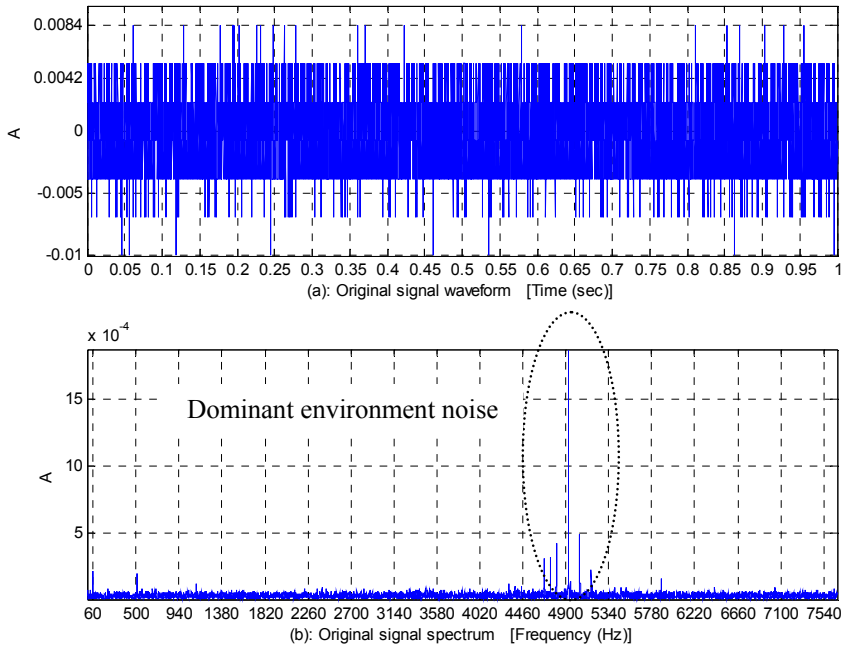


Fig. 3.11: Noise realization and corresponding signal spectrum

The noise structure can also be viewed in the time-scale domain. The wavelet decompositions of the signal in Fig. 3.11 before denoising are shown in Fig. 3.12, where fourth order Daubechies wavelet (Db4) at level five was used. For each level, there is one approximation and one detail signal; $a1$ through $a5$ represent the signal approximations and $d1$ through $d5$ denote the signal details, respectively. As seen, the noise signal possesses different energies at each level; thus, it requires that a level dependant threshold be applied for perfect suppression. Furthermore, the noise details at each level are deemed covariance stationary. Accordingly, this observation suggested that the basic model with the non-white noise structure would be an appropriate model for the denoising operation. The proper denoising approach can be achieved by evaluating the performance of the denoising operation using an exhaustive approach.

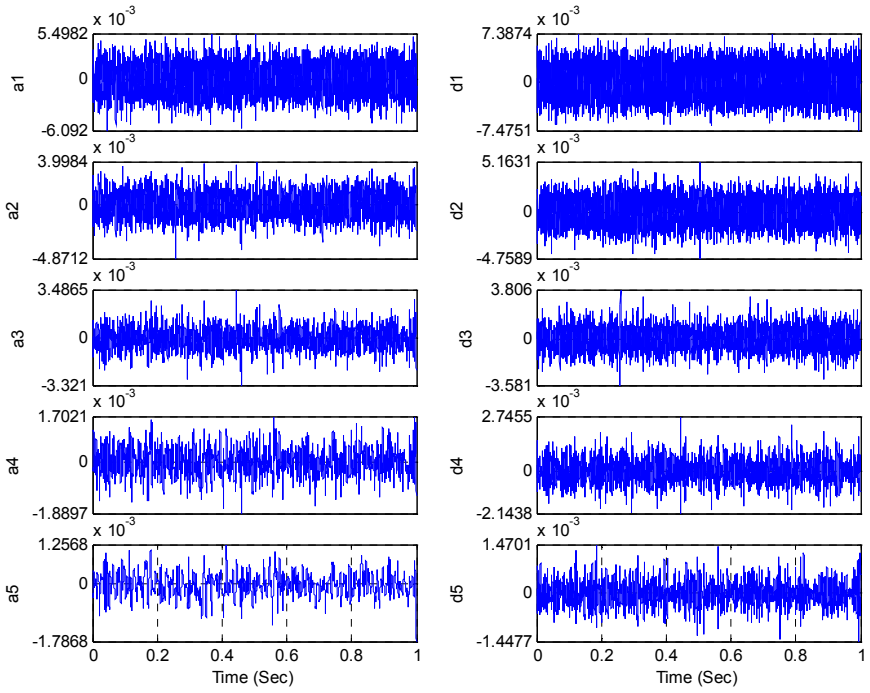


Fig. 3.12: Noise wavelet decompositions using Db4 at level 5

Given the noise model, the selection of an appropriate threshold value can be accomplished using a number of rules. Some of the proposed methods include fixed threshold selection rule, Stein's unbiased risk estimate selection rule and mixture rule [72][65]. According to the fixed threshold selection rule used in this dissertation, for a signal of length N, a universal threshold proportional to $\log N$ is selected. The threshold is calculated by (3-8) where σ is the noise standard deviation that needs to be estimated.

$$Threshold = \sigma\sqrt{2\log N} \tag{3-8}$$

The standard method to calculate the noise standard deviation involves computing the median absolute deviation (MAD) of the coefficients on the finest scale. The detail coefficients cD1 mostly contain noise energy and any small signal coefficients do not affect the median. Thus, it provides a robust estimate of σ. Once MAD is found, the standard deviation can be estimated by (3-9) [73].

$$\sigma = \frac{AMD}{0.6745} \tag{3-9}$$

The importance of the robust estimation of σ addresses two issues. One is to avoid signal edge effects producing pure artifacts when computations are performed on the edges. The second is to exclude level one details of $f(n)$ that are usually concentrated in a few coefficients provided that the function is sufficiently regular [65]. In case of Gaussian white noise, applying this rule along with soft thresholding method yields a noise free signal, sometimes at the expense of shrinkage of the signal features. On the other hand, hard thresholding retains the features better, but results in a wigglier denoised signal [72].

An exhaustive search strategy for best denoising parameters showed that the Haar or Bior1.1 wavelets at level 1 with the hard thresholding method and non-white noise model results in the minimum MSE. However, this was not acceptable since this measure tends to underestimate the noise energy. Thus, a trade-off was made to maintain low MSE and remove all noise components at the same time. The result was that Db4 wavelet at level 5 with level dependant fixed thresholds and hard thresholding method, using non-white noise model performed superior [51].

The thresholding results on a pure noise signal are shown in Fig. 3.13 where scaled, level dependant fixed thresholds were applied to the detail decompositions. As for the thresholding

method, the hard thresholding rule was used. Through the examination of the detail compositions, it is seen that the denoised signal is effectively noise-suppressed and possesses no significant high frequency components. The same thresholding methodology was applied on the noisy signal shown earlier in Fig. 3.8. The wavelet decompositions of this signal are shown in Fig. 3.14, which demonstrates an important advantage of the denoising operation. Similar to Fig. 3.9, spikes are explicitly present in d1 through d3 that shows their contribution to the spike in the original signal. Moreover, detail four is also a contributor, which is effectively revealed by the denoising action. Additionally, detail five that had very small noise energy is set to zero which is preferred from the computation point of view. When extracting signal features from these details, the denoised signal would require no further processing in detail five. Accordingly, the denoising of the signal would effectively eliminate a number of unnecessary arithmetic operations.

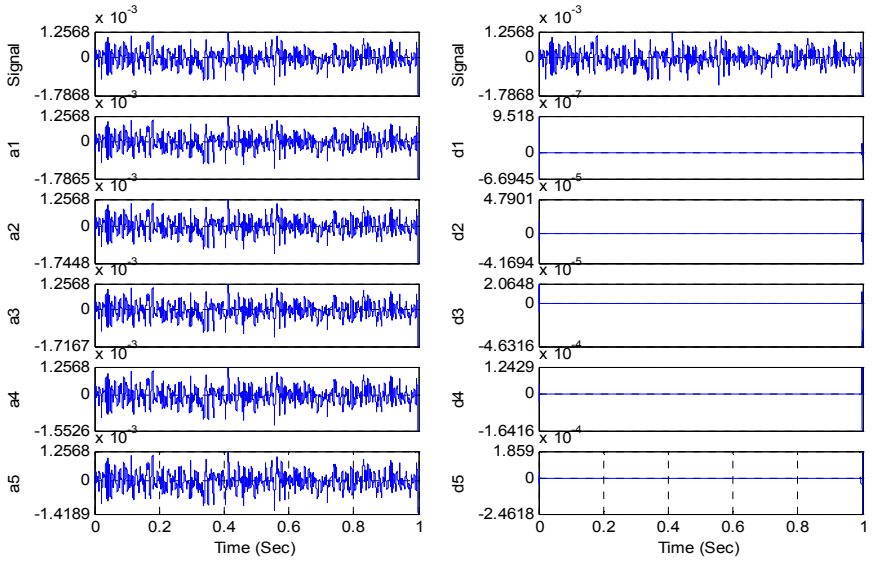


Fig. 3.13: Wavelet decompositions for noise signal after denoising

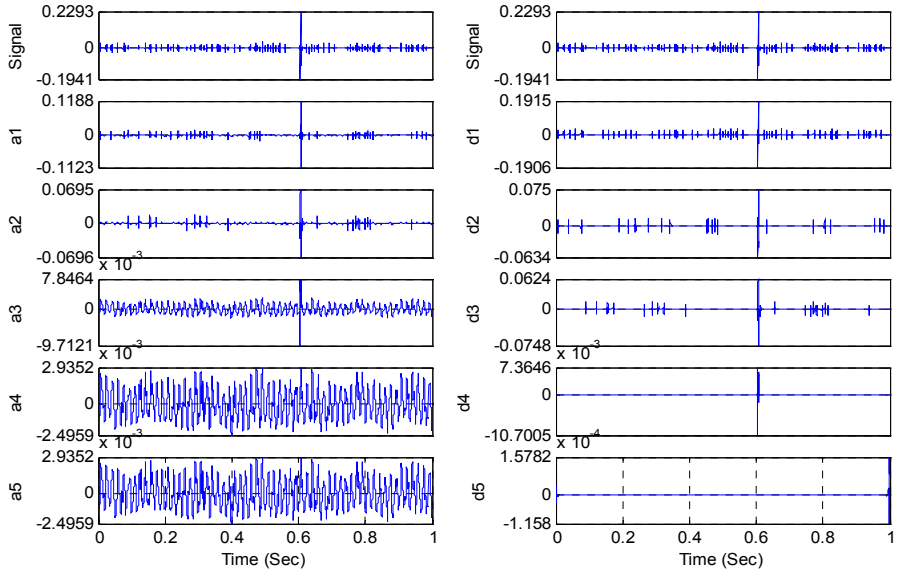


Fig. 3.14: Denoised phase notch high frequency signal in db4 wavelet domain

3.4.4.3 Reconstruction

Once the thresholding stage is completed, the thresholded coefficients are used in the synthesis process to reconstruct the signal in time domain. For the pure noise signal shown in Fig. 3.11, the reconstructed denoised signal is shown in Fig. 3.15. The effectiveness of the denoising operation can be verified in the frequency domain by examining the denoised signal frequency spectrum shown in the same figure. Similar result for the noisy signal is depicted in Fig. 3.16. It can be observed that most of the noise components that, by definition, possessed a small magnitude and high frequencies are removed. Meanwhile, the important features of the signal are preserved for feature extraction and classification purposes. Through the denoising operation, many low magnitude signal samples are set to zero, which in turn reduces the computation complexity on the later analysis procedures.

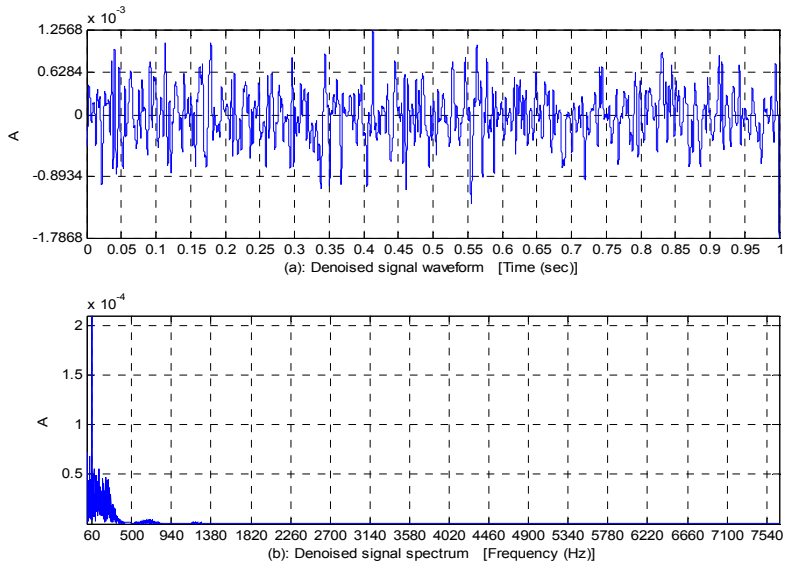


Fig. 3.15: Denoised noise signal and corresponding frequency spectrum

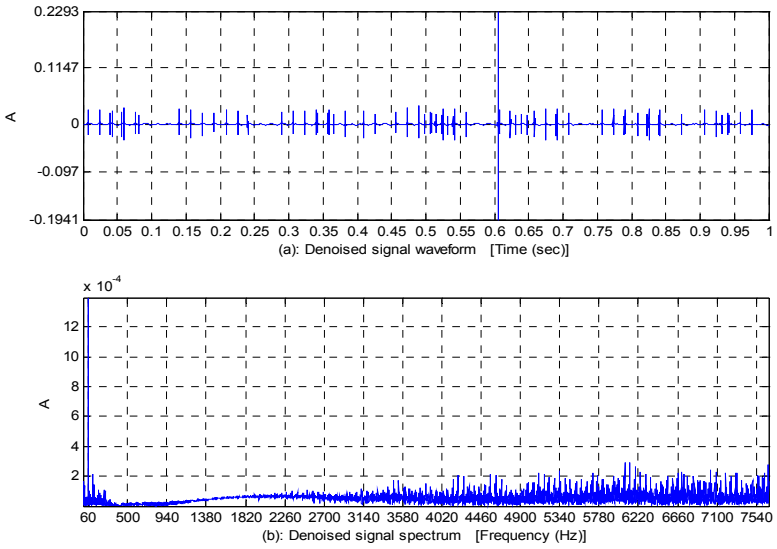


Fig. 3.16: Denoised phase notch high frequency signal and corresponding frequency spectrum

Finally, it is emphasized that in the denoising of incipient abnormalities characterized by spikes in the notch high frequency signal, the choice of the thresholding method is crucial. The spike magnitude and its energy are among important features that should be preserved during denoising operations. In the soft thresholding approach, in addition to setting to zero the elements whose absolute values are lower than the threshold, the other nonzero elements including spike magnitudes are shrunk towards zero. However, in the hard thresholding method all the signal elements whose absolute values are higher than the threshold remain unchanged. Accordingly, the hard thresholding approach was used to preserve important spike features. The soft and hard thresholding performance can be compared using Fig. 3.17. This plot illustrates the resulting denoised signal when the soft and hard thresholding methods were applied, respectively. As can be observed, the spike magnitude that is a discriminating feature among recordings was shrunk during the soft thresholding.

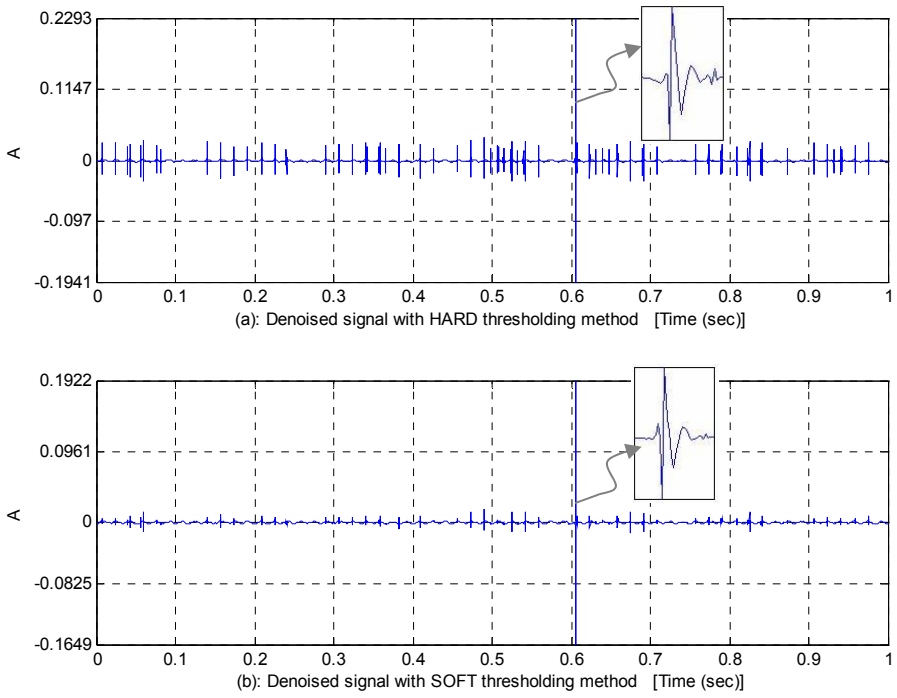


Fig. 3.17 : Comparison between hard and soft thresholding results in denoising

3.5 SUMMARY

In this chapter, the long-term monitoring system used to collect data from an underground distribution lateral was introduced. Also the type and structure of the collected data used in this research work for various design, testing, and characterization purposes were discussed. Emphasizing the necessity of preprocessing tasks, three important preprocessing operations namely, resampling, DC removal and denoising were introduced and the related procedures for each task were discussed. Resampling is performed to eliminate the potential redundancy in the signals by decimating or in general resampling the signals at appropriate factor to yield an effective sampling rate. The removal of redundant signal samples help reduce the computational complexity and thus improve the processing speed. DC removal aims at suppressing measurement DC offset introduced inevitably in field recorded data. It was found that an estimate based on a long signal observation window provided satisfactory results. As an important preprocessing step, the denoising operation was introduced and implemented. The common noise models were discussed and an estimate model was proposed. It was found that Db4 wavelet at level 5, with the level dependant fixed threshold selection rule, and the non-white noise model along with the hard thresholding method performed superior. Denoising results were illustrated by a typical notch high frequency signal. It is concluded that not only does the denoising increase the signal to noise ratio but it also reduce computation complexity, overall processing time, and enhance the performance of successive signal analysis steps by removing unwanted noise components. Once the preprocessing operations are performed, the data become ready for the classification and detection phases.

The next chapter presents the classification method and discusses the various aspects associated with this important functionality of the IFDS.

CHAPTER IV

INCIPIENT FAULT CLASSIFICATION METHODS AND RESULTS

4.1 INTRODUCTION

A pattern consists of a pair of given variables (χ, ω) where χ is the feature vector and ω is the concept behind the observation i.e. the label. Given a set of patterns for a set of observations, the goal of classification is to perform pattern assignments as to which label is the most relevant for a particular observation abstracted by its feature vector. Labels either symbolic or numeric describe the entity of an observation and just like names given to objects convey high level information about the observation. Assigning an observation to a category and determining the label is accomplished by the designed classifier. The type of the classifier used in a particular application is greatly problem-dependant and the right classifier or set of classifiers need to be selected from a pool of available classifiers.

Designing a classifier quite often incorporates information from training observations. Using training samples or observations in the design process of a classifier is called learning. Depend upon the learning method, three broad approaches can be adopted which include supervised, unsupervised, and reinforcement learning. Supervised learning is applied when a library of labeled data is already available to expose to the classifier. Such pattern analysis problems are called supervised in that both the feature vector and the correct answer are provided to the classifier. On the contrary, the unsupervised learning concerns pattern recognition problems in which the feature vector is the only given information and the goal of these methods is to cluster the data and find a model that can best explain the structure of the data and reveal natural groupings of the input patterns. This type of learning is called unsupervised since there is no teacher to provide the right answer. The data modeling principle through the unsupervised learning method constitutes the backbone of the developed detection scheme.

The third learning method is the type of learning with a critic [53]. It involves the process of training a classifier through presenting an input, computing a tentative label, and using the known target label to improve the classifier by providing feedback. Unlike the supervised

learning method, the target label is indirectly used to reinforce the learning and establish right or wrong assessments on the output of the classifier. In this research work, both supervised and unsupervised learning methods have been employed in the IFDS.

Classification process in supervised mode involves a two-step process, which includes prototype selection and classifier design. In prototype selection, a subset of given data is selected to train the classifier. Once a feature selection or extraction procedure finds a proper representation, a classifier can then be designed using a number of possible approaches. The question whether any classifier is better than the other can barely be answered in practice unless we try to answer the question "Does this classifier solve the problem or not?"

There are two types of classification operations in the IFDS. The first classifier belongs to the family of rule-based classifiers in which the prior knowledge about the behavior of patterns is expressed in the form of *if-then* statements. The classifier evaluates the rules and assigns patterns to categories described by the relevant rules. The second set of classifiers used in the IFDS is responsible for more complex relationships among the data that can not be expressed with rules only. Regardless of the type of the classifier, there are two distinct steps to design a classification scheme in terms of pattern analysis formulation: feature extraction and model selection / classifier design.

In the IFDS, the rule-based classifier recognizes the patterns of DC classes, various switching type events including load changes, normal, and potential incipient abnormality categories. The patterns classified to be incipient abnormalities are further fed to a set of supervised classifiers and categorized into three classes based upon severity degrees. SVM, SOM, KNN, and DIS are supervised classifiers that operate in parallel on the incipient abnormality patterns. However, only the SVM decisions are used in the detection phase.

4.2 RULE-BASED CLASSIFICATION

Rule-based classifiers are attractive solutions when classes of data can be characterized by general relationships among entities and there is no need for training samples [53]. The utilized rules, integral to expert systems in artificial intelligence, can be deduced automatically from the data and expressed in a broader sense as a set of crisp or fuzzy *if-then* logical statements. The biggest advantages of rule-based classifiers are that the reasons for their decisions are readily verified and the comprehension and interpretation of the results generated by these classifiers are

straightforward. Despite the relative simplicity, the rule-based classifiers have been shown to perform well on most commonly used datasets [60].

Although logical rules are the core components of the rule-based classifiers, the establishment of these rules for the distribution system data requires prior knowledge about the system events, disturbances and their characteristics. Therefore, in the next sections first a general overview of power system events is given and then the rule-based classifier developed for this application is presented. Specific rules are defined to categorize DC and normal classes, a number of switching related events, and classes of potential incipient abnormalities manifested as spikes in the high frequency signal.

4.2.1 Events, Variations and Disturbances

Events in power systems are referred to the excursions of the recorded voltage or current signals that lie outside the predetermined monitoring equipment thresholds. The term event is typically used to describe significant and sudden deviations of voltage or current from its natural or normal waveform [74]. The terms event and variation might sometimes be used interchangeably. Although both terms describe a deviation in a waveform from its nominal value, events typically convey a significant and abrupt change from normal limits and variation describes small deviations [75]. The classification of events or variations can be performed in terms of their underlying causes or in terms of disturbances [76]. When the cause of an event or variation in a signal is not known, the observed behavior is preferably called a disturbance. In this application, incipient fault-based events and variations are not classified in terms of their causes but there are other events in the system that can be classified according to their source of origination. Specifically, the rules developed for classification purposes involve DC distortion events, various load and switching type events and disturbances, potential incipient abnormalities, and normal classes.

Broadly speaking, events in power systems can be categorized by their causes. They can be introduced by loads, faults, or outside factors. Load induced events can be related to either switching actions or normal operations of nonlinear devices and power electronics. For instance, switching an induction motor might cause a voltage dip and a transient in the current that is a switching event, but harmonic distortion of current waveforms due to the normal function of a typical adjustable speed drive is a non-switching load induced event. While load-induced events are considered expected phenomena and non-destructive, the abnormalities due to faults and

external factors are damaging in nature and aggressively treated by the appropriate protection mechanisms to prevent extensive damages to susceptible power equipment.

One of the common types of events is associated with the steady state deviations from an ideal sinusoidal waveform due to the presence of frequency components other than the fundamental component (60 Hz). This type of long-term deviation is referred to as waveform distortion and can best be characterized by performing a spectrum analysis on the recorded event. Waveform distortion can be a result of the contribution of DC signal, harmonics, interharmonics, noise, or high frequency components of power electronic loads. Each of these factors shapes the distorted signal differently yet the common characteristic is that the distortion is visible in the steady state spectral content of the signal. In Fig. 4.1, distortions introduced in the voltage and current signals are shown. Circles and arrows pinpoint some instances where the distortions are seen conspicuously.

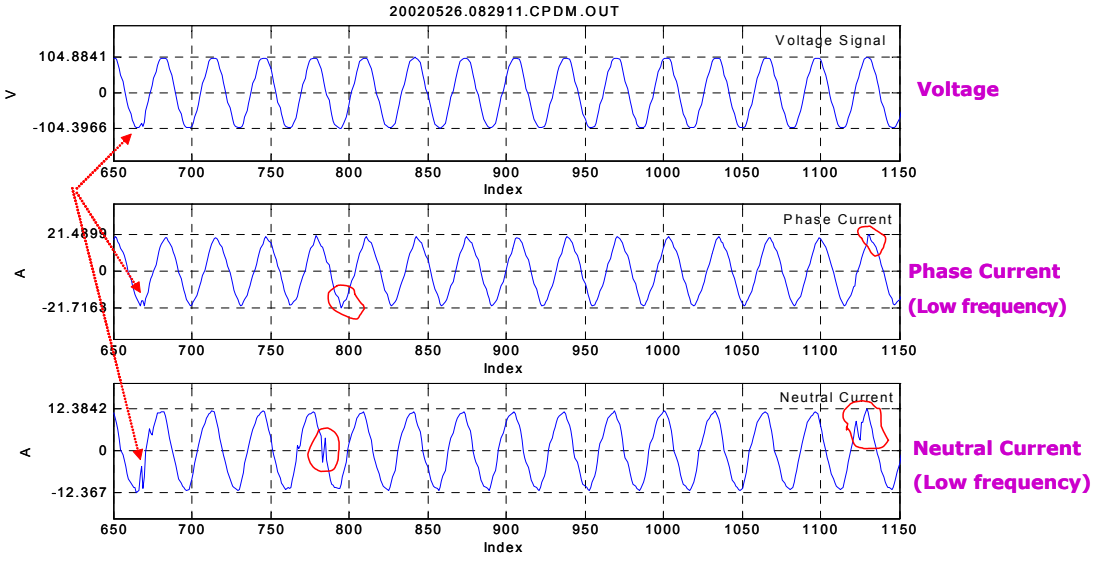


Fig. 4.1: Distortions in voltage and current signals

Another distortion event is depicted in Fig. 4.2 where the low frequency phase current signal is distorted with excessive DC components. The mean signal calculated over one cycle

shows the variation of the DC elements that are increasing over time. Ref. [77] summarizes power system events using the electromagnetic compatibility approach in which the events are categorized in terms of the frequency components, the duration of the phenomena and the typical magnitude.

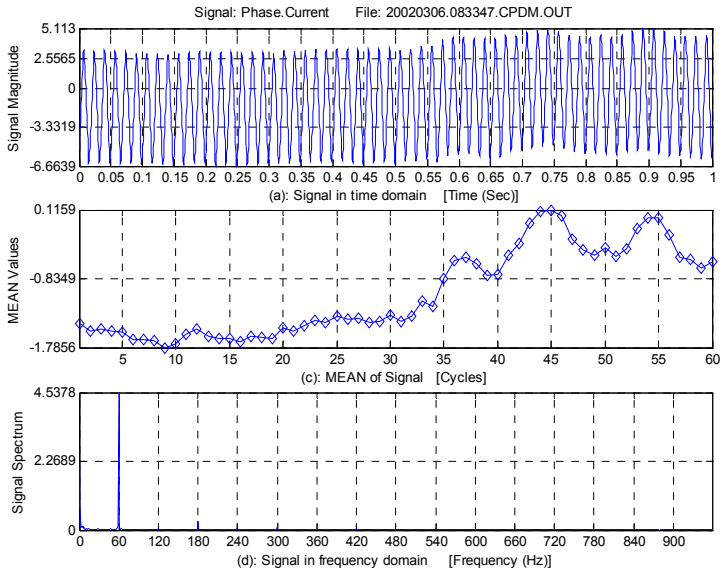


Fig. 4.2: DC distortion in phase current

4.2.2 Overall Flowchart of the Rule-based Classifier

The flowchart of the rule-based classifier is depicted in Fig. 4.3. As an unsupervised, automatic, rule-based classifier, it takes the input signals in and classifies them into a number of designated classes. The classification routine is a series of sequential operations consisting of a number of data processing modules and decision-making blocks. The input signals are consecutively passed through the processing modules, reshaped, modified, and prepared for the decision-making blocks. At every decision node, the classifier utilizes its embedded expertise to classify the capture.

The analyzers comprise three main modules: the DC/trend analyzer, the RMS shape analyzer, and the spike analyzer. In the DC/trend module, classes of data whose DC values or trends are excessively beyond the thresholds are classified as DC class data. This module incorporates a trend analysis routine that takes the low frequency current signal and classifies it into the DC class by comparing the ratio of the signal energy before and after detrending. For non-DC class signal, this ratio is very close to one, but for others it deviates from one.

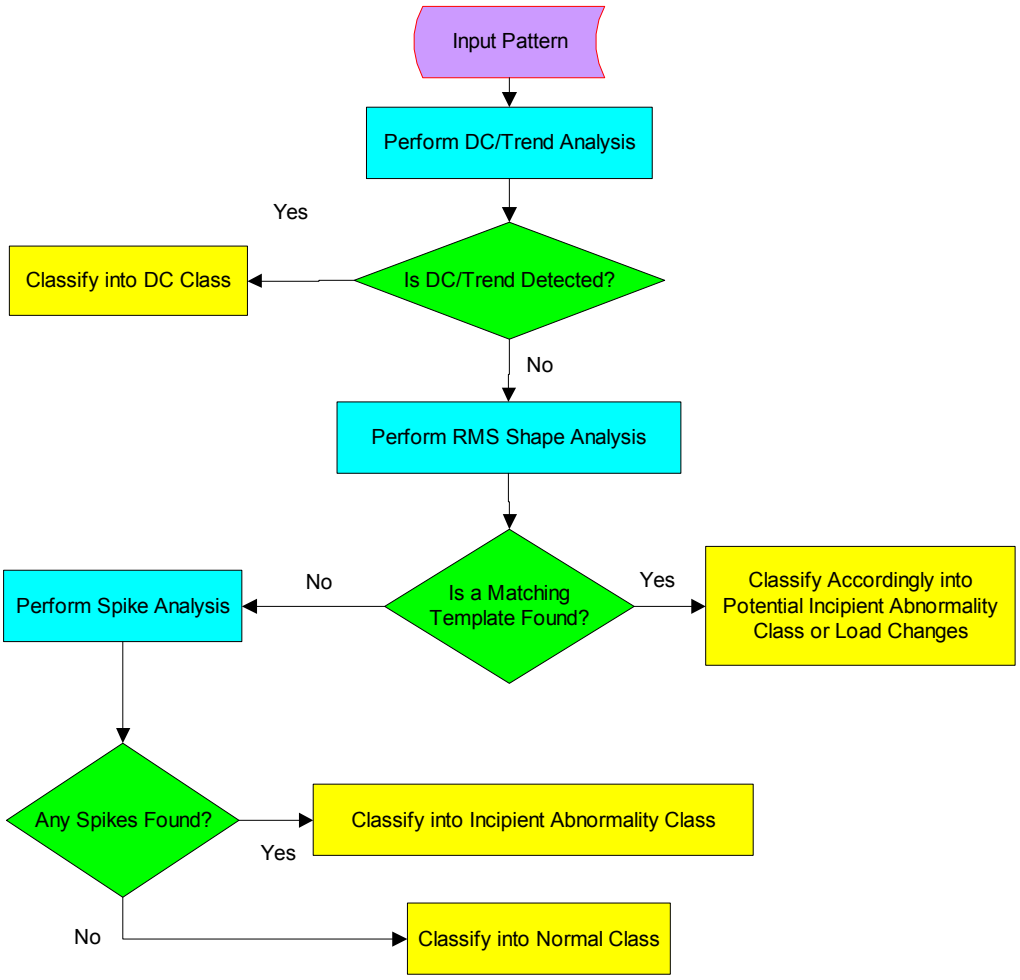


Fig. 4.3: Flowchart of the rule-based classification procedure

The RMS shape analyzer is an extensive processing module, responsible for classification of a great number of captured data. It utilizes an embedded RMS template-matching algorithm to find the best matching template from a given library of templates. Among the identified shapes, four templates capture the data, most likely conveying incipient abnormalities. These patterns are classified into incipient classes and further analyzed through supervised classification. The RMS shape analysis uses the voltage and low frequency current signals to classify the patterns in terms of the shape of the RMS signal calculated over one power cycle for one-second capture.

The spike analysis module operates on the high frequency current signal to capture the incipient abnormalities that are characterized by the spikes and not identified through previous modules. This module uses a detection algorithm based on the theory of outliers to identify the spikes in the entire recording.

This classifier was implemented as a computer code in MATLAB Release 14. Customized functions were composed for various functionalities of the classifier including DC/trend analysis, RMS shape analysis, and spike analysis. These analyses are further discussed next.

