Machine Learning Enhanced Radio Location of Partial Discharge

Ephraim Tersoo Iorkyase

BEng, MEng

A thesis submitted for the degree of Doctor of Philosophy to the

Centre for Intelligent and Dynamic Communications,

University of Strathclyde.

June 2019

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

Examination Committee:

Prof Craig Michie, Convener

Prof Brian G. Stewart, Internal Examiner

Prof Gordon Morison, External Examiner

Dedication

To my beloved wife, Anita, two boys: SenaTer and SefaTer.

And in memory of my Dad and Mum

Acknowledgements

First and foremost I would like to thank the Almighty God, the giver and sustainer of life for the grace to complete my PhD thesis.

I would like to express my profound gratitude to my supervisor Dr. Robert Atkinson, for the opportunity to undertake this work and for his advice, direction and comments when I needed them the most. Thank you for your patience and kind disposition towards me. I am indebted to you. I would also like to thank Dr. Christos Tachtatzis for the discussions, the numerous advices, and his constructive criticisms during the period of my PhD. I would like to thank Prof. Ian Glover for all his in-depth contributions during my studies. Thank you Dr. Alison Cleary for all the encouragement and advice.

Special thanks to my darling wife, Anita and my two wonderful boys, Sena and Sefa who had to put up with me being away for extended periods. Thank you so much for your patience, prayers and support throughout my studies. I am also grateful to my brothers and sisters for standing with me during this period. Thank you all for those calls. They mean so much to me.

To my family in Glasgow, Ps Alistair Matheson, Ps Ola Epemolu, Ps Ken Whitelaw and the entire members of GCC, I thank you all for making my stay in Glasgow worthwhile. I would also like to appreciate my Nigerian family in UK, Prof John Igoli and family, Dr Abimbola Famuyiwa, Dr Luter Leke, Dr Nathaniel Tsendzul, TY Dang, Steve Mbalaha, Paul Ager, Akura Mise, Attai Abubakar, Paul Kpochi, Anne Elogie, Emmanuel Salifu, Rev Idowu Obolo, for believing in me. Special thanks to Dr Raphael Amokaha and Dr Sesugh Ande who welcome me in Glasgow.

I am also grateful for my friends and colleagues past and present in Mobile Communication group for their unconditional friendship. Thank you Dr. Andrew Hamilton, Elike Hodo, Dr Xavier Bellenkens, Dr Jakub Konka, for creating a pleasant and convivial environment for research.

I would also like to thank my internal and external examiners, Prof. Brian Stewart and Prof. Gordon Morison for their valuable suggestions and advice which further helped enhance my thesis.

Finally, I would like to thank the University of Agriculture Makurdi and Tertiary Education Trust Fund (TETFUND) Nigeria, for their financial support without which my purpose of doing a PhD at the University of Strathclyde could hardly be realise.

Abstract

Partial Discharge (PD) is a well-known indicator of plant failure in electricity facilities. A considerable proportion of assets including transformers, switch gears, and power lines are susceptible to PD due to incipient weakness of their dielectric components. These discharges may cause further degradation of the insulation, which in turn may lead to subsequent catastrophic failure. The damage that results from PD activity is worth millions of pounds and endangers the lives of personnel. PD emits electrical pulses in the form of Radio Frequency (RF) signals which propagate as a travelling wave in the vicinity of the discharge site and can be detected using dedicated sensors. This has motivated the use of an enhanced radio-based technique to detect its occurrence at early stage. Early detection of PD helps utility operators to initiate an emergency maintenance outside the scheduled times when it is most cost-effective and before the equipment loses performance or suffers catastrophic failure, hence improving asset management.

Therefore this thesis presents an investigation of an enhanced machine learning approach to continuous PD localisation using a network of radio sensors. The approach being investigated relies on location dependent parameters which will be extracted from

PD measurements. This thesis demonstrates RF-based fingerprinting technique for locating PD sources using Received Signal Strength (RSS). Furthermore, Signal Strength Ratios (SSR) between pairs of sensor nodes are used as robust fingerprints given that the energy emitted by each PD event may be different due to progressive nature of PD severity as deterioration continues and the fact that different types of PD occur in nature. Sophisticated machine learning techniques are investigated and used to develop PD localisation models.

This work also investigates the plausibility of using other PD received signal parameters for locating PD sources. It has been found that the statistical characterisation of the received RF signals produces manifold PD features beside RSS. The developed localisation approach based on the analysis of these statistical features assumes that PDs generate unique RF spatial patterns due to the complexities and non-linearities of RF propagation. This approach exploits two distinct frequency bands which hold different PD information. PD location features are extracted from the main PD signal and the two sub-band signals. These features are then used to infer PD location. Moreover, due to the increased dimensionality of data that may result from PD feature generation, feature selection algorithm; Correlation Based Feature Selection (CFS) is employed for feature selection and dimensionality reduction. The use of statistical PD features improves localisation accuracy.

This study further presents a novel method for RF-based PD localisation. The technique uses Wavelet Packet Transform (WPT) and machine learning ensemble methods to locate PDs. More specifically, the received signals are decomposed by the Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) version of wavelet packet and analysed in order to identify localised PD signal patterns. The Regression Tree algorithm, Bootstrap Aggregating method and Regression Random Forest (RRF) are used to develop PD localisation models based on the wavelet PD features. The proposed PD

localisation scheme has been found to successfully locate PD with negligible error.

Additionally, the principle of the developed PD localisation system has been validated using a separate test dataset. This approach is based on purely practical reasons, given the enormity of separate experiments to be carried out. The data required is collated over an extended time period.

The results of the investigation presented in this thesis show that an autonomous and efficient substation-wide RF-based continuous PD localisation system is possible.

Contents

Acronyms

AC Alternating Current

AIS Air Insulated Substation

ANN Artificial Neural Network

AoA Angle of Arrival

CFS Correlation based Feature Selection

DGA Dissolved Gas Analysis

DoA Direction of Arrival

EM Electromagnetic

EMI Electromagnetic Interference

GIS Gas Insulated Substation

GRNN Generalised Regression Neural Network

GS Giga Samples

HFCT Highl Frequency Current Transformer

HPLC High Performance Liquid Chromatography

HV High Voltage

IEC International Electro-technical Commission

KNN K-Nearest Neighbour

LF Location Fingerprint

LOS Line Of Sight

LSSVR Least Squared Support Vector Regression

MLP Multilayer Perceptron

MODWPT Maximal Overlap Discrete Wavelet Packet Transform

MSE Mean Squared Error

MUSIC Multiple Signal Classification

NLOS No Line Of Sight

NN Nearest Neighbour

PCA Principal Component Analysis

PD Partial Discharge

QoS Quality of Service

RF Radio Frequency

RMSE Root Mean Square Error

RRF Regression Random Forest

RSS Receive Signal Strength

SNR Signal-to-Noise Ratio

SSR Signal Strength Ratio

SVR Support Vector Regression

TDE Time Delay Estimation

TDoA Time Difference of Arrival

ToA Time of Arrival

UV Ultraviolet

UHF Ultra High Frequency

VHF Very High Frequency

WKNN Weighted K-Nearest Neighbour

WPT Wavelet Packet Transform

WSN Wireless Sensor Network

List of Symbols

A	area
C_a	capacitance a
C_b	capacitance b
C_c	capacitance c
d	thickness of insulation
t	thickness of void
ε_o	permittivity of vacuum
ε_r	relative permittivity
E_{cb}	breakdown strength of cavity
V_c	voltage across the void
V_s	source voltage
V_{si}	voltage across the insulation
X	measured signal
S	signal of interest
e	noise
ϕ	scaling function

$\psi_{j,k}$	discrete wavelet function
$\psi_{a,b}$	continuous wavelet function
$a_{j,k}$	approximate coefficients
$d_{j,k}$	detail coefficients
τ	threshold
N	number of samples
η_{soft}	universal fixed soft thresholding
l	location
$\mathbf{L}$	location vector
r	receive signal strength
$\mathbf{R}$	receive signal strength matrix
D_{euc}	euclidean distance
D_{mal}	mahalanobis distance
Σ	covariance matrix
T	transpose
x	x-location coordinate
$\hat{x}$	estimated x-coordinate
y	y-location coordinate
$\hat{y}$	estimated y-coordinate
k	nearest neighbours
w	weight
$f(r)$	function of RSS
θ	grnn activation function
σ	spread factor
D	distance in signal space
W	weight matrix
b	bias vector
ϵ	range within which errors are allowed

ζ	slack variable
T_d	training dataset
$k(r, r_i)$	kernel function
α	lagrange multipliers
γ	regularisation constant
Ψ	mapping function
C	box constraint
∂	partial derivative
SV	number of support vectors
$L(W, b, e, \alpha)$	Lagrangian
ω	representing kernel function
$h(k)$	coefficient of the low-pass filter
$g(k)$	coefficient of the high-pass filter
f_s	sampling frequency
$e_{j,n}$	wavelet energy of coefficient
SS	sum of squares
T_m	number of trees

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Background

The electricity supply industry has grown over the years to become very critical to human and societal development all over the world. Unfortunately, the ever increasing demand on electricity for both home and industrial usage leads to more and more generation, transmission and distribution of electricity. This naturally increases the stress on electrical assets such as transformers, cables, and switch gear in electricity substations leading to increased probability of equipment deterioration. Consequently, these assets could fail without advance warning. Such failures are sometimes catastrophic and may include irreversible damage leading to high cost of replacement or repair. The economic losses due to plant breakdown and unplanned power outage could be substantial. Moreover, energy regulators such as Ofgem in UK often impose fines on companies for outages/power cuts. There are also Quality of Service (QoS) incentives available for power operators

to deliver an acceptable level of performance with respect to number and duration of interruptions per 100 customers and the number of customer minutes lost [1][2]. As a result, uninterrupted functioning of these assets is crucial to facilitate efficient and reliable operations of power systems.

With the high cost associated with building new infrastructure, power utilities now face the challenge to maximally utilise their ageing electrical equipment which may be prone to failure to meet the high energy demand whilst minimising operation and maintenance cost. This has increased interest in predictive condition-based maintenance over the traditional time-based strategy, which is not only expensive and labour intensive but also ineffective in identifying problems that develop between scheduled inspections. Predictive maintenance techniques are able to evaluate the health condition of equipment in real-time by performing continuous monitoring. This helps utility operators to perform maintenance at scheduled times to achieve an acceptable trade-off between the value of asset, economic burden of failure and cost of maintenance strategy.

With the shift towards condition-based maintenance, various techniques have been researched that can assess the condition of electrical assets and assist in maintenance and management. One recognised technique is through monitoring of Partial Discharge (PD) activity [3][4]. The term PD refers to an electrical discharge that only partially bridges the insulation between conductors. PD phenomenon is caused by highly localised insulation defects such as the existence of voids in solid insulation, at conductor-dielectric interfaces, gas bubbles, or impurities in liquid insulation and along the surface of insulators [5]. Regardless of the underlying cause, PD is indicative of degraded insulation. The discharges themselves further deteriorate the quality of the insulation, thereby giving rise to a vicious circle of atrophy until total failure [6]. Therefore, instituting a continuous monitoring system permit PD activity to be detected at an early stage and proactive maintenance can be employed to avoid catastrophic failure of assets. Thus, unplanned

outages and expensive repair costs can be significantly ameliorated.

The concept of PD monitoring has become a dominant tool for assessing the condition of electrical equipment [7]. This has fostered growing interest in developing efficient and practical substation-wide PD based condition monitoring systems. One important task in building such a system is to develop techniques that can accurately locate the point of discharge. Locating the point of discharge helps in easy and quick identification of the piece of equipment that needs repair/ or replacement and hence facilitate maintenance activity. The occurrence of PD can be determined using protection equipment that monitors changes in the electric current [8]. The discharge produces acoustic emissions [9]; consequently, ultrasonic detectors have been used to determine their location. This is particularly useful in small indoor installations. Another approach is to monitor the radio spectrum for Radio-Frequency (RF) pulses emitted by the discharges [4]. Due to the suitability of RF method for larger substations, much effort has been focused on the development of RF-based PD localisation technique [10][3]. The conventional RF techniques for locating PD integrate Time-Difference-of-Arrival (TDoA) or Direction-of-Arrival (DoA) methods in the monitoring system to trilaterate or triangulate PD location [6][3]. The contact-sensor methods used to locate PDs can be prohibitively expensive and resource intensive. Each sensor is dedicated to one item of plant. This has motivated an investigation into the development of an enhanced approach to PD localisation. To this end the work presented in this thesis investigates the feasibility of implementing a relatively inexpensive RF-based autonomous PD localisation system using a network of sensors in conjunction with flexible models based on machine learning technique.

1.2 Research Objectives

This investigation is driven by the motivation to develop an enhanced PD localisation system that can be deployed to monitor an entire substation. The work reported in this

thesis explores the feasibility of deploying a matrix of radio sensors in the form of a wireless sensor network using commercially off-the-shelf components to locate PD. The hypothesis explored in this work is that PD location can be inferred from parameters of the RF signal emitted by PD using sophisticated machine learning algorithm. With recent technological advancement towards smart grid, autonomous uninterrupted monitoring of electrical facilities is becoming more desirable. Future smart grids can benefit from the relatively inexpensive but efficient localisation of PD. It is for this reason that this thesis explores the problem of machine learning enhanced radio location of PD. More specifically, the main objectives of this thesis are:

1. To understand partial discharge phenomena that occur in electrical equipment;
2. To investigate the feasibility of deploying a network of sensor for substation-wide PD localisation;
3. To investigate the implementation of received signal strength (RSS) based fingerprinting technique for PD localisation;
4. To investigate the possibility of using a wide range of PD parameters of the received RF signals as a base for current and future work on PD localisation;
5. To model the PD location problem using sophisticated machine learning techniques for efficient PD localisation;
6. To develop robust methods based on feature engineering and machine learning algorithms to improve the performance of the proposed PD localisation system.