4.2.3 Recognition of DC Classes

DC offset is the 0Hz component of signal spectrum that manifests itself in the steady state conditions in the low frequency current signal. Temporary DC components due to faults either disappear in a few cycles as the transient goes away, or are interrupted in a few milliseconds as the protection system clears the fault. However, other factors superimpose a steady state DC offset on the power signals. The well-known factors include geomagnetic disturbance and half-wave rectification [77]. DC components in an ac power system are not welcomed and should be avoided if possible. While the presence of DC offset is rarely evitable, it can potentially elevate the electrical stress level on the insulation, saturate transformers operating near their knee point, and cause other detrimental effects.

As shown in section 3.3.3, the measurement DC component is removed from the original input signals through the preprocessing step by estimating its level over the entire recording. It is observed that in case of a non-DC offset signal, the processed signal contained almost no significant DC components. However, when the original input signal is a true DC offset, trends and low frequency components are still present in the output signal even after the preprocessing task is performed. This characteristic of DC offset signal is used as a guide in the detrending procedure. In this analysis, linear trends and mean values are computed for a preprocessed input

signal. This computation involves the least squares fit of a straight line or composite line for piecewise linear trends to the data. By subtracting the resulting function from the data, the output signal called detrended signal (ds) is obtained. If there are trends and DC offset, the energy of the detrended signal defined in (4-1) differs noticeably from that of the input data. However, for a non-DC offset signal, there are no appreciable trends so that the energy of the data before and after the detrending procedure is approximately preserved. Based on the preceding observations, the classification rule is defined by (4-2).

$$E = \sum_{i=1}^{N} s(i)^2 = \|s\|_2^2 \tag{4-1}$$

$$If \ \frac{E(s)}{E(ds)} > T_{dc} \Rightarrow DC\,Flag = 1 \tag{4-2}$$

where s and ds denote the input signal and detrended signal, respectively. $\| \ \|_2^2$ denotes the squared 2-norm of the signal which is an equivalent term used for the energy. The threshold value (T_{dc}) is 0.9999.

The trend analysis is demonstrated in Fig. 4.4 where the input DC offset current is detrended using the least squares fit of composite lines in each cycle. The detrended signal is shown in (b) and the corresponding trend signal is depicted in (c), respectively. As seen, significant amount of trends and DC components are removed in each cycle and the detrended signal appears trend-free and covariance-stationary. The norm ratio for this particular case is 0.9831.

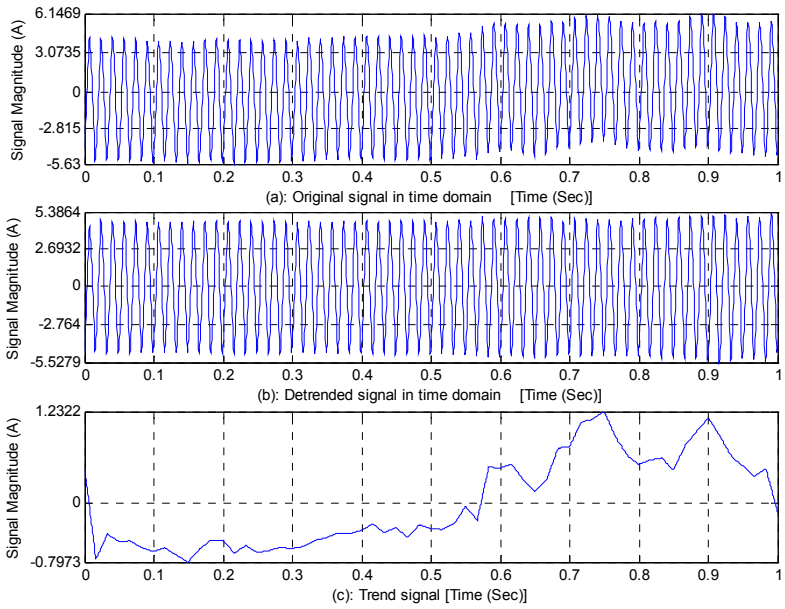


Fig. 4.4: Input signal and corresponding detrended and trend signal

4.2.4 Recognition of Switching Disturbances and Events

The method used to classify switching disturbances and various events is based on the RMS signal shape analysis of the voltage and phase current signals. In disturbance classification, we are mainly concerned with classification of the recorded measurements in terms of the identified disturbances and not in terms of the underlying events. Some observed power system events are neither known nor well characterized. On the other hand, there are disturbances that cannot be associated with unique events.

4.2.4.1 RMS Shape Analysis Method

In this method, an exact match is searched for in the library of ten already-defined templates. If the match is found, the corresponding signal is classified into the class represented

by that template. If not, the signal is passed through the successive classification modules. The RMS shape analysis for a given input signal involves five steps as follows:

- Calculate RMS signal over non-overlapping, fixed windows
- Smooth and normalize RMS signal to improve the shape resolution if necessary
- Detect change points within the snapshot (60 cycles)
- Find the matching template
- Classify the signal with respect to the matched template

4.2.4.1.1 Calculation of RMS Values

For a digital signal s of length N, the RMS values is calculated by (4-3) where $s(i)$ denotes the i^{th} sample of the signal. Typically, the RMS values can be calculated over a window of two cycles, one cycle or half a cycle resulting in a series of numbers termed as the RMS signal. As in many signal processing applications, the selection of the window size is a time/frequency resolution trade-off in which the shorter the window size, the better the ability to track the fast changes. The RMS signal obtained using the one-cycle window does not significantly differ from the calculated RMS value using a one-cycle window in Short Time Fourier Transform (STFT) analysis. Technically, the difference is zero for an ideal sinusoidal waveform with no harmonics [74]. In this analysis, one-cycle non-overlapping windows were chosen.

$$RMS = \left(\frac{\sum_{i=1}^{N} s(i)^2}{N} \right)^{\frac{1}{2}} \tag{4-3}$$

4.2.4.1.2 Template Matching

As mentioned, the classifier looks for an exact match in the library of ten already-defined templates to accomplish its task. These templates are depicted in Fig. 4.5. It is assumed that each snapshot is taken when a disturbance takes place. If there is more than one disturbance within a recording, a segmentation procedure is required so that each represents a single disturbance. It is important to point out that the real RMS signal calculated over one cycle deforms from these ideal cases. Nevertheless, the essence of the signal shape may resemble one of the ideal prototypes.

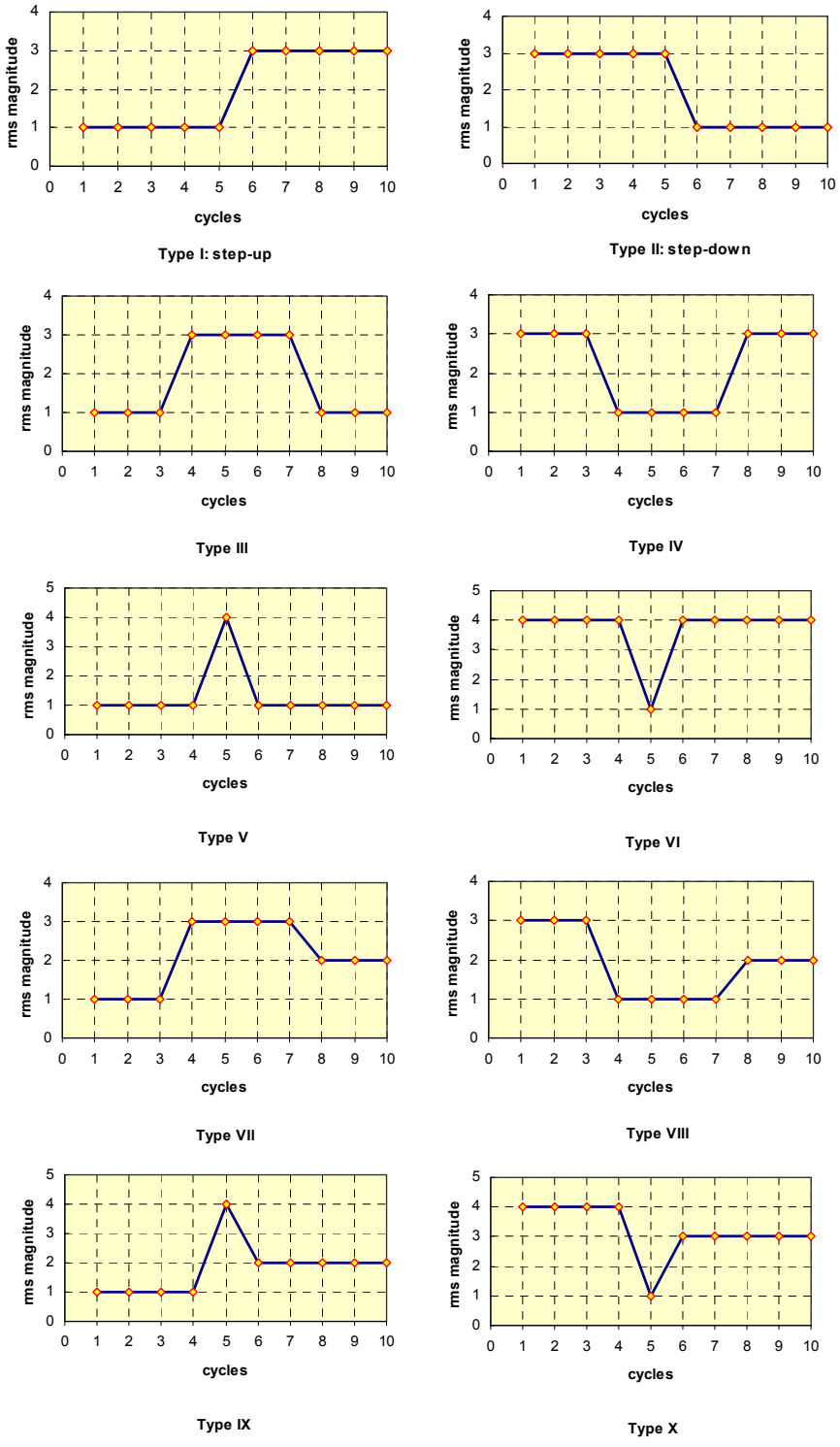


Fig. 4.5: Prototypes of ideal RMS shapes

As mentioned earlier, ten individual templates are considered in the RMS shape analysis. Out of the ten templates, types III through VI are flagged as potential incipient abnormality patterns and the corresponding signals are used in the supervised classification stage. The other templates namely I, II, and VII through X represent load changes and/or non-incipient patterns. Therefore, no further classification is performed on patterns matching these templates. If no match is found, the input signal is passed through the subsequent spike analysis module for a check on the presence of spikes. From the field recorded data, a few examples of RMS signals matching prototypes I, VII, and II are presented in Fig. 4.6.

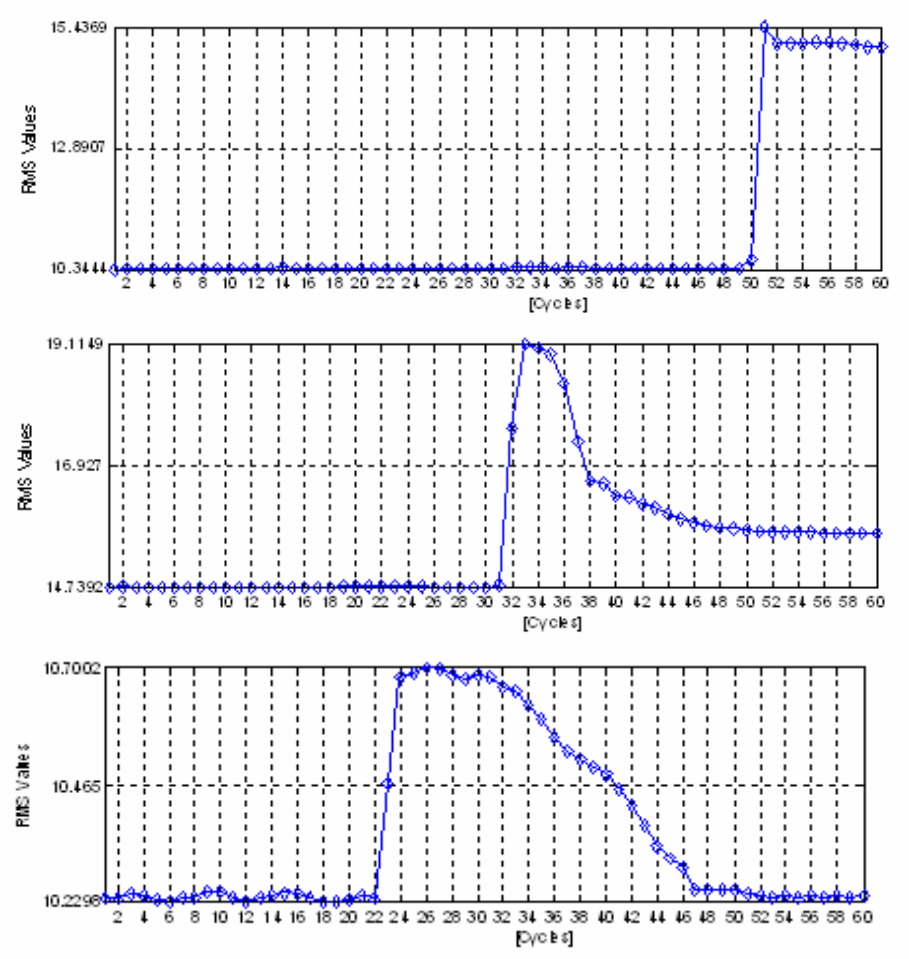


Fig. 4.6: Examples of RMS signals recognized by the rule-based classifier

The features that provide the necessary and sufficient information to recognize the prototype shapes include: Number of change points (NCP), Most Frequent Number of Horizontal Crossing Points (MFNHCP), Direction of Change (DC), Disturbance Duration (DD), and Ratio of the average RMS values before and after the disturbance (RAV). These features are listed in Table 4-1, and the corresponding values are assigned for each type of the prototypes.

The change point, expressed in units of cycles, is defined as the minimum of the two successive cycle indices, where a meaningful change is observed. The meaningful change occurs when a non-zero difference between two successive RMS values is obtained. Practically, it is indicated whenever the difference between two successive RMS values becomes greater than the specified threshold percentage change. For a normalized RMS signal, the plausible choice for the threshold is 0.5. For example, in prototype I, a meaningful change takes place between the fifth and the sixth cycle; thus, by definition, the change point is five. In case III, however, there are two change points, which include the 3^{rd} and 7^{th} cycles, respectively.

Table 4-1:
Characteristics of ideal prototypes

Type No.	NCP	MFNHCP	DC	DD	RAV
I	1	1	I	< 1 cycle	> 1
II	1	1	D	< 1 cycle	< 1
III	2	2	ID	> 1 cycle	1
IV	2	2	DI	> 1 cycle	1
V	2	2	ID	1 cycle	1
VI	2	2	DI	1 cycle	1
VII	2	2 $\vee$ 1	ID	> 1 cycle	>1
IIX	2	2 $\vee$ 1	DI	> 1 cycle	<1
IX	2	2 $\vee$ 1	ID	1 cycle	>1
X	2	2 $\vee$ 1	DI	1 cycle	<1

The most frequent number of the horizontal crossing points is the most informative feature that can be obtained using the histogram analysis of the number of horizontal crossing points. This feature for prototype III, for example, equals two, and for prototype II, it amounts to one. The direction of change is determined by comparing the two RMS values at the change point. The increasing or decreasing directions are coded by I and D, respectively, in the table. When there is more than one change point, a combined code shown in column four of the table is used.

4.2.5 Recognition of Spikes

In the context of the present work, spikes correspond to the signal values that occur over a very short duration and are associated with large gradient values. Usually, these components are observed in the notch high frequency current signal. From a statistical point of view, a spike is defined to be a data point whose first difference is an outlier value. An outlier is a data point, which deems to be inconsistent with the remaining data in a signal [78]. Outliers are determined statistically from the signal. The presence of spikes in the notch high frequency signal is a symptom of swift changes in the signal components. Previous studies suggest that incipient abnormalities manifested in the notch high frequency signal can be characterized by the spikes [4][5]. Some load switching actions can also introduce spikes in these signals at the time of switching. Fig. 4.7 shows an example in which spikes with different duration and magnitude are infested in the high frequency signal. The time domain characteristics associated with each spike are demonstrated on the same figure. Spikes are characterized by the duration, magnitude, and peak value of the elements comprising them. For detection purposes, the positive and negative elements are treated equally. The energy of a spike is determined by the magnitude of its elements as defined in (4-4), where x_i denotes the i^{th} element of a N -element spike.

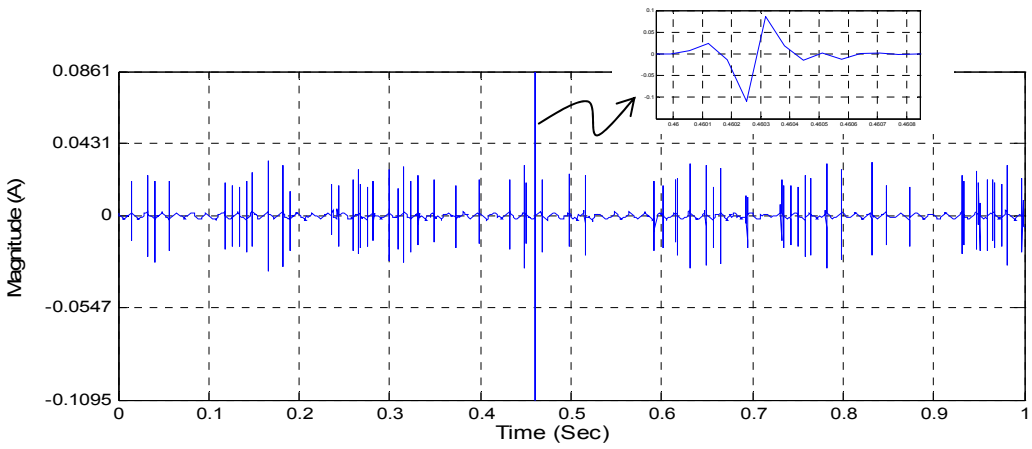
$$E_s = \sum_{i=1}^{N} x_i^2 \tag{4-4}$$

The analysis of spikes manifested in time series has been an interesting problem for many engineering disciplines such as biomedical and electrical engineering. Spike detection in electroencephalograms (EEG) plays an important rule in the diagnosis of epilepsy [79]. Many researchers in this field have discussed the problem from different points of view specific to that application [80][81][82][83]. In the realm of signal processing, the detection of spikes is important because such algorithms generally produce biased outputs in the presence of spikes [78][84]. In analyzing underground power system incipient failures, the detection and characterization of incipient abnormalities based on the spikes are the motivations to perform spike analysis.

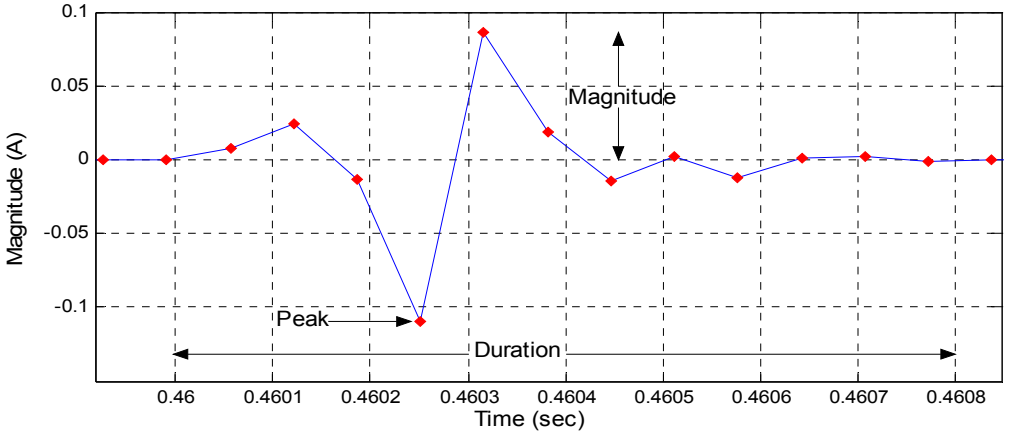
Usually a spike detection algorithm involves two steps namely, spike improvement and detection. In the spike improvement stage, which is highly recommended for spikes of small magnitudes, the visibility and detectability of a spike within a signal is augmented such that it

becomes more detectable and inconsistent with the remaining data. In the detection step, the spike features are used to identify it. Broadly speaking, the spike detection algorithms can be divided into three main categories in terms of the strategy utilized in the first step as follows.

- Time domain parametric methods
- Transform domain parametric techniques
- Non-parametric methods



(a) Macroscopic view of spikes



(b) Microscopic view of a spike

Fig. 4.7: Illustration of spikes and related terminologies

Time domain parametric methods enhance the distinct characteristics of spike samples using a time-domain approach. Gradient [85], matched filter [86], high pass filter and median filtering [87] based algorithms are examples that fall into this category. In the transform domain parametric approaches, the signal is first transformed into other domains, and then the detection algorithm is applied on the transformed signal, which mainly contains the high frequency components of the original signal. The Wavelet and discrete cosine transforms (DCT) [88] are suitable applications for such purposes. Finally, non-parametric techniques view the spike detection problem as a linear predictive modeling of the data in which the prediction residuals are used as indications of spike locations. When a linear model such as autoregressive or moving average is fit to the data, the prediction residuals typically possess large values at the spike locations. A comparison study on the above spike detection techniques indicate that the methods based on the high-pass filter, wavelets, and DCT outperform the other techniques for small spikes in data [84].

4.2.5.1 Spike Detection Method

The spike detection algorithm deployed in the IFDS involves two steps: spike improvement and detection. In the spike improvement stage, the visibility and detectability of a spike within the signal is enhanced so that it becomes detectable and inconsistent with the remaining data. This stage is critical for spikes of small magnitudes. In the detection step, the spike features are used to identify it. The detection algorithm uses a parametric method to enhance the distinct characteristics of spike samples in the time domain. Fig. 4.8 presents the overall flowchart of the spike detection algorithm. The spike-no-spike decision criterion is a multi-term binary statement defined based on the skewness, kurtosis, and crest factor measures of the high frequency signal.

The skewness of a distribution is a measure of asymmetry around the mean. For any symmetrically distributed data, the skewness is zero; otherwise, it is a non-zero value. If the distribution of data has a longer tail out to the right of the mean, the skewness becomes positive; if the distribution has a longer tail to the left of the mean, the skewness takes negative numbers [73]. Presence of spikes in the signal distorts the symmetry of data points; thus, large deviation of skewness from zero is an indication of potential spikes. The skewness of a distribution is measured by (4-5). s is a sample from the distribution or the high frequency signal in this

application. E, μ, and σ denote the expected value, mean, and standard deviation of the signal, respectively.

$$skewness = \frac{E(s-\mu)^3}{\sigma^3} \tag{4-5}$$

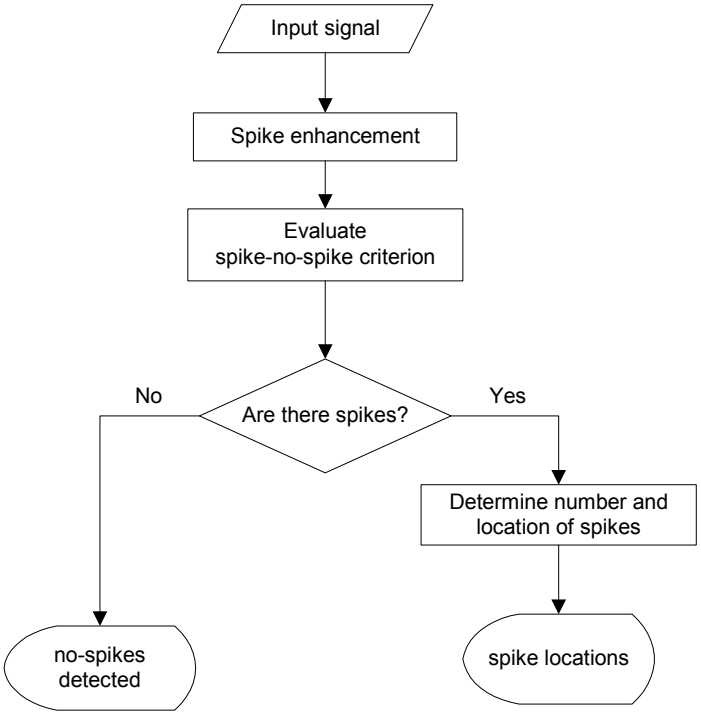


Fig. 4.8: Flowchart of spike detection algorithm

The kurtosis of a distribution is a measure of presence of outliers. A high value of kurtosis implies a sharply humped or peaked distribution, and vice versa. Flat-topped distributions maintain a lower kurtosis value. For normally distributed data, the kurtosis is 3, but it takes large values if the sample contains inconsistent data, i.e., spikes. Thus, if the calculated kurtosis for the signal differs significantly from 3, the signal is highly likely to contain spikes. The mathematical expression for kurtosis is given in (4-6).

$$kurtosis = \frac{E(s-\mu)^4}{\sigma^4} \tag{4-6}$$

It should be noted that kurtosis is not necessarily a measure of peakedness. As pointed out in [73][89], a flat-topped distribution might have infinite kurtosis, and a sharp-peaked distribution may possess a low value of kurtosis. Thus, this measure alone is not sufficient to be used in the process of making a spike-no-spike decision.

The crest factor of sequence is another measure used to indicate potential presence of spikes. This measure is mathematically defined in (4-7), where $s_{\max}$ and s_{rms} denote the maximum and RMS values of the signal, respectively. When the crest factor is significantly greater than an empirical threshold, the likelihood of spikes is higher.

$$crest\ factor = \frac{s_{\max}}{s_{rms}} \tag{4-7}$$

Generally speaking, the selection of thresholds employed in the detection algorithm is a compromise between probability of failure and false indication. Failure is realized when there is a spike in the signal, which is not being detected by the algorithm. False indication happens when there is no spike in the signal by visual examination, but the spike-no-spike decision leads to the yes-there-is outcome. The solution to this type of problem is provided by introducing a cost function. Intuitively, the cost of a spiky signal being flagged as a spike-free signal is significantly higher than that of falsely indicating a spike-infested signal. Formally stated, the thresholds are set to be inclined more to the false positive decision region than the other. Hence, the flag is raised indicating potential spikes, but there may not be spikes in the signal as examined by an expert.

In this section, the rules used by the rule-based classifier were presented and discussed. An important task that was implicitly accomplished in the rule-based classifier is the computation of feature values. As noticed, rules were defined by the variables or features and hence the evaluation of each rule required that the corresponding feature values be known. Having said that, the set of informative features for the supervised classifiers is determined through the feature extraction process, discussed next.

4.3 FEATURE EXTRACTION FOR SUPERVISED CLASSIFICATION

The goal of feature extraction is to obtain a set of sufficient and necessary features that discriminate classes with a high degree of accuracy. This important procedure is a key step for the success of any classifier. The feature extraction process involves two steps: feature determination and dimensionality reduction.

First, a number of raw features are determined. Second, the raw features are projected onto a lower dimensional space by means of multivariate statistical techniques in order to reduce the dimensionality while preserving a good classification rate [52][53]. Raw features of interest can be categorized into two main groups which include time-domain and time-scale or rather wavelet domain features. While in most applications only one type of features is used, classification of incipient abnormality data requires features from time domain as well as wavelet domain.

To mitigate the redundancy and eliminate potential correlations among wavelet features which are inherent characteristics of this type of features, three dimensionality reduction techniques were employed. These techniques include Principal Component Analysis, Linear Discriminant Analysis, and feature subset selection techniques.

4.3.1 Feature Determination

Representing a signal with its features is essentially a way of encoding the signal into a number of numeric or symbolic values whose combination models signal characteristics in an effective way. For example, a time series composed of a sinusoidal 60 HZ waveform can be encoded by just two features, namely amplitude and initial phase angle. Similarly, the goal of feature determination for an incipient fault signal is to find a number of few features that can model the signal exclusively. Although there might be a linear or nonlinear relationship among the raw features - as is the case for features from the wavelet domain- this step is not concerned with such dependencies or correlations but to include all relevant features. The next step i.e. dimensionality reduction determines a subset of a few informative features from the set of initial features.

4.3.1.1 Features from Wavelet Packet Analysis

The first class of features utilized in the IFDS includes normalized energy features from wavelet packet decompositions of the notch high frequency current signal. It has been shown that wavelet analysis performs very well when working with essentially non-stationary power signals [54][55][56]. Since, the incipient fault high frequency data are non-stationary and transient in nature, utilizing powerful analysis techniques such as wavelets would be more descriptive than traditional Fourier analysis methods. To extend the capabilities of the wavelet transform, the wavelet packet analysis were used.

The wavelet packet analysis is an extension to DWT, which provides a richer range of possibilities for the analysis of the signal. In every level of DWT, the signal is split into a detail and approximation and in the subsequent scales, the approximation is further decomposed into another approximation and detail signal. For an n level decomposition, there are $n+1$ possible ways to encode the signal. However, in wavelet packet decompositions, the details as well as approximations are split yielding more than $2^{2^{n-1}}$ different ways to decompose the signal [65]. The result is the wavelet packet decomposition tree shown in shown in Fig. 4.9.

From previous studies, it was determined that Db4 wavelets provide sufficient frequency resolution and satisfactory decomposition results for the notch high frequency signal. Fig. 4.10 and Fig. 4.11 show an example signal along with its 16 decompositions at the fourth level. The packet numbers are printed along the Y axis on each plot.

The wavelet packet analysis was conducted on a sample length of one-second at a time. To get a better frequency range in the 4^{th} level, a symmetrization technique was used to achieve dyadic sample length of 16384. Considering this modification, the frequency range for each of the 4^{th} level details is depicted in Table 4-2. From this table, it can be observed that the harmonic content of the fundamental frequency is associated with a single and unique detail and the frequency range of each level 4 details neatly covers a frequency band of 0.5 KHz, facilitating the interpretation of extracted features in terms of the covered frequency range.

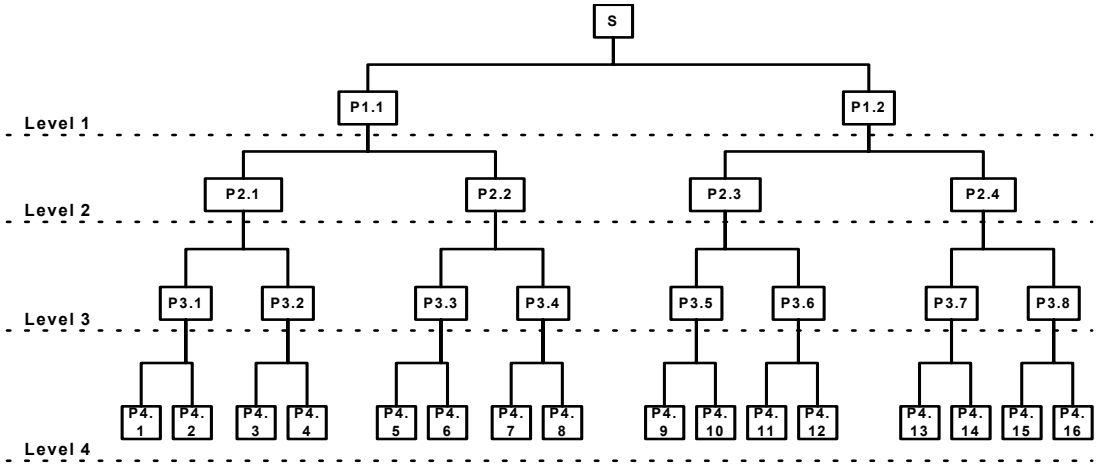


Fig. 4.9: Wavelet packet binary tree

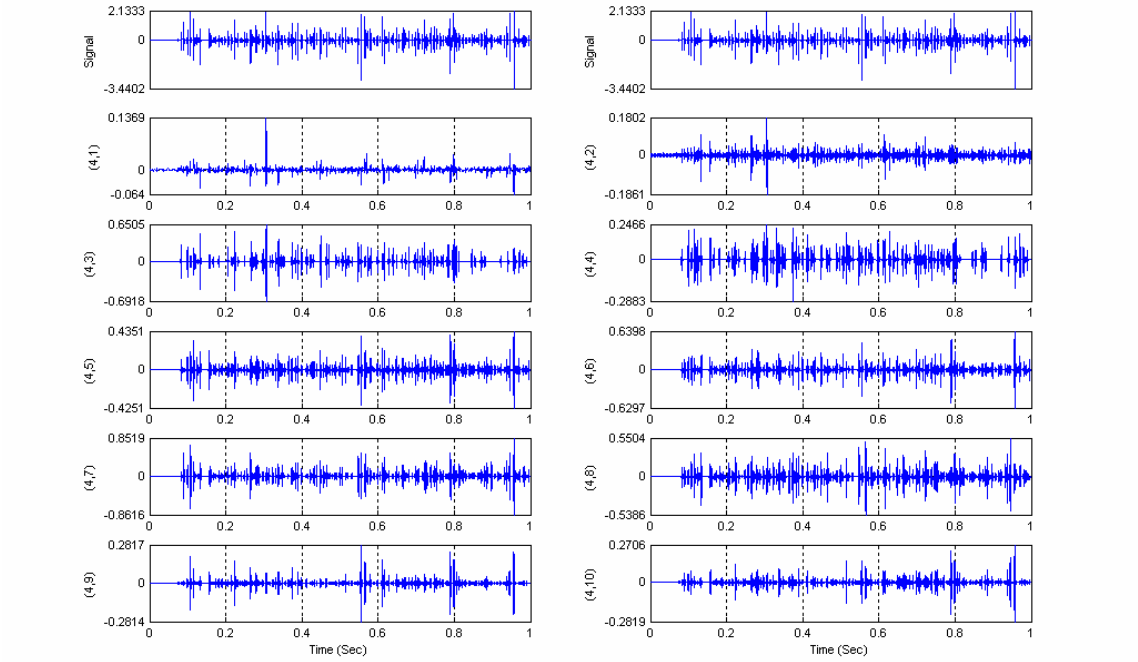


Fig. 4.10: Wavelet packet decompositions of a notch high frequency signal (1)

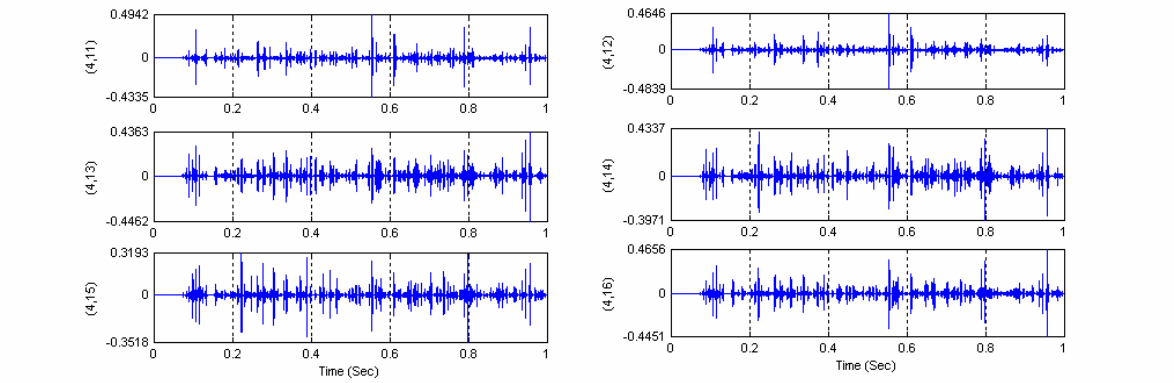


Fig. 4.11: Wavelet packet decompositions of a notch high frequency signal (2)

Table 4-2:
Frequency ranges for the level 4 decompositions

Packet No.	Frequency Range KHz (approx.)
1	0 – 0.5
2	0.5 – 1
3	1 – 1.5
4	1.5 – 2
5	2 – 2.5
6	2.5 – 3
7	3 – 3.5
8	3.5 – 4.0
9	4.0 – 4.5
10	4.5 – 5
11	5 – 5.5
12	5.5 – 6
13	6 – 6.5
14	6.5 – 7
15	7 – 7.5
16	7.5 – 8

There are a number ways to extract features from wavelet transform among which the average energy and standard deviation features are widely used in various applications [90][91][92]. The energy features consists of 16 values computed from the signal decompositions at the 4th level. (4-8) provides an expression for the normalized energy of the i^{th} packet for a signal of length N.

$$F_i^P = \frac{E_i^P}{E_s} \tag{4-8}$$

where $E_s = \sum_{i=1}^{N} x_i^2$ is the total energy of the signal and E_i^P is the energy of the i^{th} packet. The energy of a signal of length N is determined by the magnitude of its elements x_i's. Using the above formula, the corresponding features were computed and plotted in Fig. 4.12. The input signals are decomposed using 'Db4' wavelets at level 4 that result in 16 coefficients [52]. Features are defined to be the percentage energy of each coefficient with respect to the energy of the original signal. As seen, the 7th packet of the signal possesses the largest portion of the signal energy – about 18%- which corresponds to the 3-3.5 KHz frequency range. Notice that the sum of all energy values equals 100%.