It is important to recognise that PD localisation requires a high level of reliability, flexibility and robustness. Therefore approaches that take advantage of signal variations due to multipath are desirable and will be key to the future of condition monitoring in electrical network. It is believed that the signal analysis and PD locator models developed in this work will lead to enhanced methods for the localisation of PD in electrical networks.

1.3 Thesis Statement

Location of PD in each item of plant in electrical substations is a costly and complex task requiring increasing processing power. This work investigates a relatively inexpensive approach using an array of sensor that can be deployed for uninterrupted monitoring of an entire substation where the discharge source can be located as accurate as possible. Moreover, sophisticated machine learning algorithms are explored to reduced computational burden and enhanced localisation accuracy.

1.4 Main Contributions

The work presented in this thesis resulted in several contributions that complement the state of the art in the field of PD monitoring. Specifically, this research has:

- Pioneered a new method of PD localisation suited for the entire substation using off-the-shelf commercial radio sensors. This system has useful characteristics such as no line-of-sight restrictions, area-wide coverage, accuracy of a few meters. Thus, this research provides a robust PD localisation system well suited for the cluttered environment of the electricity substation.
- Developed an enhanced PD localisation model using fingerprinting technique. The received RF signals captured by the receiver sensors have been explored. Both absolute received signal strength and the ratios of the signal strength by sensor pairs are used as PD signal features. The fingerprinting technique provides a good alternative for PD localisation to signal propagation model where the transmit power as well as the characteristics of the environment are expected to be known a priori.
- Produced enhanced PD features to improve PD localisation. Both time-domain

and frequency domain features are extracted from the RF signals emitted by PD sources. This provides robust PD fingerprints for PD localisation.

- Developed new PD localisation models based on machine learning algorithms. The sophisticated machine learning algorithms provide a robust means of estimating PD location.
- Developed techniques to improve the performance of the PD localisation scheme. The use of feature engineering and feature selection algorithms enhanced accuracy and reduce the dimensionality of the feature space for low computational cost.

1.5 Thesis Outline

This thesis consists of seven chapters. Chapter one provides the background of the study, research aims and objectives, research contributions and the structure of the thesis.

Chapter two provides a general overview on the concept of PD mechanism and definitions of PD. This chapter also provides the causes and effect of PD in electrical equipment and a general description of the types of PD and characteristics nature of PD waveforms. A general description of PD detection technique is presented. Furthermore, relevant previous research work on the application of RF PD monitoring and localisation in electrical substation are reviewed.

In chapter three, the PD localisation problem is formulated. Experimental procedure undertaken to demonstrate substation-wide RF PD localisation is presented. The laboratory experiment involves the design of a measurement grid in the laboratory and the use of pulse generator as PD emulated source. The pulse generator is set to produce repeatable and characteristic (unique) PD pulse signals at each grid point. Three sensors (monopole antennas) are deployed at arbitrary positions in the laboratory to capture the

RF pulses emitted by the PD source. This chapter also provides the description of the PD data.

Chapter four presents an investigation into the possibilities of implementing fingerprinting technique for PD localisation using Received Signal Strength (RSS). To determine PD location, the RSS of the RF signals are captured and used as PD input features into the localisation algorithm. This chapter also introduce a centralized method for PD localisation. The method consists in defining a model, whose inputs and outputs are respectively the RSS fingerprints and the PD locations. k-Nearest Neighbour (KNN) and Artificial Neural Networks (ANN) are used to model PD localisation problem. The chapter concludes with performance analysis and evaluation of the localisation models on the PD data collected.

Chapter five focuses on localisation methods that combine radio-location fingerprinting and kernel-based learning technique. Beginning with the original idea of Support Vector Machine (SVM), a Support Vector Regression (SVR) model is developed for PD localisation. A low complexity variant of the proposed SVR technique; Least Square Support Vector Regression (LSSVR) is formulated and implemented for PD localisation. More robust PD features are computed from RSS. Instead of absolute RSS values, their ratios are used as input to the kernel-based models developed. Different kernels are used to evaluate the performance of the models with focus on accuracy.

Chapter six introduces technique for improved PD source localisation. Feature engineering and selection tools are explore to develop hybrid machine learning methods for PD localisation. New PD features are extracted from the RF signals. The PD data analysis leads to a rich feature database that facilitate PD location. The Correlation Based Feature Selection (CFS) algorithm is discussed and implemented for feature reduction/selection. Also Wavelet signal analysis is used in conjunction with robust ma-

chine learning techniques to improve PD localisation. Performance evaluation of these techniques is presented.

Finally, chapter seven is a summary of the thrust of the findings of all research aspects proposed in this thesis, conclusion, direction and recommendations for future work.

1.6 Research publications

The work described in this thesis has been published in the following:

Peer Reviewed Journals:

- **E.T Iorkyase**, C. Tachtatzis, I. A. Glover, P. Lazaridis, D. Upton, B. Saeed, R. C. Atkinson., "*Improving RF-based partial discharge localisation via machine learning ensemble method*", IEEE Transactions on Power delivery, vol. 34, no. 4, pp. 1478 1489, 2019.

- **E.T Iorkyase**, C. Tachtatzis, P. Lazaridis, I. A. Glover, R. C. Atkinson., "*Low Complexity Wireless Sensor System for Partial Discharge Localisation*", IET Wireless Sensor Systems, vol. 9, no. 3, pp. 158-165, 2019.

- **E.T Iorkyase**, C. Tachtatzis, I. A. Glover, R. C. Atkinson., "*RF-based location of partial discharge sources using received signal features*", IET High Volt., 2019, vol. 4, iss. 1, pp. 28–32.

- **E.T Iorkyase**, C. Tachtatzis, P. Lazaridis, I. A. Glover, R. C. Atkinson., "*Radio location of partial discharge sources: A support vector regression approach*", IET Science, Measurement and Technology, vol. 12, no. 2, pp. 230–236, 2018.

Peer Reviewed Conferences:

- **E.T Iorkyase**, C. Tachtatzis, I. A. Glover, R. C. Atkinson., "*Localisation of partial discharge sources using radio fingerprinting technique*", Loughborough Antennas

Propagation Conference (LAPC), Loughborough, 2015.

CHAPTER 2

Literature Review

2.1 Introduction

A significant cause of plant failure in electrical substations is attributable to insulation degradation. PD is a result of electrical insulation breakdown irrespective of the causal mechanism. PD activity is also a predominant cause for further insulation degradation which may eventually lead to catastrophic failure of electrical equipment with significant societal and economic consequences. Since PD usually occurs before complete breakdown, PD monitoring provides a forewarning to de-energise plant before catastrophic failure occurs. As a result, PD monitoring is considered as one of the most valuable non-destructive means of assessing the insulation condition of assets, thereby condition-based maintenance strategies can be formulated. In this chapter, the fundamentals of PD are presented. The definitions and mechanisms that leads to the occurrence of PD are described. Methods of PD detection are outlined. This includes the conventional methods

and the non conventional modern techniques. This chapter also present previous works on radiometric methods for PD measurement and source localisation.

2.2 Partial Discharge

Partial discharge has been recognised as a pernicious ageing process and a major factor that accelerates degradation of electrical insulation from the beginning of last century when High Voltage (HV) technology was introduced in the utility industry for the generation and long-distance transmission of electrical power [11]. PD monitoring has received considerable attention over the last decade as majority of breakdown in electrical equipment is caused by PD [12][13][14]. According to the International Electro-technical Commission (IEC) 60270 standard on high voltage test technique - partial discharge measurement, PD is defined as a localised electrical discharge that only partially bridges the insulation between conductors which may or may not occur adjacent to a conductor [15]. PD generally occurs whenever there is a local electrical stress concentration in the insulation or on the surface of the insulation as a result of dielectric imperfection, such as sharp edges in ambient air, floating particles in gas insulation and gaseous voids or inclusions in liquid and solid insulation. The presence of any defect in the insulation structure can cause this local field enhancement and consequently local perturbation in the insulator giving rise to impulsive pulses otherwise known as partial discharge. Here, a PD pulse is a voltage or current pulse resulting from a discharge event and can be measured by suitable detector circuits [15]. It is important to note that PD activity can occur anywhere within the dielectric, where the applied electric field strength exceeds the breakdown strength of that portion of the material. As PD is a small electrical avalanche caused by locally disrupted electric fields in dielectric materials, it is seen as both a symptom of insulation degradation and a cause for further deterioration of the insulating material [16][17]. Therefore, continuous monitoring of equipment in electricity substations is essential in detecting and locating PD before it grows into complete breakdown of insulation and

causes unexpected power failure.

2.3 Partial Discharge Occurrence and Mechanisms

PD usually begins for example within voids, cracks or at conductor-dielectric interfaces in insulation system, in bubbles within liquid insulations or inclusions within solid dielectrics. PD can also occur along the surface of solid insulators or boundary between different insulating materials. The presence of PDs may indicate defects in the insulation materials.

Generally the occurrence of PD is based on two conditions [18][19]; the first is the availability of a starting (free) electron to initiate an electron avalanche within the volume in which discharge is expected and the second is that there must be a sufficiently high electric field within the volume to start an avalanche by ionisation and create a breakdown path. The electron avalanche related to the ionisation process leads to the deployment of free charges on the void surface [20]. The resulting charges move towards the surface of the void in the direction of the applied field. These charges are then trapped at the void surface resulting in the apparition of an induced electric field which opposes the applied electric field. This phenomenon occurs within a very short time (in the range of microseconds or less). The total field across the cavity drops rapidly and it leads to the extinction of the discharge. Following this transient event, the trapped charges may decay through recombination processes of positive and negative charges. After some time, the field is again large enough to start a new discharge. Since PD are associated with the ionisation of molecules, the fundamental mechanisms of PD include Townsend and streamer discharge mechanism [21][18]. In Townsend mechanism, the gap current grows as a result of ionisation by electron impact and electron emission at the cathode by positive ion impact [21]. The electrons and ions create a small space charge compared to the external field and can be neglected, thus the ionisation process is mainly a direct

ionisation. The Townsend discharge occurs within a short-gap and has the following forms: rapid and slow rise time spark-type pulses, true pulse-less glows or pseudo-glow discharges [18] which are all cathode emission sustained discharges. The rise time of these pulses can be as long as several tens of nanoseconds and its duration can last several hundreds of nanoseconds [22]. Streamer discharges are independent of cathode emission, but dependent upon the photoionisation process in the gas volume [18]. The streamer discharge occurs in longer gaps and involves ionisation wave propagation in a very high field region where ionisation and influx of electrons at the discharge head is produced by a space charge field due to the separation of positive and negative charges [18]. The space charge fields have an important role in the corona and spark discharge in non-uniform field gaps [23]. Streamer discharges are commonly of shorter duration than Townsend discharges; the streamer discharge pulse length can vary from 1 ns to 10 ns [18].

In general, the occurrence of PD can be analysed by considering the well known electrical "a-b-c" model [24] of cavity in a dielectric material. The cavity is usually filled with gas. The "a-b-c" model assumes that the dielectric components can be replaced with capacitances. Figure 2.1 shows a potential PD source (gas-filled cavity) in solid or fluid dielectric material between two electrodes X and Y and its equivalent electrical circuit (a-b-c model) represented by a simple capacitor arrangement. The gas-filled cavity in Figure 2.1 (a) is represented as C_c in Figure 2.1 (b), and the two parts of the insulation above and below the cavity form the two capacitances $C_b^{'}$ and $C_b^{''}$ [25]. The remaining insulation outside the cavity is represented by $C_a = C_a^{'} + C_a^{''}$ [25]. The total capacitance of the insulation in series with the cavity is given as in Equation 2.1;

$$C_b = (C_b^{'} * C_b^{''})/(C_b^{'} + C_b^{''}). \tag{2.1}$$

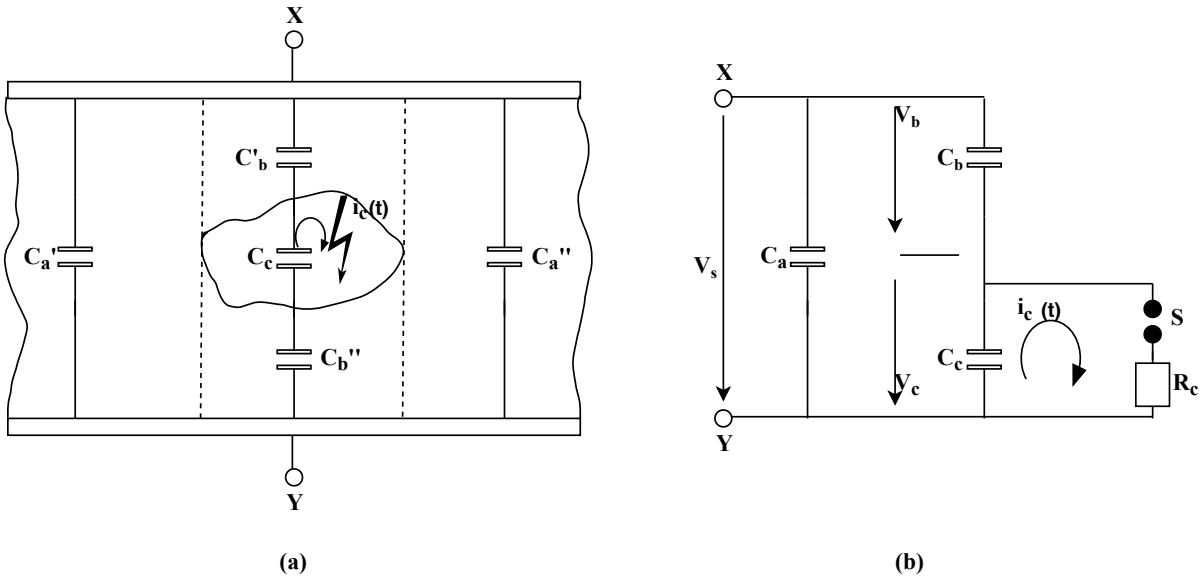


Figure 2.1: Basic PD circuit, an insulation system with a cavity defect and Electrical equivalent circuit, [26][27]

From Figure 2.1, let's assume the thickness of the insulation is d and that of the void is t with area A. Here, the values of the the capacitances C_a, C_b and C_c depend on the geometry or dimensions and dielectric constants of respective portions a, b and c. Assuming the breakdown strength of the cavity is E_{cb}, the PD inception voltage can be derived in terms of the void as follows; Let the capacitance of C_b and C_c be expressed as in Equations 2.2 and 2.3;

$$C_b = (\varepsilon_o * \varepsilon_r * A)/(d - t) \tag{2.2}$$

and

$$C_c = \varepsilon_o * A/t \tag{2.3}$$

where ϵ_o is the permittivity of vacuum and ϵ_r is the relative permittivity of the solid dielectric. Then the voltage across the void can be calculated as shown in Equation 2.4;

$$V_c = (C_b/C_b + C_c)V_s. \tag{2.4}$$

Substituting C_b and C_c into Vc gives Equation 2.5;

$$V_c = V_s/[1 + (1/\varepsilon_r(d/t - 1))]. \tag{2.5}$$

Thus the voltage across the insulation which initiates PD in the void can be expressed as in Equation 2.6;

$$V_{si} = E_{cb}.t(1 + 1/\varepsilon_r(d/t - 1)) \tag{2.6}$$

where V_{si}=voltage across insulation which causes the void to start to discharge. E_{cb}=breakdown strength of the void material.

PD disrupted voltage associated with the void and the dielectric represented by capacitance in Figure 2.1 is linked to recurrence of discharges. When an alternating current (AC) voltage is applied, the field gradient in the void begins to increase due to the difference in permittivity caused by inter-facial polarisation as well as by the shape of the cavity. The capacitance is charged and the voltage across the void will continue to increase until the breakdown strength of the gas in the void is exceeded. This causes the void gas to ionise, causing a spark or a discharge to appear. The first discharge will appear at the crest or rising part of a half-cycle [28]. When breakdown occurs, charge is transferred across the void, reducing the voltage across the the void. Once the field across the void has been sufficiently reduced, the breakdown channel extinguishes. This breakdown process is shown in Figure 2.2, where V_c representing the voltage in the void is just a fraction of V_s, the applied voltage. When V_c reaches V^+, which is the gas breakdown voltage inside the void, discharge will occur and the voltage in the void will drop to V^- at which point the discharge stops. After the discharge occurs, the voltage again builds across the void, as the applied voltage continuous along its normal cycle. At this point, the critical breakdown strength within the void is not reached before the AC voltage reaches its peak value, resulting in a single discharge per half cycle. When the applied AC voltage increase, the critical breakdown strength in the void may again be reached before the AC voltage reaches its peak value. This results in multiple discharges per half cycle as shown

in Figure 2.2. This will occur several times. The discharges in the void cause impulses to concentrate in areas where the applied voltage increases or decreases.

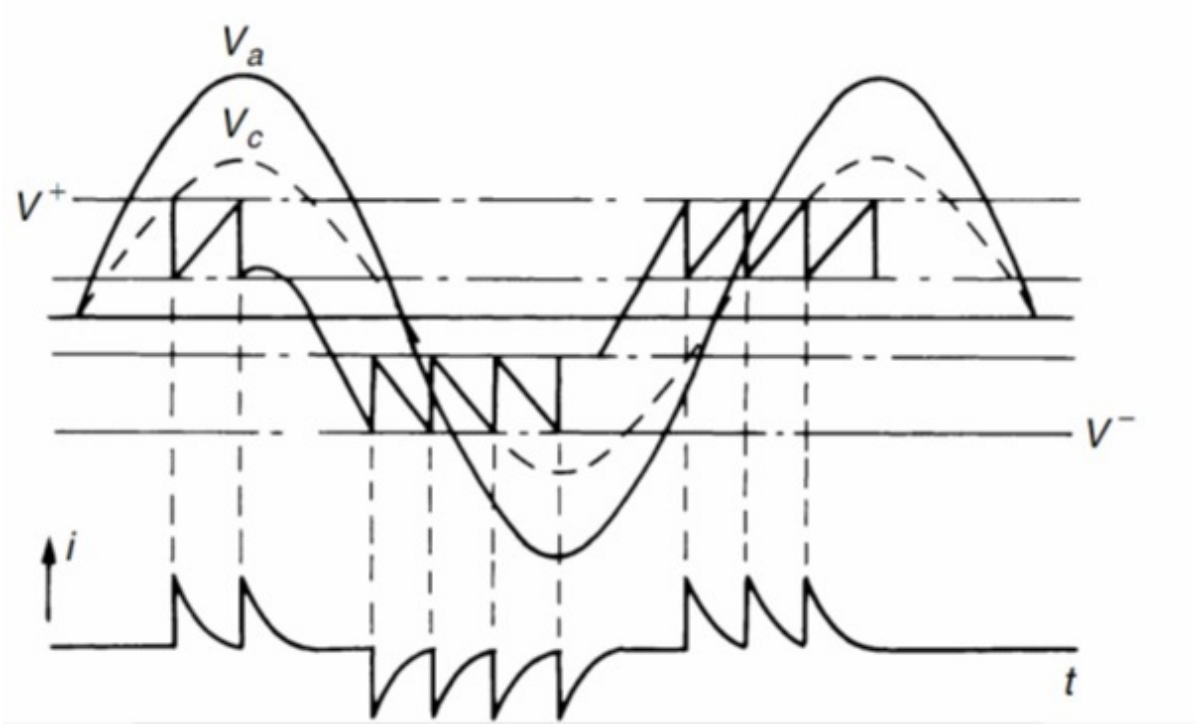


Figure 2.2: Sequence of PD occurrence, [26]

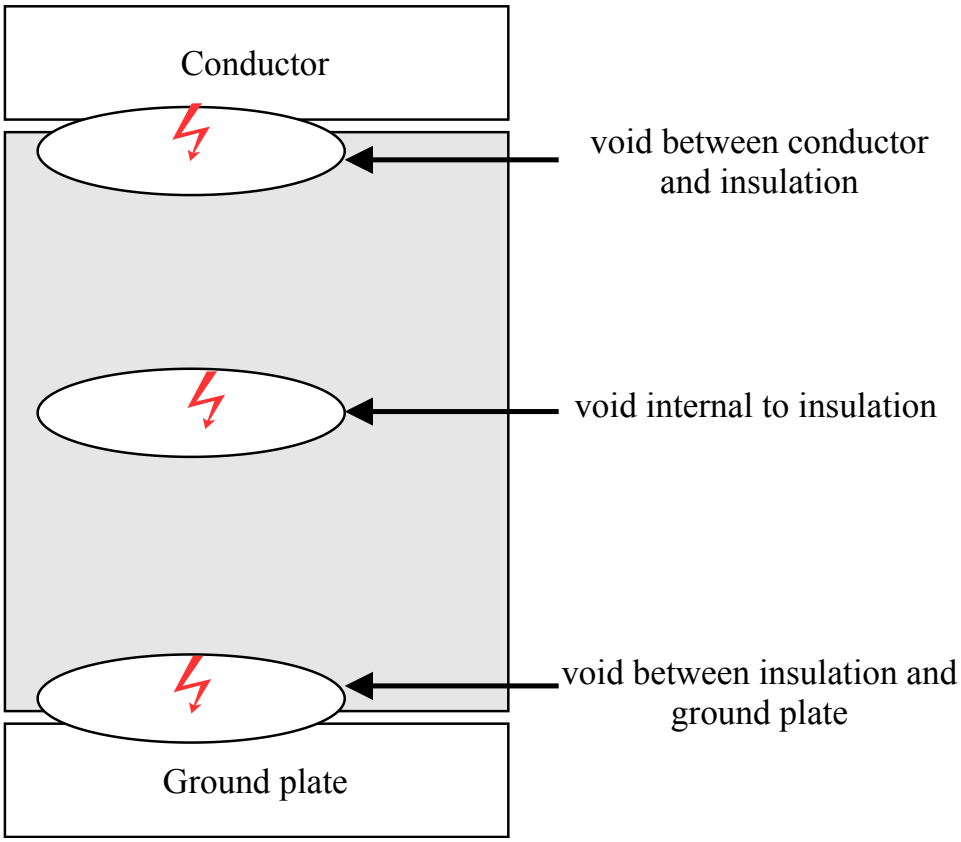
2.4 Partial Discharge Types

Based on the origins and the mechanisms that lead to discharges, PDs can be classified into four categories, namely; surface discharges, internal discharges, corona discharge, and electrical trees, [21] [23][29][18].

2.4.1 Internal Discharge

Internal discharges usually occur inside the insulation due to the presence of voids or inclusion of particles or near insulation interfaces. Figure 2.3a is a representation of a void internal discharge. It shows the possible locations of voids within the insulation system. The void may be located between the conductor and insulation wall, or internal

to the insulation itself, or between the insulation wall and the grounded. These voids are regarded as weak points within the insulation caused by poor manufacturing or ageing [30]. They are filled with gas or liquid materials. The filler materials are usually of lower permittivity and dielectric strength such that the void section is of higher electric field compared to the surrounding sections of the insulation. PD activity is normally initiated in the void with less voltage than the breakdown voltage of the whole insulation. This explains the reason for the random nature of PDs. The discharges are effectively small sparks occurring within the insulation system, thereby deteriorating the insulation, and can result in ultimate complete failure. Figure 2.3b shows the effect of these internal discharges on an epoxy resin busbar [30]. This can result into the so called insulation treeing.



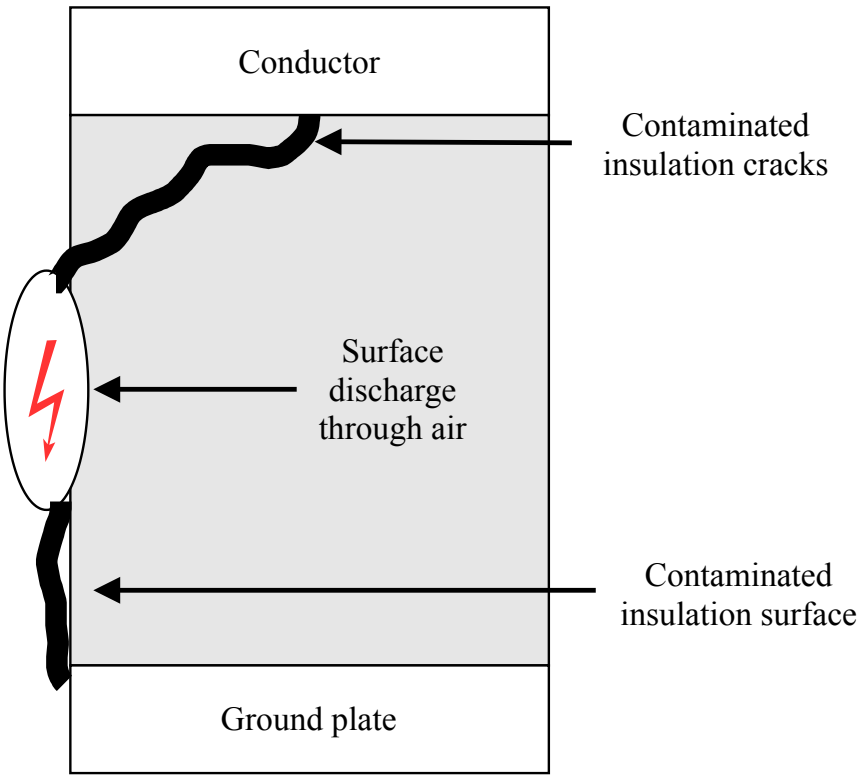
(a) Schema of internal discharge

(b) Effect of internal discharge, [30]

Figure 2.3: Internal discharge and its effect

2.4.2 Surface Discharge

Surface discharge also known as surface tracking [31] usually occur along the interface between insulation and air. Figure 2.4a is a representation of surface discharge on an insulation system. At the interface, any irregular protrusion will have a stronger electric field than the other portions. This stronger electric field ionises the surrounding air, thereby causing it to become conductive. PD may begin to happen in this region. The presence of moisture, water pockets or contaminations can form a leakage path to ground. As a result, a leakage current is generated [32]. The heat generated by the leakage current due to ohmic loss combine with the area of the ionised air causes vaporisation of the moisture component of the leakage path, consequently creating cracks and gaps on the surface. This results in the interruption of current flow at some point, and since the entire surface is still conductive, most of the voltage drops across the dry surface creating an arc across it. The heat from the arcs result in the burning of the insulation surface leading to the production of permanent carbonised paths on the surface otherwise known as surface tracking. When the moisture has been completely dried, the cavities created between carbonised paths become new sources of PD. This process continues across the insulation until failure results. This type of discharge is known to be common in paper insulation but seldom occurs on polymer or ceramic insulation types [23]. An example of electrical device under surface tracking is shown in Figure 2.4b. The damage caused by surface tracking results in permanent electrically conductive path that reduces the effective voltage the material can withstand and may eventually result in complete breakdown.