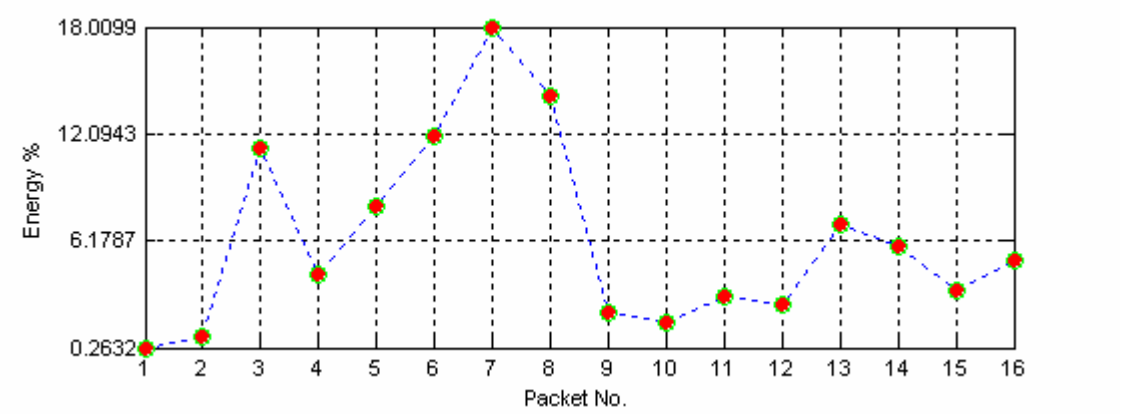


Fig. 4.12: Normalized energy features

4.3.1.2 Features from Time Domain Spike Analysis: VSI and HSI

In addition to the wavelet energy features, two new features from the time domain spike analysis were defined and computed for incipient data. The new features are termed Vertical Severity Index (VSI) and Horizontal Severity Index (HSI) and computed for the notch high frequency signal. VSI and HSI measure the severity of spikes in the signal from two dimensions. VSI takes into account the Y dimension i.e. the magnitude of each individual spike and HSI addresses the severity along the X axis by considering the time between successive spikes in the signal.

The *Vertical Severity Index* of a signal is defined as the normalized energy of spikes with respect to the total energy of the signal. This is mathematically expressed in (4-9) where n is the number of spikes detected in the signal, E_{s_i} denotes the energy of the i^{th} spike, and E_s is the energy of the signal including spike components.

$$VSI = \frac{\sum_{i=1}^{n} E_{s_i}}{E_s} \tag{4-9}$$

The definition of the energy of a spike is given in (4-4). It can be seen that $0 \le VSI < 1$. In other words, if no spikes are present, the vertical severity index is zero. The maximum value of VSI, on the other hand, is always less than one. According to this definition, the contribution of the spike energy to the total energy of the signal is the measure of severity with respect to the spike magnitudes. The name "vertical" implies the fact that the spike magnitudes are taken into account in determining this type of severity measure. One can see that the higher the magnitude of the spikes, the higher the VSI and the higher the perceived severity. As mentioned, VSI features are computed from the notch high frequency signal.

Another time domain feature used in the supervised classifiers is called *Horizontal Severity Index* (HSI). This index takes into account the times at which spikes take place in the one-second high frequency signal. HSI measures the frequency of occurrence of spikes and their local arrival times within the signal. Intuitively, as the number of spikes increases, the interarrival times decreases which in turn implies a frequent arcing activity within the signal. HSI quantifies the severity of local arcing activities using a mathematical expression defined in (4-10).

$$HSI = \frac{\frac{\sum_{i=1}^{n} t_i}{n}}{t_m \times \sqrt{\frac{1}{12 \times n}}} \tag{4-10}$$

In this formula, $t_1, t_2, ..., t_n$ denotes the local arrival times of n spikes detected in the high frequency signal. Notice that all t_i s and the respective HSI are non-negative real numbers. The theoretical boundary for HSI is the range of [0, 214.66]; however, the empirical upper bound is observed to stay below 45.0.

A typical behavior of this index using simulated local arrival times is depicted in Fig. 4.13. Three fundamental scenarios were considered which include decreasing, constant, and increasing rate of local arrival times. Each column corresponds to one scenario and the HSI under each scenario is shown inside parentheses on top of the first row of each column. Each graph in the first row shows the assumed arrival times under each scenario. The corresponding interarrival times i.e. the times between successive arrival times are plotted in the second row and finally the last row graphs represent the partial HSI values calculated under each scenario at each arrival time. The final HSI value amounts to the last partial HSI value.

As seen, if the spikes happen to be concentrated toward the end of the interval (column three), the severity index is higher than the case in which they are distributed at the beginning of the interval (column two) or uniformly across the interval (column one). Fig. 4.14 shows an example signal for which the HSI amounts to 15.83.

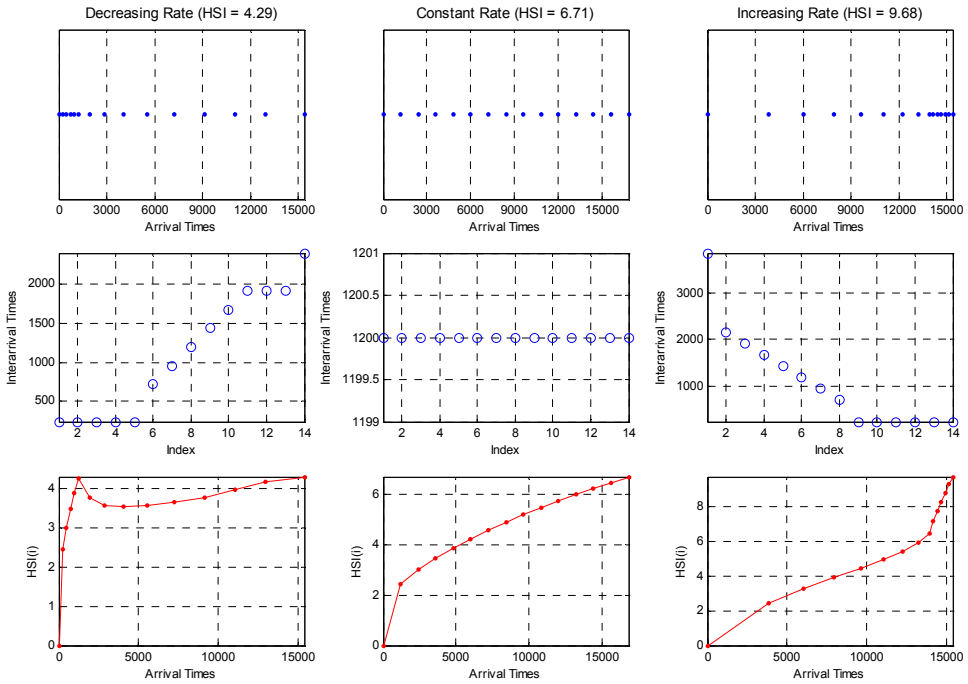


Fig. 4.13: Response of the horizontal severity index under decreasing, constant, and increasing rate of local arrival times

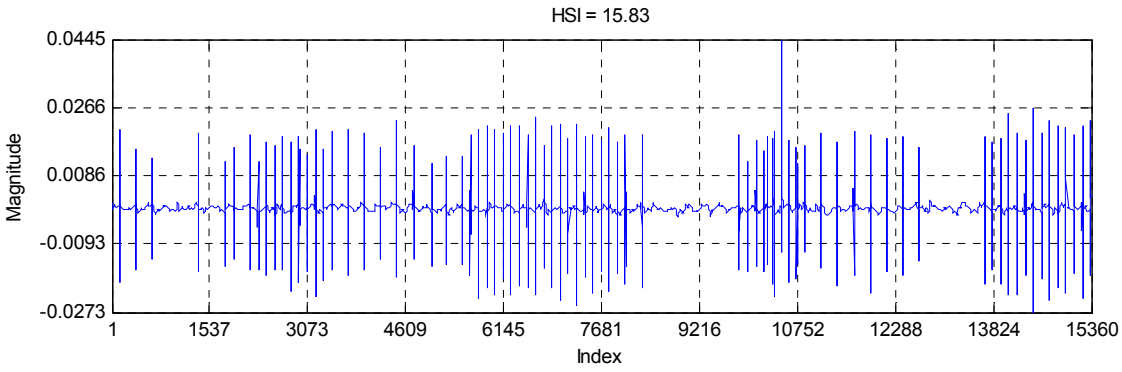


Fig. 4.14: HSI corresponding to a field recorded notch high frequency signal

4.3.2 Dimensionality Reduction

The objective of dimensionality reduction is to keep the dimensionality of the pattern recognition problem (i.e. the number of features) as small as possible while preserving good classification accuracy. Dimensionality reduction can be accomplished by means of feature selection or feature extraction. The term feature selection refers to techniques that select the best subset of the input features set. Methods that create new features based on transformations and combinations of the original feature set are called feature extraction methods. The choice between feature selection and extraction depends on the application domain [52]. Feature extraction can be accomplished using a number of methods. The utilized techniques are Principal Component Analysis (PCA) and Linear Discriminante Analysis (LDA), which are based on the transformation of the original feature matrix.

4.3.2.1 Transformation Techniques

Principal Component Analysis or PCA is a linear unsupervised feature extraction method [57]. The linear transformation is defined by the eigenvectors of the covariance matrix, which leads to vectors that are uncorrelated regardless of the form of the distribution. If the distribution happens to be Gaussian, then the transformed vectors will be statistically independent. The objective of PCA is to perform dimensionality reduction while preserving as much as the randomness (variance) in the high-dimensional space as possible. PCA performs a coordinate rotation that aligns the transformed axes with the directions of maximum variance. The main limitation, however, is that as an unsupervised method, it does not consider class separability information. There is no guarantee that the direction of maximum variance will contain good features for discrimination. This is graphically illustrated in Fig. 4.15. There are two classes of data represented by ones and twos in the original two-dimensional space along x and y axes. PCA finds new features along X and Y axes. If the data points are projected onto the new X axis, the data can be represented with a reduced number of features i.e. only one feature and successful classification is achievable; however, projection onto the direction of the maximum variance i.e. Y axis results in losing the discriminatory information and hence poor classification accuracy.

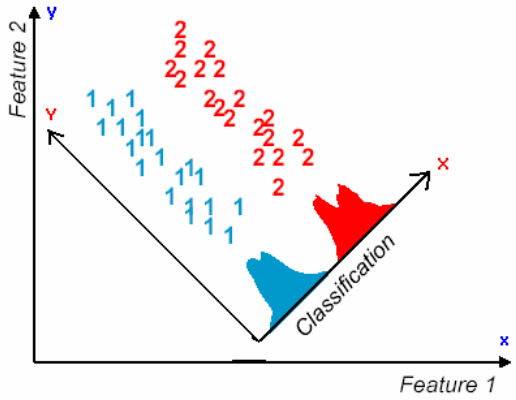


Fig. 4.15: Illustration of PCA [59]

Linear Discriminante Analysis (LDA) is another linear feature extraction method, but unlike PCA it is supervised [57][58]. The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. In LDA, interclass separation is measured by Fisher criterion, which finds the eigenvalues of the between-class scatter matrix to the within-class scatter matrix. The within-class scatter matrix is defined by:

$$S_w = \sum_{i=1}^{c} S_i \tag{4-11}$$

where,

$$S_i = \sum_{x \in w_i} (x - \mu_i)(x - \mu_i)^T \tag{4-12}$$

x denotes the data, c is the number of classes and μ_i is the mean vector of class w_i. The between-class scatter is defined by (4-13) where N_i is the number of patterns of class i, and μ is the mean of the entire distribution. Fig. 4.16 demonstrates the within-class and between-class scatters defined for a three-class problem.

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{4-13}$$

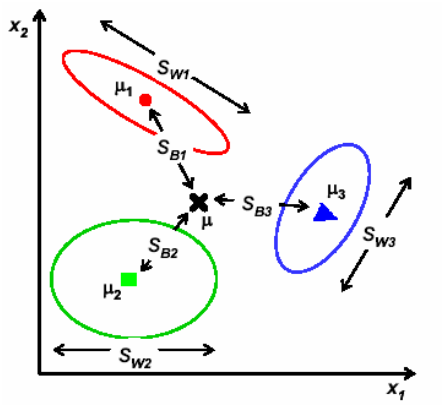


Fig. 4.16: LDA within-class and between-class scatter representation [59]

The solution proposed by Fisher is to maximize the function that represents the difference between the means of the classes (between-class scatter) normalized by a measure of the within-class scatter. The projections with maximum class-separability are the eigenvectors corresponding to the largest eigenvalues of $S_w^{-1}S_B$. This method produces as many projections as the number of classes minus one. If the classification error estimates establish that more features are needed, some other methods must be employed to provide additional features. LDA will fail when the discriminatory information is not in the mean of the data but rather in the variance.

4.3.2.2 Selection Techniques

Feature subset selection was accomplished using Sequential Forward Selection search strategy with wrappers objective functions. Unlike feature extraction techniques where the final set of features is a linear or non-linear combination of initial features, feature subset selection approaches aim at selecting a number of raw features without performing any transformations. In general the need for feature selection arises when the raw features are expensive to obtain and not numeric. Also, in some applications it may be important to extract meaningful rules from the classifier results. In such situations, feature extraction methods do not work. Hence, feature subset selection (FSS) methods must be employed. Feature subset selection requires a search strategy to select candidate subsets and an objective function to evaluate these candidates. There

are a large number of search strategies among which Sequential Forward Selection (SFS) was chosen.

Objective functions are divided into two groups, filters and wrappers. Filters evaluate feature subsets by their information content; typically interclass distance, statistical dependence or information-theoretic measures. Wrappers are essentially pattern classifiers, which evaluate feature subsets by their predictive accuracy by statistical resampling or cross- validation. Filters are fast to be executed and their results exhibit more generality. However, they tend to select the full feature set as the optimal solution. On the other hand, wrappers generally achieve better classification rates than filters and have the mechanism to avoid overfitting. The main disadvantage is slow execution. In this application the classification accuracy outweighs the speed; thus, wrappers were used to evaluate the selected set of features. Fig. 4.17 illustrates the feature subset selection procedure using wrapper objective function.

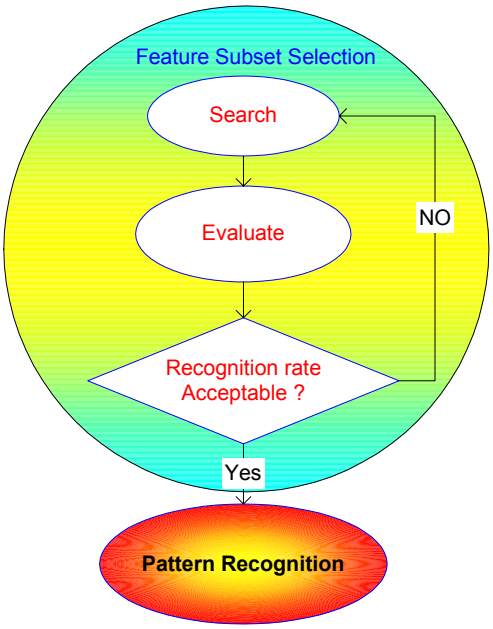


Fig. 4.17: Feature subset selection procedure using wrappers

4.3.3 Formation of Feature Vector

Through dimensionality reduction methods applied on the original feature matrix, it was found that the feature subset selection approach using the sequential forward search strategy performed superior. Accordingly, 14 informative features were determined and feature vector was formed for the classification purposes. The features included the first eight features at level 4, last four features at level 3 of the wavelet packet decomposition tree, and two time-domain features termed as VSI and HSI. The feature vector used as the input vector for all the supervised classifiers is shown in Fig. 4.18. In this figure, nE denotes the normalized energy and the following number represents the feature number.

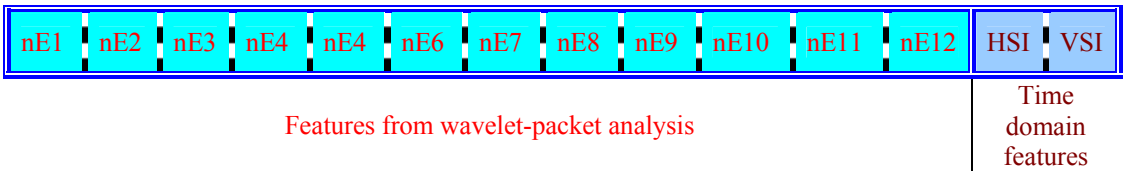


Fig. 4.18: Feature vector used in supervised classification

4.4 SUPERVISED CLASSIFICATION METHODS

The if-then expressed information used by the rule-based classifier can be extracted for non-incipient data but the patterns manifested in the incipient data are not expressible by linear, single level, and crisp relationships used in rule-based classifier. In fact, the complexity involved in these patterns is the actual challenge of the classification task. Furthermore, rule-based classifiers using crisp rules identify one class as the result and thus can only provide a black-and-white picture, whereas incipient data vary according to some gradation defined by the severity. Therefore, a set of other classifiers known as supervised classifiers were designed for the purpose of recognizing the complex patterns in the incipient data.

Classification using any of the supervised classifiers involves a model selection step which should be specifically defined for this application. The incipient abnormalities are classified into

one of three classes in terms of their local severity degrees. The severity degrees are expressed in terms of three fuzzy words, namely low, medium, and high. These classes are designated L, M, and H, as depicted in Fig. 4.19. Examples of incipient abnormalities representing each class are presented in Fig. 4.20, Fig. 4.21, and Fig. 4.22. The function of the supervised classifiers is to make decisions as to the degree of severity of an incipient abnormality conveyed by its feature vector. Classifiers used in the system include Support Vector Machines (SVM), Self-Organizing Map (SOM), K Nearest Neighbors (KNN), and Discriminant classifiers.

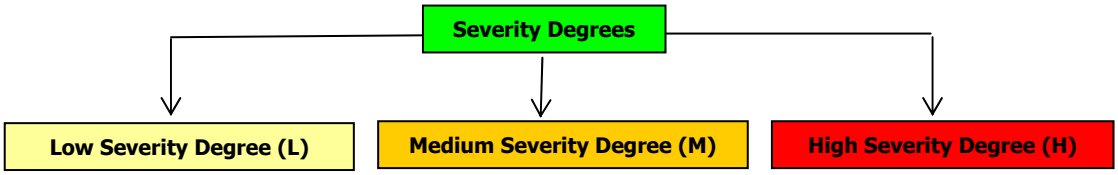


Fig. 4.19: Classification of incipient abnormalities in terms of severity degrees

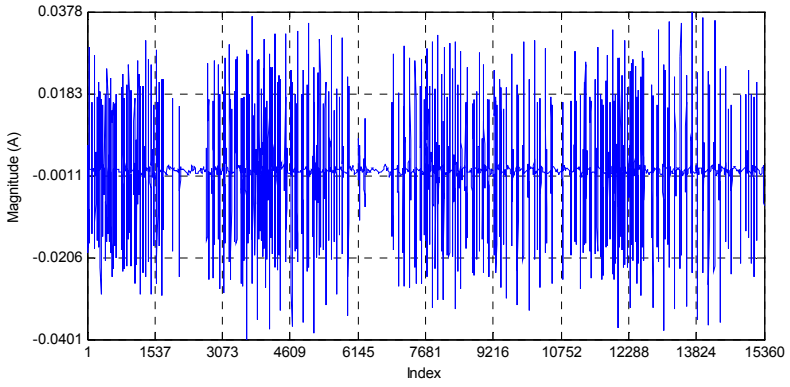


Fig. 4.20: Notch high frequency signal classified into high severity degree class

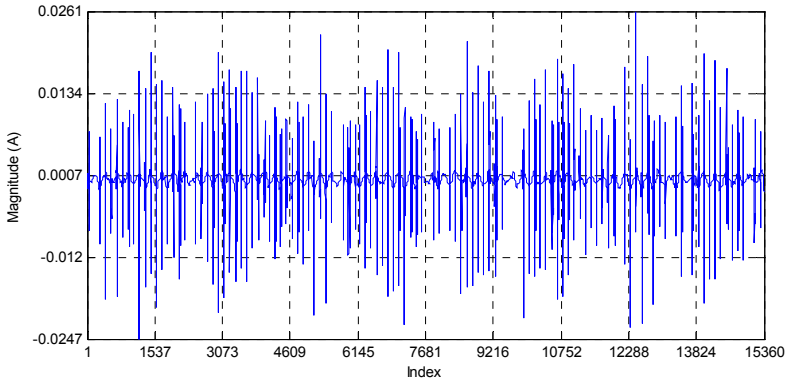


Fig. 4.21: Notch high frequency signal classified into medium severity degree class

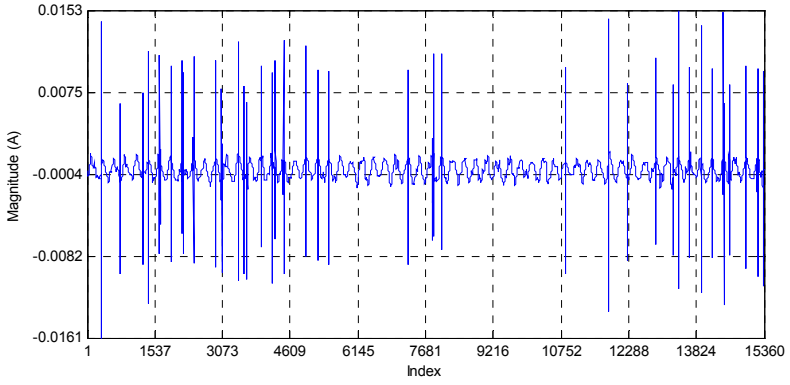


Fig. 4.22: Notch high frequency signal classified into low severity degree class

4.4.1 Support Vector Machines

Pattern recognition based on Support Vector Machines (SVM) is a relatively new technique that makes use of the statistical learning theory [93][94][95]. Given a set of training examples, the goal is to induce a function from these examples that acts as a classifier with good generalization property. Without loss of generality, consider the binary classification problem with linearly separable classes presented in [59]. Intuitively, one can imagine many linear

classifiers to separate the data, but only one function is optimal in separating the data points and maintaining best generalizing capabilities. The optimal hyperplane is the only classifier that will generalize well by maximizing the distance between it and the closest data point of each class. Other hyperplanes, although can separate the training data, will be unlikely to work well on the data outside of the training set.

In Fig. 4.23, a simple scatter plot of two features is given to illustrate SVM terminologies including the maximum margin, support vectors and separating hyperplanes. The largest margin is the minimum distance of an observation to the decision surface i.e. the oblique green line on the plot. The training samples that define the largest margin are known as the support vectors and the separating hyperplane is the one that is farthest from all training samples. The support vectors are the only points that contribute to defining the optimal hyperplane and thus the complete dataset is shrunk to include only those samples once they become known at the end of the training phase.

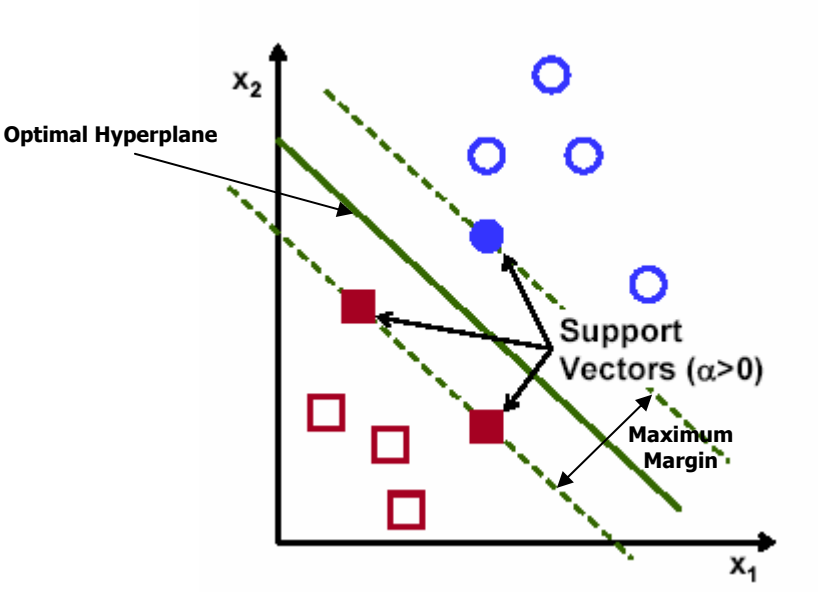


Fig. 4.23: Illustration of SVM for a binary pattern recognition problem [59]

As one can speculate, the procedure to construct the optimal hyperplane from the training data gives rise to an optimization problem. Given a training set of patterns that consist of feature vectors and numerical labels, SVMs resolve the generalized optimization problem [96] defined in (4-14).

$$\begin{aligned} &Min\, J(W,\xi) = \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\xi_i \\ &Subject\ \ to\ \begin{cases} y_i(W^T X_i + b) \geq 1 - \xi_i & i = 1,...,N \\ \xi_i \geq 0 & i = 1,...,N \end{cases} \end{aligned} \tag{4-14}$$

where W denotes the weight vector of the separating hyperplanes, b is their bias term, $C > 0$ is the penalty parameter of the error term, ξ is a slack variable to relax the constraints of the canonical hyperplane equation for non-separable case, N is the number of observations in the training set, and y_i is the label designated for the i^{th} observation. The penalty parameter is a trade-off between misclassification and complexity of the solution. Large values give rise to a better classification rate but increase the complexity of the solution as well [59].

Using the classical Lagrangian optimization techniques and Kuhn-Tucker theorem, the optimization problem boils down to the following dual problem in (4-15) [96].

$$\begin{aligned} &Maximize: \\ &L_D(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j X_i^T X_j \\ &subject\ to: \\ &\begin{cases} \sum_{i=1}^{N}\alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, \quad i = 1,2,...,N \end{cases} \end{aligned} \tag{4-15}$$

α_i s are the Lagrange multipliers which are zero for all the training examples except for the support vectors. For the linearly separable problem, the upper bound of C is removed for α_i constraint. Notice that in the dual problem, the slack variables and their associated Lagrange multipliers are disappeared and the goal is to maximize $L_D(\alpha)$ with respect to the Lagrange multipliers and not W and b. In addition, the training data appear only as dot products.

SVMs have shown superior generalization capabilities and yield appealing results. The power of SVMs originates from the principle in Cover's theorem on the separability of patterns. According to this principle, "*A complex pattern-classification problem cast in a high-dimensional space nonlinearly is* more *likely to be linearly separable than in a low-dimensional space*" [59]. Incorporating this principal, pattern classification based on SVMs proceeds in two stages. First, the feature vector is non-linearly mapped onto a high-dimensional space using a specific function ϕ called the kernel function. Then, an optimal separating hyperplane is constructed that separates the classes of data with maximum margin in the high-dimensional space. At the first glance, this contradicts with the feature extraction objectives that seek to minimize the dimensionality of the problem. The astonishing trick though is that the mapping is only implicit. The reason lies in the observation that the training data always appear as dot products i.e. $X_i.X_j$ in the SVM solution. Thus, in the high-dimensional space the mapping function operations would always appear as $\phi(X_i).\phi(X_j)$. Accordingly, one only needs to define a function K such that $K(X_i, X_j) \equiv \varphi(X_i).\phi(X_j)$ and use only that function in the training algorithm without explicitly performing the required mapping. Such a function is called the kernel function in SVM terminology and it must satisfy the Mercer's condition to be considered as a kernel [93][97]. Three basic kernel functions have been investigated, namely linear, polynomial, and radial basis function kernels.

- Linear: $K(X_i, X_j) = X_i^T X_j$
- Polynomial: $K(X_i, X_j) = (\gamma X_i^T X_j + r)^d, \gamma > 0$
- Radial Basis Function (RBF): $K(X_i, X_j) = \exp(-\gamma \left\| X_i - X_j \right\|^2), \gamma > 0$

$\gamma, r,$ and d are the kernel parameters that are typically determined through cross-validation. In this work, a non-linear polynomial kernel function was utilized whose parameters are as follows: $\gamma = 2$, $r = 3$, and $d = 6$. The SVM was implemented in MATLAB using functions from the LIBSVM toolbox [98]. The SVM classifier uses the feature vector presented in section 4.3.3.

The dual Lagrangian problem for the non-linear SVM can be expressed as a general quadratic programming problem (QP) defined in (4-16).

$$J(\alpha) = 1^T \alpha - \frac{1}{2}\alpha^T H \alpha$$
$$\textit{Subject to} \tag{4-16}$$
$$\begin{cases} \sum_{i=1}^{N} \alpha_i y_i = 0 \\ 0 \le \alpha_i \le C, \quad i = 1,2,...,N \end{cases}$$

where $H_{ij} = y_i y_j K(X_i, X_j)$. Notice that H is a quadratic matrix of N^2 elements, which consequently limits the size of the training examples. A number of alternative approaches have been proposed to cope with the curse of complexity and reduce the computational burden of SVM solution. Chunking, decomposition, and sequential minimal optimization methods are just a few examples [59].

Besides excellent generalization capabilities, other advantages make SVMs appealing classifiers. Since the SVM solution is a QP problem, there is no local minima problem. The final results are robust and repeatable since there is no random initialization phase. Support vectors are the only information needed to define the decision surfaces, thus the SVM solution is sparse. In addition, complexity and dimensionality of SVM solution are independently controllable. From implementation point of view, there are relatively few model parameters to select as opposed to that of other classification techniques [59].

4.4.2 Self-Organizing Maps

The Self Organizing Map (SOM) is a novel, widely used artificial neural network model introduced by Kohonen in 1981 [99]. SOM is generally used for data visualization and abstraction achieved by converting the complex and nonlinear statistical relationships between high dimensional data into a low dimensional grid of neurons called output map. It is also recently used for a number of power applications involving power transformers, tap-changers , and paper-insulated cables [100][101][102][103][104].

SOM in basic terms maps the high dimensional data in the input space into a low dimensional display of an output space while preserving the most important topological and metric relationships of the primary data points. In this sense, it is a non-linear vector projection and quantization method. Vector projection algorithms map the high dimensional data into low-dimensional coordinates preserving the shape of the original data cloud as much as possible. They are typically used for dimensionality reduction. Principal component analysis, curvilinear

component analysis [105] and Sammon's mapping projection [106] algorithms are some examples.

SOM can also be thought of a data classifier that helps find the natural groupings amongst inputs through a nonparametric, recursive regression process. These aspects can be utilized in a number of ways in complex tasks such as data clustering, numerical modeling, classification, computer-aided learning, estimation, machine perception, control, and communication. There are several variants to the basic SOM, which aim to either improve the quantization capabilities of SOM or make it a better classifier.