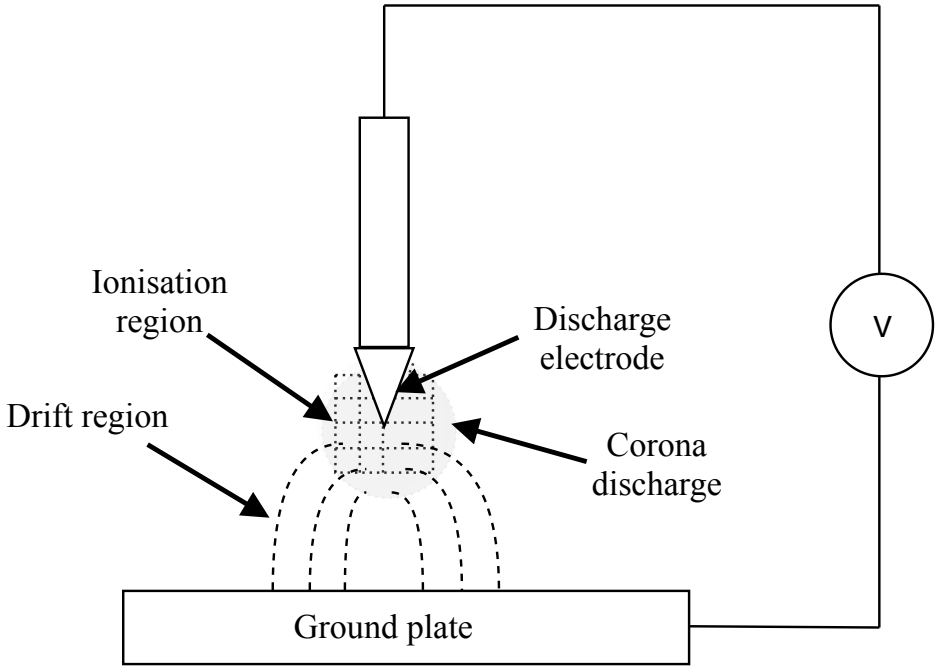
(a) Schema of surface discharge

(b) Effect of surface discharge, [30]

Figure 2.4: Surface discharge and its effect

2.4.3 Corona Discharge

Corona discharge is a form of electrical discharge that occurs around sharp points or edges of electrodes placed near an insulating surface [21]. A schema of corona discharge produced by a sharp pointed electrode is shown in Figure 2.5a. The sharp edges enhance the field around that region, thus increasing the stress. This high stress can initiate an ionisation process of the medium (air) surrounding a conductor thus producing the discharge call corona. The energies such as heat and light produce during corona are of low intensity and could be considered harmless at the initial stage, however can become damaging to the insulation over time due to their cumulative effect and failure may occur without warning. Corona is often see as a glow in the air adjacent to a pointed conductor carrying high voltage. An example of a glow discharge on an insulator string of an overhead power line is as shown in Figure 2.5b. Corona discharges constitute economically significant waste of energy for electric utilities.



(a) Schema of corona discharge

(b) Effect of corona discharge, [30]

Figure 2.5: Corona discharge and its effect

2.4.4 Electrical treeing

Electrical treeing is a type of discharge known as a pre-breakdown discharge in that it occurs just before catastrophic failure. It is so named for it tree-like structure [33].

Electrical treeing emanates from weak spots within the insulation such as gas voids, sharp electrode edges, or particle inclusion, where the electric field is high. These defective spots can grow progressively over time forming trees inside the insulation. There are generally three stages of tree development, namely; initiation, propagation and runaway stages [34]. The initiation stage is the starting point of tree growth where electrons are released due to the excitation of the molecules within the weak spot between two electrodes. The propagation stage represents the phase between the tree initiation and when the first branch of the tree reaches the opposite electrode. The runaway stage is what is known as the acceleration phase. It occurs at the point where the tree is almost crossing over the material to the other electrode.

2.5 Partial Discharge Detection Techniques

The condition of insulation system in high voltage equipment determines to a large extend the reliability of such equipment. As a result, early detection of any symptom of insulation deterioration becomes an important component of HV equipment monitoring and maintenance. PD is considered an effect of dielectric defects and a cause for further insulation degradation that can be harnessed for effective condition monitoring of electrical equipment in substations. Even though there is no direct correlation between occurrence of PD and HV equipment failure, it has been established that the presence of PD may lead to complete breakdown of the equipment overtime [20]. The presence of a defect in an insulation enhances the local electric field such that it exceeds the dielectric strength of the rest of the insulating medium. This triggers ionisation within the insulation [35][36]. The liberated electrons that result of the ionisation process are accelerated in a flow of electric current [15]. The current pulse that flows in the insulating medium is what is known as PD. This current is quenched once the local electric field falls below the level required to sustain further ionisation [37][38]. Generally, the occurrence of PD often involves the exchange and dissipation of energy caused by the build up of

electric field intensity in a local region of the insulator. These energy exchanges manifest into macroscopic-physical effects such as chemical reactions in the insulation composition, electrical pulses, acoustic emissions in the audible and ultrasonic ranges due to the rapid thermal expansion and contraction of the discharge channel [39]. PD also produces electromagnetic emissions in the form of radio waves, light and heat. Monitoring experts leverage on the released energies to detect the presence of PD using appropriate sensors or measuring device.

This section describes a variety of measurement techniques that exploit these fundamental physical and chemical processes for the detection of PD. Based on these processes that occur during a discharge, PD detection methods can be broadly classified into two groups; namely conventional and non-conventional methods as shown in Figure 2.6.

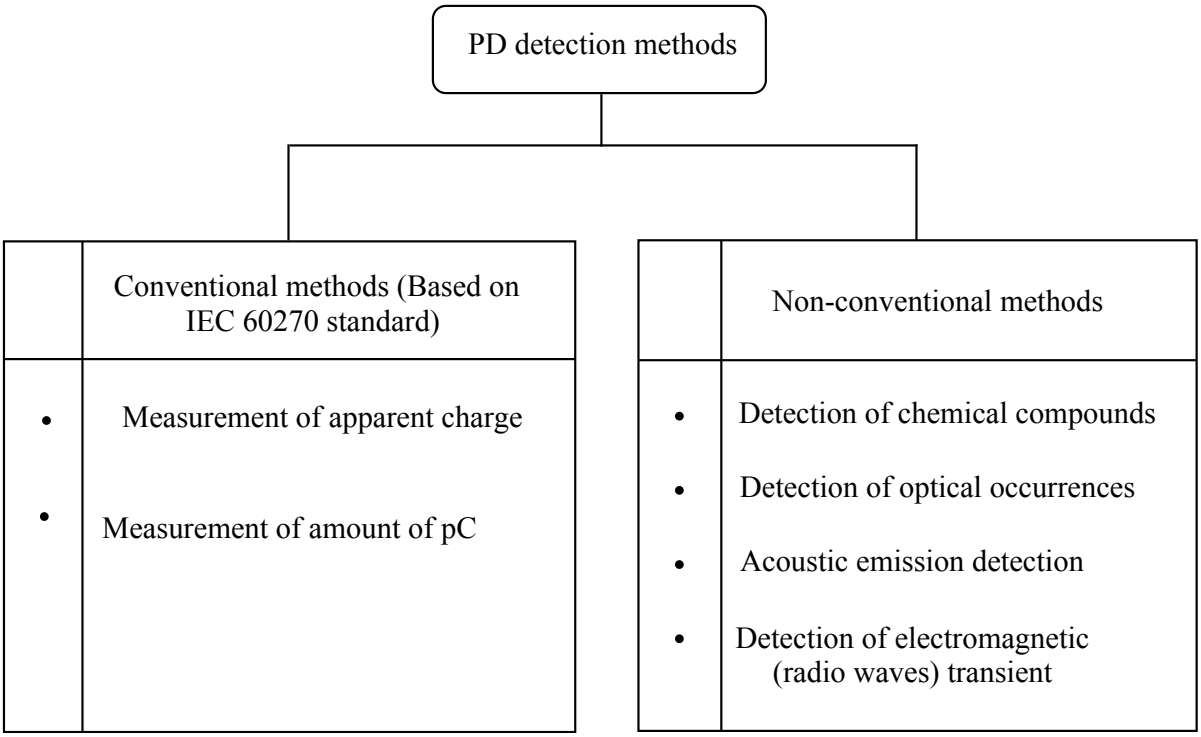


Figure 2.6: PD detection methods

2.5.1 Conventional Partial Discharge Detection Methods

This section describe the conventional PD detection techniques that is based on the measurement of apparent charge;.

2.5.1.1 IEC 60270-based Technique

The standardised electrical method for detecting PD pulses has been described in IEC 60270 standard publication [15]. This technique comprises of a coupling capacitor and is based on measuring the apparent charge q of the pulse usually expressed in pico-coulombs (pC). Apparent charge q of a PD pulse is that charge which, if injected within a very short time between the terminals of the test object in a specific test circuit, would give the same reading on the measuring instrument as the PD current pulse itself. Due to the inaccessibility of the discharge site where PD pulses originate, the apparent charge is not a direct measure of the PD current but is obtained from the voltage drop across the terminals of the test object and is not equal to the actual quantity of charge released at the discharge site. Phase-resolved PD (PRPD) patterns are often obtained using conventional IEC 60270 technique. The PRPD consists of PD measurement recorded over a period of time represented on a 3-dimensional $\phi - q - n$ plot, where ϕ is the phase angle for the PD pulse, q is the apparent charge and n is the number of discharge [40]. A significant advantage of IEC 60270 measurement technique is that the PRPD correlate with the extent of insulation damage and are widely available for different discharge defects and for different HV components. Other quantities that can be measured include; PD pulse amplitude, pulse repetition rate, and average discharge current [41][42]. Although the conventional electrical method is an accurate method for measuring sudden short current pulse generated during PD activity [42] and is widely used in PD detection, it has several limitations. For example, this method cannot provide the location of the discharge source and the true PD waveform are not recorded during measurement. Moreover, the

application of the IEC 60270 method in some electrical apparatus such as Gas Insulated Substation (GIS) can be difficult or impracticable due to the requirement of external coupling capacitor for individual sources. Therefore, the electrical technique is not ideal for the proposed substation-wide PD monitoring and localisation system.

2.5.2 Non-conventional Partial Discharge Detection Methods

Several non-conventional techniques such as electromagnetic, optical, and acoustic methods have been developed over the years for PD detection [43]. These methods are known as non-conventional because they do not conform to the IEC 60270 standard since they measure different PD quantities compare to apparent charges set out in IEC 60270 standard. These methods are discussed as follows.

2.5.2.1 Chemical Detection Technique

PD accelerated current streamer across a void or cavity can influence chemical decomposition of the surrounding insulation material into different components. These chemical changes can be exploited for the detection of PD in electrical equipment. For example in SF6 (sulfur hexafluoride) and oil insulation, the chemical reactions will produce a combination of constituent gases such as ozone and nitrous oxide which will diffuse freely or become dissolved in the liquid. The relative concentrations of by-products produced from the degradation of the insulation can be related to the type of fault and the rate of reaction can be indicative of the severity of the fault [44]. The most commonly used chemical detection methods are High Performance Liquid Chromatography (HPLC) and Dissolved Gas Analysis (DGA) [45]. HPLC measures by-products such as glucose or degraded forms of glucose, produced from the breakdown of the insulator. The test is carried out in the laboratory by evaluating samples collected from the insulation. In DGA the volume or level of the gas produce is indicative of the presences of PD. Different levels of each of the gases can be correlated to a specific type of fault.

Although Chemical detection technique have the advantage that measurement are immune to spurious events such as electromagnetic interference, it still has some limitation. chemical detection technique provides little or no insight into the location of PD [46]. Sufficient by-products or gases have to be collected and taken to laboratory before analysis can begin. Chemical detection methods are also considered less sensitive for general PD monitoring. They are often tailored to specific applications and can not be generalised for different equipment. Chemical detection technique is therefore not suitable for real-time PD continuous monitoring and hence proposed PD localisation technique considered in this thesis.

2.5.2.2 Acoustic Detection Technique

PD activity is also accompanied by sound generation and this provides an alternative for PD detection and localisation. The acoustic signals are pressure waves generated by the rapid thermal expansion and contraction of the discharge channel. This signals can range from audible to ultrasonic frequencies. The acoustic waves can propagate from the discharge site in a narrowband range, approximately 10kHz to 300kHz and can be picked up by pressure transducers and converted to electrical signals for PD analysis [47]. This technique is one of the most common technique that has been effectively used for locating PD in power transformers and Gas Insulated Substations (GIS)[47][48][49][50][51]. The significant advantage of this method lies in it ability to provides a level of immunity to external noise sources such as Electromagnetic Interference (EMI), which pose a great challenge to other methods. However, acoustic emission sensing is less popular with equipment such as cables and enclosed devices due to the complex nature of the acoustic wave propagation [47]. Also the intensity of the received acoustic wave is inversely proportional to the distance from the discharge source. Therefore acoustic sensor positions required can be overwhelming, increasing cost. This makes acoustic detection unsuitable for the PD localisation scheme proposed in this thesis.

2.5.2.3 Optical Detection Technique

PD emits electromagnetic radiation during discharge as a result of various processes such as excitation, ionization and recombination and optical sensors can be used to detect the electro-luminescence effect produced. Though all electrical discharges emit electromagnetic radiation, the optical spectrum depends on the chemical composition of the insulation medium and the nature of discharge. The optical spectrum ranges from the ultraviolet to the infrared range. For example the wavelength of faint corona discharge is between 280nm and 410nm, appearing as purple, about 95 percent is in the ultraviolet (UV) region. Strong flash discharge emits radiation in the 400nm - 700nm spectral range [52]. The optical sensor used for PD detection is a direct function of the optical spectrum. The advantage of optical detection is it high immunity against electromagnetic interference and sensitivity compared to conventional method especially for on-site measurement. However, this technique depend on the medium within which PD is expected. The insulation medium is required to be optically transparent. This limit the used of optical detection technique to optically transparent medium such as gas and liquid since most solid dielectrics are not optically transparent. The optical technique is insensitive to any form of internal PD and can be highly directional. The measuring equipment can be very costly. This makes it difficult to consider optical technique for the proposed PD monitoring and localisation scheme.

2.5.2.4 Electromagnetic wave detection Technique

To overcome the limitations of the other detection methods mentioned in this thesis, electromagnetic wave method for PD measurement was developed. PD usually occur due to the displacement of charges [53]. This gives rise to a rapid rate of change of current pulse that occurs during the breakdown process. As the pulse rise time become shorter due to rapid rate of change, EM waves with energy spectra extending into the radio frequency region are produced [54]. As a result, a percentage of the PD energy is

radiated into free-space in the vicinity of the discharge site and can travel to long distances with minimal attenuation in different medium. This makes radiometric PD measurement possible without interrupting a facility's electrical service. It can also be applicable to all HV power apparatus. PD measurement in higher frequency range (30MHz - 3GHz) was first proposed for gas-insulated substations [55][56]. This technique has also been proven for PD diagnostics of HV cable accessories [57] and later for power transformers as reported in [58]. Researchers have conducted extensive investigations into the capability of PD detection in the VHF/UHF range for on-site application [59][60][61].