The inputs to SOM are highly complex multidimensional numerical features usually arranged in a data matrix. Each row of the matrix is an observation or data sample and each column corresponds to a feature. It is important that every sample have the same set of features. The output usually consist of a set of nodes (neurons) arranged on a two-dimensional grid called the map.

As an example, a 20×11 map is depicted in Fig. 4.24. There are more than 220 hexagons, the number of neurons, on this plot because the distances between adjacent neurons are also shown by hexagons. The normalized distances among the neighboring neurons are shown using a color map and the vertical bar on the right hand side of the map represents the numeric value corresponding to each color. Higher dimension grids are possible but their visualization is much more problematic. The number of neurons on the map is determined during the training process.

SOM neurons have in fact two positions; one in the input space and another in the output space. In the input space, each neuron is associated with a prototype vector[7] whose dimension is the same as the input vectors. In the output space, the neurons positions are specified by the map. The neurons are connected to the adjacent neurons by a function called the neighborhood function that dictates the map topology.

In this dissertation, SOM was used for two main purposes; Supervised SOM [99] –a variant of SOM- was used for classification purposes while unsupervised SOM was utilized for numerical modeling of the incipient data and development of a detection index. Application of SOM for the classification problem is introduced and discussed in this chapter while chapter V presents the data modeling principle and the resulting detection index.

[7] Prototype vector may also be called weight vector, reference vector, codebook, or model.

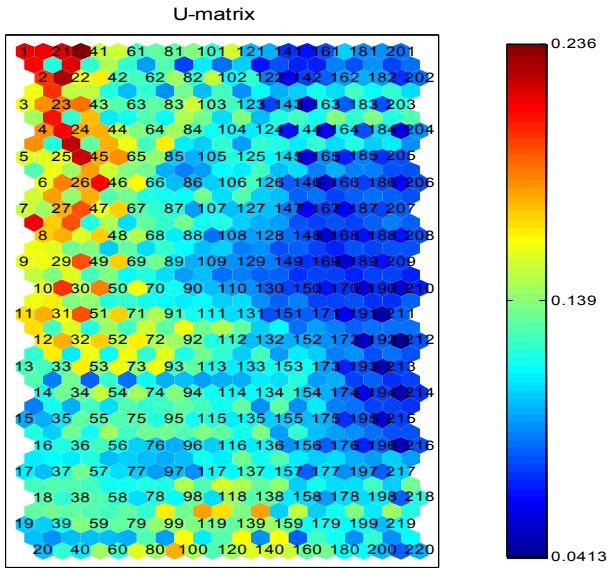


Fig. 4.24: SOM output map

4.4.2.1 SOM Design and Training Procedure

The design stages of a SOM usually involve the selection of the map lattice, map shape, number of neurons, and neighborhood function. On a commonly used two-dimensional map, the neurons can be arranged on a rectangular or hexagonal lattice, but the latter preferably provides smoother and more eye-appealing map. Thus, the map lattice was selected as hexagonal. There are three different map shapes namely, sheet, cylinder, and toroid. In this work, the sheet shape map was used as recommended in [107]. The map grid sidelights was automatically determined by the ratio between eigenvalues of the training data.

The general practice in selecting the number of neurons is to choose as many as possible. This improves the resulting map granularity and results in a more flexible map. Nevertheless, it impacts the computation complexity of the training phase of a SOM application. Hence, an appropriate selection should be considered to meet the present trade-off. We used the recommended number of neurons is $5\times\sqrt{N}$, where N is the number of training samples [107].

The neighborhood function describes the distance relationships between neighboring neurons. It is a decreasing function of the distance between every two neurons on the map grid. The choice of the function and the number of neurons are the determinants of SOM accuracy and

generalization capabilities. The neighborhood function utilized for the SOM design was a Gaussian function.

The learning process of unsupervised SOM requires no a-priori knowledge. It is based on a dissimilarity measure between the input data and map units measured in terms of distance, typically Euclidian distance. The greatest similarity is achieved when a neuron is a minimum distance from the input sample. Such a neuron is termed the best matching unit (BMU). Accordingly, in each learning step, one sample from the input data set is randomly chosen and the BMU is searched. Once it is found, the prototype vectors of the BMU and its topological neighbors are moved closer to the input vector in the input space as graphically shown in Fig. 4.25. The training procedure is performed in two stages. First, the SOM is approximately tuned to the input data, and then through fine-tuning in the second stage, the final map topology is achieved. The training parameters are relatively large in the first stage but they possess small values right from the beginning in the second stage [107].

The unsupervised version of SOM does not utilize the class information; however with some modifications, the network can take the class into account. This is achieved by augmenting the original feature vector where the class information is treated as an extra feature. This classifier in either mode was implemented in MATLAB with the help of a great SOM toolbox generously provided by a group of researchers in Finland [107]. The original feature vector used by this classifier was discussed in section 4.3.3.

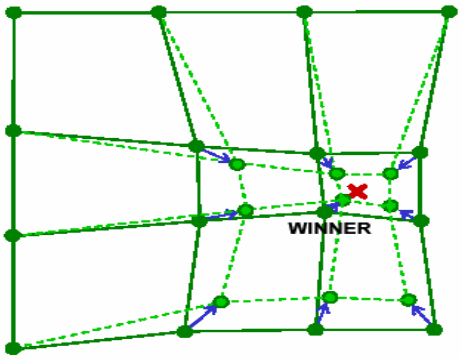


Fig. 4.25: SOM training phase

4.4.3 K-Nearest Neighbors

In KNN classifiers, graphically shown in Fig. 4.26, the K closest examples in the training data set are found, and the majority class is determined and assigned to the unlabeled example [108]. In the case shown, the unknown input is assigned to class 1. The term "closest" implies utilization of a metric to measure closeness. Most often, this metric is a dissimilarity measure of Euclidean type distance. In addition to the need for a metric, KNN classifiers require an integer K and a set of labeled examples, i.e., training data. In this dissertation, $K = 8$, which was determined through cross-validation. The 8NN classifier uses the feature vector presented in section 4.3.3 and was implemented in MATLAB using a built-in function called *knn*.

KNN classifiers are easy to implement and nearly optimal in the large sample limit, They yield highly adaptive behavior using the local information. However, they are computationally intensive and highly susceptible to the curse of dimensionality. Furthermore, KNN classifiers impose large storage requirements since the entire training data set must be employed for classification of every new pattern.

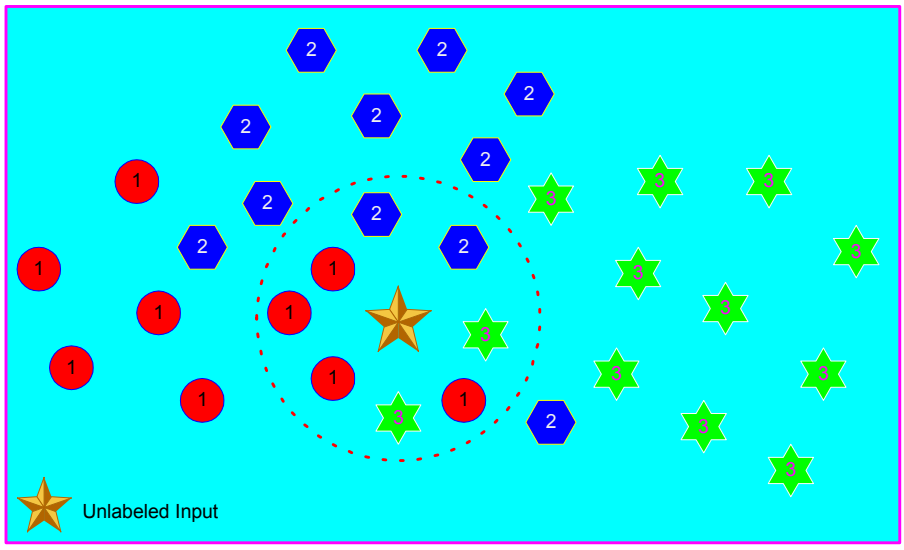


Fig. 4.26: KNN classification with K= 8

4.4.4 Discriminant Functions

The discriminant classifier can be viewed as a network of functions called discriminant functions $g_i(X)$ that operate on each input vector X to determine its category. A discriminant classifier selects the category corresponding to the largest discriminant [53]. Fig. 4.27 shows the network structure of a discriminant classifier. With c discriminant functions, the feature space is divided into c decision regions, and every function responses to a specific region. The linear discriminant function used in this classifier fits a multivariate normal density for each group, with a pooled estimate of covariance. The linear DIS classifier uses the feature vector presented in section 4.3.3. This classifier was implemented in MATLAB as a function.

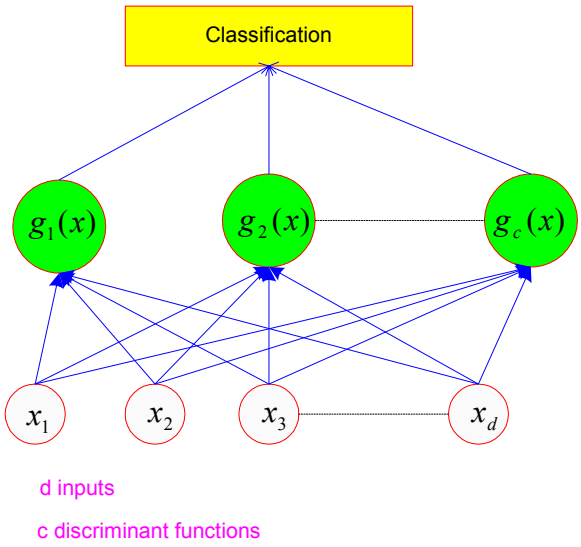


Fig. 4.27: Discriminant classification structure

4.5 PERFORMANCE ANALYSIS OF CLASSIFICATION FUNCTIONALITY OF THE IFDS

In this section, the performance of the classification functionality of the IFDS is studied. The objective of this analysis is to assess operational aspects of the classifiers in terms of the classification rate. The performance of the classifiers is also evaluated using the false positive (FP) and false negative (FN) measures. The false positive exists when the classifier identifies a pattern incorrectly where no such pattern exists in reality. Detection algorithms of all kinds have the tendency to create such false alarms. The false negative occurs when the classifier fails to reveal a pattern that actually exists in the data. Specifically, if the IFDS labels a non-incipient data as an incipient abnormality, this decision is deemed a false positive decision. On the other hand, if the IFDS fails to indicate an incipient abnormality and makes a no-incipient decision, it is considered a false negative result. Evidently, false positive decisions are more tolerable than the false negative results. Thus, the overall goal of the classification is to make the classification rate as high as possible and if there are false decisions, false positive decisions are more welcomed than false negative ones.

In addition to the classification accuracy measures, an analysis was conducted to determine the class composition of measurements before and after actual faults that occurred in the underground cable. This study uses the labeled outputs of the system to provide some insights into the emerging patterns of the data preceding and following a failure.

4.5.1 Classification Rate

The overall classification rate of the IFDS is determined by the performance of the individual classifiers, including the rule-based and supervised ones. The performance of a classifier can be measured in terms of its classification rate (CR) defined by (4-17). Most of the time, the objective is to minimize the error rate by maximizing the classification rate.

$$CR = \frac{number\ of\ correct\ \ assignments}{number\ of\ total\ \ assignments} \tag{4-17}$$

For the ensemble of classifiers operating in series in the system, when a classifier performs well, the overall classification rate is high. On the contrary, a less accurate classifier has a

negative impact on the performance of the entire system. In addition to the performance of individual classifiers, the composition of the input data is also a determining factor of the classification rate. If a classifier performs poorly, its adverse influence on the system classification rate is exacerbated when it needs to make a decision on the majority class. Specifically speaking, the supervised classifiers of the system respond to the majority class of the inputs. Thus, their classification rate has a major impact on the overall classification rate and must be improved as much as possible.

The classification rate of the system can be evaluated using the training and test data. The system should perform well on the training data because it has seen the patterns comprising the input space. However, caution must be taken to avoid the over-fitting problem in which the system memorizes the patterns rather than learning them. An over-fitted classifier presents superior classification accuracy on the training data but possesses poor generalization capabilities and does not perform well on the data that it has not seen before. Using the test data, the system should respond adequately well to meet the prescribed classification rate.

To analyze the classification rate of the supervised classifiers and that of the IFDS, studies were conducted using field recorded data. Specifically, one study was conducted using 3001 measurements captured from June 16, 2002 until July 17, 2002. For supervised classifiers, 2163 captures were incipient abnormalities at various degrees of severity. The function of classifiers was to classify data in terms of the severity degrees. In all studies, the training and test set were formed using a 75/25 blend. The results are as follows.

4.5.1.1 Classification Rate of the Supervised Classifiers

Each classifier was trained using the training data and relevant training algorithm. When trained, the classifiers were tested individually against two sets of data to evaluate their performance. The results are presented in Fig. 4.28 where the classification rates of the individual supervised classifiers are reported. In this study, there were 1622 samples in the training set and 541 samples in the test set randomly selected from the incipient data. As can be seen, for this set of data, the SVM classifier outperforms the other classifiers, achieving an average rate of 98.52% on the test data. The other classifiers perform fairly well on the test data. It should be mentioned that this result was observed to be true in all studies performed with different random samples from the data.

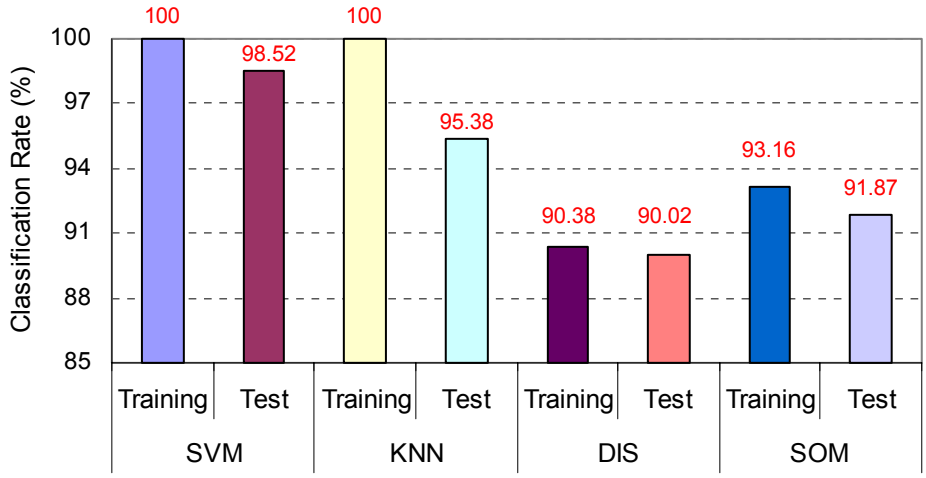


Fig. 4.28: Classification rate of supervised classifiers on the training and test data

4.5.1.2 Classification Performance of the IFDS

To assess the overall classification rate, both incipient and non-incipient data totaling 3001 measurements were used. The incipient data were composed of all kinds of incipient abnormalities at different degrees, and non-incipient data were comprehensive and composed of all classes. The classification problem was defined as a 4-class classification problem where the IFDS was supposed to make decisions as to the label of the data. The four classes included the three incipient classes and one cumulative class for the non-incipient data. The idea was to see how well the measurements were classified by the combined classifiers including the rule-based (RB) and supervised classifiers. The measurements were ordered chronologically as they were measured and fed to the IFDS. A few separate runs were conducted employing a different classifier in the supervised classification stage. Once the classification task ended, the overall classification rate was computed by counting the number of correct assignments taking the average of all the resultant numbers for classification rate. The overall classification rates using different classifiers are presented in Fig. 4.29. As seen, an overall classification rate of 97.17% was achieved with RB-SVM classifier, and that of RB-KNN, RB-DIS, and RB-SOM were similar, varying around 94% on the average.

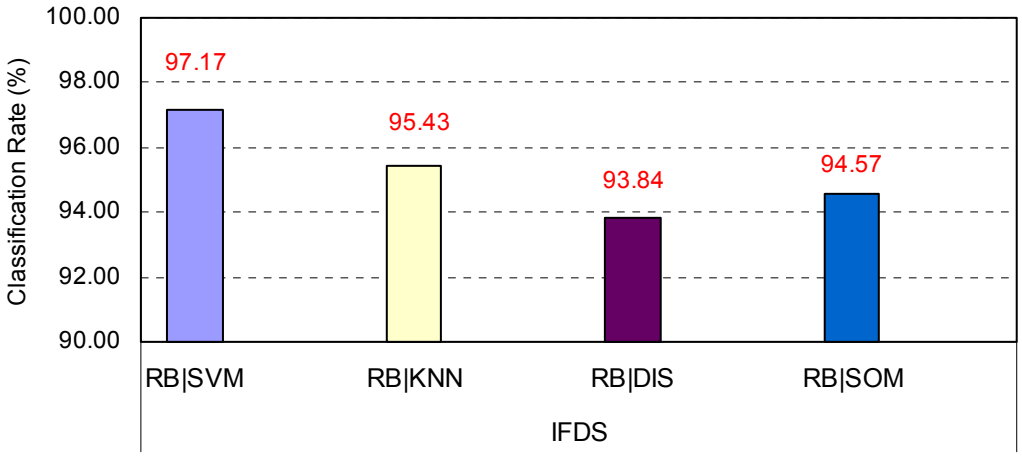


Fig. 4.29: Overall classification rates

The performance of the classifiers was also assessed using the false positive and false negative measures. A false negative classification occurred when an incipient abnormality measurement was incorrectly labeled as non-incipient or it was assigned to a lower severity incipient class. On the other hand, a false positive resulted when a non-incipient capture was assigned to the class of incipient abnormalities. The resulting measures using each classifier combination in the system are shown in Fig. 4.30. The presented numbers are the FP and FN measures obtained on the 1380 test data. It can be concluded that while all classifier combinations are fairly skewed toward the false positive region, the RB-DIS classifier provided the most wrong assignments that are mostly false positive. From this point of view, the RB-DIS classifier combination appears to be the least appealing choice to be used in the system.

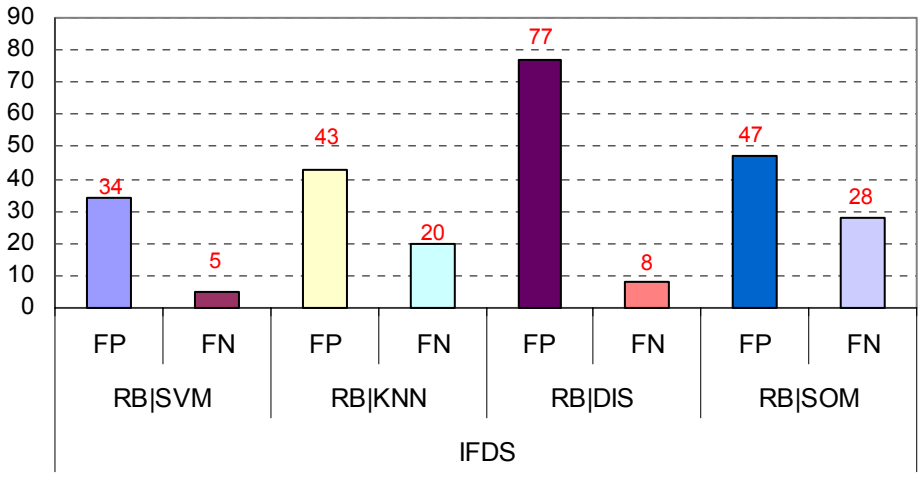


Fig. 4.30: False positive and false negative measures

4.5.2 Illustration of Classification Results through Class Compositions

Determining the class composition of measurements is another quantitative performance analysis approach that was conducted to observe the classification results before and after actual faults in the underground cable. An example of such analysis is provided in Fig. 4.31. This plot presents the percentage of the data assigned to various classes before and after a cable fault that took place on 12/24/02. From this graph, the following observations can be made.

High-class incipient abnormalities compose 45.8% of the data before the fault, but this proportion significantly decreases to 3.4% after the fault, indicating that a point of transition has passed. Medium-class incipient abnormalities make up 20.3% of the data before the fault and 34.3% after the fault, respectively. This may be an indication that the incipient abnormalities have lessened in severity, but they have not completely disappeared after the fault time has passed. The composition of low-class incipient abnormalities has sharply risen from 1.8% before the fault to 34.3% afterwards. This confirms that the incipient abnormalities have been reduced in severity after the failure. The aggregate class of non-incipient data shows a slight decrease in its composition percentage, dropping 7.5 points to 24.7%.

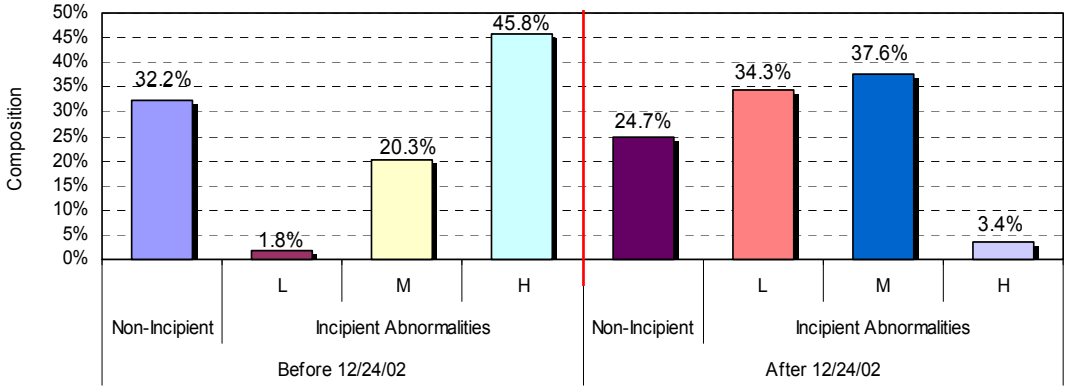


Fig. 4.31: Class composition of measurements for case I

Another similar study was accomplished for another cable failure on 06/11/03. The class composition of the measurements preceding the failure is given in Fig. 4.32. As can be seen, the three classes of incipient abnormalities make up 46.1%, 23.1%, and 1.3% of the data, respectively, and 29.5% of the data are labeled non-incipient. As expected, the high-class incipient abnormalities compose the majority class before the fault time. In terms of the actual numbers, a good deal of similarity is observed in comparing the pre-fault compositions of the data in the two cases as conveyed by Fig. 4.31 and Fig. 4.32.

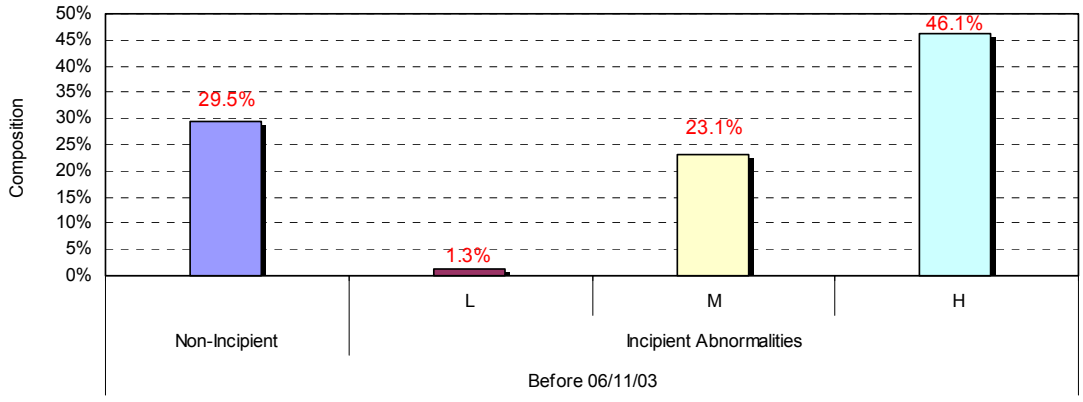


Fig. 4.32: Class composition of measurements for case II

4.6 VISUALIZATION OF CLASSIFICATION LABELS

The purpose of this study is to visually illustrate the behavior of the class labels and the trend before and after two faults recorded at the monitoring site. For a fault on 12/24, the classification labels before and after a fault are presented in Fig. 4.33. The data included captures two days before and three days after the fault time. Four types of labels are shown which include three incipient classes: low, medium, and high intensity incipient data, and another aggregate class for non-incipient data. As can be seen, the high intensity incipient data is the most frequent class observed before the fault time and the least frequent class afterwards. The opposite is true for low intensity incipient data. Since these classes are mutually exclusive, every capture can be assigned to only one class, and the decreasing rate of occurrence for the high intensity incipient data naturally results in an increasing rate of occurrence for the other classes. This is seen for the medium intensity incipient data where the frequency of occurrence shows an increase after the fault time. The presence of reduced intensity incipient abnormalities might be an indication of persisting system deterioration that will eventually lead to another failure. In fact, six months later another fault occurred in a different section of the cable lateral.

Similar behavior can be observed for a later fault on 06/11 that took place approximately six months after the first fault. The corresponding plot for this fault is shown in Fig. 4.34. The post-fault data were not available due to a monitoring system failure. The trend of arrival frequency conforms to the pattern observed in Fig. 4.33. As the fault time approaches, the high intensity incipient abnormalities tend to occur more frequently than before and dominate the other classes. In both scenarios, the increasing rate of occurrence of high intensity incipient abnormalities was consistent and may be used to indicate the progressive deterioration of the system toward an imminent fault. It should be noted that the observed behavior can not be simply generalized from these two observations. As pointed out in Chapter II, there are many different failure mechanisms in cables, which may or may not present the same pattern. Nonetheless, what remains valid is that the classifiers will capture a pattern if it exists.

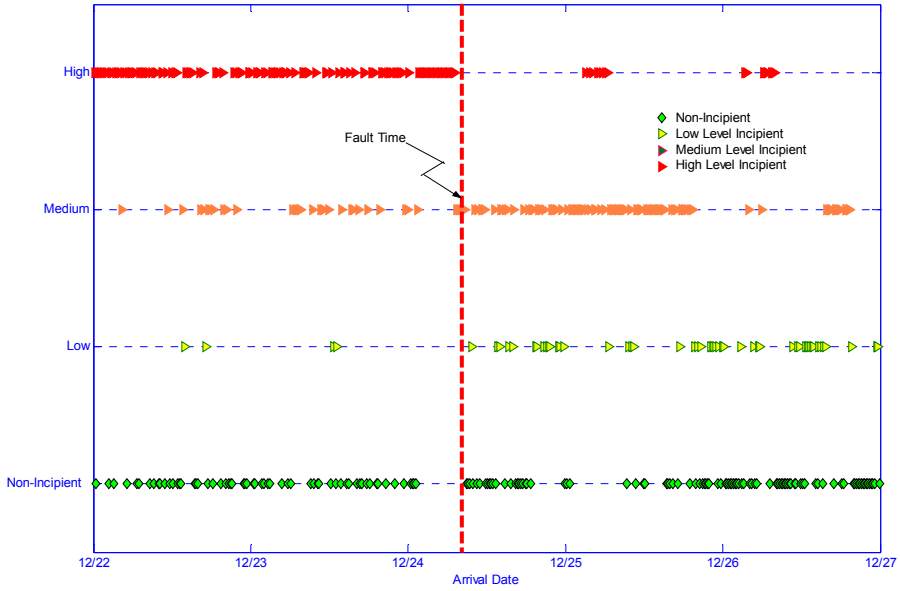


Fig. 4.33: Class labels before and after fault on 12/24

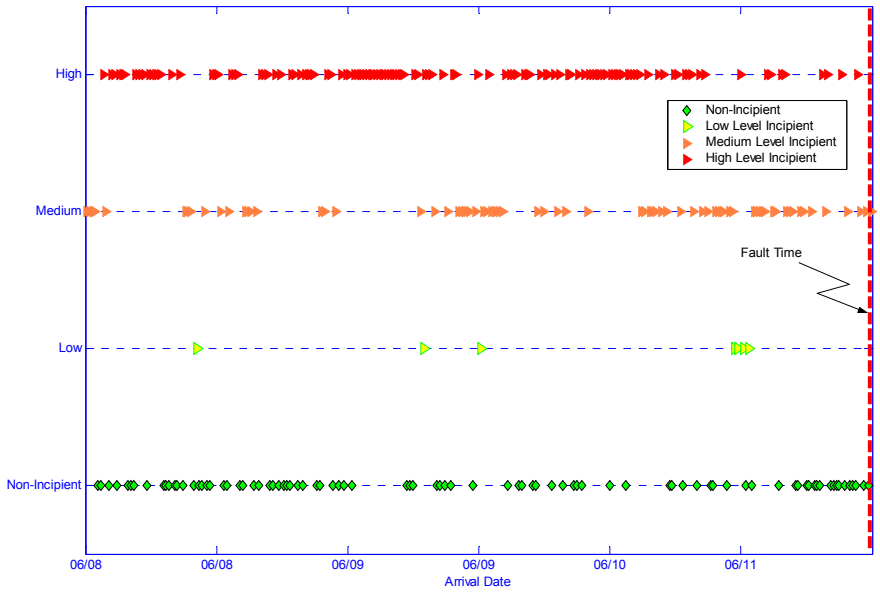


Fig. 4.34: Class labels before fault on 06/11

4.7 SUMMARY

In this chapter, the classification functionality of the IFDS using rule-based and supervised classifiers was presented in detail. The rule-based classifier designed for various switching type events and DC classes was introduced and its operational aspects were discussed. Also, spike analysis and detection algorithms were developed and discussed. The supervised classifiers included SVM, SOM, KNN, and DIS classifiers, which were designed and configured for classification of complex incipient abnormality data. Based upon the defined severity degree measures, three classes of low, medium, and high severity classes were emerged and the classifiers were trained to solve the classification problem. Evaluation of the performance of individual and combined classifiers showed satisfactory classification accuracy. In comparing various classifier combinations used in the IFDS, the RB-SVM combination was found to perform superior over other classifiers. Therefore, in the detection phase, the IFDS uses the class labels that are determined by the SVM classifier. The results from the other classifiers are used for evaluation and performance studies.

The next chapter brings up the other important functionality of the IFDS which is the detection of near failure. The relevant methods and developed detection approaches constitute the major subject matter of the next chapter.

CHAPTER V

INCIPIENT FAULT DETECTION METHOD

5.1 INTRODUCTION

This chapter discusses the detection functionality of the incipient fault diagnosis system (IFDS). First, a temporal severity measure is introduced. This measure provides a descriptive trajectory of incipient behavior as the severity degree exacerbates over time. The expectation is that the severity of an imminent incipient fault shall increase monotonically, as the fault draws near. The severity measure is based on a new severity index called the *Global Severity Index* (GSI) computed from the classification results. GSI serves as supplemental information for detection purpose.

Although GSI provides insights into the trajectory of the deterioration process, it is not designed to give quantitative information such as the near failure detection. This important functionality of the IFDS is accomplished by a quantitative detection approach which is based on the numerical modeling of measurements using a SOM. First, a numerical model of the present incipient state of the underground cable is established. Then, the SOM model is used to determine and compute gradual deviations of the incipient behavior from its present status. If there is an ongoing incipient activity in the cable, the accuracy of the model is expected to worsen over time and the resulting modeling errors are utilized to compute the detection index. The detection index values are computed from the modeling errors by taking the natural logarithm of the minimum modeling errors. Once the detection index is established, determination of the alarm time(s) translates into a change detection problem where the goal is to detect when the detection index instantaneous values cross a specified threshold. The threshold crossing represents the system deviation from its initial state captured by the SOM model.

Three detection algorithms namely, Cumulative Sum (CUSUM), Exponentially Weighted Moving Averages (EWMA), and Generalized Likelihood Ratio (GLR) are discussed and designed to solve the change detection problem. These algorithms were selected so as to conduct a performance analysis and compare the results from each algorithm.

5.2 TEMPORAL SEVERITY MEASURE

Through the classification functionality, the IFDS is able to automatically recognize input patterns and classify them into designated classes. Once classified, each measurement is assigned a prescribed label that describes its class and information content. An analysis of interest is to use present classification labels in conjunction with the past information to draw meaningful conclusions about the frequency/trend of occurrence of particular class of measurements, namely, the class of incipient abnormalities. By defining a new index called the *Global Severity Index* (GSI), a temporal severity measure was developed. This section provides details about this measure and its application.

5.2.1 Global Severity Index

The supervised classifiers label the incipient abnormality data in terms of their severity degrees where the input capture is classified based upon its local content, regardless of the history of previous captures. In other words, it is a static classification approach. However, the real scenario in which incipient faults evolve and lead to a catastrophic failure may involve historical information as well. When evaluating the severity of an incipient abnormality, it is important to take into account the past behavior of the system and consider the severity changes over time. This temporal severity analysis is accomplished by a measure referred to as the *Global Severity Index (GSI)*. The two-stage severity analysis involving GSI is shown in Fig. 5.1. The GSI is defined for every capture that is an incipient abnormality. It provides information as to how the current incipient abnormality is related to the past behavior of the cable.

It is worth mentioning that in temporal analysis, the incipient information is viewed as survival data in which the time of a particular event is of a concern [109]. Survival data encountered in reliability and applied statistics refer to the time of an event. In the context of this work, an event is defined as the arrival of an incipient abnormality. Thus, the analysis includes monitoring the system for a definite period and recording the time at which an incipient abnormality takes place. During the observation interval, events i.e. incipient abnormalities may or may not occur. If they occur, the attributed arrival times are recorded and the events are considered uncensored. If the events do not occur, they are considered censored and the associated arrival times are discarded.

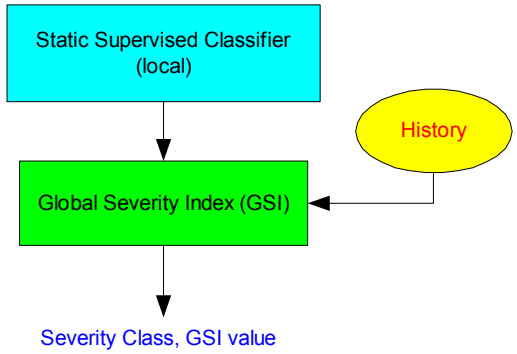


Fig. 5.1: Stages of temporal severity measure

5.2.2 Computation of GSI

The Laplace Test Statistic (LTS) plays an important role in the computation of GSI. This test uses a set of points that are chronologically ordered based upon the arrival times of events of interest [62]. Let $T_1, T_2, ..., T_m$ be a sequence of m arrival times in their chronological order. The times between arrival times are called interarrival times and are denoted by the sequence of $X_1, X_2, ..., X_{m-1}$. The Laplace Test Statistic (LTS) for the set of m incipient events is defined by (5-1) assuming that the observation stops at the last m^{th} recorded incipient event.

$$LTS = \frac{\frac{\sum_{i=1}^{m-1} T_i}{m-1} - \frac{T_m}{2}}{T_m \times \sqrt{\frac{1}{12(m-1)}}} \qquad (5\text{-}1)$$

Having all the arrival times for incipient events available ($T_1 = 0$), the GSI is the LTS calculated for each event using a fixed-size sliding window as given in (5-2). In this formula, W denotes the window number and m is the size of the window. Fig. 5.2 symbolically shows the method of assessing global severity of incipient abnormalities using a sliding fixed window size of seven events. The span of calculation m is a parameter that defines how much past

information accounts for current incipient incidence. Intuitively, GSI is not defined for the events that lie in the first window because there is not enough information to calculate it.

$$GSI = LTS\big|_{W\ m=const.} \qquad (5\text{-}2)$$

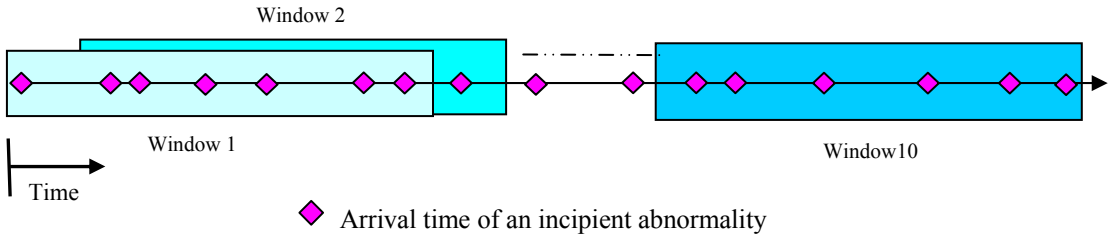


Fig. 5.2: Illustration of GSI computation using a sliding window of size 7

5.2.3 Behavior and Significance of GSI

Three scenarios under which the incipient behavior might progress are considered. Incipient abnormalities may arrive at a constant, increasing, or decreasing rate. Equivalently, the interarrival times may remain constant, decrease, or increase over time. In order to understand the usefulness of this measure in assessing the global severity of an incipient event, it is informative to examine the behavior of the test statistics under the preceding scenarios using simulated arrival times. This is graphically illustrated in Fig. 5.3. In this figure, there are nine plots in three columns. Each column corresponds to each scenario printed on top of the plots in the first row. This row shows the assumed arrival times under each scenario and the three graphs in the middle row depict the corresponding interarrival times. Using 15 simulated arrival times for each scenario, the LTS was computed using a growing window and the resulting values were plotted for each case on separate graph in the third row. The GSI corresponding to each scenario is the final value of LTS.

Under a constant rate of occurrence or equivalently constant interarrival times, the mean arrival times are approximately $\frac{T_m}{2}$, and the numerator in the Laplace test becomes a small number close to zero. Hence, as shown in the figure, the Laplace test asymptotically approaches zero as time continues. If the arrival times convey an increasing trend toward the end of the interval, the mean grows and becomes larger than the half interval time. Accordingly, the test statistics increases toward large positive numbers, indicating decreasing interarrival times. Finally, if the rate of occurrence is decreasing, the test statistics possess negative values, implying that the interarrival times are becoming larger and larger. Accordingly, when applied to the incipient abnormality data, the temporal severity can be measured with respect to the past observations as a numeric value. This measure quantifies the behavior of successive incipient events during the observation interval.

As discussed earlier in the IFDS, the incipient data are partitioned into three classes in terms of their severity degrees. A GSI value is computed for each class, but the GSI computed for the high-severity class data naturally carries a considerable significance over the other two classes. The GSI values are calculated for each measurement and chronologically ordered as a time series. Thus, the global severity index is a function of time and an aggregate measure of severity of an incipient abnormality with respect to its present local and past severity degree. The significance of this global index is that it can signify the deterioration trajectory of the underground system empirically observed to be a monotonic process. When GSI presents an increasing trend, it naturally implies that a fault could be imminent. On the other hand, a decreasing trend could be an indication of system improvement over time because the severity of the incipient abnormalities deceases. The applicability and usefulness of this temporal severity measure in depicting the progressive severity trend of incipient abnormalities will be illustrated in chapter V using field recorded data from the monitoring site.

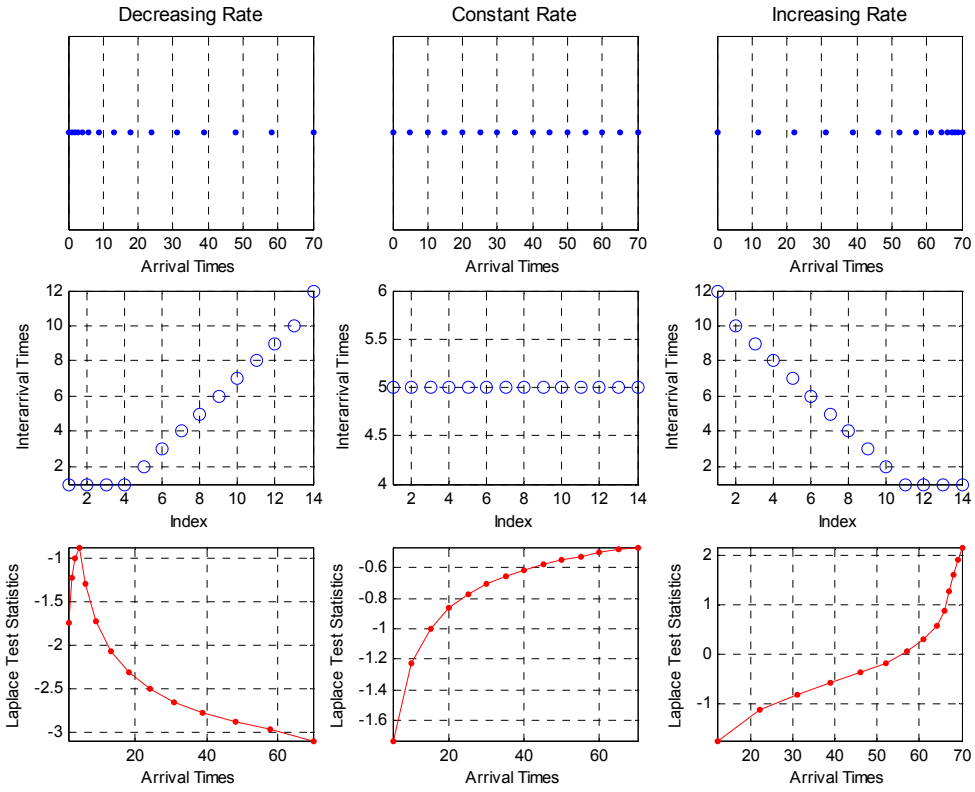


Fig. 5.3: Response of global severity index under decreasing, constant, and increasing rate of arrival times

5.3 SOM-BASED INCIPIENT FAULT DETECTION METHOD

The SOM based incipient fault detection method provides a quantitative measure to statistically predict near failure for the cable being monitored. This method is based on the numerical modeling of the cable's present state and uses the modeling error to determine the degradation path over time. The minimum modeling errors (MME) constitute the fundamental component of the detection index (DI). This index is defined to be the natural logarithm of the minimum modeling errors. The natural logarithm modification is necessary as it facilitates the subsequent statistical formulation of the change detection problem.

To numerically model the data, a SOM representing the present status of the cable is designed and the resulting prototype vectors are used to serve as the codebook. When a new pattern becomes available, it is exposed to the SOM model and the minimum modeling error is computed. Ideally, if the error is within the tolerance boundary, the new pattern is deemed similar to the model and labeled as an "ally". Once the error limit is crossed, the corresponding input vector is flagged and a structural change point is detected. Thus, when chronologically ordered, the minimum modeling errors serve as the fundamental unit of the detection index. It should be noted that the presence of outliers in data can also give rise to substantial spurious errors and thus it is imperative that potential outliers be removed from the input patterns before training the SOM and exposing the new patterns to the SOM model. The formal definition of the detection index and its characteristics are discussed next.

5.3.1 Detection Index (DI)

As mentioned, the detection index is in fact the natural logarithm of the minimum modeling errors. When new input vectors are exposed to a trained SOM, the minimum modeling errors can be used to quantitatively describe the modeling accuracy. Naturally, smaller errors indicate a higher modeling accuracy and that is highly desired in clustering applications. On the contrary, the prototype vectors and the incoming input patterns are highly dissimilar when large modeling errors emerge. Further, the higher the computed modeling error, the greater the dissimilarity. Substantial modeling errors indicate that the prototype vectors are no longer good representatives of the input patterns and thus a change in the structure of the data has occurred. This is the principle used in the incipient fault detection method based on SOM numerical modeling. Kang and Birtwhistle applied this principle in condition assessment of power transformer on-load tap changers [101][102][103].

As discussed in section 4.4.2, SOM maps high dimensional data in the input space into a low dimensional display of an output space while preserving the most important topological and metric relationships of the primary data points. The inputs to SOM are multidimensional numerical data arranged in a data matrix. Each row of the matrix is one observation or input vector and each column is a variable or feature. The output usually consists of a set of nodes (neurons) arranged on a two-dimensional grid called the map.

SOM neurons have in fact two positions. In the input space, each neuron is associated with a prototype vector whose dimension is the same as the input vectors. In the output space, the

neurons positions are specified by the map. In a trained SOM, each input vector is represented by its Best Matching Unit (BMU) on the map. The representation error or quantization error between data vectors and their BMUs is a measure of data mapping accuracy. During training procedure, the objective is to minimize the average quantization error. Once a SOM is trained and fine-tuned, the map units represent the input vectors with minimum quantization error. The resulting prototype vectors associated with BMUs represent the numerical models of the input vectors. Using these numerical models, the response of the SOM to a new input can be described by computing the minimum quantization error or minimum modeling error.

The minimum modeling error is based on the proximity measure between the input vector F and its BMU as defined by (5-3). In this formula, the weight vector W is the same size as F and associated with the BMU. The distance metric is the Euclidean norm that provides a measure of dissimilarity between two D-dimensional vectors W and F and happens to be the most commonly used proximity measure. This metric is expressed in (5-4). Once MME is computed, the detection index is established by taking the natural logarithm of the MME as given in (5-5). DI values must be in chronological order. The dimension D is determined by the number of features used in the modeling, which is $D = 16$. The features are the sixteen normalized energy values computed from the wavelet packet analysis at the fourth level. These features are defined and computed in section 4.3.1.1.

$$MME = \min\left\{\left\|W - F\right\|\right\} \tag{5-3}$$

$$\left\|W - F\right\|_e = \left(\sum_{i=1}^{D}(W_i - F_i)^2\right)^{1/2} \tag{5-4}$$

$$DI = Ln(MME) \tag{5-5}$$

5.3.2 Statistical Characteristics of the MME and DI

Although the ideal behavior of the modeling errors discussed in the beginning of this section makes the change detection problem less challenging, it is not the case in practice. Typical behavior of the MME during the training phase of a SOM is depicted in Fig. 5.4. This figure is obtained by using classified field data and treated as an example in the rest of this section. As seen, the average modeling error is minimized but not the variance of the errors (bias-variance tradeoff). The chronologically ordered modeling errors average around 5.48,

however the resulting variance is large. In fact, there is an inherent statistical variability in the modeling errors in the training phase even after removing potential outliers.

The resulting challenges are twofold; on one hand, the existing statistical variability makes it difficult to determine a simple global threshold for the change detection problem and on the other hand, the sole application of modeling errors for detection purposes is noise and outlier prone, and thus highly sensitive. Controlled sensitivity is a desired feature of the IFDS, but uncontrolled and unexpected sensitivity give rise to a high probability of false alarm and thereby indifferent response of maintenance crew.

To overcome the challenges involved, the detection index was developed and the corresponding change detection problem was solved using statistical methods. In order to implement the change detection method, the probability distribution of the minimum modeling errors and the detection index are needed, which are presented next. In order to specifically discuss the results, the MME values shown in Fig. 5.4 are used as a benchmark. Therefore, the values determined hereafter are associated with these data.

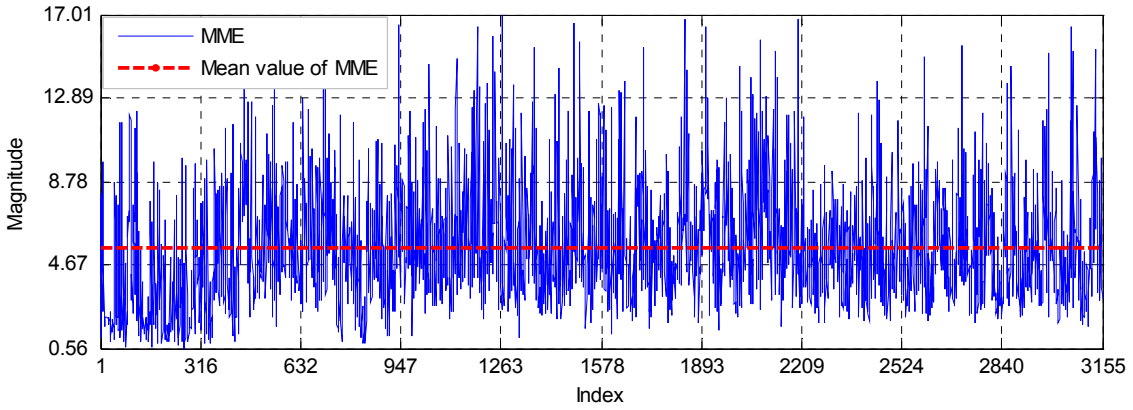


Fig. 5.4: Typical behavior of the modeling errors during the training phase

The probability distribution function of the minimum modeling errors (MME) was investigated through a number of statistical studies. The results showed that the probability distribution function approximately follows a log-normal distribution. In other words, the

logarithm of MME is approximately a normal distribution. This result makes sense because the log-normal distribution is applicable when the quantity of interest is positive since the logarithm of a negative number is not defined. The log-normal distribution typically results whenever the random variable is the product of a large number of independent, identically-distributed variables. The normal and lognormal distributions are closely related. If x is distributed lognormal with parameters μ and σ, then $\ln x$ is normally distributed with the same parameters [110]. Since the DI is the logarithm of MME, this implies that DI is normally distributed. The log-normal probability density function for a random variable x with parameters μ and σ is defined in (5-6).

$$pdf(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}} \tag{5-6}$$

The probability distribution of the MME and DI can be demonstrated using a number of plots as shown in Fig. 5.5. These plots were obtained from the data used in Fig. 5.4. The upper left histogram shows the distribution of the MME that approximate a log-normal distribution. The next graph (upper right) is the histogram of the logarithm of MME, i.e. DI. This graph shows that the data are reasonably symmetric, there do not appear to be significant outliers in the tails, and that it is reasonable to assume that the data are from a normal distribution.

The normal probability plot shown in the lower left can be used to verify that the DI values come from a normal distribution. The plot has three graphical elements. The plus signs show the empirical probability versus the data value for each point in the sample. The solid line connects the 25th and 75th percentiles of the data and represents a robust linear fit. The dashed line extends the solid line to the ends of the sample. The normality assumption holds since all the data points fall near the line. If the data were non-normal, the plus signs would follow a curve instead of the line. Notice that the y-axis values are probabilities ranging from zero to one.

The parameters of the normal distribution associated with DI can be estimated from the data. The empirical cumulative distribution function (CDF) for the DI along with its estimated CDF is given in the lower right graph. The empirical CDF, $F(X)$ is defined as the proportion of error values less than or equal to x. This plot is useful for examining the distribution of the error values. The overlaid theoretical CDF on the same plot compares the empirical distribution of the sample to the theoretical distribution. Alternatively, one can utilize other normality tests to verify

the probability distribution [111]. By estimating the distribution parameters, it was determined that $DI \approx N(\mu,\sigma)$ where $\mu = 1.59$ and $\sigma = 0.564$. The equation for DI is given in (5-5). Knowing the DI values and the distribution parameters, it is time to formulate the change detection problem and establish the alarm values through a statistical approach. The formulation of the change detection problem follows and the latter is discussed later in section 5.3.7.

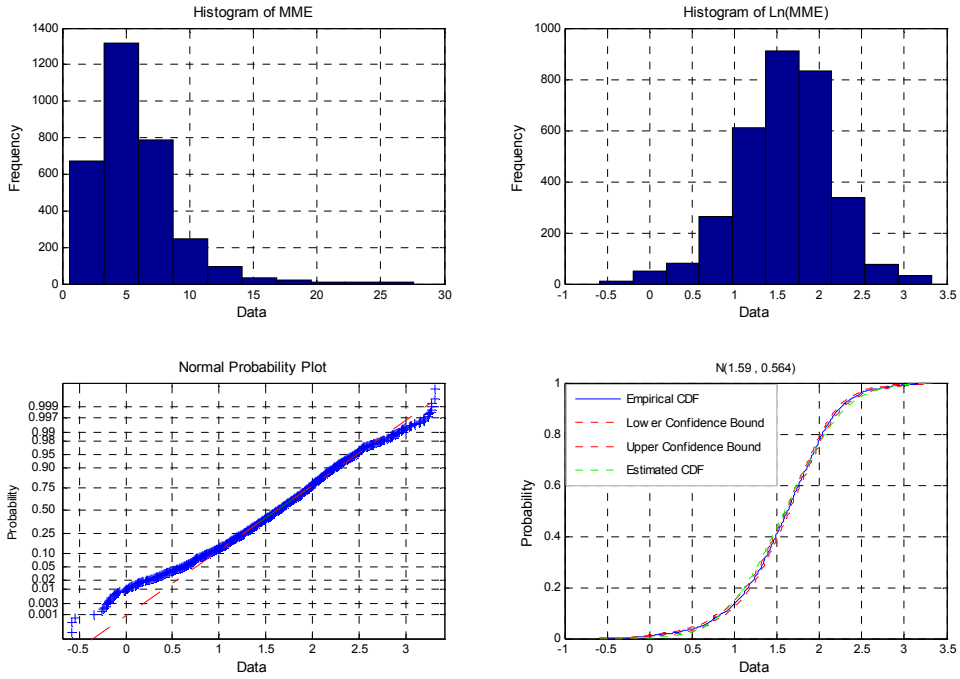


Fig. 5.5: Probability distribution of the MME and DI

5.3.3 Formulation of the Detection Problem

As discussed in the characterization of incipient faults, the nature of degradation phenomena is normally gradual and random. This nature of incipient faults causes the detection index values to gradually shift upwards and go through a monotonic process. When there is an

active degradation process in the underground cable, the new patterns exposed to the SOM model show less similarity to the prototype vectors. As the degradation process moves into different stages and deviates from its present status incrementally, the dissimilarity increases translating into a monotonic upward drift in the mean values. The non-decreasing mean path of the detection index captures the trend of changes in the states of cable.

Formally speaking, assume that the mean values ($\mu_1, \mu_2, ..., \mu_i, ..., \mu_n$) of the detection index undergo a non-decreasing process where $\mu_1 \le \mu_2 \le ... \le \mu_i \le ... \le \mu_n$. Estimation of the mean path from DI values amounts to finding the least square fit for the restricted regression problem defined in (5-7).

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\| DI_{ij} - \mu_i \right\| \qquad \text{(5-7)}$$
$$\textit{subject to } \mu_1 \le \mu_2 \le ... \le \mu_n$$

where n is the number of subgroups of the DI values and m_i is the number of DI values in subgroup i. This estimation problem can be solved using a numerical procedure in restricted regression called the Pool-Adjacent-Violators Algorithm (PAVA) [112]. This algorithm provides the maximum likelihood estimate (MLE) of the mean path under monotonicity restrictions.

Fig. 5.6 illustrates this procedure using the field data. Fig. 5.6 (a) shows the DI values in chronological order along with the MLE of the detection index. The SOM model used for this study was established through the data used in Fig. 5.5. The non-decreasing MLE of the mean path was computed using the PAVA. For clarity, Fig. 5.6 (b) depicts the same estimated mean path on a separate graph. As can be seen, the mean path is a monotonic and non-decreasing process undergoing gradual drifts.

Using the detection index values, the incipient fault detection problem is translated into a change detection problem. The detection index i.e. the natural logarithm of the modeling errors defined in (5-5) is viewed as a sequence of independent random numbers $X_1, X_2, ...$ with $X_i \approx N(\mu_i, \sigma)$ in which the means $\mu_1 \le \mu_2 \le \mu_3 \le ...$ are monotonic and nondecreasing. The objective is to detect when the mean of the DI crosses a specified threshold δ. This important step is accomplished through modified versions of three change detection algorithms, namely

CUSUM, EWMA, and GLR. Before discussing each method, a general overview of the statistical change detection terminologies and methods seem necessary.

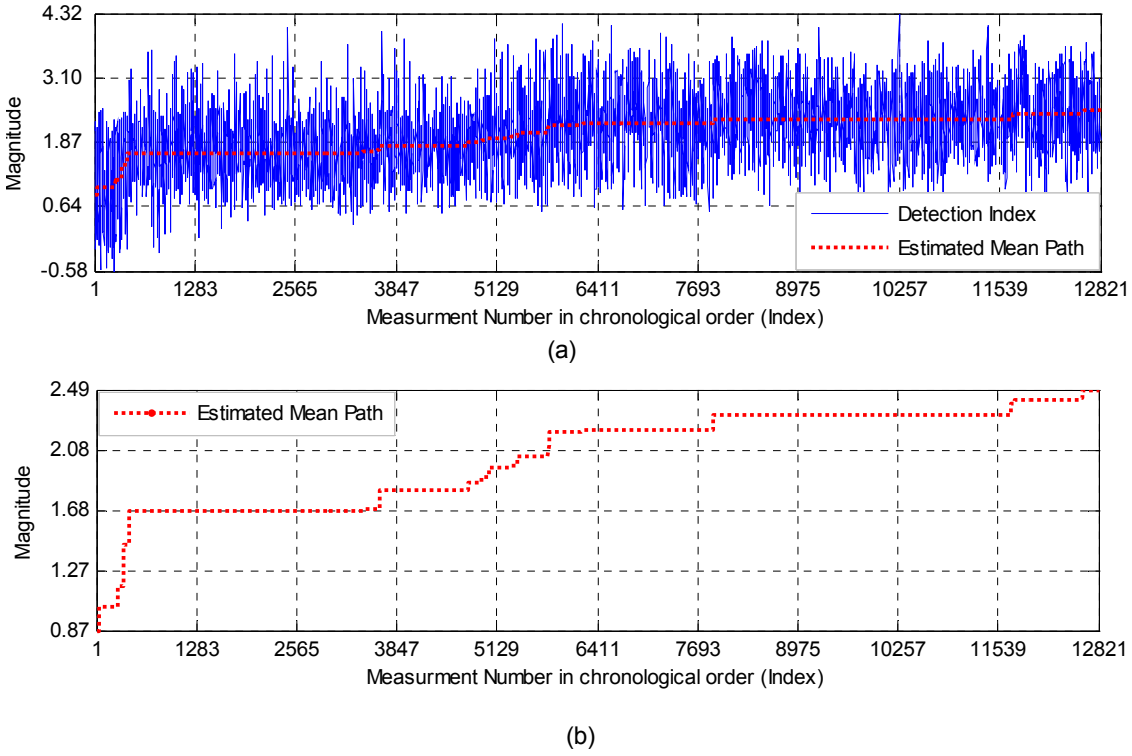


Fig. 5.6: Detection index and corresponding estimated mean path

5.3.4 Statistical Change Detection

Change detection problems are encountered in many applications such as quality control, signal segmentation, fault detection, and monitoring. Unlike the time series analysis in which the characteristics of the monitored data are either constant or slowly time-varying, the change detection methods model the data with parameters that are subject to change at unknown time instants. The basic idea of interest in change detection problems is, therefore, to detect whether there is a change in the characteristics of the data and estimate the change time.

The complexity of a change detection problem varies by the number of parameters and the need for estimation of the change time. In some applications, the model parameter is multidimensional and the change time needs to be estimated. These types of problems utilize various modeling methodologies such as state space models or multidimensional statistical models to overcome the complexities involved. On the other hand, in a number of applications, the changing parameter is a scalar and the change time need not be estimated. The parametric statistical tools are instrumental in such problems that are mostly encountered in industrial applications [113].

In addition to the dimension of the parameter, the nature of the change is of great importance. Two main problems concern additive and nonadditive changes. Additive changes assume that the change is in the mean value of the sequence of observations, whereas nonadditive changes occur in the variance, correlations, spectral characteristics, and dynamics of the signal. The complexity of change detection problem escalates if the nature of changes is of nonadditive type. The underlying concept in change detection methods based on mathematical statistics is to view the samples of measurements as a realization of a random process. The change detection can be technically performed in an online or offline fashion, however, this work is primarily concerned with online detection methods since the principle goal is to anticipate the fault. Historically, the online change detection was applied to the area of quality control, where control charts were introduced in 1931 and then cumulative sums charts in 1954 [113].

5.3.5 Online Change Detection Terminologies

The online change detection problem applied to this work treats the detection index values as a sequence of independent random variables with a probability density function that has one scalar parameter denoted by θ. Let $(y_k)_{1\le k\le n}$ be such a sequence with conditional density $p_\theta(y_k|y_{k-1},...,y_1)$. At the unknown change time t_0, the conditional density parameter θ migrates from its constant value of θ_0 to a different value of θ_1 and the objective is to detect the change occurrence at t_a with minimum delay $t_a - t_0$ where t_a is the alarm time at which the change is detected [113].

Ordinarily, the estimation of parameters θ_0, θ_1, and knowledge of the actual change time t_0 are not required. In some cases, one or all of these parameters are known and in that case such prior information simplifies the detection algorithm, relatively. The situation where none of the parameters are known is the most interesting from a practical point of view, but also the most complex one from the design and implementation point of view. It is implicitly assumed that only one change at a time is considered and multiple change times are sufficiently away from each other so that the assumption is held. The detection is accomplished by a stopping rule generally defined in (5-8) where *sup* represents the extrema with respect to a real value, δ is the threshold, and $(g_n)_{n\geq 1}$ is a family of functions of n coordinates [113].

$$t_a = \sup\{n : g_n(y_1,...,y_n) \geq \delta\} \tag{5-8}$$

A common change detection problem encountered in industrial plants is known as the online quality control. The measurements are assumed to follow a normal distribution with μ_0 and σ_0 parameters under normal operating conditions. During faulty conditions, two basic types of changes can take place: a change in the mean or a change in the variance. Fig. 5.7 shows a change in the mean value of normally distributed measurements. Fig. 5.8 shows another case where there is a change in the variance at constant mean. Evidently, the incipient fault detection problem is of type I, namely a change in the mean. The problem is to develop a decision rule that detects these changes and issues an alarm signal notifying the operator of the faulty conditions. The decision rules are provided through the change detection algorithms.

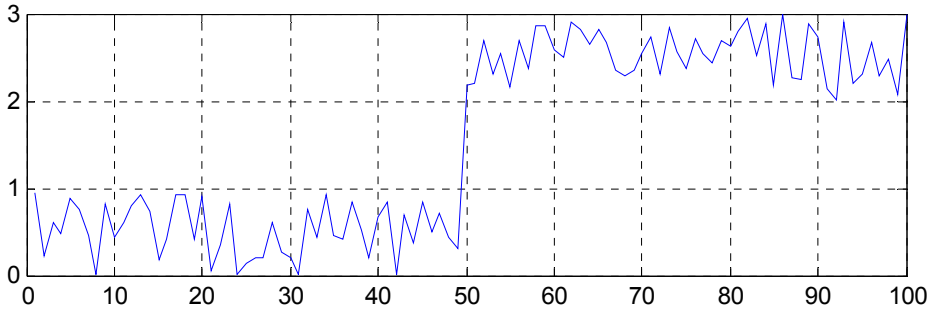


Fig. 5.7: Increase in the mean at constant variance

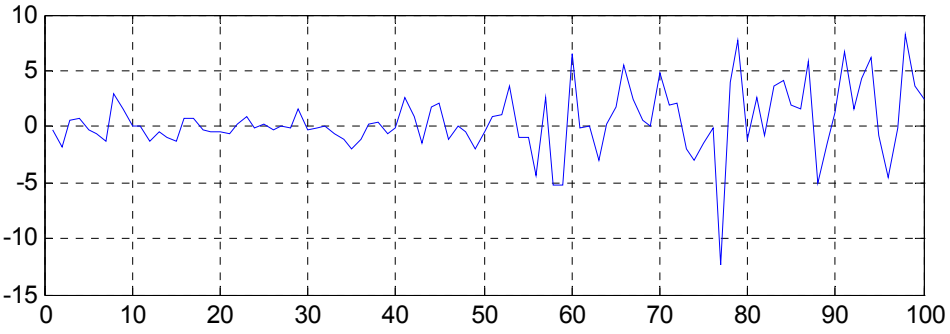


Fig. 5.8: Increase in the variance at constant mean

The change detection algorithms make use of an important property of the log-likelihood ratio defined in (5-9). The key property of the log-likelihood ratio is that a change in the parameter translates into a change in the sign of the mean value of the log-likelihood ratio [113]. Formally speaking, if E_{θ_0} and E_{θ_1} denote the expectations of the random variables under the two distributions p_{θ_0} and p_{θ_1}, respectively; then (5-10) holds true.

$$s(y) = Ln\frac{p_{\theta_1}(y)}{p_{\theta_0}(y)} \tag{5-9}$$

$$\mathrm{E}_{\theta_0}(s) < 0 \;\; and \;\; \mathrm{E}_{\theta_1}(s) > 0 \tag{5-10}$$

The performance of change detection algorithms can be evaluated using a number of indices such as the probability of false alarm and the mean delay of detection. Assume that t_0 is the actual change time and that a detection algorithm indicates the crossing at a stopping point t_a such that $t_a = \sup\{n : g_n \geq \delta\}$, where g_n denotes the response of the detection algorithm. Depending upon the relative values of the two times, three possible scenarios are possible.

- If $t_a < t_0$, then the detection algorithm has made a false decision and $P\{t_a < t_0\}$ represents the probability of false alarm.
- If $t_a > t_0$, then the detection algorithm has detected the change with a delay equal to $t_a - t_0$. Thus, the expected delay is represented by $E(t_a - t_0 \mid t_a > t_0)$.

- If $t_a = t_0$, then the detection is accomplished just on time, and hence the probability of false alarm and expected delay are both zero.

Intuitively, a change detection algorithm with a small probability of error and minimum detection delay is desired. The best solution is the quickest detection of the change with as few false alarms as possible but minimizing the delay increases the false alarm rate. Frequent false alarms are inconvenient due to the costs associated with stopping the service and pinpointing the origin of fault. Also, from a psychological point of view, frequent false alarms make the operator less sensitive about the true alarms over time. Therefore, the optimal solution is usually a tradeoff between fast detection and few false alarms considering the costs associated with the two objectives. Another desired characteristic of change detection algorithms is the robustness in the presence of noise conditions and modeling errors. The detection algorithms that are robust and easy to tune on a new signal are practically preferred.

5.3.6 Change Detection Algorithms

The algorithms applied to the change detection problem include the cumulative sum (CUSUM), exponentially weighted moving average (EWMA), and generalized likelihood ratio (GLR) [113][114][115]. The original versions of these algorithms are designed to indicate a change in the mean when it jumps from one level to another level during which it crosses the threshold. The change detection problem in this work does have a different nature as follows.

As discussed earlier, degradation and development of incipient faults are gradual in nature and do not occur suddenly. The incipient activity might persist in the cable for a few days, months, or even years before it actually develops into a short circuit fault. Realistically, the monitored degradation parameter may experience a gradual incremental increase over the course of incipient activity before a catastrophic failure initiates. The parameter may jump to one level, stay constant for a while, and jump to another level, and so on. The challenges involved is that the original change detection algorithms that are designed to detect a single jump from one level to another need to be modified so that the process continues its monotonic trend until a prescribed threshold is crossed.

The mission of the IFDS is then to monitor the condition of the system as long as the monitored parameter is below a specified threshold. Once the threshold is passed, the system must detect this violation in time so that corrective and preventive actions can take place. In the

following sections, the detection algorithms investigated in this work are introduced and tailored to solve the detection problem using the instantaneous values of the DI.