Because of the extremely wide electromagnetic spectrum of PD pulses, which covers the radio frequency ranges of high frequency (HF 3-30 MHZ), very high frequency (VHF 30-300 MHz) and ultra-high frequency (UHF 300-3 GHz), various kinds of PD couplers have been developed in the past, which are commonly classified as capacitive, inductive such as Rogowski coils and high frequency current transformer (HFCT) electromagnetic sensors [62][63][64]. Recently, much research effort have been directed towards designs of RF remote sensing units [65][10] for PD detection and computer analysis of the PD signals obtained. However, one of the most important benefit obtained from this technique is the ability to locate PD sources. Methods that use remote sensors to detect the propagated electromagnetic wave at different frequencies such as VHF and UHF [58][66] are among the simplest ones. The summary of the comparison between different PD detection and measurement methods is presented in Table 2.1. The RF technique in comparison with other PD detection techniques is particularly convenient and appropriate PD detection method for the proposed substation-wide PD localisation system.

2.6 Localisation of Partial Discharge sources

The basic principle of PD monitoring of electrical apparatus is to detect energy release during PD activity due to the deteriorating condition of the insulation system prior to

Table 2.1: Comparison of PD detection methods.

Methods	Electrical	Chemical	Optical	Acoustic	EM/RF
Description	Electrical circuit that measures the pulse current generated during PD event as a result of charge transfer.	Analysis of compounds produced during chemical changes that occur in the composition of dielectric materials.	Measures the electro-luminescence effect produced during PD.	Measures the acoustic waves generated by a PD event.	Measures electromagnetic waves in the radio frequency range produced by PD.
Merit	Good sensitivity, high precision measurements, very low signal attenuation, standard for all HV equipment.	Well established and easy to measure, immune against noise, highly sensitive.	Immune against EMI, non-contact, applicable to all voltages.	Non-contact, applicable to all voltages, allows measurement on equipment under operating conditions, location of PD is possible.	Non-contact, applicable to all voltages, allows measurement on equipment under operating conditions, location of PD is possible, can be calibrated.
Demerit	Sensitive to electrical noise, impracticable for equipment under operating condition, affected by EMI.	Cannot provide location of PD, amount/level of dissolved gas has no correlation with type of PD, create uncertainty.	Insulator must be optically transparent, highly directional, impossible to calibrate.	Sensitive to other acoustic emissions, signals cannot always propagate through insulation.	Interference of signals by environmental noise, EMI emissions can prevent PD detection.
Sensor types	Capacitive, inductive	DGA, HPLC	Fiber optics (UV) sensors	Acoustic sensor, piezoelectric sensors, fiber optic acoustic sensors.	RF antennas, HFCT, TEV, radiometers, UHF sensor.
Area of application	Can be applied to all HV equipment	Power transformers, GIS	Power transformers, GIS and cables	Power transformer and switchgear	Cab be applied to all HV equipment

failure. An important aspect of PD monitoring is the possibility to reliably and accurately determine the location of a PD source and hence identify the particular item of plant experiencing PD. As a result of which maintenance can be carried out in the part of plant most likely to be at risk with high accuracy at significant cost saving.

PD source localisation comes with it challenges such as cost and interference. Given that PD can occur inside insulator, across the surface of insulating material or around an electrode and that most practical PD localisation methods depend on the detection of the energy radiated by PD event, the localisation process is affected by the structure of the equipment, the nature of the propagation environment and external interference. The metallic portion of the equipment can form a radiating structure that would influence the radiation of energy from the PD site. Due to resonance, the specific structure of these conductors leads to frequency selective radiation and as a consequence, the PD waveform is distorted. The signal received is not exactly what is been produced at the discharge site. Other metallic structures in substations may also influence the nature of the PD signal received due to attenuation, scattering, reflection and multipath propagation. Background noise/interference from internal or external sources which are not related to the detected PD signals may occur. These make the noise environment to be cluttered and challenging for accurate PD localisation.

Nevertheless, the drive towards a national target of 0 % breakdown events culminate into increased interest to investigate robust and cost effective technique for PD source localisation in order to enhance continuous monitoring of electrical substations. Radio-location presents itself as an effective solution for PD localisation problem.

There are several approaches to RF signal based PD source localisation. One of the technique used is based on TDoA [67][68][69][10][70] from pairs of synchronised receivers. The Time of Arrival (ToA) of the PD pulses at each sensor can be used to calculate

the TDoA between each sensor pair. This information can then be used to estimate the distance from the source to a sensor and thereby trilaterate the location of PD source. In order to locate PD sources using TDoA, a minimum of three sensors are required to record PD pulses emanating from the discharge source. RF antennas have been used to detect and locate PD sources in GIS [71][72][73], AIS [3][74] and power transformers [17][75] by measuring the TDoA of the RF signals. A TDoA-based method was proposed by [76] for localisation of PD sources via detection of EM waves generated by the PD. Here, four time-synchronised antennas were installed with a movable simulated GIS in a utility's 500kV substation. The TDoA that provides the basis for the localisation of PD was determined via first peak technique [67]. Due to the presence of noise, a smoothing process (moving average) was employed to discard noisy peaks. In [3] the authors describe an RF-based PD early warning system for an air-insulated substation where four antenna array were mounted on top of a building. They employed the TDoA method to trilaterate PD location. Here, the cross-correlation method is used to determine the TDoA based on wave-front. A radiometric PD localisation method based on TDoA proposed in [10], consists of four element radio antenna array used to detect the radio frequency interference generated by PD. The locating equipment is considered suitable for use in the vicinity of energised HV plant and can locate sources up to 15m from the array. In [77] and [78], the time delay method based on energy accumulation was employed to estimate the location of a PD source in three dimensions (3-D). This is done by finding the knee point where there is a sudden change from the energy curve. This method depends on human judgement to decide the knee point and this can be a source of error. The setup is composed of four omnidirectional micro-strip antennas and four omnidirectional discrete disk-cone antennas. In another investigation [79], the fourth order cumulant and bispectrum estimation method were used to calculate TDoA for improved PD source localisation.

Another technique that has been employ for radiolocation of PD sources is based on Direction of Arrival (DoA) or Angle of Arrival (AoA) [80] of received PD signals. The usual approach for calculating DoA is to use four receivers in the VHF and UHF bands forming an array of 2 x 2 elements. This array is virtually expanded to increase the resolution and then, the direction of the emission of the PD is computed using a multiple signal classification (MUSIC) algorithm [81]. A remote detection of PD in power transformers has been demonstrated using mobile wideband radio-frequency (RF) receiving equipment [6]. The antenna array mounted on the roof of a vehicle was moved in steps of 2m around the transformer in order to monitor the radiated field from all possible positions on the transformer. The bearing of PD source was determined by calculating the DoA of the impulse with respect to each pair of antennas in the array. DoA PD localisation method based on an L-shaped antenna array proposed in [82], is composed of four UHF omnidirectional antennas. The DoAs of PD signals were computed using the principle of estimation of signal parameter via rotational invariance techniques. A vehicle, such as a van, furnished with the necessary equipment has been used to periodically monitor substations for the presence of PD [83]. In this investigation, the bearing of the discharge source is determined by computing the DoA of each possible pair of antennas. The authors in [84] developed an autonomous PD source localisation inside large power transformer winding under laboratory conditions. This is based on the analysis of measured current energies associated with different frequencies of the signal. They measured the currents at the bushing tap point and neutral to earth connection. The assumption used in this work is that different PD source locations will generate unique signal profiles in both frequency and time domain. Wavelet transform algorithm and Principal Component Analysis (PCA) were applied to PD signals to extract features and reduce the dimensionality of the feature vectors. They represented their system using a series of finite impulse response filters for practical purposes. A TDoA method for locating multiple sources of PD using only two antennas is presented in [85]. One of the receiver antennas is installed in a fixed position, while the second receiver is made movable in the circumference of the first one. The

two receivers are mechanically linked by a rod. The TDoA from the sources collected at different angles is used to locate the sources. The authors aim at reducing the number of sensors, hence reducing the cost of instrumentation. Authors in [86] deployed four RF antennas with different positioning arrangement (square, trapezoidal and pyramidal) in a volume of cube to detect the radiated electromagnet energy from PD and evaluate the location of PD sources based on TDoA. In [87], the authors proposed a methodology for locating single and multiple PD sources using the acoustic signals that emanate from such sources. The localisation model is based on sequence of arrival and levels of peak amplitude of the acoustic waves captured by different acoustic sensors mounted on the outside walls of the cubical box used in the experiment. Another method of PD source localisation on transformers is presented in [88]. The method is based on singular value elimination with planar arrangement of UHF sensors. This method uses the gap leakage EM wave signals generated during PD event. The method maps the 3-dimensional coordinate sample set of PD source to 1 dimensional coordinate and uses the improved box-plot diagram method to eliminate the singular values. The least square method is then used to find the optimal coordinate of the real PD source. A general survey on the performance of source localisation algorithms was conducted by authors in [89]. These include multilateration algorithms; Standard Least Squares (SLS)and Hyperbolic Least Squares (HLS), Particle Swarm Optimisation (PSO) algorithm, Maximum Likelihood Estimator (MLE), Bancroft algorithm, Hyperbolic Positioning Algorithm (HPA) and a combination of these algorithms. The analysis was undertaken with three different sensor layouts. This was carried out using time differences and errors due to the effect of digital time sampling.

Previous work on radiolocation of PD sources are based on TDoA/Time Delay Estimation (TDE) and DoA [74][90][91][70] of the RF signal. However, time based methods require accurate synchronisation, making them computationally complex and energy hungry, hence are not economically viable solution for the proposal substation-wide PD

localisation system. DoA requires an array of directional antennas and relies on Line Of Sight (LOS) path which is difficult to achieve in the cluttered substation propagation environment where PD is expected. It also brings extra computational cost. Both TDoA and DoA are seen as uneconomic.

The cost/complexity of the aforementioned methods motivates an investigation into methods based on the use of RSS due to its simplicity, low-power consumption and cost effectiveness. However, the common RSS triangulation approach requires detailed models of RF propagation and does not account for existing variation. It is practically impossible to implement in situations where the transmit power is unknown which is the case under study. An alternative approach considered in this thesis is to use signatures of received RF signals such as RSS to infer PD location. The key challenge is how to effectively and efficiently model RSS/location relation and hence derive PD location from RSS. The complexity of this inverse problem motivates the use of flexible models based on machine learning. This approach not only obviates the need for a propagation model but can also improve localisation precision. Also in the proposed method, sensor array does not require any physical/electrical contact with equipment, and can be permanently installed for continuous monitoring of energised item of plant. The machine learning based technique turns the multipath phenomenon to surprisingly good use: by combining the multipath pattern with other signal characteristics, it creates a signature unique to a given location. To this end, this thesis focuses on non invasive radio PD monitoring to develop an enhanced PD localisation system using off the shelf commercially available sensors.

2.7 Summary

In this chapter, the background of partial discharge phenomena is introduced. Five common PD detection methods have been compared with each other, including electrical

detection, chemical detection, optical detection, acoustic detection and electromagnetic detection. The EM detection method that measures EM waves in the radio frequency range has been identify as the most suitable PD detection technique for the proposed PD localisation system. A review of RF-based PD localisation techniques is presented. The drawbacks of the commonly used PD localisation techniques; TDoA and DoA are discussed. These techniques are often deploy to monitor one electrical apparatus at a time and can be prohibitively expensive. The approach employed in this thesis which is the subject of the next chapter is to deploy inexpensive commercially off-the-shelf sensors at strategic locations as a proof-of-concept for a substation-wide PD localisation.

CHAPTER 3

Partial Discharge Measurement and Signal Characterisation for Source Localisation

3.1 Introduction

The use of radio signals emitted during PD events for localisation is beneficial for many reasons, as discussed in the previous chapter. This chapter leverages the RF sensing technology to investigate the feasibility of deploying a matrix of radio sensors in an electrical substation to monitor the radio spectrum for RF pulses emitted by the discharges and ultimately determine PD location. The propagation characteristics and properties of the PD received signals as they relate to localisation are studied. This analysis is needed to understand the underlying features of location-dependent patterns in the received signals. The received signal strength is used to demonstrate the location dependence of PD received signal features. An understanding of the location properties of received signal

features for fingerprinting can provide help in improving PD localisation system. Measured PD signals and their derivatives are used in further analysis as input data to train, and validate proposed PD localisation techniques. To this end, this chapter reports a PD measurement campaign undertaken in this study as a proof of concept for the proposed substation-wide PD localisation. A set of mathematical assumptions is proposed at the end of this chapter based on the analysis. These assumptions will be used in the later chapters for developing PD localisation models.

3.2 Experimental Setup

In order to evaluate the PD localisation system proposed in this thesis, a systematic measurement campaign was developed to simulate/emulate PD events in electrical substation and hence produce measurement data for further analysis. The experiment was conducted in a 19.20 m x 8.40 m laboratory at the University of Strathclyde. The laboratory contains chairs, cubicles, computers, antennas, robot arms and different metallic objects on the tables, walls and ceiling. There was also human movement/activities in the laboratory. The radio environment within this laboratory is not necessarily similar to that in a substation. However, the salient point is that the presence of clutter in both environments is likely to give rise to a complex, multipath-rich, propagation environments. The algorithms to be applied herein demonstrate an ability to tune themselves to the underlying radio environment. Furthermore, a more complex (multipath-rich) environment is likely to give rise to more distinguishable feature and hence facilitates the proposed PD localisation technique. Therefore, the radio environment within this laboratory is sufficiently complex to enable evaluation of the PD localisation techniques being investigated. In real-life deployment, a substation may be subject to occasional changes in topography that affect the radio environment: temporary or permanent changes to the layout. The proposed system will undertake periodic re-training of the model to accommodate for these factors.