5.3.6.1 Cumulative Sum Algorithm

The Cumulative Sum (CUSUM) algorithm was first proposed by Page in 1954 for quality control applications when the parameter θ experiences a single jump from one level to another [113]. The decision function g_k is the difference between the current value of the cumulative sum of the log-likelihood ratio S_k and its current minimum m_k [113]. Mathematically, $g_k = S_k - m_k \geq h$ where h is a specified alarm value and

$$\begin{aligned} S_k &= \sum_{i=1}^{k} s_i \\ s(i) &= Ln\frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)} \\ m_k &= \min_{1\leq j\leq k} S_j \end{aligned} \tag{5-11}$$

The stopping time given in (5-12) is reached when the decision function S_k exceeds the alarm value h or equivalently the cumulative sum crosses an adaptive threshold of $m_k + h$. Although h is a constant, the sum of $m_k + h$ varies at every step. m_k is updated online and thereby the information manifested in the past observations is preserved.

$$t_a = \min\{k : S_k \geq m_k + h\} \tag{5-12}$$

Relevant to the change detection problem, assume that a Gaussian sample of size N is given and the mean of the underlying Gaussian distribution μ changes from μ_0 to μ_1 at t_0 while the variance σ^2 stays constant. The probability density and the log-likelihood functions are respectively given by (5-13) and (5-14).

$$p_\mu(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \tag{5-13}$$

$$s_i = \frac{\mu_1 - \mu_0}{\sigma^2}\left(y_i - \frac{\mu_0 + \mu_1}{2}\right) \tag{5-14}$$

Hence,

$$S_1^N = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^{N} (y_i - \frac{\mu_0 + \mu_1}{2}) \tag{5-15}$$

The cumulative sum function typically shows a decreasing trend before change and an increasing one after change. In one of the classic well-known quality control algorithms called Shewhart Control Charts [113], samples of fixed size N is taken and a decision is made to choose one of the two potential hypotheses about the parameter using the cumulative sum function. The null hypothesis H_0 holds when there is no change and the alternative hypothesis H_1 prevails when the parameter changes.

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu = \mu_1 \end{aligned} \tag{5-16}$$

The process is called "in-control" as long as the decision is made in favor of H_0. Once the decision rule indicates that H_1 holds true, the monitoring stops and a change point is detected. The optimal decision rule using the alarm value h is expressed in (5-17).

$$d = \begin{cases} 0 & if\ S_1^N < h \qquad H_0\ is\ chosen \\ 1 & if\ S_1^N \geq h; \qquad H_1\ is\ chosen \end{cases} \tag{5-17}$$

The original CUSUM algorithm designed for a single jump is not directly applicable to the incipient fault detection problem and needs to be revisited and modified. The goal is to keep the process somewhere below a threshold or "in-control". In this problem, there is no target value for the change parameter. This algorithm in more general situations is applied by using the following recursion form [114].

$$W_n = Max(0, W_{n-1} + X_n - r) \tag{5-18}$$

where W_n is the CUSUM computed at sample n, X_n is the n^{th} observation, and r is a user-defined reference value. In the incipient fault change detection problem, the natural selection of r amounts to δ, the starting point $W_0 = 0$, and X_n is the n^{th} instantaneous value of the DI.

The response of the CUSUM applied to the detection index values computed from data used in Fig. 5.6 is depicted in Fig. 5.9. As seen, unlike the original CUSUM algorithm, the modified

function remains around zero as long as the mean of the detection index is below the threshold and starts growing once the threshold is passed. An alarm signal is issued as soon as the CUSUM response crosses a pre-specified alarm value which is 19.77 in this case.

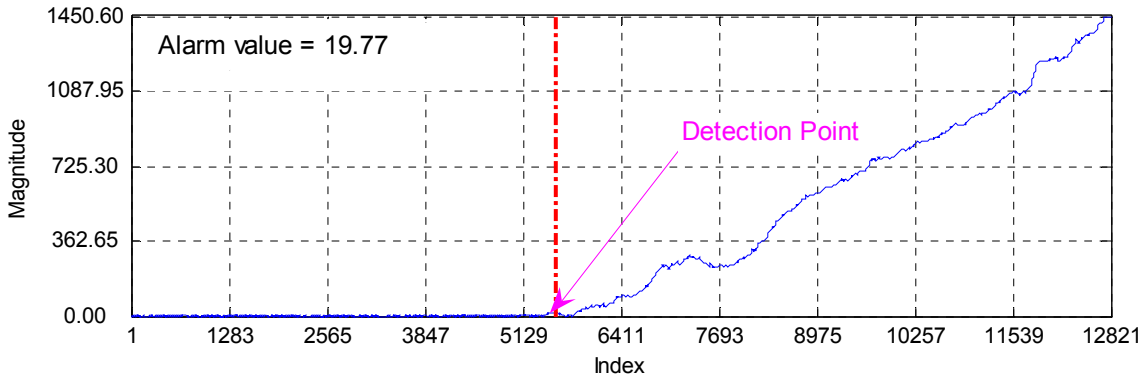


Fig. 5.9: CUSUM response

5.3.6.2 Exponentially Weighted Moving Average Algorithm

The Exponentially Weighted Moving Average (EWMA) algorithm uses a finite set of weights to emphasize differently on past and present observations. A variant of this statistic formulated for the incipient fault change detection problem in recursive form is given in (5-19) [114].

$$E_n = Max\{\beta, (1-\lambda)E_{n-1} + \lambda X_n\} \tag{5-19}$$

where β is a reflecting barrier for the change in one direction and λ is the smoothing parameter lying between 0 and 1. The higher the smoothing parameter, the greater emphasis is placed on the recent observations. An alarm is raised at time $t_a = Min\{n : E_n \geq h\}$ for some specified h. There are three parameters to specify in the EWMA algorithm, namely $\lambda, \beta,$ and E_0. Unlike CUSUM, there is no unique choice for these parameters by statistical theory, but the findings are as follows.

The values of λ in the range of 0.005 ± 0.001 was found to perform well. The choice of β is determined more subjectively using two standard deviations below the threshold led to good performance. The natural choice for the starting point E_0 in the original quality control applications is the specified target value. As in the incipient fault change detection problem, there is no notion of a single target value, E_0 was intuitively set equal to β. With these settings, the response of the EWMA to the sequence shown in Fig. 5.11 is depicted in Fig. 5.10 in which $\lambda = 0.0052$ and β is two standard deviations below the threshold of 2.1. Unlike the other two change detection algorithms, in EWMA the response is a smoothed version of the detection index values. As soon as the smoothed sequence passes the threshold, a change point is detected.

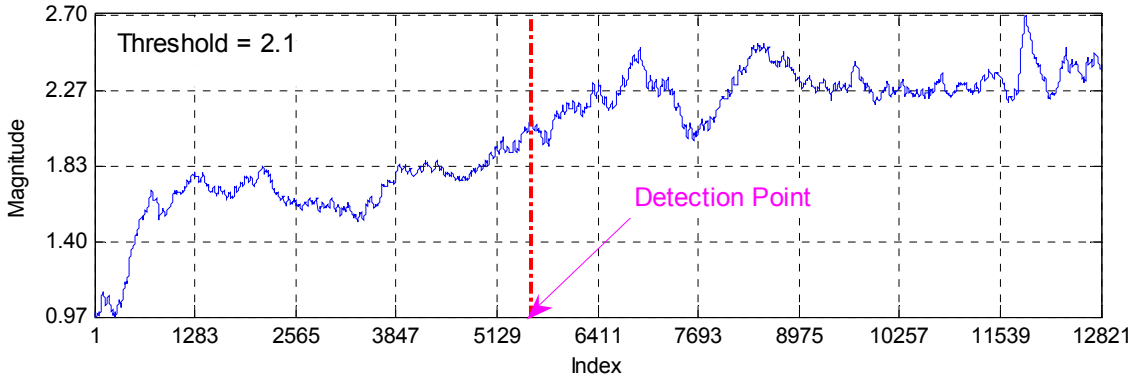


Fig. 5.10: EWMA response

5.3.6.3 Generalized Likelihood Ratio Test

The generalized likelihood ratio test is a general method used extensively in detection problems. For the incipient fault change detection problem, the null hypotheses is $H_0 = \{\mu_1 \le \mu_2 \le ... \le \mu_n \le \delta\}$ and the alternative is $H_1 = \{\mu_1 \le \mu_2 \le ... \le \mu_n, \mu_n \ge \delta\}$. Therefore, using the GLR algorithm, hypothesis testing is performed at every step and the

detection is completed once the decision is made in favor of the alternative [114]. The log-likelihood ratio statistics is given in (5-20).

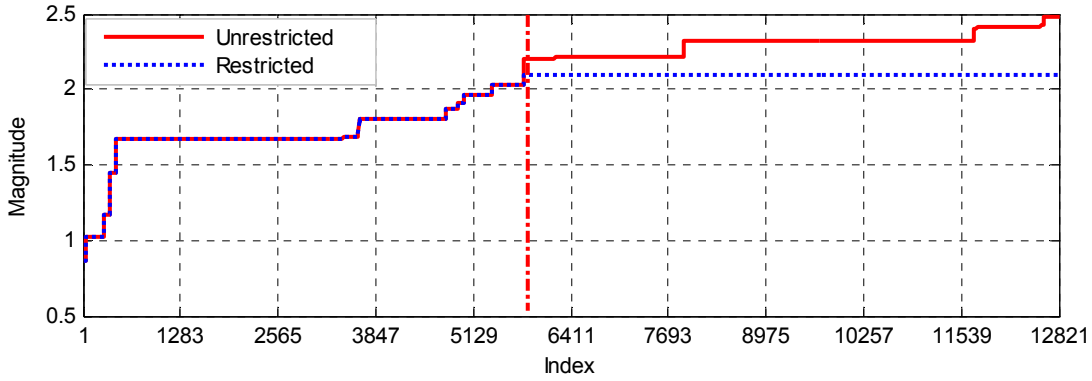
$$\begin{aligned}&\underset{\mu\in H_0\cup H_A}{Max}\left\{Ln(p(X|\mu)\right\}-\underset{\mu\in H_0}{Max}\{Ln(p(X|\mu)\}=\\&Min\left\{\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu_i)^2\middle|\mu_1\le\mu_2\le\ldots\le\mu_n\le\delta\right\}-Min\left\{\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu_i)^2\middle|\mu_1\le\mu_2\le\ldots\le\mu_n\right\}\end{aligned} \tag{5-20}$$

This optimization problem gives rise to two least-square regressions under monotonicity restrictions, a problem known as isotonic regression [112]. The restricted regression Z corresponds to hypothesis H_0 and can increase up to δ. The unrestricted regression Y corresponds to the $H_0 \cup H_A$ hypothesis. These two regressions provide the maximum likelihood estimates under the two hypotheses.

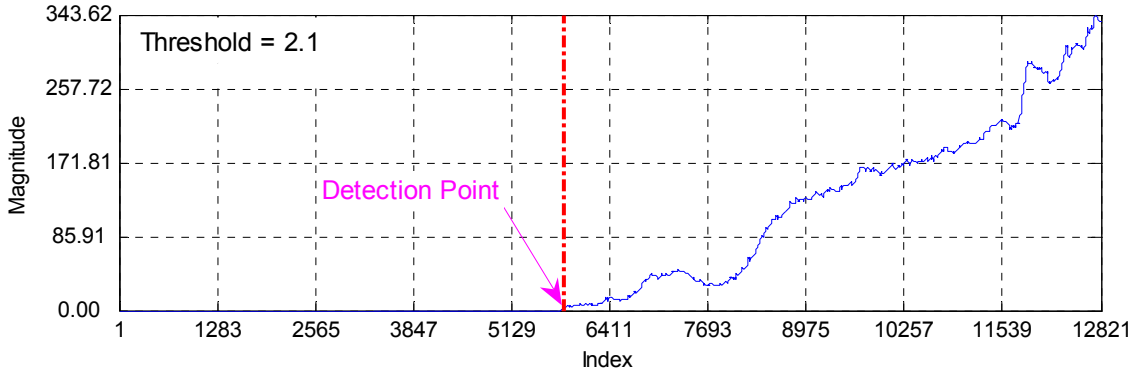
The solution to find the unrestricted regression Y is provided by the pool-adjacent-violators algorithm (PAVA) [112]. Having computed Y at every step, the computation of Z is trivial as $Z_i = \min(Y_i, \delta)$. The GLR test corresponding to the normally distributed DI values is provided by evaluating M_n which is defined in (5-21) [114].

$$M_n=\sum_{i=1}^{n}(X_i-Z_i)^2-\sum_{i=1}^{n}(X_i-Y_i)^2=\begin{cases}0 & if\ Y_n\le\delta\\ \sum_{i=J}^{n}(X_i-\delta)^2-\sum_{i=J}^{n}(X_i-Y_i)^2 & if\ Y_n>\delta\end{cases} \tag{5-21}$$

where $J=\min\{i: Y_i>\delta\}$. Once M_n crosses the alarm value h, the process stops and the violation is detected at $t_a=\min\{n: M_n\ge h\}$. The two regressions and the response of the GLR algorithm corresponding to the detection index values in Fig. 5.6 are shown in Fig. 5.11 . The threshold for this experiment was set at 2.1. As seen, once the threshold is passed the difference between the two regressions grow and indicate that a change has taken place.



(a) two isotonic regressions



(b) GLR response

Fig. 5.11: GLR regressions and response

5.3.7 Determining Alarm Values through a Probabilistic Approach

The change detection problem requires that an alarm value (h) be specified given an assumed probability of false alarm. This can be carried out using the following probabilistic approach. If the probability of false alarm is assumed as 5%, then it follows that $P\{t_a < t_0\} = P\{\max_{1\le k\le t_0-1} g_k \ge h\} = 0.05$. In this formula, g_k represents the output of each detection algorithm. For CUSUM algorithm, $g_k = W_k$. By knowing the distribution of $\max_{1\le k\le t_0-1} g_k$,

the alarm value h may be estimated by finding the 95th percentile of that distribution. If the mean behavior is known, one can produce a large number m of mean paths through computer simulations and the corresponding detection index values, and calculate $Z_j = \max\{g_1, g_2, ..., g_{t_0-1}\}$ for $j = 1, ..., m$. The alarm value is the order statistic $Z_{(0.95m)}$, i.e. $h = Z_{(0.95m)}$ [114].

In order to generate a large number of simulated detection index values, the probability of a jump and the distribution of jump heights and residuals are needed. Assuming a geometric distribution for the times between jumps, the probability of a jump can be calculated by (5-22).

$$\Pr(Jump) = \frac{\text{Total number of jumps in the PAVA means}}{\text{Total number of time periods}} \tag{5-22}$$

The geometric distribution defined by (5-23) is a discrete distribution defined for nonnegative integers [110]. It is useful for modeling the runs of consecutive successes - a jump in this application - in repeated independent trials of a system.

$$pdf(x) = p(1-p)^n = pq^n \tag{5-23}$$

where $0 < p < 1, \quad q = p - 1$.

A good approximation for the jump heights is an exponential distribution. The exponential distribution is the only memoryless random distribution, and is defined in (5-24) where μ is the mean parameter of the distribution. For the example detection index values depicted in Fig. 5.6, the parameter was determined to be 0.062. The corresponding jump values are shown in Fig. 5.12(a). The empirical and theoretical CDF curves are given in Fig. 5.12(b). The dotted lines specify the upper and lower confidence bounds which contain the CDF curves indicating the acceptable goodness of the fit.

$$pdf(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \tag{5-24}$$

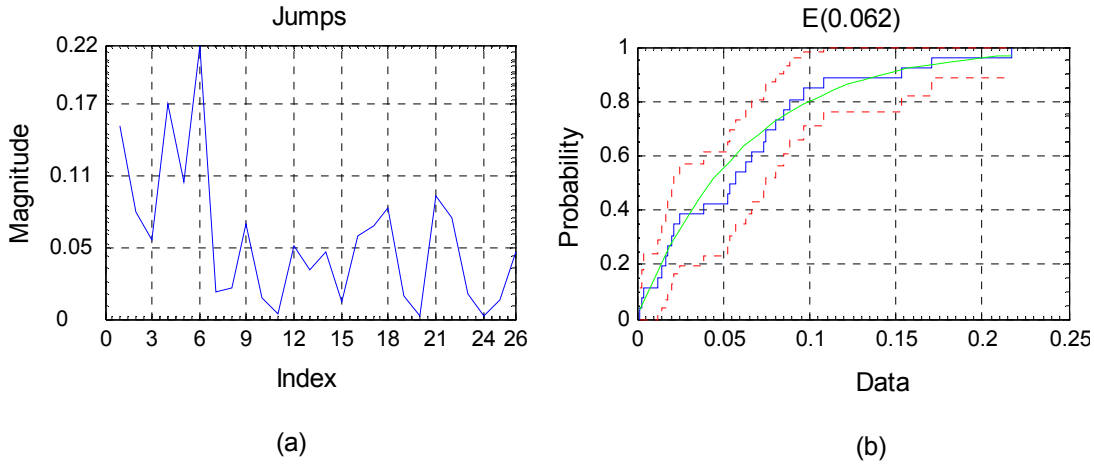


Fig. 5.12: Jumps and the distribution of jumps

Residuals are defined as $r_i = X_i - \hat{\mu}_i$, where X_i is the detection index value at time i and $\hat{\mu}_i$ is the corresponding estimated PAVA mean at time i. For the example detection index values, the statistical analysis results of the residuals are presented in Fig. 5.13. The normality property of the residuals can be confirmed by the histogram, probability plot, and the q-q- plot. A normal distribution fit results in a zero mean and a standard deviation of 0.545. Thus, the detection index values are modeled as $X_i \approx N(\mu_i, \hat{\sigma})$ in which the means (μ_i) are unknown but non-decreasing and $\hat{\sigma}$ was estimated from the residuals. Using this model, the simulated mean paths and corresponding detection index values were computed. Fig. 5.14 shows the actual PAVA path along with 10 simulated mean paths. By examination, it is seen that the adopted model simulates the original PAVA paths well.

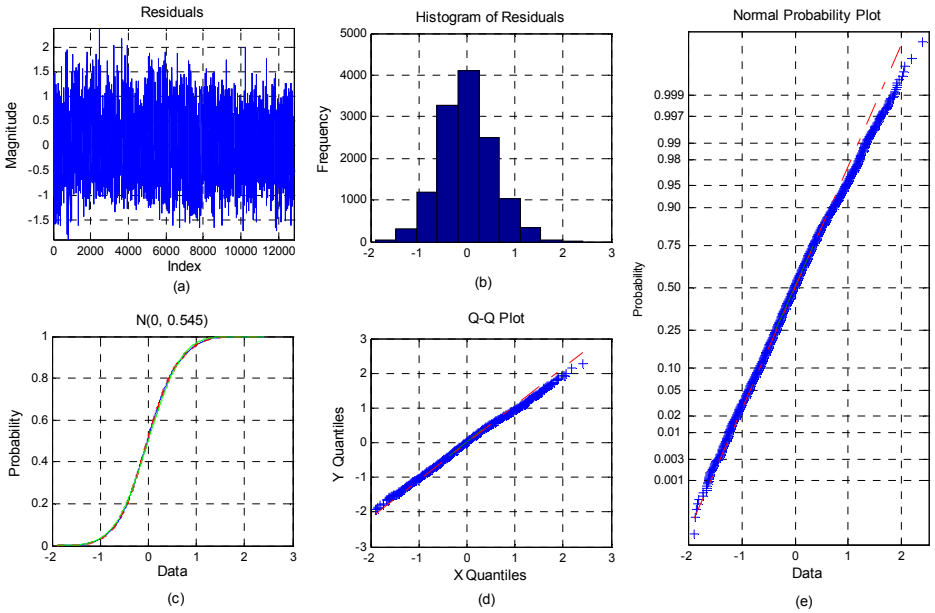


Fig. 5.13: Distribution of residuals

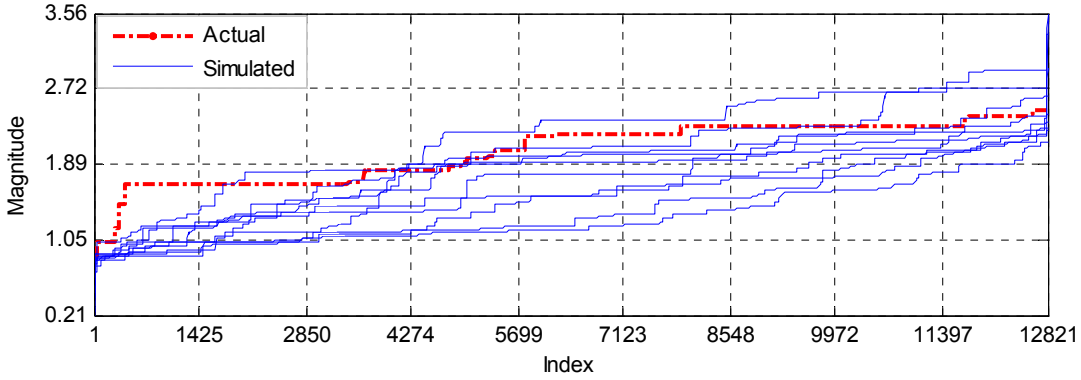


Fig. 5.14: Estimated mean path and 10 simulated mean paths from the model

5.4 SUMMARY

In this chapter, the foundation and principles of the detection functionality of the IFDS were discussed. The temporal severity measure referred to as the GSI was introduced and discussed. The detection method using SOM minimum modeling errors were presented and pertaining operational aspects of this method were discussed. Using the detection index values, the change detection problem was formulated. Three detection algorithms namely, Cumulative Sum (CUSUM), Exponentially Weighted Moving Averages (EWMA), and Generalized Likelihood Ratio (GLR) were discussed and specifically designed to solve the change detection problem.

In the next chapter, after introduction of the IFDS, its performance in detecting near failure time will be presented and discussed using data recorded from the monitoring site.

CHAPTER VI

INCIPIENT FAULT DIAGNOSIS SYSTEM AND PERFORMANCE ANALYSIS

6.1 INTRODUCTION

In this chapter, the Incipient Fault Diagnosis System (IFDS) is reviewed and its operational components are discussed. Using field recorded data, the detection functionality of the IFDS is tested and its performance is evaluated. The detection results also include the GSI application on the field data.

6.2 OVERVIEW OF THE IFDS

The developed incipient fault diagnosis system is a multi-module system implemented using MATLAB release 14. In addition to the embedded MATLAB toolboxes including the statistical toolbox, a SOM toolbox generously provided by a group of researchers in Finland [107] was utilized in the implementation of the IFDS. Additionally, the LIBSVM toolbox developed by Chang and Lin was used [98].

The overall flowchart of the IFDS and its main modules are presented in Fig. 6.1. The IFDS uses voltage and current measurements – collectively called data - as the input information, and utilizes advanced signal processing and pattern analysis techniques to classify data into various categories. In the detection phase, the classified incipient abnormality data are used to compute detection index values and through modified change detection algorithms, the threshold crossing at the specified threshold is indicated.

The data is first fed to the preprocessing module. As discussed in Chapter III, the preprocessing operations include DC removal, resampling, and denoising. The preprocessed data are then passed through the rule-based classifier. In this module as discussed in section 4.2, a set of time-domain features is computed and used by the rule-based classifier. The rule-based classifier utilizes its embedded knowledge to evaluate each rule and assign its input to one of the predetermined classes. At this stage, categories of data belonging to DC classes, various

switching type events, normal classes, and potential incipient abnormalities are distinguished. If the classification at this stage does not indicate the presence of an incipient abnormality, the capture is directly passed to the post-processing module. Otherwise, the feature extraction module operates on the data, and another set of features in time and wavelet domain is extracted. The horizontal and vertical severity index values and normalized energy features are the products of this module. Details about this module can be found in section 4.3. In the following classification module, the features are used to classify the event into one of the three classes in terms of the severity degrees, namely low, medium, and high degrees. These classes and the utilized classifiers were discussed in section 4.4.

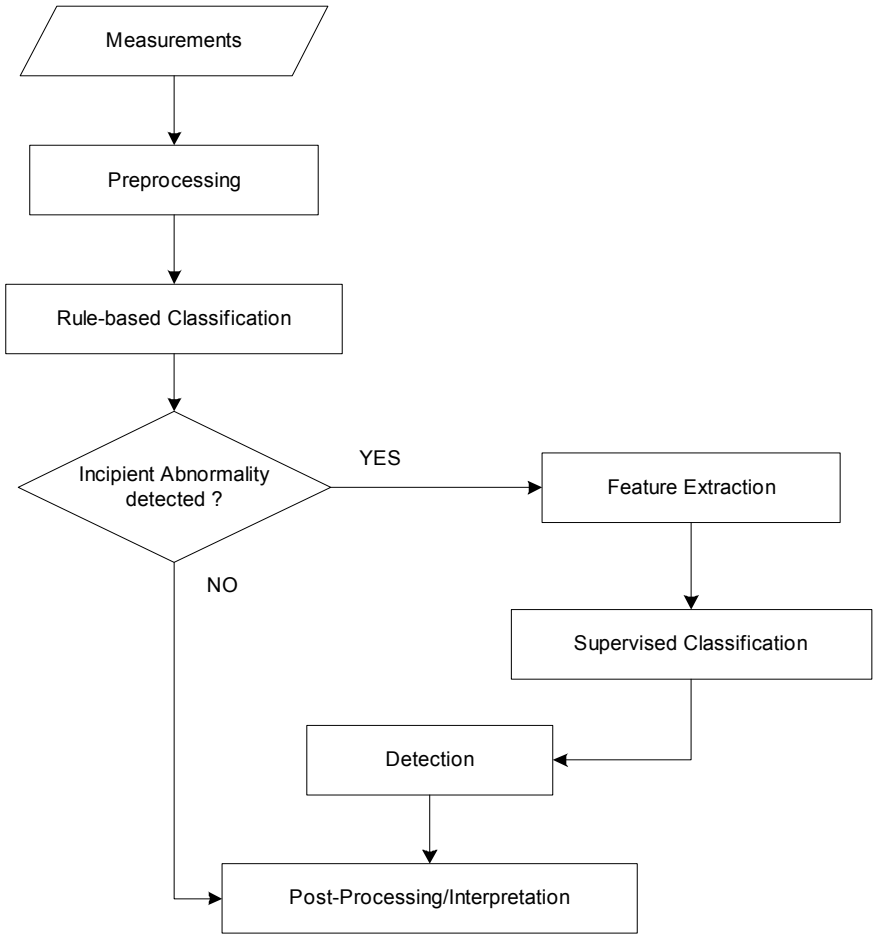


Fig. 6.1: Overall flowchart of the IFDS

In the detection phase, as presented in section 5.2, the class of incipient abnormally is used to compute the GSI values corresponding to each severity degree. The GSI graphs are used in the detection module as a complementary piece of information that indicates the trend of incipient severity in the cable at three severity level. More specifically, the graphs show the operator whether the severity degree increases over time, remains constant, or decreases. Further, the data classified to be potential incipient abnormality are fed to the detection method. At this stage, the detection index values are computed and the detection algorithms are utilized to solve the change detection problem. The algorithms include CUSUM, EWMA, and GLR. Further information about the detection method can be found in section 5.3.

After all computations in the detection module are complete or the rule-based classifier does not classify the event as an incipient abnormality, the post-processing operations are initiated. The results are combined and interpreted in the post-processing module and the system outputs its results to the user. For each capture, the important outputs include the date and time of the capture, the classification label from each classifier, the global severity index for each severity degree, the current value of the detection index, and lastly, whether or not a threshold crossing has occurred. A graphical representation of the IFDS output is shown in Fig. 6.2.

Notice that the GSI value shown corresponds to the high severity class and the threshold-crossing output corresponds to the CUSUM algorithm. Essentially, there are three threshold-crossing outputs for each algorithm and three GSI values for each severity degree. The IFDS output includes a number of other operational quantities that provide supplemental information. The IFDS operation is sequentially repeated for each capture and can be continued as long as the detection module indicates that the cable is operating in its green zone. As soon as a threshold crossing takes places, an alarm is issued to warn the operator about the change.

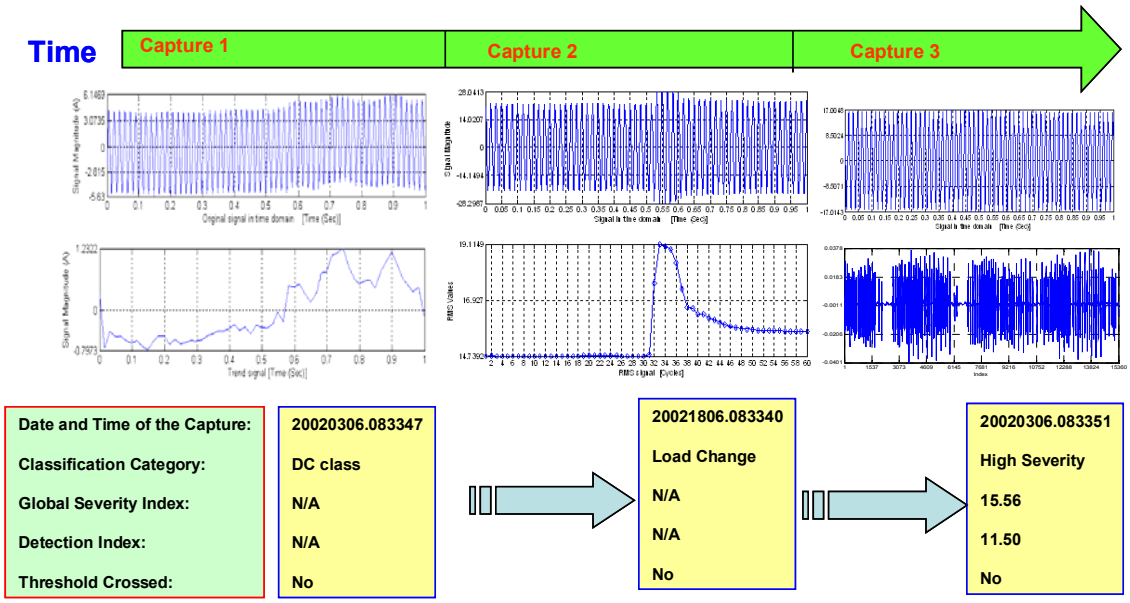


Fig. 6.2: Graphical representation of the IFDS output

The ideal operation cycle of the IFDS in three warning levels is shown in Fig. 6.3. Initially, the system needs to be calibrated. This includes training classifiers, setting alarm values and thresholds, and establishing the numerical SOM model. Then the system is ready to put into operation and make decisions once a capture becomes available. While in operation, the system uses the detection index to prompt the user upon violating the detection zones defined by the threshold values. As soon as a threshold is crossed, the IFDS indicates the violation at the alarm times defined by the corresponding alarm values at levels I, II, and III. Once alarm level III is hit, the near failure point is reached and corrective actions should be taken to prevent the imminent catastrophic failure. When the required corrective tasks are completed, the IFDS needs to be recalibrated before it resumes its operation for a new monitoring cycle. Availability of three levels of detection points helps to prioritize the maintenance tasks as the least benefit. Nonetheless, it requires that enough information i.e. a number of fault cases be available to properly set the threshold values for each level.

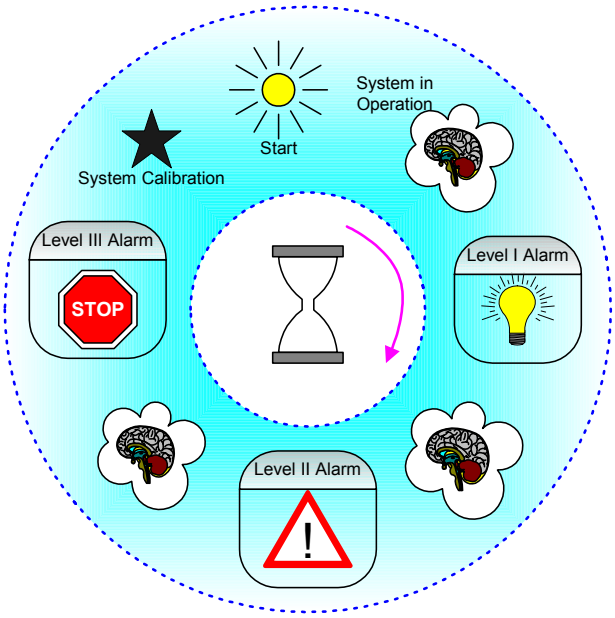


Fig. 6.3: Ideal IFDS operation cycle

6.3 DETECTION RESULTS USING FIELD DATA

In this section, the performance of the IFDS is evaluated for the faults that actually happened during the long-term monitoring. There have been several cable failures in the monitored lateral among which seven scenarios are discussed in this dissertation. These scenarios include fault cases for which the pre-fault data were available. The system was run in sequential mode to make decisions and output its detection results at three warning levels. For all fault cases, the detection results from three detection algorithms were determined and it was observed that at least one method was able to produce the desired alarm signal. The results are thoroughly presented for two scenarios namely faults on 12/24 and 06/11; for all other faults, a summary of the detection results is given and a discussion follows. The two fault cases are referred to as scenario I and scenario II, henceforth.

6.3.1 Introduction of Case Studies

The performance analysis of the IFDS is accomplished using field data recorded from the residential cable lateral shown in Fig. 6.4. Section 3.2 describes the cable lateral and provides details about the monitoring system and recorded signals. The data recorded over three years were divided into seven sets where each set concluded to a cable failure, either on the monitored lateral or on the adjacent lateral. Each set defines a case study. Table 6-1 gives information about the faults and number of measurements for each case. As indicated, two out of seven faults included in the table happened in the adjacent laterals under normal weather condition.

Fig. 6.5, Fig. 6.6, Fig. 6.7, and Fig. 6.8 show plots of phase current signal on the fault day for cases I through VII, respectively. The values shown in the plots were obtained by averaging the phase current measurements over a 15-minute window. As it can be seen, when the fault was on the monitored lateral, the phase current goes down to zero and stays around zero for the duration of the fault. For faults on the adjacent laterals, cases V and VI, however, the phase current signal undergoes a change but not a radical one.

To prepare and calibrate the IFDS, the supervised classifiers were trained using 4000 data patterns. The SOM models were established for each case according to the principles described in section 5.3 using 25% of the available incipient data samples. Notice that the models need to be updated every time there is a change in the cable lateral or a fault occurs. Furthermore, the GSI computation for each case requires measurements at fixed intervals and no gaps should exist in the data.

For each case, adaptive thresholds were determined consistently as follows. The minimum threshold or the warning level 1 threshold was established through the UMP test [116] at a significance level of 1% and found to perform well. The threshold for the third warning level or the maximum threshold was computed by establishing a significance level of 4% in rejecting the null hypothesis i.e. the hypothesis that assumes the data come from the SOM model distribution. Finally, the second warning level threshold was simply computed by taking the average of the minimum and maximum threshold values

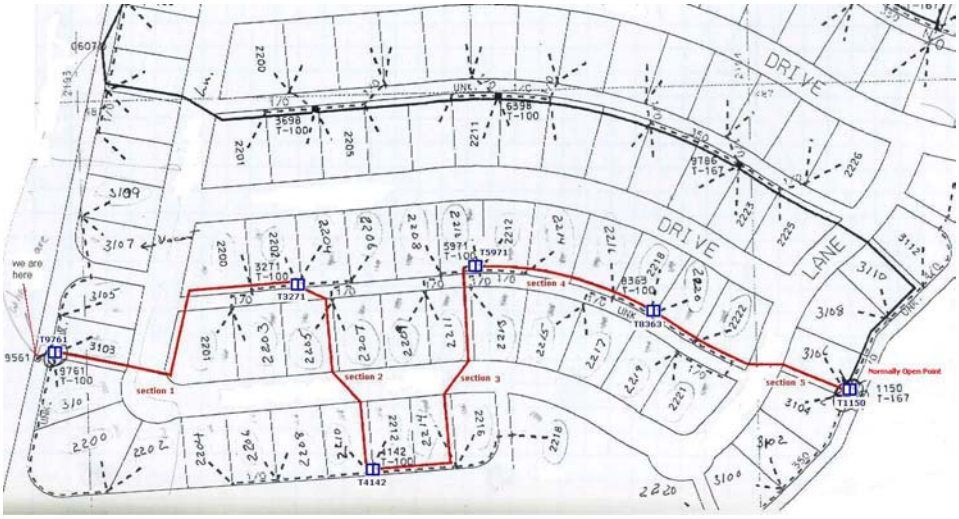
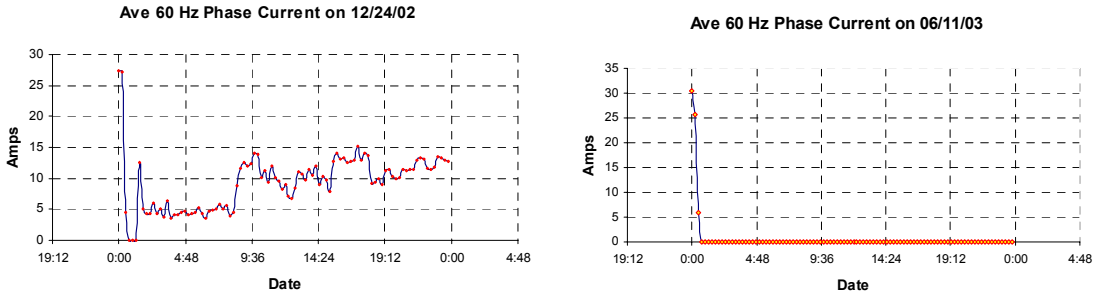


Fig. 6.4: Monitored underground cable lateral

Table 6-1:
IFDS case studies

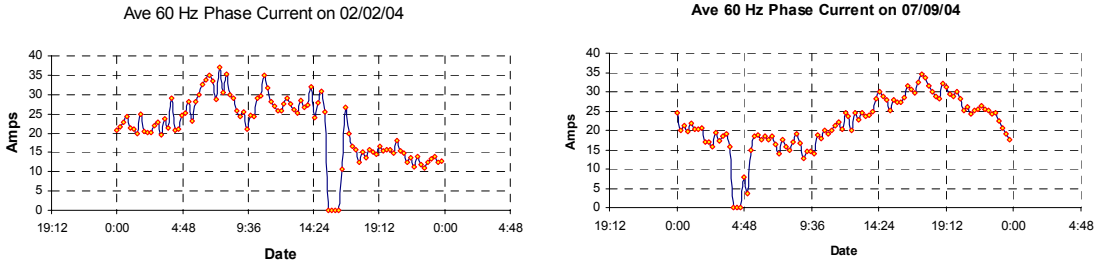
Case	Fault Date	Duration (min)	Weather	Location	Number of Measurements
I	2002-12-24	68	Adverse	ML	18278
II	2003-06-11	79	Adverse	ML	16109
III	2004-02-04	68	Normal	ML	9162
IV	2004-07-09	65	Normal	ML	2632
V*	2004-09-26	78	Normal	AL	7492
VI*	2004-10-05	79	Normal	AL	762
VII	2004-10-12	375	Normal	ML	571

ML: Monitored lateral
AL: Adjacent laterals
* : fault on the adjacent laterals



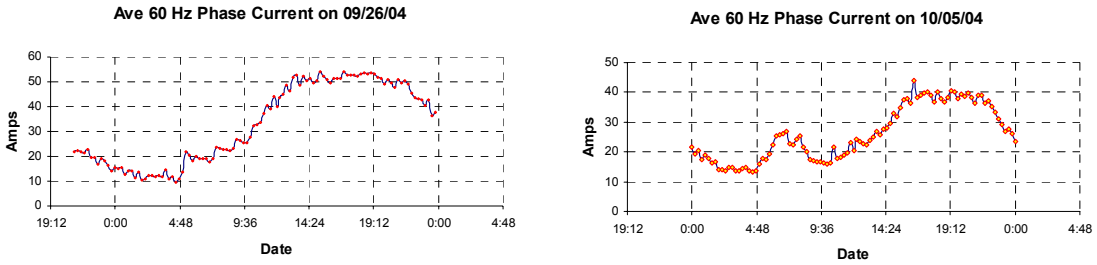
(a) case I (b) case II

Fig. 6.5: Average phase current in fault cases I and II



(a) case III (b) case IV

Fig. 6.6: Average phase current in fault cases III and IV



(a) case V (b) case VI

Fig. 6.7: Average phase current in fault case V and VI

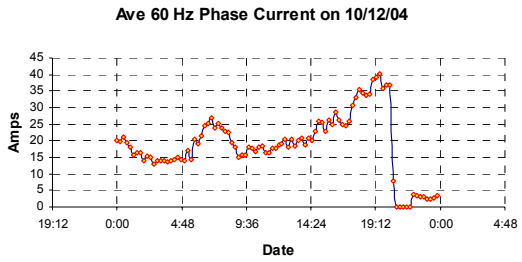


Fig. 6.8: Average phase current in fault case VII

6.3.2 Application of the GSI as a Temporal Severity Measure

To illustrate the application of the GSI in depicting the trend of incipient faults, the IFDS was used to classify data captured before and after faults in the system. Using the classified high-level incipient abnormalities along with their global severity indices in chronological order, the following illustrations were completed. The results for scenario I and II are presented as follows. Note that the post-fault data for scenario II were not available due to damage to the monitoring system, and thus only the plots of pre-fault interval are presented.

6.3.2.1 Scenario I

The first case pertains to a fault, which occurred on December 24 in the monitored cable lateral. The resulting GSI and its PAVA mean path for high severity degree over a 27.73-day interval before and after the fault are presented in Fig. 6.9 and Fig. 6.10, respectively. This index uses a span of 50 samples. As can be seen, there is a non-decreasing trend toward the fault time and a non-increasing one afterwards. This behavior of the GSI indicates that the high severity of captured incipient abnormalities escalates as the underground system approaches the failure time. After the failure, however, the severity of high severity incipient abnormalities lessens sharply in terms of the frequency and intensity.

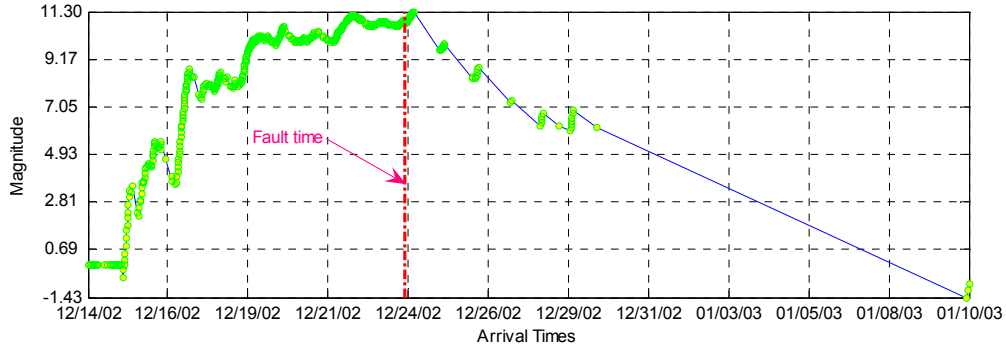


Fig. 6.9: Global severity index values in chronological order for scenario I

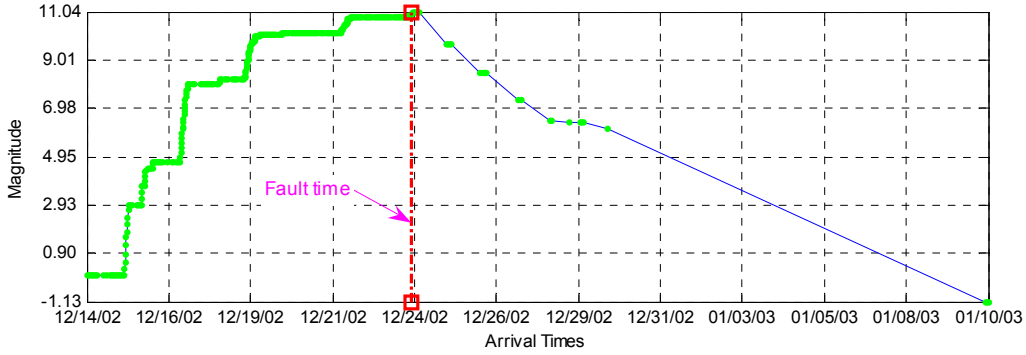


Fig. 6.10: Mean path of the global severity index in scenario I

In order to explain the behavior of the GSI during the observation interval, arrival times, interarrival times, and cumulative plots shown in Fig. 6.11 are used. The fault time is indicated by a colored dotted line terminated by squares at both ends. From the arrival times plot, it is seen that 499 incidences of high-class incipient abnormalities have been captured over a 27.73-day interval. The sharp increase of the arrival times a few days after the fault time is unique and interesting. It corresponds to the post-fault portion of the GSI plot where a decreasing trend is observable.

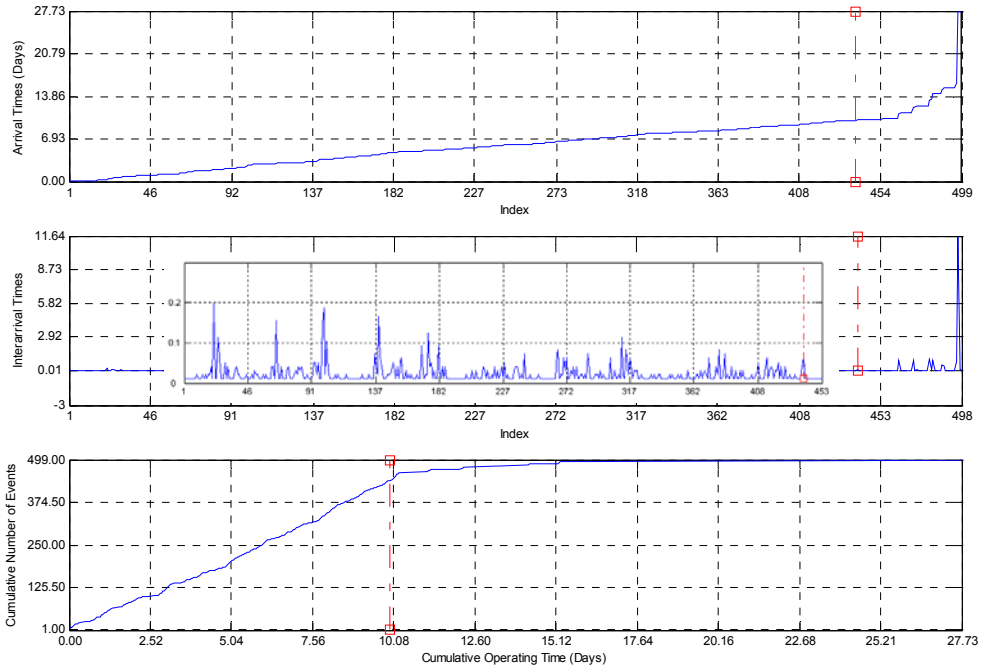


Fig. 6.11: Arrival times and the associated plots for scenario I

A better picture of the frequency of high incipient abnormalities can be observed through the plot of interarrival times. Although the dominant value of 11.64 days masks the other data points on the original plot, the behavior of interarrival times is apparent from the zoomed plot. As the fault time approaches, the spikes that represent the time between successive incipient abnormalities decease in magnitude. This behavior of the interarrival times corresponds to the increasing trend of the GSI. When the time between arrival times decreases in magnitude, the incipient abnormalities happen more frequently, and the time length of the observation window shrinks. The GSI conveys this characteristic as a trend that monotonically increases.

The lower subplot shown in Fig. 6.11 depicts the cumulative number of incipient abnormalities versus cumulative operating time expressed in days. When interarrival times tend to become larger, the system is said to be improving. Conversely, if the interarrival times become smaller, the system is deteriorating over time [62]. In other words, the cumulative plot of the number of events versus cumulative operating time concaves up when the system is

degrading and vice versa; hence, this graphical technique corresponds to the trajectory of GSI, as it tends to slope upward before the failure and bend down immediately after.

6.3.2.2 Scenario II

This case is related to another fault that occurred on June 11th of the following year. Following the same procedure as in the previous case, the resulting GSI and its PAVA mean path are depicted in Fig. 6.12 and Fig. 6.13, respectively. Similar to scenario I, the trajectory of GSI presents a non-decreasing trend toward the fault time.

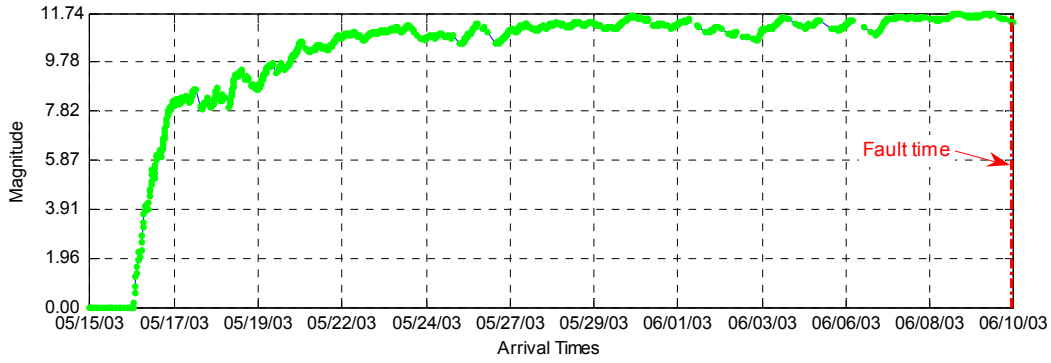


Fig. 6.12: Global severity index values in chronological order for scenario II

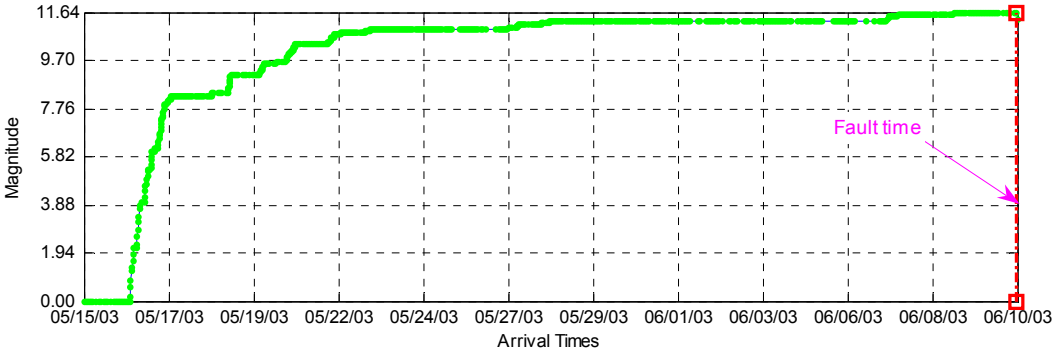


Fig. 6.13: Mean path of the global severity index in scenario II

The corresponding arrival times, interarrival times, and cumulative plots are shown in Fig. 6.14. The fault time is indicated by the colored dotted line at the end of the observation period. From the plot of the arrival times, it is seen that 459 incidences of high-class incipient abnormalities have been captured over a 11.90-day pre-fault interval. The rate of increase, i.e., the slope of the arrival times, is not constant and occasionally becomes steep. From the plot of interarrival times, it is seen that as the fault time draws near, the incipient data tends to happen more frequently, and the time between successive abnormalities shrinks. This behavior of the interarrival times corresponds to the increasing trend of the GSI. The cumulative number of incipient abnormalities versus cumulative operating time conveys a similar message as in scenario 1. When it concaves up, the underground system undergoes deterioration due to severe incipient abnormalities, and when it concaves down, the system is improving. As implied from the figure, the sequence of degrading/improving is not consistent and does not necessarily follow a special pattern. This observation is in harmony with the characteristics of incipient faults believed to be random in nature.

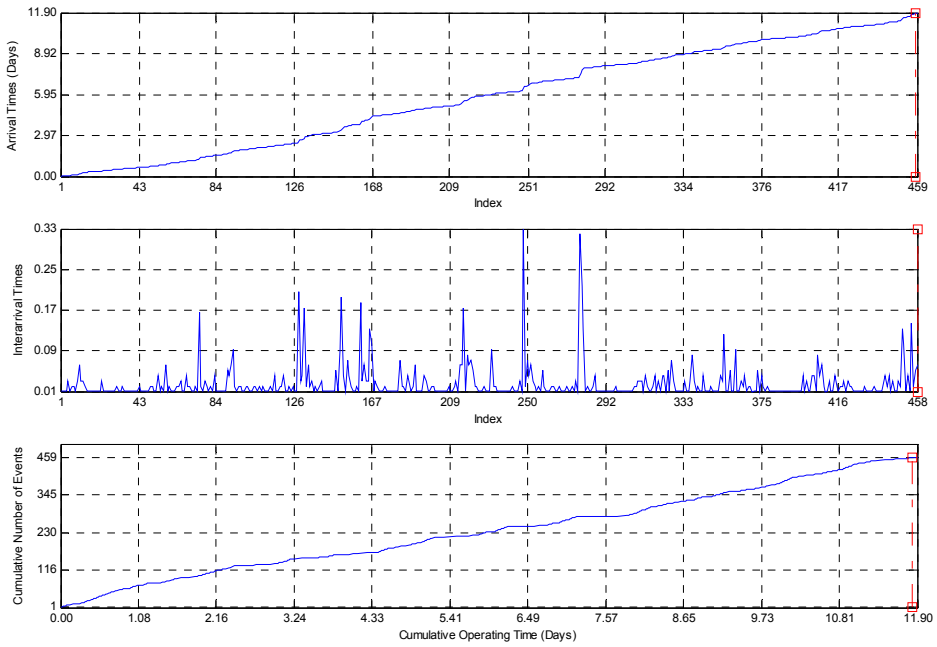


Fig. 6.14: Arrival times and the associated plots for scenario II

6.3.3 Detection Results Using SOM Based Detection Index

In this section, the detection results using the SOM-based detection index is given. The results for two scenarios are provided in detail, but for all other fault cases, a summary of the detection results from the three algorithms is provided.

6.3.3.1 Scenario I

The resulting detection index values for this scenario are graphed in Fig. 6.15 along with the plot of the mean path as a dotted line. The mean path was obtained using the PAVA algorithm. From this graph, it is seen that the mean value of the detection index is a non-decreasing monotonic process stepping up at some irregular times before the fault time. Using the methodology described in chapter V, the alarm value was set at 19.77 and the threshold was 2.1. As shown in Fig. 6.16, all three detection algorithms indicated the change point approximately around the same time. DP on this graph stands for detection point. Since this is the actual data and not a simulation, the real change point is not known. Notice that the IFDS detected the crossing point before the actual fault time. For the specified threshold value at warning level 2, the time-to-failure (TTF) was about 110.71, 110.66, and 106.74 days using CUSUM, EWMA, and GLR algorithms, respectively. It is noted that the general response of the detection algorithms were presented in Chapter V, Fig. 5.9, Fig. 5.11, Fig. 5.10, respectively.

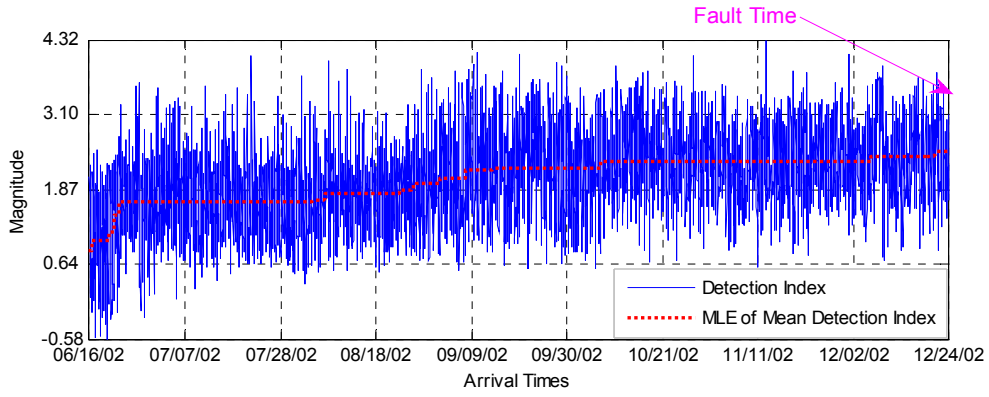


Fig. 6.15: Detection index and MLE of the mean path

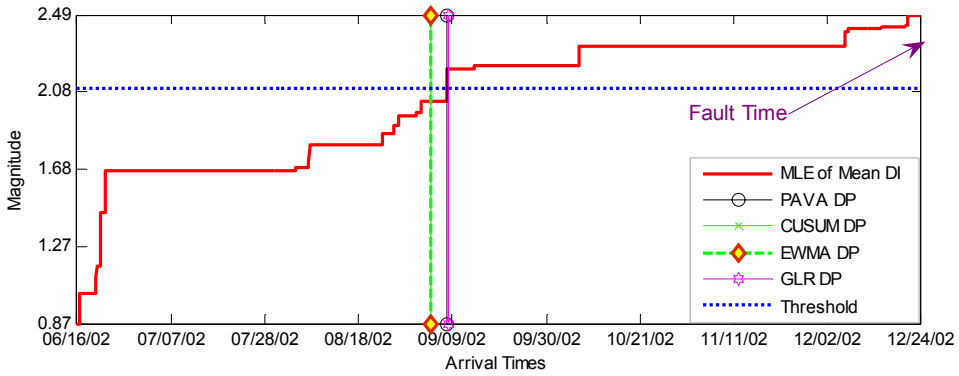


Fig. 6.16: Detection results of three algorithms

6.3.3.2 Scenario II

The resulting detection index values for this scenario are depicted in Fig. 6.17. From this graph, it is seen that the mean value of the detection index stays around the same value for a few weeks in between but rises monotonically preceding the fault time. The message conveyed by this behavior of the detection index may be that the incipient behavior of the cable was almost stable for a few weeks, but worsened over time leading to a failure in June. This graph in conjunction with the graph of the global severity index depicted in Fig. 6.13 indicates that the cable is migrating from its current state toward a more deteriorated state as the severity degrees show an increasing trend. Therefore, there are twice as many evidences as before that a fault is imminent.

The CUSUM response and detection result is given in Fig. 6.18. Setting the threshold at 2.59, alarm value at 20.28 with 5% probability of error, CUSUM predicted the fault 48.06 days prior to the actual fault time. The time-to-failure is a function of the alarm value. Increasing the alarm value results in a decrease in TTF. However, unreasonably chosen alarm values may cause CUSUM to fail to indicate the change.

The detection results and the response of the EWMA algorithm are provided in Fig. 6.19. In this case, the TTF was 45.99 days, approximately two days less than that of CUSUM. The response of the algorithm was computed using the same parameters reported in Chapter V.

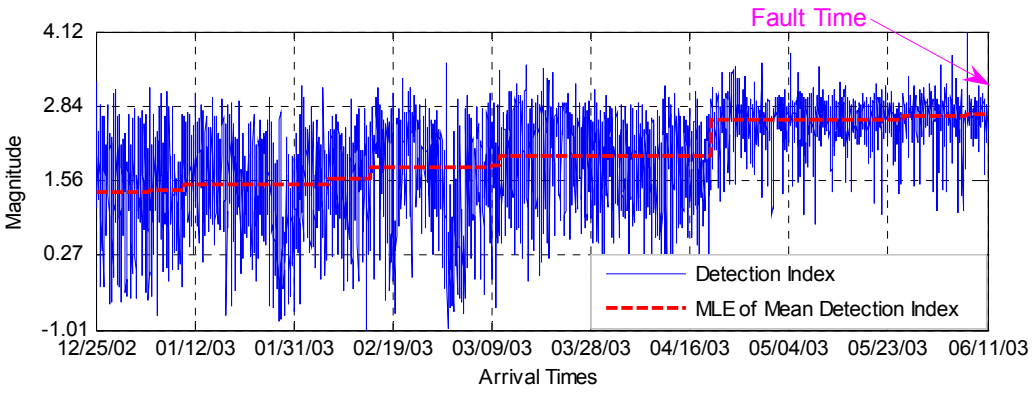


Fig. 6.17: Detection index and MLE of the mean path

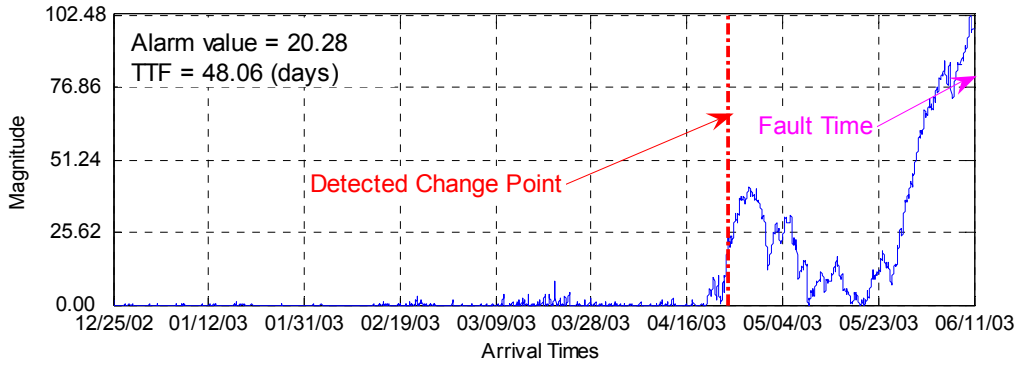


Fig. 6.18: Detection using CUSUM

Finally, Fig. 6.20 presents the detection result and the response of the third algorithm i.e. the GLR method. In this case, the TTF was 14.35 days. Fig. 6.21 depicts the detection results from the three algorithms on a single graph for comparison. From this graph, it is observed that the change time was indicated by all three techniques before the fault time. This result along with the complementary graph of the global severity index shown earlier in Fig. 6.13 provides sufficient evidences that a fault is imminent.

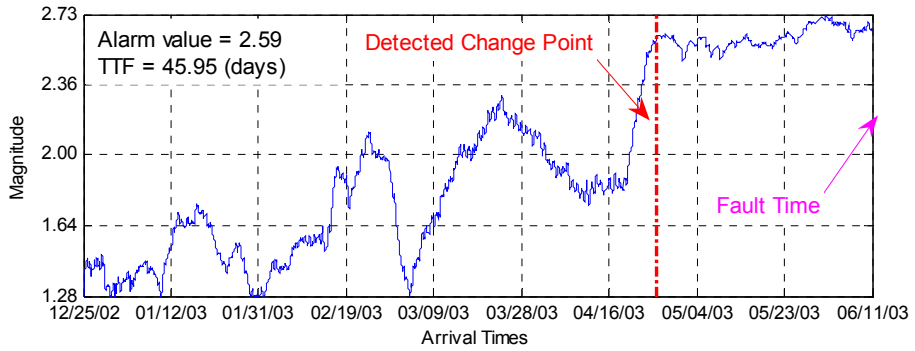


Fig. 6.19: Detection using EWMA

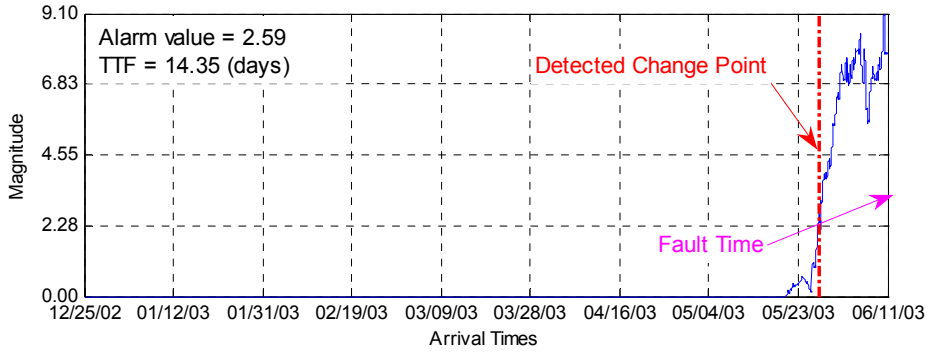


Fig. 6.20: Detection using GLR

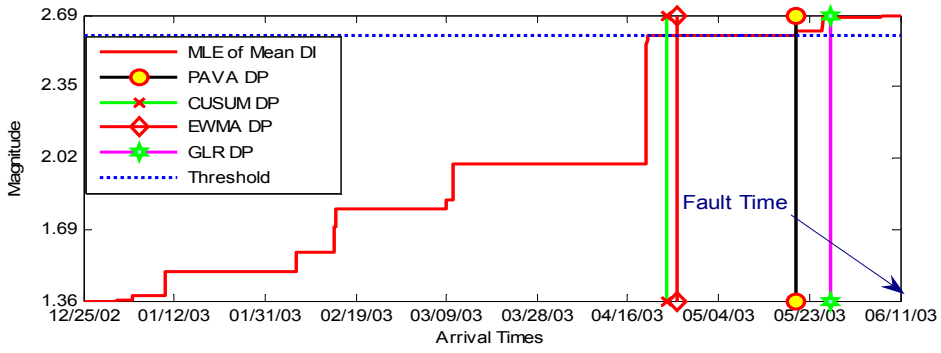


Fig. 6.21: Detection results from three algorithms

6.3.3.3 Summary of Detection Results for All Cases

The cumulative results of the IFDS for all fault cases at warning level 1, 2, and 3 are presented in Table 6-2, Table 6-3, and Table 6-4, respectively. For each case, the tables include the Time-To-Failure (TTF) values for each detection algorithm obtained from running the IFDS in sequential mode. The TTF values are expressed in days.

Table 6-2 presents the performance of the IFDS at the first warning level. This warning level indicates the first time the underground cable deviates from the norm established through the SOM model. Intuitively, the resulting TTF values provide the maximum time window available prior to the failure. The IFDS detected all incipient faults by at least one of its detection algorithms. The minimum TTF at this warning level was 3.66 days achieved for case VII through the GLR and CUSUM algorithms.

It should be noted that in view of variable number of measurements, the TTF values for each case are specific to that case and should not be compared across all faults. Comparing the detection algorithms, the CUSUM algorithm was the most successful, giving 100% success rate for all the faults, but there are more parameters to specify in this algorithm. The EWMA and GLR performed similarly at this warning level.

Table 6-2:
Summary of the IFDS results at warning level 1

Case	TTF (days)		
	CUSUM	**EWMA**	**GLR**
I	182.00	180.00	183.00
II	167.70	150.02	149.80
III	87.90	87.10	87.10
IV	22.80	16.95	18.16
V*	17.78	17.08	
VI*	4.54		
VII	3.66		3.66

* fault on adjacent lateral
TTF: time-to-failure

The detection results of IFDS at warning level 2 were presented in Table 6-3. At this warning level, the IFDS detected all incipient faults that were on the monitored lateral, but the faults on the adjacent laterals were not seen. For most of the cases, CUSUM and GLR successfully detected incipient faults; however, the CUSUM algorithm was the most successful, giving 100% success rate for all faults that were on the monitored lateral. The minimum TTF of 3.62 days was achieved in case VII through the CUSUM algorithm.

Table 6-3:
Summary of the IFDS results at warning level 2

Case	TTF (days)		
	CUSUM	**EWMA**	**GLR**
I	110.71	110.66	106.74
II	114.82	108.75	52.06
III	39.93		1.07
IV	17.15	13.48	16.87
V*			
VI*			
VII	3.62		

* fault on adjacent lateral
TTF: time-to-failure

The detection results of IFDS at warning level 3 were presented in Table 6-4. This warning level is the most critical level which provides the shortest time window prior to a fault. From the table, it is seen that the IFDS detected all incipient faults that were on the monitored lateral, but similar to warning level 2, the faults on the adjacent laterals were not seen. The CUSUM algorithm was the most successful, giving 100% success rate for all faults that were on the monitored lateral. The minimum TTF of 3.57 days was achieved in case VII through the CUSUM algorithm.