The acquisition system for the measurement campaign consists of a pulse generator, three inexpensive omnidirectional radio frequency (RF) antennas and a high speed four (4)-channel LeCroy SDA900 serial data analyser with 40GS per sec maximum sample rate. The pulse generator is capable of generating sub-nanosecond current pulse which is fed to a monopole antenna. The monopole antenna has a wide bandwidth of 930 MHz (70 - 1000 MHz) with an omnidirectional radiation pattern. The radiating monopole antenna therefore represents an artificial PD source. During the experiment, the pulse generator was set in order to generate pulses which have a duration of $10ns$, repetition period of $10\mu s$ and amplitude of $10V$. A sample of the PD emulated pulse is as shown in Figure 3.1.

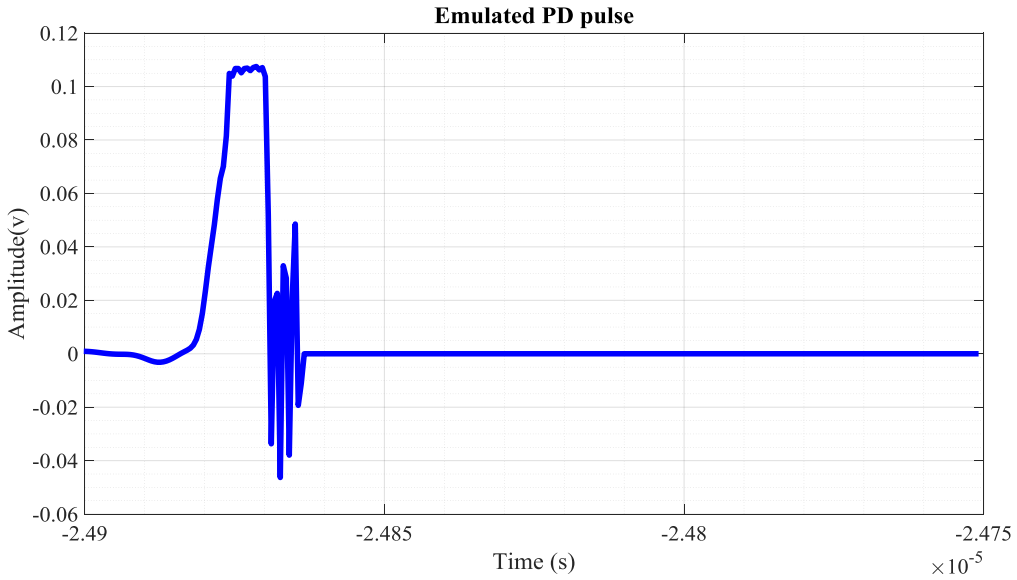


Figure 3.1: Emulated PD pulse.

Figure 3.2 shows the arrangement of the floor map of the laboratory for the experiment. In this experiment, two distinct data sets were collected: a training set and a test set, all the training sampling points are indicated by the blue circles which we refer to as PD training locations and the test points indicated by red squared marks referred to as PD test locations.

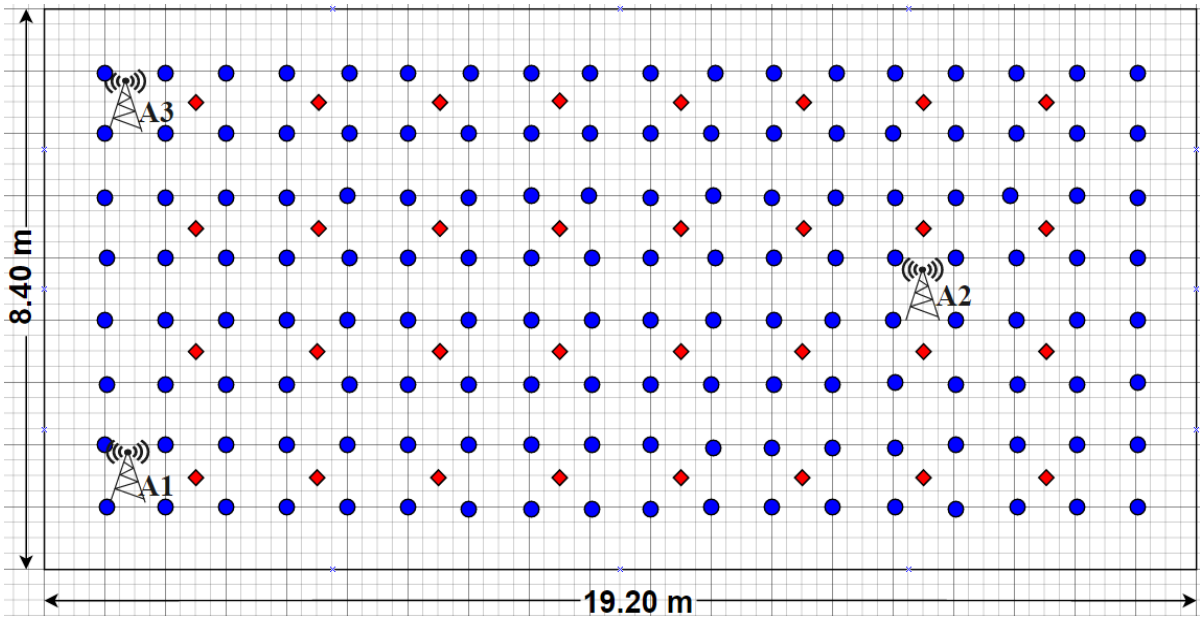


Figure 3.2: Grid for measurement campaign.

These training and test locations simulate PD sources and are so called given that they represent different sets of data to be used in this work. There are 144 training sampling points located on a uniform grid with spacing of $1m$. The artificial PD source elevated at a height of 1.9m was made movable and used to generate emulated PD traces at each of the 144 sampling points. This implies that the height was constant throughout the experiment. Therefore, this work will be limited to 2-dimensional localisation problem. However, the concept used in this thesis can be easily extended to a 3-dimensional problem. Another measurement was taken at the 32 distinct test points in the same way but on different days. Both the 144 training and 32 test locations thus represent sources of discharge. As a proof of concept, three RF antennas positioned at predefined locations in the laboratory as shown in Figure 3.2, with antennas A1, A2 and A3 stationed at (1.5m, 1m, 1.62m), (15m, 4.5m, 1.62m) and (1.5m, 7m, 1.62m) respectively were connected to the multichannel digital oscilloscope to capture and record PD measurements. The antennas are positioned in a triangular configuration. The experiment utilised commercially-off-the-shelf antennas: 1/4 wavelength antennas operating at 173 MHz to sense radiated pulses. The maximum gain of the antenna is in fact 200.7 MHz. This is shown in sub-

sequent chapters. 50 ohms BNC coaxial cables were used to interconnect the antennas to the oscilloscope terminal in order to minimise the reflections of the signal from the antennas. The risetime of the pulse is sufficiently fast () to extend into RF spectrum. The oscilloscope records two separate information; timestamp data and waveform data (three blocks of time-series data relating to RF signals recorded on each of the three antennas. All pulses were sampled at 2 GS/s to ensure adequate resolution of the signal in the time domain. The PD measurements acquired will be used in subsequent analysis and their derivatives as inputs to the proposed localisation algorithms. This set-up ensures that the training and testing RF signal samples measured are disjoint enough to provide realistic results.

3.3 Data Collection

For the purposes of the work reported in this thesis, a training data set of PD measurement was gathered from the experimental set-up described in section 3.2 via the three RF sensors at each of the 144 grid points (locations). A total of 2880 PD measurements, corresponding to 20 consecutive RF measurements per grid point per antenna were collected. This data is used to build PD radio map for the purpose of training the PD localisation models developed in subsequent chapters. A second, independent, test set was gathered in the same way, 20 consecutive RF measurements but at 32 distinct locations (the inter-grid red squares in Figure 3.2). The first dataset was further divided into training and validation subsets, which are used to train and optimize the developed PD localisation models. This will be discussed further in subsequent chapters. The independent test set was subsequently used to benchmark the performance of the models. PD features (e.g RSS) extracted from the RF measurement collected are used as fingerprint input vectors to the localisation algorithms. Such features are expected to be location dependent.

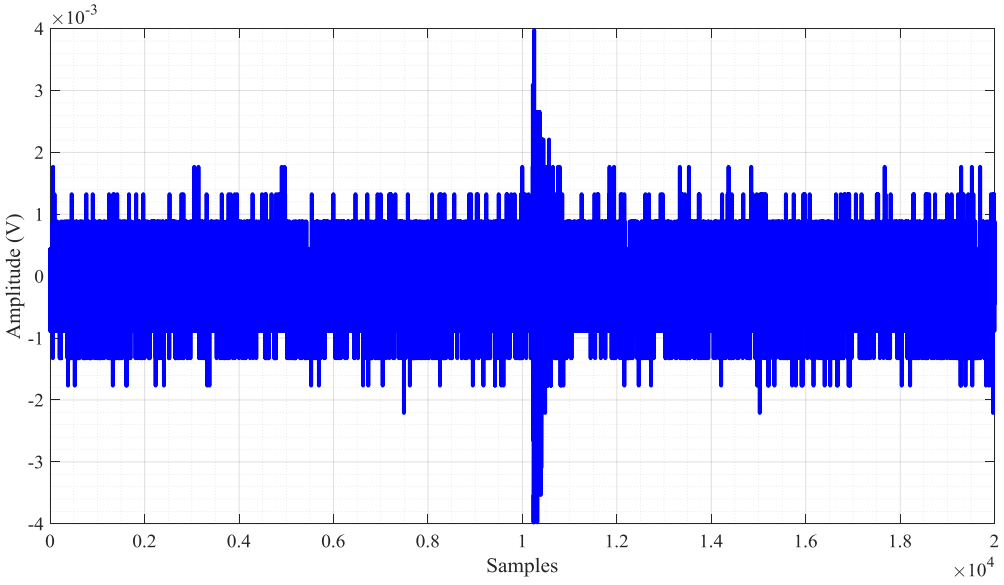
3.4 PD-Signals De-noising

The RF signals that result from PD measurements are often contaminated by noise from the environment. This could be from communication systems, switching operations, lightning and/or random noise. Such noisy measurements with low signal-to-noise ratio (SNR) present difficulty in PD detection, feature extraction and by extension PD localisation [92]. The PD measurements collected during experiment are expected to have different SNRs. Visualisation and SNR of the signals help to highlight the effect of noise on the measured signal. Samples of the RF signals recorded during the measurement campaign are as shown in Figures 3.3 (a), (b), (c) and (d). The definition of SNR used in this thesis is base on the amplitude of the signal. This definition is given in Equation 3.1 [93]:

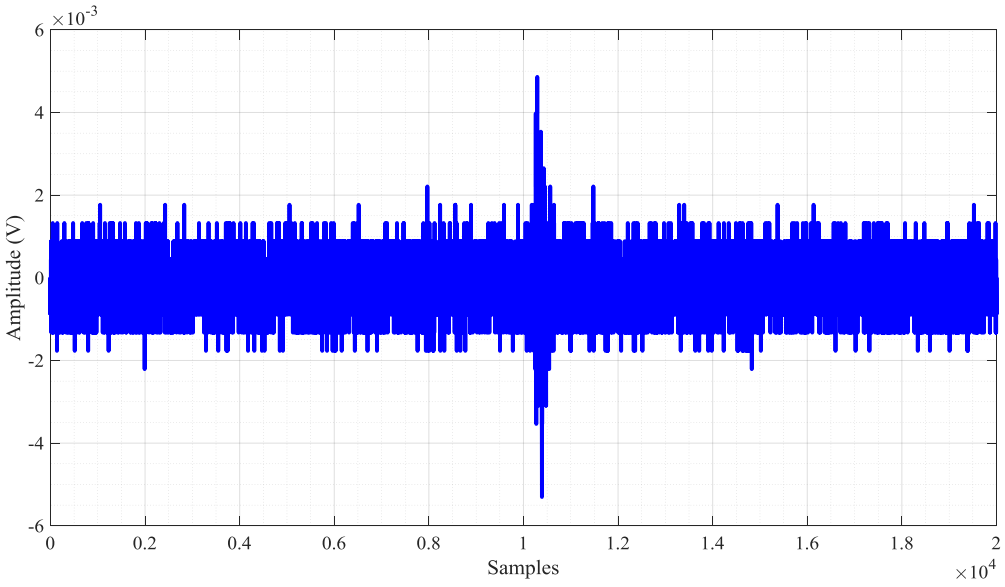
$$SNR = 10log(\sum_{n=1}^{N}[X(n)]^2 / \sum_{n=1}^{N}[X(n) - S(n)]^2) \tag{3.1}$$

where $X(n)$ is the measured (original) signal, $S(n)$ represent the de-noised signal, and N is the length of the signal. The values of SNR (-7.87, -7.20, -4.68 and -3.22 dB) of the four measured signal samples shown in Figure 3.3 is an indication of the low SNR of the PD measurements in the reported experiment. The PD signals are totally corrupted by the presence of noise. As a result, the desired time-domain characteristics of the signals are not clearly identified. The noise is modelled as white Gaussian noise (which include quantisation noise and random disturbances), wherein the frequency components are distributed over all frequency range while the signal of interest lies within a specific range of frequency. Extracting PD pulses from such noisy measurements before further analysis is therefore crucial and will be the first step towards the proposed PD localisation system. It is also important that this is done in such a way that the features of the PD pulse are preserved as much as possible. The noise component of the measured RF signals can be removed by de-noising process. The stochastic and transient nature of PD signals demands sophisticated techniques for its analysis. Wavelet transform technique which

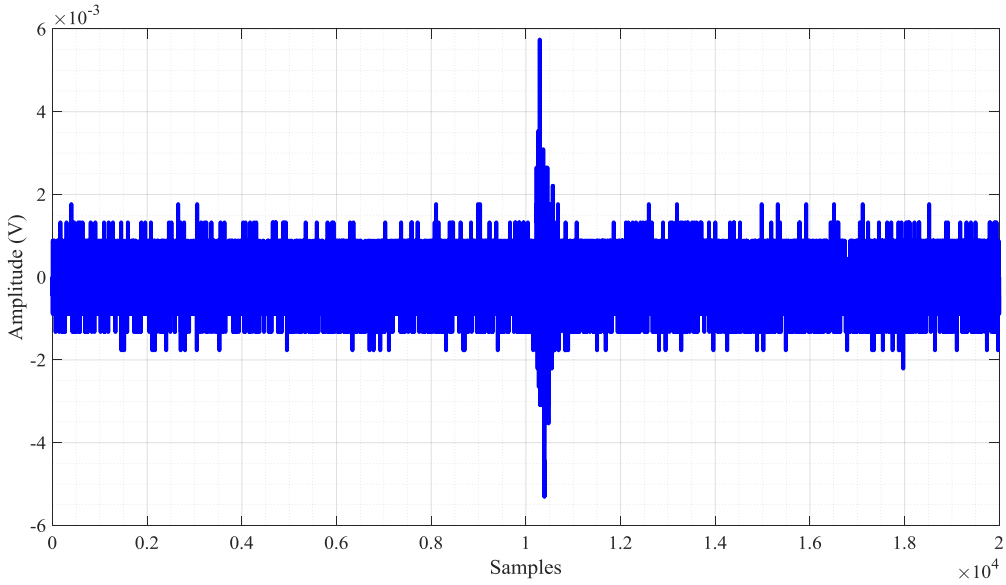
involves decomposition of measured data and thresholding is usually used for de-noising PD measured data [94][95]. In this thesis, a de-noising technique [93][96][9] that combines Wavelets [94][95][97] and Principal Component Analysis (PCA) [98][99] is used to de-noise the received RF measurements under low SNR without assuming any a priori knowledge about the PD features.