Table 6-4:
Summary of the IFDS results at warning level 3

Case	TTF (days)		
	CUSUM	EWMA	GLR
I	15.99	15.19	
II	48.06	45.95	14.35
III	39.24		
IV	6.73		
V*			
VI*			
VII	3.57		

* fault on adjacent lateral
TTF: time-to-failure

Through a number of studies, it was observed that the TTF values are greatly influenced by how the mean values of the detection index grow and reach a maximum before a fault. The slope or rates of increase in the mean values determine the best achievable TTF. When relatively rapid changes take place in a shorter interval within a few days, as in case VII, the best achievable TTF is highly sensitive to the selected warning level. On the other hand, for a gradual increase as in case I and II, the best achievable TTF is not greatly affected by the warning level. It was concluded that it is always feasible to consistently achieve a risk-free TTF which provides the longest time window prior to a fault, but the minimum TTF determination depends upon the allowable risk of missing the detection. As expected, higher risks result in smaller TTF values and vice versa. Formulation and quantitative assessment of this optimization problem require that more data and more fault scenarios be studied.

6.4 SUMMARY

In this chapter, the IFDS and its operational components were reviewed. Using field recorded data, the detection functionality of the IFDS was evaluated. The detection functionality of the IFDS is achieved through a temporal severity measure termed GSI and a detection method based on the SOM modeling of the incipient abnormality data. The GSI is based on the temporal analysis of arrival times of incipient abnormalities. The results using this measure provided compelling picture of the deterioration trajectory of the cable preceding each failure.

The detection approach uses the numerical modeling capabilities of SOM and statistical change detection techniques. The SOM based quantitative detection approach uses the natural logarithm of the minimum modeling errors in chronological order as the detection index. Three modified change detection algorithms namely Cumulative Sum (CUSUM), Exponentially Weighted Moving Averages (EWMA), and Generalized Likelihood Ratio (GLR) were introduced and tailored to each fault scenario. Using these algorithms, the near failure time was determined at three warning levels. For each warning level, an adaptive threshold was determined in a consistent manner across all fault cases. The detection results showed satisfactory operation and specifically showed that at least one detection technique provides an early warning that a fault is imminent. It was observed that CUSUM algorithm was the most successful, giving 100% success rate for all fault cases on the monitored lateral. However, there are more parameters to specify in this algorithm, as opposed to EWMA and GLR.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

7.1 CONCLUSIONS

In this dissertation, a methodology for an efficient, non-destructive, and online incipient fault diagnosis system (IFDS) was developed. The system detected underground cable incipient faults before they become catastrophic and provided vital information to help the operator with the decision-making process regarding the condition assessment of the underground cable. The IFDS used voltage and current measurements as the input information and utilized a comprehensive data processing modules to achieve its two major functionalities i.e. classification and detection. Field recorded data over a long period from monitoring sites were used to evaluate, test, and study the performance of the IFDS.

Through advanced digital signal processing and pattern analysis techniques, crucial information manifested in data was extracted and classifiers were designed to classify patterns into a number of designated categories. In the IFDS, the data was first fed to the preprocessing module to perform the necessary preprocessing operations, namely DC removal, resampling, and denoising. The preprocessed data were then passed through the rule-based classifier. In this module, a set of time-domain features was extracted and used to evaluate a set of rules defined for the designated categories of the data. At this stage, DC classes, various switching type events (load changes), normal, and potential incipient abnormality patterns were recognized. If the classification at this stage did not result in an incipient abnormality pattern, the measurement was directly passed to the post-processing module. Otherwise, the feature extraction module operated on the data, and another set of features in time and wavelet domain was extracted.

Applying the dimensionality reduction principles, 14 informative features were determined and used in the supervised classifiers. These features included the first eight features at level 4 and last four features at level 3 of the wavelet packet decomposition tree as well as and the two time-domain features termed as VSI and HSI. The supervised classifiers deployed in the system were Support Vector Machines (SVM), Self-Organizing Map (SOM), K Nearest Neighbors (KNN), and Discriminant classifiers. The classification results of the IFDS using these classifiers were satisfactory. Overall, an average classification rate of 97.17% was achieved during a

testing. For the given test data, the SVM classifier outperformed the other classifiers in terms of the classification rate and false positive/negative measures. Each classifier made decisions on the input patterns; however the system classification labels used in the detection stage were determined by the SVM classifier.

For the detection module, a novel severity measure called the GSI was developed. This temporal severity measure not only takes into account the local content of the measured incipient abnormality, it also incorporates the past observations in assessing the current severity degree of incipient abnormalities. The Laplace Test Statistic played an important rule in computation of the GSI. The deterioration trajectory of the underground system studied for two case studies showed that this index was in fact applicable and successful in portraying the incipient behavior of the system.

The severity measure based on the GSI provided the severity path of ongoing incipient behavior over time, but it was not designed to indicate the warning time. This desired feature of the IFDS was achieved through the SOM based detection approach utilizing the natural logarithm of the minimum modeling errors in chronological order as the detection index. The detection approach incorporated the numerical modeling capabilities of SOM and statistical change detection techniques. Three modified change detection algorithms namely Cumulative Sum (CUSUM), Exponentially Weighted Moving Averages (EWMA), and Generalized Likelihood Ratio (GLR) were introduced, and tailored to this specific application. Using these algorithms, the near failure time was determined from the instantaneous values of the detection index at three warning levels. For each warning level, an adaptive threshold was determined in a consistent manner across all fault cases.

The performance studies and detection results of the integrated incipient fault diagnosis system were conducted and presented using the field recorded data. The performance studies included an assessment of the IFDS capability in predicting the faults that actually occurred in the monitored lateral. The detection results showed highly satisfactory operation and specifically showed that at least one detection technique with proper settings successfully provided an early warning that a fault was imminent. It was observed that the CUSUM algorithm was the most successful one giving 100% success rate for all the fault cases on the monitored lateral. It was concluded that it is always feasible to consistently achieve a risk-free TTF which provides the maximum amount of time prior to fault, but the minimum TTF determination depends upon the

allowable risk of missing the detection. As expected, taking the higher risks result in the smaller TTF values and vice versa.

7.2 COMMENTS AND RECOMMENDATIONS

The developed incipient fault diagnosis system was shown to perform as expected on the field data and available fault scenarios. Nevertheless, its operation and performance are subject to a number of considerations. These conditions and requirements can be discussed and commented in terms of the classification and detection functionalities as follows.

Speaking of the classification functionality of the system, three points must be made. The rule-based classifier, although it uses a set of rules, is very effective in reducing the computation complexity of the supervised classifiers. It is possible to design the system with supervised classifiers only but this will not be the best possible solution. The good classification rate, acceptable training time and generalization capability of such a system requires much more data and logistics than the current design using an ensemble of classifiers. Furthermore, the settings and thresholds used in the decision rules are either automatically induced from the data or selected by the user in a subjective manner. It is important to realize that the non-incipient data are not used in the supervised classifiers. Yet, if a non-incipient pattern is directed to this module, the consequences are more tolerable by the system than that of the case where an incipient pattern is spuriously labeled as non-incipient. This is helpful in setting the user-defined thresholds.

The second comment about the classification is about the supervised classifiers. Although, implicitly mentioned, the first challenge in utilizing any classification method in supervised mode is to have access to a set of training data that are already labeled. This time-consuming task was accomplished manually on the field data and the labeling was made through a subjective judgment of the captures. This requires some degree of familiarity with the incipient abnormality data and their characteristics which was achieved through studying and examination of a large number of field data from various sites. Application of the IFDS may require that the training phase be repeated over time. Currently, there is no automatic mechanism to update and modify the trained classifiers.

As discussed, the IFDS classification functionality includes four supervised classifiers. The SVM classifier was found to perform superior over the other three classifiers and thus the class labels needed for the GSI computations are the SVM decisions. Nevertheless, extensive studies

might be needed to possibly improve the system operation by combining the classification results from all supervised classifiers included in the system.

The detection functionality of the IFDS requires a number of important attentions. The temporal severity measure used to quantify the degradation path of the cable, although very effective as shown, is solely a subjective measure and should not be expected to behave similarly at all times. The performance and usefulness of this measure is influenced by the result of the supervised classifiers. As it uses the arrival times of the incipient data, it is important to make sure that the classification results are as high as possible. False classification results may cause the GSI to indicate spurious deterioration paths. On the other hand, this measure is not highly susceptible to a single wrong classification. This is due to the fact that its computation involves a set of arrival times located within the window and not an individual arrival time. Therefore, a single wrong arrival time should not drastically influence the GSI values as long as the right arrival times in the same window are highly dominant. Another point regarding this measure is the size of the window. A good rule of thumb is to remember that the bias-variance trade off holds true in this case as well. The longer the size of the window, the smoother the GSI values and vise versa. The optimum choice of the window size can be obtained through a number of runs with different values of the span.

The measurements and monitoring must be conducted at regular uninterrupted intervals for computation of the GSI. The interrupt times where no capture is made are seen as elapsing times with no incipient data. The GSI will spuriously take the interrupt times into account once the first incipient data becomes available after the interruption is over. The best practice to avoid false calculations is to start over every time there is an interruption in data. The SOM based detection index does not use the arrival times of incipient abnormalities, but depending upon the cause of the interruption, it may become necessary to modify the model and update the settings. The best examples of such situations are the interruptions imposed by a change in the configuration of the underground cable or a catastrophic failure.

Finally, the IFDS detection functionality includes three algorithms, indicating the change point at different times. The final decision of the IFDS regarding an imminent incipient fault may be improved by integrating the results from each detection algorithm. However, this improvement necessitates extensive studies and a number of fault cases.

7.3 FUTURE WORK

The future work may investigate the following aspects. The developed diagnosis system was used to make decisions in sequential mode on a number of actual fault scenarios and the results were highly acceptable. Nevertheless, the generalization capability of the IFDS is a major topic for future work. More data from different sites and more fault cases are needed to assess the IFDS adaptation and generalization capabilities. Also, new rules and trained classifiers need to be developed for novel patterns that neither observed nor emerged in the data.

The IFDS classification and detection functionalities include four supervised classifiers and three detection algorithms. It was found that the SVM classifier and the CUSUM detection algorithm performed well; nevertheless, future work may include extensive studies to possibly improve the system operation by integrating the classification and detection results provided by the other classifiers and detection algorithms.

Hidden Markov Model (HMM) is a statistical approach that has been predominantly used in automatic speech recognition, yet it has become the gold standard for time series analysis. HMM has a great potential for classification problems involving temporal or sequential patterns. In the IFDS, the features used in the classification of incipient abnormalities are treated as static values and the temporal nature of the patterns is absorbed in the GSI. One can investigate application of HMMs to combine the static classification and GSI to produce a unified statistical detection index addressing the temporal nature of incipient faults.

Finally, the IFDS primarily relies on the voltage, phase current, and notch high frequency signals to perform its operation. The notch low frequency current signal is not an information resource for the system and application of this signal can be investigated for possible improvements.

REFERENCES

[1] *Underground Distribution System Design and Installation Guide*, Washington D.C.: National Rural Electric Cooperative Association, 1993.

[2] W. F. Buddy, "The basics of power cables", *IEEE Transactions on Industry Applications*, vol. 30, no. 3, pp. 506 – 509, May – June 1994.

[3] J. Cardoso, "Characterization of Underground Cable Incipient Failures from On-Line Monitoring of Underground Distribution Power Systems", M.S. thesis, Texas A&M University, College Station, TX, 1999.

[4] R. Chaturbedi, "Characterization and Detection of Incipient Underground Cable Failures", M.S. thesis, Texas A&M University, College Station, TX, 2002.

[5] K. L. Butler-Purry, B. Don Russell, R. Chaturbedi, and M. J. Mousavi, "Characterization of Underground Distribution Cable Failures", Years 3-5 Progress Report, Dallas, TX, Feb. 2003.

[6] K. L. Butler-Purry, B. Don Russell, and M. J. Mousavi, "Online Detection of Underground Distribution Cable Failure", Report 056933 to EPRI and TXU Electric Delivery, Palo Alto, CA and College Station, TX, Jan 2005.

[7] Rosevear R. D., "Overview of cable engineering", *IEE Colloquium on Supertension*, Nov. 1995, pp. 1/1-1/3.

[8] S. Y. King and N. A. Halfter, *Underground Power Cables*, New York: Longman Inc., 1982.

[9] M. J. Mousavi and K. L Butler-Purry, "Study of thermal aging effects on distribution transformer solid insulation", in *Proceedings of 34th North American Power Symposium*, Tempe, AZ, Oct. 2002, pp. 160-167.

[10] J. Densley, "Ageing mechanisms and diagnostics for power cables – an overview", *IEEE Electrical Insulation Magazine*, vol. 17, no. 1, January – February, 2001, pp. 14 - 22.

[11] *Evaluation and Qualification of Electrical Insulation Systems*, IEC Standard 60505, 1999.

[12] N. H. Malik, A. A. Al-Arainy, and M. I. Qureshi, *Electrical Insulation in Power Systems,* New York: Marcel Dekker, 1998.

[13] N. Srinivas, and N. Ahmed, "Condition assessment of distribution and transmission class voltage cable systems", in *2003 IEEE 10th International Conference on Transmission and*

Distribution Construction, Operation and Live-Line Maintenance, Orlando, FL, Apr. 2003, pp. 194-201.

[14] G. J. Paoletti and A. Golubev, "Partial discharge theory and technologies related to medium-voltage electrical equipment", *IEEE Transactions on Industry Applications*, vol. 37, no. 1, pp. 90 – 103, Jan. –Feb. 2001.

[15] N. Ahmed and N. Srinivas, "On-line versus off-line partial discharge testing in power cables", in *2001 IEEE/PES Transmission and Distribution Conference and Exposition*, Atlanta, GA, Oct. 2001, vol. 2, pp. 865-870.

[16] M. S. Mashikian, "Partial discharge location as a diagnostic tool for power cables", presented at *IEEE Power Engineering Society Winter Meeting*, Singapore, Jan. 23-27, 2000, pp.1604 -1608.

[17] N. Ahmed and N. Srinivas, "Extruded insulated cable deterioration mechanism and PD", *2002 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, Cancun, Mexico, Oct. 2002, pp. 744-747.

[18] H. E. Orton, "Diagnostic testing of in-situ power cables: an overview", *IEEE/PES Transmission and Distribution Conference and Exhibition 2002: Asia Pacific Conference*, Yokohama, Japan, Oct. 2002, vol. 2, pp. 1420-1425.

[19] Q. Su and K. Sack, "New techniques for on-line partial discharge measurements", in *Proceedings of IEEE International Multi Topic Conference (INMIC 2001)*, Lahore, Pakistan, Dec. 2001, pp. 49-53.

[20] N. Ahmed and N. Srinivas, "Can the operating conditions of the cable system affect the data of the field PD testing?" *2001 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, Ontario, Canada, Oct. 2001, pp. 311 – 314.

[21] N. Srinivas, N. Ahmed, and G. Raju, "On line measurement of partial discharges", *2000 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, British Colombia, Canada, Oct. 2000, vol. 2, pp. 678 – 681.

[22] N. H. Ahmed and N. N. Srinivas, "On-line partial discharge detection in cables", *1997 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, Minneapolis, MN, Oct. 1997, vol. 1, pp. 214-217.

[23] Y. Lim, J. Lee, W. Kang, and J. Koo, "Chaotic analysis of partial discharge (CAPD) as a novel approach to investigate insulation degradation caused by the various defects*",* in

Proceedings of IEEE International Symposium on Industrial Electronics, Pusan, South Korea, June 2001, vol. 1, pp. 413-416.

[24] X. Ma, C. Zhou, and I. J. Kemp, "Wavelets for the analysis and compression of partial discharge data", *2001 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, Ontario, Canada, Oct. 2001, pp. 329-334.

[25] E. Gulski, B.R. Hamerling, F.J. Wester, J. J. Smit, E. Groot, and P. Schikarski, "Insulation condition assessment of medium voltage power cables using on-site PD detection and analysis techniques", in *16th International Conference and Exhibition on Electricity Distribution*, Amsterdam, The Netherlands, June 2001, vol. 1, pp. 1-5.

[26] B. R. Hamerling, F. J. Wester, E. Gulski, J. J. Smit, and E. R. S. Groot, "Fundamental aspects of on-line PD measurements on distribution power cables", in *Proceedings of the 2001 IEEE 7th International Conference on Solid Dielectrics*, Eindhoven, The Netherlands, Jun. 2001, pp. 408-411.

[27] I. Shim, J. J. Soraghan, W. H. Siew, F. McPherson, and P. F. Gale, "Partial discharge location on high voltage cables", in *11th International Symposium on High Voltage Engineering*, London, UK, Aug. 1999, vol. 5, pp. 248-251.

[28] B. Dean,"Condition monitoring of 11-132 kV cable accessories", in *IEE Colloquium on Ensuring the Reliability of 11-132 kV Cable Accessories*, London, UK, Feb. 1998, pp. 5/1-5/4.

[29] W. Wei, Z. Sen, C. Bin, and L. En-Heng, "The study of on-line PD detector in power cable", in *Proceedings of the 6th International Conference on Properties and Applications of Dielectric Materials*, Xian, China, Jun. 2000, vol. 1 , pp. 177-180.

[30] J. J. Smit, "Insulation condition estimation of distribution power cables by on-line and off-line diagnostics", in *Proceedings of 2001 International Symposium on Electrical Insulating Materials*, Himeji, Japan, Nov. 2001, pp. 513-517.

[31] M. S. Mashikian, R. Luther, Sr., J. C. McIver, J. Jurcisin, Jr., and P. W. Spencer, "Evaluation of field aged crosslinked polyethylene cables by partial discharge location", *IEEE Transactions on Power Delivery*, vol. 9, no. 2, pp. 620 -628, Apr. 1994.

[32] D. Zhifang, M. S. Mashikian, and F. Palmieri, "Partial discharge propagation model and location estimate", presented at the *IEEE International Symposium on Electrical Insulation*, Pittsburgh, PA, Jun. 1994, pp.99 -102.

[33] D. Zhifang, P. K. Willett, and M. S. Mashikian, "Performance limits of PD location based on time-domain reflectometry", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 4, no. 2, pp. 182 -188, Apr. 1997.

[34] S. M. Miri and A. Privette, "A survey of incipient fault detection and location techniques for extruded shielded power cables", presented at the *26th Annual Southeastern Symposium on System Theory*, Athens, OH, Mar. 20-22, 1994, pp.402 – 405.

[35] N. Ahmed and N. Srinivas, "The noise effect in conducting on-line partial discharge testing in distribution-class cables", *1998 Annual Report Conference on Electrical Insulation and Dielectric Phenomena*, Atlanta, GA, Oct. 1998, vol. 1, pp. 319 – 322.

[36] H. Oonishi, F. Urano, T. Mochizuki, K. Soma, K. Kotani, and K. Karnio, "Development of new diagnostic method for hot-line XLPE cables with water trees", *IEEE Transactions on Power Delivery*, vol. 2, no. 1 , pp. 1-7, Jan. 1987.

[37] S. Yamaguchi, S. Soda, and N. Takada, "Development of new type insulation diagnostic method for hot-line XLPE cables", *IEEE Transactions on Power Delivery*, vol. 4, no. 3, pp. 1513-1520, July 1989.

[38] Y. Zhang, J. Lewiner, C. Alquie, and N. Hampton, "Evidence of strong correlation between space-charge buildup and breakdown in cable insulation", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 3, no. 6, pp. 778-783, Dec. 1996.

[39] J. Lewiner, "Direct determination of space charge distributions in dielectrics: The pressure wave propagation method", presented at the *Third International Conference on Conduction Breakdown in Solid Dielectrics*, Trondheim, Norway, July 3-6, 1989.

[40] W. L. Weeks and J. P. Steiner, "Instrumentation for the detection and location of incipient faults on power cables", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 7, pp. 2328 – 2335, July 1982.

[41] W. L. Weeks and J. P. Steiner, "Improvement in the instrumentation for partial discharge location in cables", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 4, pp. 754 – 760, Apr. 1985.

[42] L. Wieringa, "Location of small discharges in plastic insulated high voltage cables", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-104, no. 1, pp. 2–8, Jan. 1985.

[43] M. S. Mashikian, R. Bansal and R. B. Northrop, "Location and characterization of partial discharge sites in shielded power cables", *IEEE Transactions on Power Delivery*, vol. 5, no. 2, pp. 833-839, Apr. 1990.

[44] A. H. Baghurst, "A new method for the location of partial discharge sites using modulated X-rays", in *Proceedings of the Conference on Electrical Insulation and Dielectric Phenomena,* Arlington, VA, Oct 1985.

[45] W. E. Anderson, J. D. Ramboz, and A. R. Ondrejka, "The detection of incipient faults in transmission cables using time domain reflectometry techniques: Technical challenges", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 7, pp. 1928–1934, July 1982.

[46] DTE Energy, "Technologies", accessed on May 2005. [Online]. Available: http://www.dtetech.com/technologies/cablewise/

[47] IMCORP, "Products", accessed on May 2005. [Online]. Available: http://www.imcorptech.com/index.htm

[48] PowerTech, "Products", accessed on May 2005. [Online]. Available: http://www.powertech.bc.ca/cfm/index.cfm

[49] EA Technology, "Products", accessed on May 2005. [Online]. Available: http://www.eatechnology.com/

[50] K. L. Butler-Purry, M. Bagriyanik, M. J. Mousavi, and P. Palmer-Buckle, "Experimental investigation of internal short circuit faults leading to advanced incipient behavior and failure of a distribution transformer", in *Proceedings of Power Systems Conference and Exposition*, New York, NY, Oct. 2004, vol. 3, pp. 1407-1416.

[51] M. J. Mousavi and K. L. Butler-Purry, "Wavelet-based Denoising of Field Recorded Power System Measurements", in *Proceedings of the 36th North American Power Symposium (NAPS)*, University of Idaho, Aug. 2004.

[52] M. J. Mousavi and K. L. Butler-Purry, "Classification of load change transients and incipient abnormalities in underground cable using pattern analysis techniques", *IEEE/PES Transmission and Distribution Conference*, Dallas, TX, Sept. 2003, vol. 1, pp. 175-180.

[53] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd Edition, New York: Wiley, 2001.

[54] A. W. Galli and O. M. Nielsen, "Wavelet analysis for power system transients", *IEEE Computer Applications in Power*, vol. 12, no. 1, pp. 16, 18, 20, 22, 24 -25, Jan. 1999.

[55] C. H. Kim and R. Aggarwal, "Wavelet transforms in power systems. I. General introduction to the wavelet transforms", *IEE Power Engineering Journal*, vol. 14, no. 2, pp. 81 -87, Apr. 2000.

[56] C. H. Kim and R. Aggarwal, "Wavelet transforms in power systems. II. Examples of application to actual power system transients", *IEE Power Engineering Journal*, vol. 15, no. 4, pp. 193 -202, Aug. 2001.

[57] A. K. Jain, R. Diun and J. Mao, "Statistical pattern recognition: A review", *IEEE Transactions on Pattern Analysis Intelligence*, vol. 22, no.1, pp. 4 –37, Jan 2000.

[58] A. Webb, *Statistical Pattern Recognition*, London: Arnold, 1999.

[59] R. Gutierrez-Osuna, Pattern Recognition Lecture Notes, 2003. [Online] Available: http://research.cs.tamu.edu/prism/lectures.htm

[60] R. C. Holte,"Very simple classification rules perform well on most commonly used datasets", *Machine Learning* , vol. 3, pp. 63 –91, 1993.

[61] W. Duch, N. Jankowski, K. Grabczewski, and R. Adamczak, "Optimization and interpretation of rule-based classifiers", in *Intelligent Information Systems*: *Advances in Soft Computing*, pp. 1-13, June 2000.

[62] H. Ascher and H. Feingold, *Repairable Systems Reliability: Modeling, Inference, Misconceptions, and Their Causes*, vol. 7, Newbury, UK: Marcer Dekker Inc, 1984.

[63] C.W. Therrien, K. L Frack, and N. Ruiz Fontes, "A short-time Wiener filter noise removal in underwater acoustic data", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)*, Munich, Germany, Apr. 1997, vol. 1, pp. 543 –546.

[64] R. J. Barsanti and M. P. Fargues, "Wavelet-based denoising of acoustic transients", in *Proceedings of 30th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1996, pp. 848-852.

[65] *MATLAB Wavelet Toolbox User's Guide*, version 2, MA : The Mathworks Inc, 2000.

[66] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed., San Diego: Academic Press, 1999.

[67] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Revised ed. , Wellesley, MA: Wellesley-Cambridge Press, 1997.

[68] D. L. Donoho and I. M. Johnstone "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, vol. 81, pp. 425-455, 1994.

[69] A. Grossmann, "Wavelet transform and edge detection", in P. Blanchard, L. Streit, and M. Hazewinkel, editors, *Stochastic Processes in Physics and Engineering*. New York: Reidel, Dordrecht, 1986.

[70] A. Bruce and H. Gao, "WaveShrink: Shrinkage functions and thresholds", Technical Report, StatSci Division of MathSoft, Inc., 1995.

[71] S. K. Nath, R.M. Vasu, and M. Pandit, "Wavelet based compression and denoising of optical tomography data", *Optics Communications*, vol. 167, no. 1, pp. 37-46, Aug. 1999.

[72] D. Donoho and I. M. Johnstone, "Threshold selection for wavelet shrinkage of noisy data", in *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Baltimore, MD, 1994, vol.1, pp. 24a-25a.

[73] S. Kenney, *Mathematics of Statistics*, New York: D Van Nostrand Company, 1942.

[74] E. Styvaktakis, "Automating Power Quality Analysis", Ph.D. dissertation, Chalmers University of Technology, Goteborg, Sweden, 2002.

[75] M. H. J. Bollen, *Understanding Power Quality Problems: Voltage Sags and Interruptions*, New York: IEEE Press, 1999.

[76] E. Styvaktakis, M. H. J. Bollen and I. Y. H. Gu, "Expert system for classification and analysis of power system events", *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 423-428, Apr. 2002.

[77] *IEEE Recommended Practice for Monitoring Electric Power Quality*, IEEE Standard 1159, 1995.

[78] V. Barnett, *Outliers in Statistical Data*, 3rd ed., New York: Wiley & Sons, 1994.

[79] L. G. Kiloh, A. J. McComas, J. W. Osselton, and A. R. M. Upton, *Clinical Electroencephalography*, 4th ed., London, U.K: Butterworths, 1981.

[80] T. Kalayci and O. Ozdamar, "Wavelet preprocessing for automated neural network detection of EEG spikes", *IEEE Engineering in Medicine and Biology*, pp. 160-166, Mar./Apr. 1995.

[81] C. C. Pang, A. R. Upton, G. Shine, M. V. Kamath, "A comparison of algorithms for detection of spikes in the electroencephalogram", *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 4, pp. 521-526, Apr. 2003.

[82] D. Sanchez, M. Adjouadi, A. Barreto, P. Jayakar, and I. Yaylali, "Application of the Walsh transform in an integrated algorithm for the detection of interracial spikes", in *Proceedings of the 23rd Annual EMBS International Conference*, Istanbul, Turkey, Oct. 2001, pp. 939-942.

[83] R. Walters, J. C. Principe, and S. Park, "Spike detection using a syntactic pattern recognition approach", *IEEE Engineering in Medicine and Biology Society 11th Annual International Conference*, Seattle, WA, 1989, pp. 1810-1811.

[84] M. Das and T. Hunt, "A comparative evaluation of various methods for detection of small spikes in data", in *Proceedings of 1998 Midwest Symposium on Circuits and Systems*, Notre Dame, IN, Aug. 1998, pp. 510-513.

[85] M. Petrou, "The differentiating filter approach to edge detection", *Advances in Electronics and Electron Physics*, vol. 88, pp. 297-345, 1994.

[86] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., New York: McGraw-Hill, 1991.

[87] M. Juhola, "Algorithms for standard median filtering", *IEEE Transactions on Acoustic Speech and Signal Processing*, vol. 39, pp. 204-208, 1991.

[88] K. R. Rao and P. Yip, *Discrete Cosine Transform*, San Diego: Academic Press, 1990.

[89] I. Kaplansky, "A common error concerning kurtosis", *Journal of Amr. Stat. Assoc.,* vol. 40, pp. 259-263, 1945.

[90] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, no. 1, pp. 83 -88, Jan. 1999.

[91] A. M. Gaouda, M. M. A. Salama, M. R. Sultan and A. Y. Chikhani, "Power quality detection and classification using wavelet multiresolution signal decomposition", *IEEE Transactions on Power Delivery*, vol. 14, no. 4, pp. 1469 -1476, Oct. 1999.

[92] Z. Chen and P. Urwin, "Power quality detection and classification using digital filters", *IEEE Porto Power Tech Conference*, Porto, Portugal, vol. 1, pp. 4, Sep. 2001.

[93] C. J. C. Burges, *A Tutorial on Support Vector Machines for Pattern Recognition*, Boston: Kluwer Academic Publishers, 1998.

[94] V. Vapnik, *Statistical Learning Theory*, New York : John Wiley and Sons Inc., 1998.

[95] S. R. Gunn, "Support Vector Machines for Classification and Regression", Dept. of Electronics and Computer Science, University of Southampton, Tech. Rep., 1998.

[96] V. Cherkassky, F. Mulier, *Learning from Data: Concepts, Theory and Methods,* New York : John Wiley & Sons, 1998.

[97] S. Kaykin, *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, NJ: Prentice Hall Inc., 1999.

[98] C. Lin, "A Library for Support Vector Machines", accessed on May 2005. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html

[99] T.Kohonen, *Self-Organizing Maps*, 2nd ed., NewYork: Springer-Verlag, 1997.

[100] K.F. Thang, R. K. Aggarwal, A. J. McGrail, and D. G. Esp, "Analysis of power transformer dissolved gas data using the self-organizing map", *IEEE Transactions on Power Delivery*, vol. 18, no. 4, pp. 1241-1248, Oct. 2003.

[101] P. Kang and D. Birtwhistle, "Condition assessment of power transformer onload tap changers using wavelet analysis and self-organizing map: Field evaluation", *IEEE Transactions on Power Delivery*, vol. 18, no. 1, pp. 78-84, Jan. 2003.

[102] P. Kang and D. Birtwhistle, "Condition monitoring of power transformer on-load tap-changers. Part 1: Automatic condition diagnosis", *IEE Proceedings, Generation, Transmission, and Distribution*, vol. 148, no. 4, pp. 301-306, July 2001.

[103] P. Kang and D. Birtwhistle, "Condition monitoring of power transformer on-load tap-changers. Part 2: Detection of ageing from vibration signatures", *IEE Proceedings, Generation, Transmission, and Distribution*, vol. 148, no. 4, pp. 307-311, July 2001.

[104] J. Rodriguez Arroya, A. J. Beddoes, and N. M. Allinson, "Condition monitoring of 11 kv paper insulated cables using self organizing maps", *Canadian Conference on Electrical and Computer Engineering,* Toronto, Canada, May 2001, vol. 1, pp. 259-264.

[105] P. Demartines and J. Herault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets", *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148-154, Jan. 1997.

[106] J. W. Sammon, Jr. "A nonlinear mapping for data structure analysis", *IEEE Transactions on Computers*, vol. 18, no. 5, pp. 401-409, May 1969.

[107] J. Vesanta, J. Himberg, E.Alhoniemi, and J. Parhankangas, "SOM toolbox for MATLAB 5", Technical report, [Online]. Available: http://www.cis.hut.fi/projects/somtoolbox/

[108] C. M. Bishop, *Neural Networks for Pattern Recognition*, New York: Oxford University Press, 1995.

[109] J. K. Taylor and C. Cihon, *Statistical Techniques for Data Analysis*, 2nd ed., Boca Raton, FL: Chapman&Hall/CRC, 2004.

[110] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 5th ed., Duxbury: Thomas Learning, 2000.

[111] W. J. Conover, *Practical Nonparametric Statistics,* New York: Wiley, 1980.

[112] T. Rebertson, F. T. Wright, and R. L. Dykstr, *Order Restricted Statistical Inference*, NewYork: John Wiley & Sons, 1988.

[113] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes*, Englewood Cliffs, N.J.: Prentice Hall, 1993.

[114] J. T. Chang and R. D. Fricker Jr., "Detecting when a monotonically increasing mean has crossed a threshold", *J. Qual. Technol*., vol. 31, no. 2, pp. 217-234, 1999.

[115] M. Basseville and A. Benveniste, *Detection of Abrupt Changes in Signals and Dynamical Systems*, New York: Springer-Verlag, 1986.

[116] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, New York: Wiley, 2001.

VITA

Mir Rasoul Jaafari Mousavi received the B.S. degree from Isfahan University of Technology in 1996 and the M.S. degree from Sharif University of Technology in 1999 all in electrical engineering. He worked for Niroo Research Institute (NRI) from 1999 to 2001 as an R&D engineer. He received his Ph.D. degree in electrical engineering from Texas A&M University in December 2005.

From January, 2001, to January, 2005, and in summer 2005, he worked as research assistant in the Power System Automation Laboratory at Texas A&M University. He was a graduate lecturer at the Department of Electrical Engineering in spring 2005.

Mr. Mousavi is a student member of IEEE and IEEE Power Engineering Society (PES). He served as president and treasurer of the IEEE-PES-IAS-PELS Joint Student Chapter at Texas A&M University. He received the Best Paper Award at the 2003 IEEE/Power Engineering Society Transmission and Distribution Conference. His research interests are related to power system modeling, simulation, and analysis, fault diagnosis, distribution automation, signal processing, and pattern analysis applications.

Mr. Mousavi may be reached at Mail stop 3128, Department of Electrical Engineering, Texas A&M University, College Station, TX 77843. His email address is mousavi@ee.tamu.edu.