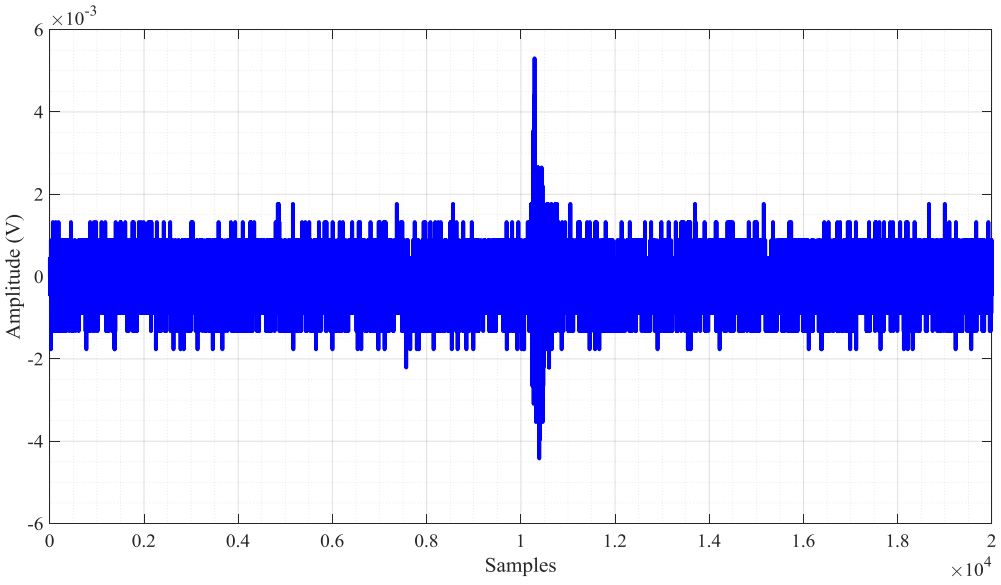
(a) Pulse 1



(b) Pulse 2



(c) Pulse 3



(d) Pulse 4

Figure 3.3: Noise corrupted PD signals.

3.4.1 Wavelet-PCA de-nosing

The PD data de-noising problem can be modelled as in Equation 3.2;

$$X(n) = S(n) + e(n) \tag{3.2}$$

where X, the measured signal is m-dimensional (m = number of data channels) which contains information on the propagation environment, S is the PD signal of interest, and e is the noise (white Gaussian) component of the measured signal. The goal is to recover the approximate signal of interest by removing the effect of noise as much as possible. The de-noising technique used in this thesis is carried out in two stages, namely; Wavelet decomposition and PCA. The resulting lower detail coefficients that contain the noise are then eliminated by thresholding. The flowchart for the Wavelet-PCA de-noising technique used in this thesis is as shown in Figure 3.4.

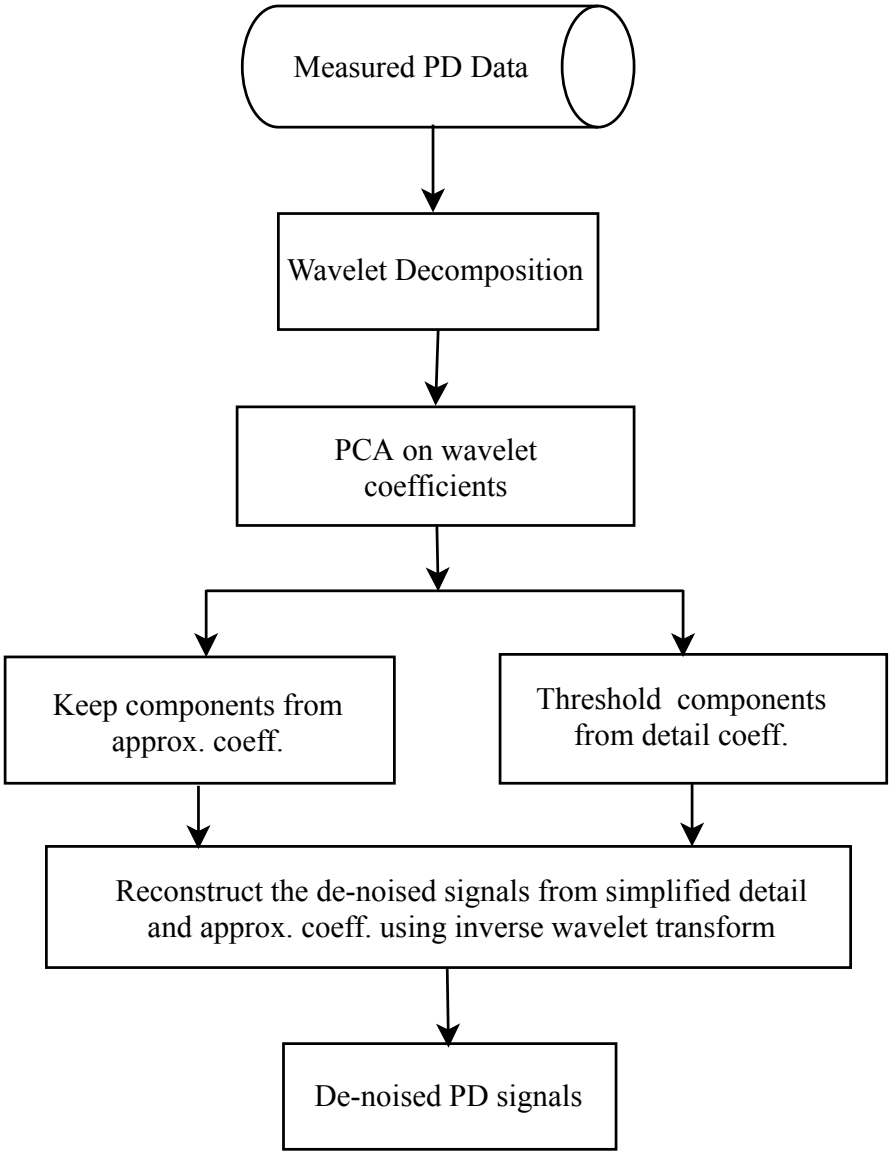


Figure 3.4: Flow chart of the wavelet-PCA de-noising.

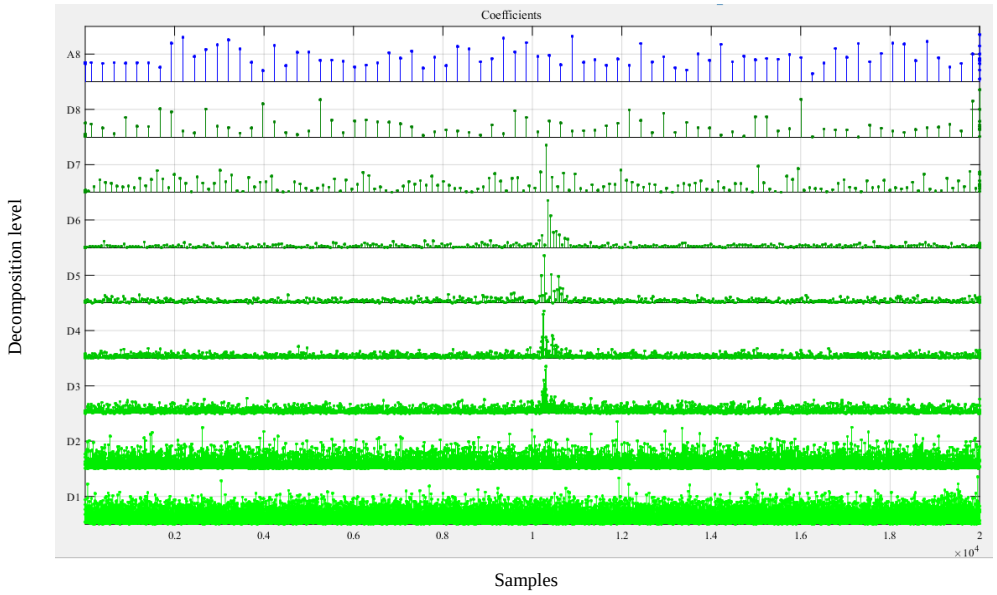
3.4.1.1 Wavelet De-noising

The wavelet-based de-noising method is the most commonly used technique for removing noise interferences from partial discharge (PD) measurements. This is based on thresholding the detail coefficients that contain the noisy component of the measured signal. The wavelet de-noising process is accomplished in three steps. Firstly, the signal is decompose into approximation and detail coefficients up to 8 levels using the Daubechies wavelet filter [95]. Figures 3.5 (a), (b), (c) and (d) show the decomposed approximation, A_8 and

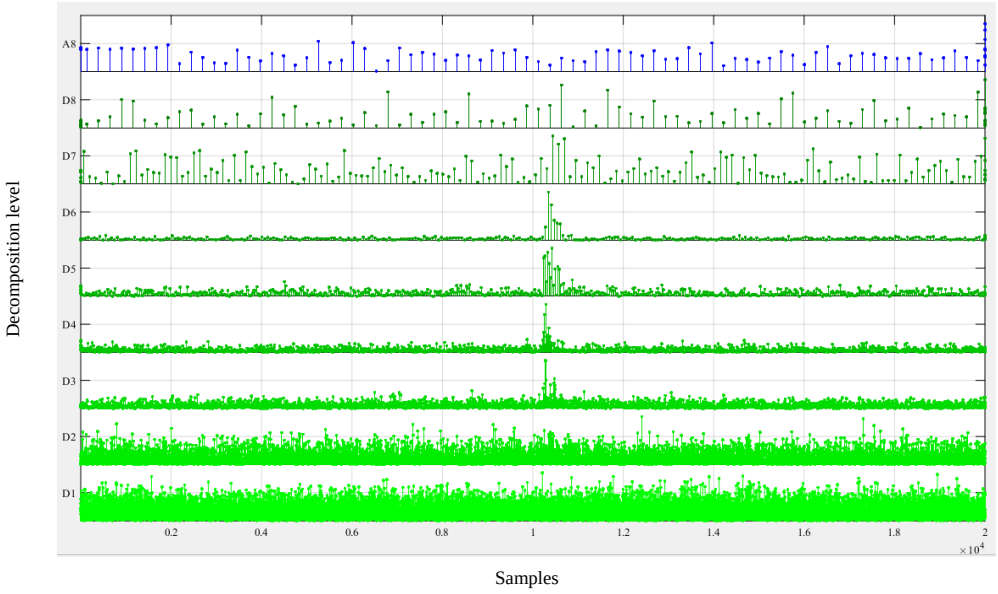
detail coefficients, $D_i, i = 1, ..., 8$ of the PD samples. The decomposed signal x can be expressed in the form of Equation 3.3;

$$\hat{x}(n) = \Sigma_{k=-\infty}^{+\infty}(x, \phi_{J,k})\phi_{J,k}(n) + \Sigma_{j \leq J}\Sigma_{k=-\infty}^{+\infty}(x, \psi_{j,k})\psi_{j,k}(n) \tag{3.3}$$

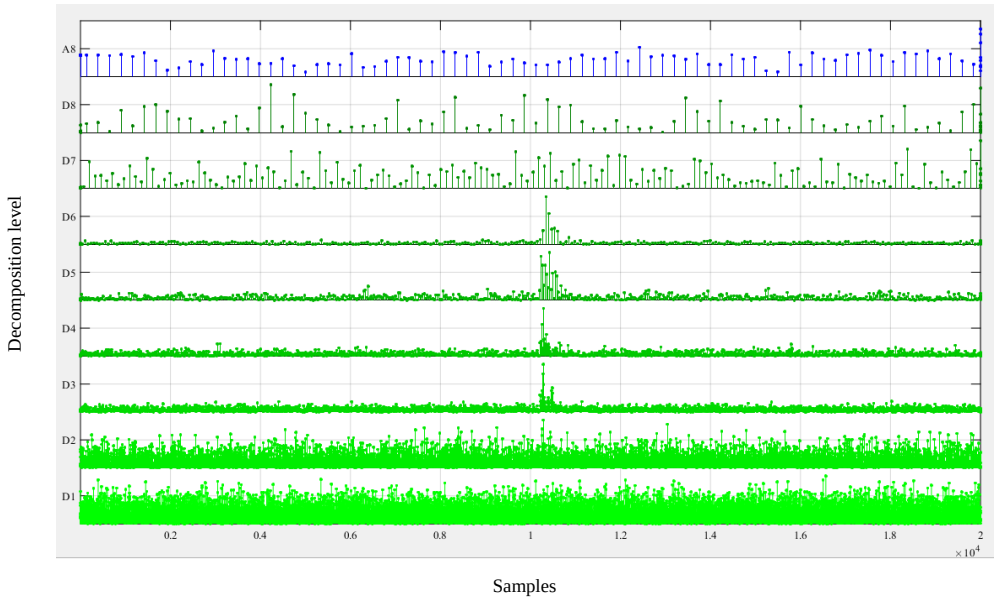
where ϕ is the scaling function and ψ is the wavelet function.



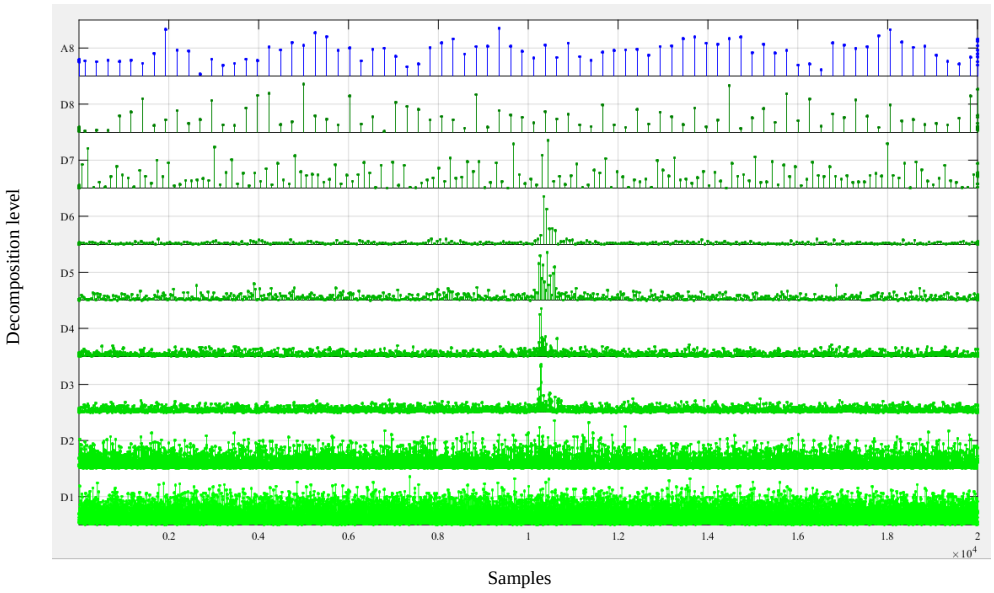
(a) Pulse 1



(b) Pulse 2



(c) Pulse 3



(d) Pulse 4

Figure 3.5: PD wavelet decomposition.

The approximate coefficients, $a_{j,k} = (x, \phi_{j,k})$, describe the overall waveform of the signal, whereas detail coefficients $d_{j,k} = (x, \psi_{j,k})$, describe the fine localise features of the signal. In other words, the energy in the signal of interest is distributed amongst the higher scale coefficients whereas the lower scale coefficients contain the energy of the noise. Hence the lower scale detail coefficients correspond to the noisy part of the signal. Thresholding

these detail coefficients via a defined threshold say τ eliminates the noise component of the signal and thus highlight the important information. This forms the second step in the wavelet de-noising technique. In PD de-noising, the most frequently used thresholding rule is soft thresholding, which shrinks all detail cofficients $d_{j,k}$ below τ to 0 [96]. In this thesis, we employ the universal fixed form of soft thresholding as defined in Equations 3.4 and 3.6 respectively. This selection is informed by the fact that PD measurements are often characterised by low signal-to-noise ratio (SNR).

$$\tau = \hat{\sigma}\sqrt{2log(N)} \tag{3.4}$$

where N is the length of the signal and $\hat{\sigma}$ the estimate of the noise standard deviation given by Equation 3.5 [96].

$$\hat{\sigma} = \sqrt{2}mad(d_1)/0.6745 \tag{3.5}$$

where $mad(d_1)$ is the median absolute deviation of the detail coefficients at level 1.

$$\eta_{soft}(x) = \begin{cases} x - \tau & x > \tau \\ 0 & x \leq \tau \\ x + \tau & x < -\tau \end{cases} \tag{3.6}$$

where x is the data and τ is the threshold. The third and final step involves the reconstruction of a de-noised version of the original signal from the threshold coefficients using the inverse wavelet transform. Wavelet de-noising method is preferred because it does not consider homogeneous noise structures and generates accurate reconstruction of the de-noised signal with respect to original signal [98]. However, the classical wavelet de-noising technique requires prior knowledge of signal characteristics and that the PD triggers above the noise level [97][100]. Therefore another layer of de-noising otherwise known as PCA is desirable.

3.4.1.2 PCA

PCA is a statistical method used in the analysis of multi-channel data through an orthonormal projection. PCA does not only have ability to decompose a signal into constituent components but also to remove uncorrelated noise in the signal. As a result, PCA has been applied to many areas of research including data compression, digital image enhancement, dimensionality reduction and de-noising. Even though PCA methods vary significantly with application, they are similar in that they deconstruct a multi-channel signal into a set of orthogonal bases of decreasing energy. Geometrically, the PD data can be seen as an $m \times n$ in an n-dimensional data space, where m is the number of channels and n is the number of locations where measurements were taken. Given the measured PD data consisting of multiple traces, $x_{i,j}, \quad i = 1, ..., n; \quad j = 1, ...m$, where the ith trace is associated with a particular location and j represent a particular channel of observation at that location. What PCA does is to generate a rotation matrix that rotates the coordinate system of the data traces onto a set of orthogonal directions called Principal Component (PC) directions such that the first axis is along the direction where the data has greatest variance and so on [101][102]. In such a rotation, the resulting components are in order of decreasing energy. The first components with most energy tend to capture the coherent data signal whereas the later components with less energy tend to represent the incoherent noise in the data. The rotation matrix can be defined by decomposing the covariance matrix $\mathbf{C}$ of the data into its corresponding eigenvalues and eigenvectors as in Equation 3.7;

$$C = VPV^T \tag{3.7}$$

where P is a diagonal matrix consisting of eigenvalues of C and V is the eigenvector matrix whose columns are the corresponding eigenvectors. The eigenvectors define the PC directions and the matrix V^T is the rotation matrix that decomposes the measured (original) signal into such components. In compact notation, this decomposition is given

in Equation 3.8.

$$\Phi = V^T X \tag{3.8}$$

And the measured PD signal can be reconstructed by Equation 3.10;

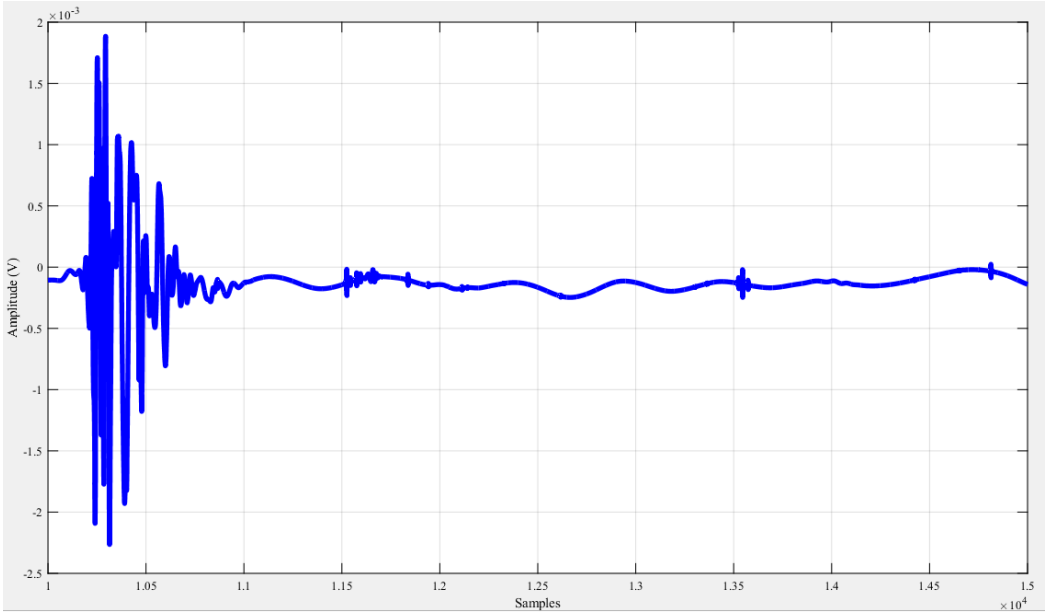
$$\hat{X} = V\Phi \tag{3.9}$$

It is important to note that the eigenvector matrix V is a unitary matrix and satisfies $VV^T = I$, where I is an identity matrix. Thus the reconstruction is exact and unique. Mathematically the reconstructed signal is exactly equal to the original measured signal when all PCs are included. Reconstructing with selected subset of PCs would yield a truncated reconstruction that is devoid of the energy captured in the discarded components. For example, incoherent noise often populates later PCs and discarding them during reconstruction would yield a cleaner signal that is minimally affected by the noise. Thus to perform de-noising of the measured PD signals, only selected subset of the PCs are used in the reconstruction stage. In matrix notation, this is given by Equation 3.10.

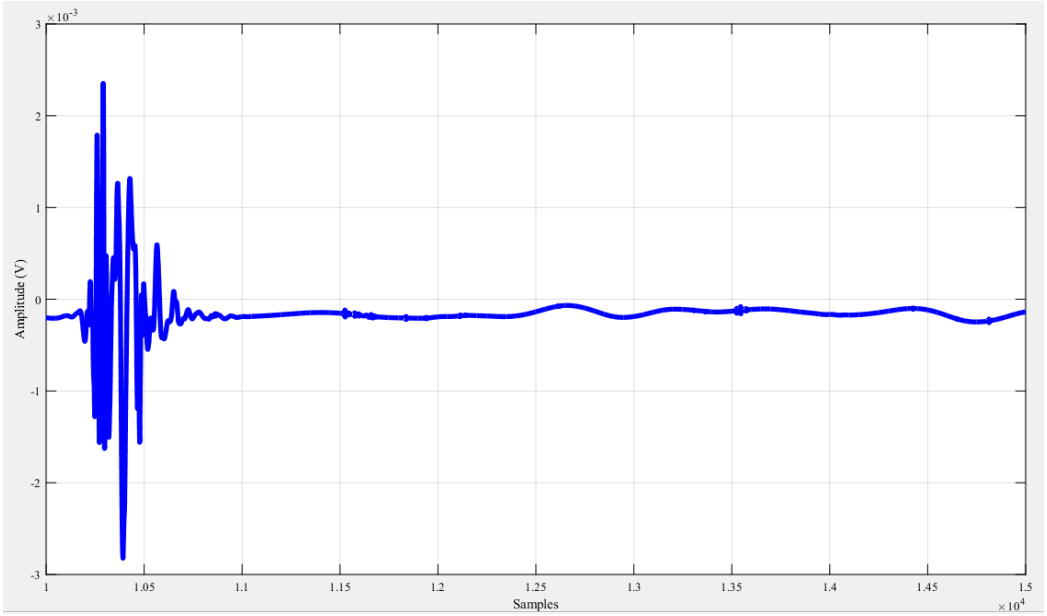
$$S = VBV^T X \tag{3.10}$$

where B is a diagonal matrix, with ones at the rows corresponding to the components used for reconstruction, and zeros everywhere else. In this work, PCA is used to offer an additional de-noising layer to the multivariate wavelet de-noising method before thresholding is carried out. PCA is performed on the wavelet detail coefficients before thresholding to eliminate noisy and redundant components of the measured signals. In like manner, it is performed on the wavelet approximation coefficients in order to keep only the most informative features of the PD signals. Kaiser's rule [103] is used to select the minimum number of retained principal components. In Kaiser's rule, components whose eigenvalues are greater than the average eigenvalue or have their 95% confidence limits encompassing the average eigenvalue are retained. This is because these components summarise more

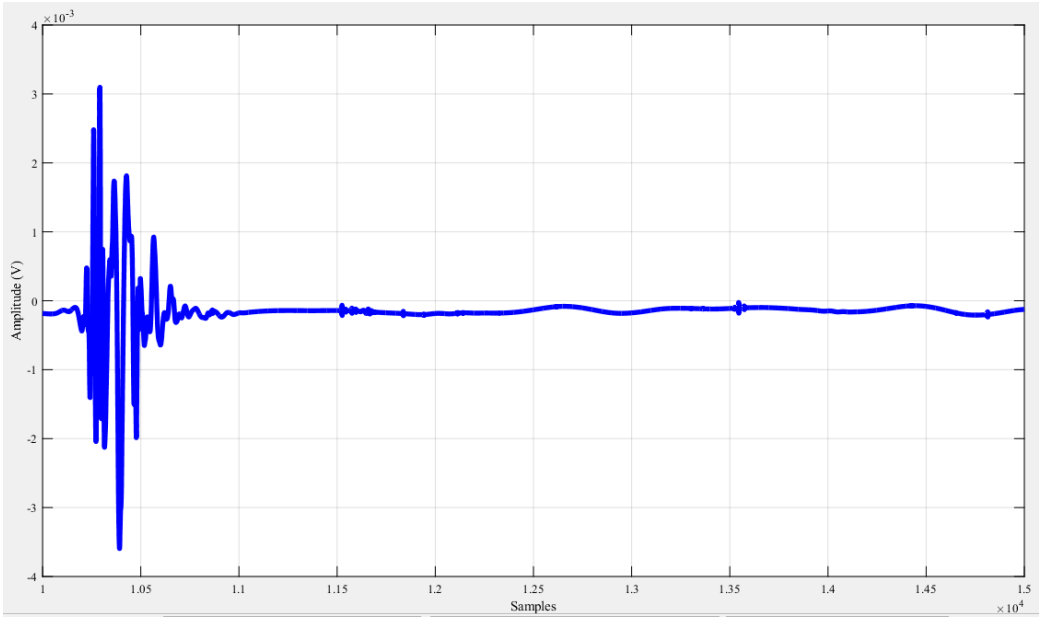
information than any of the original variable.



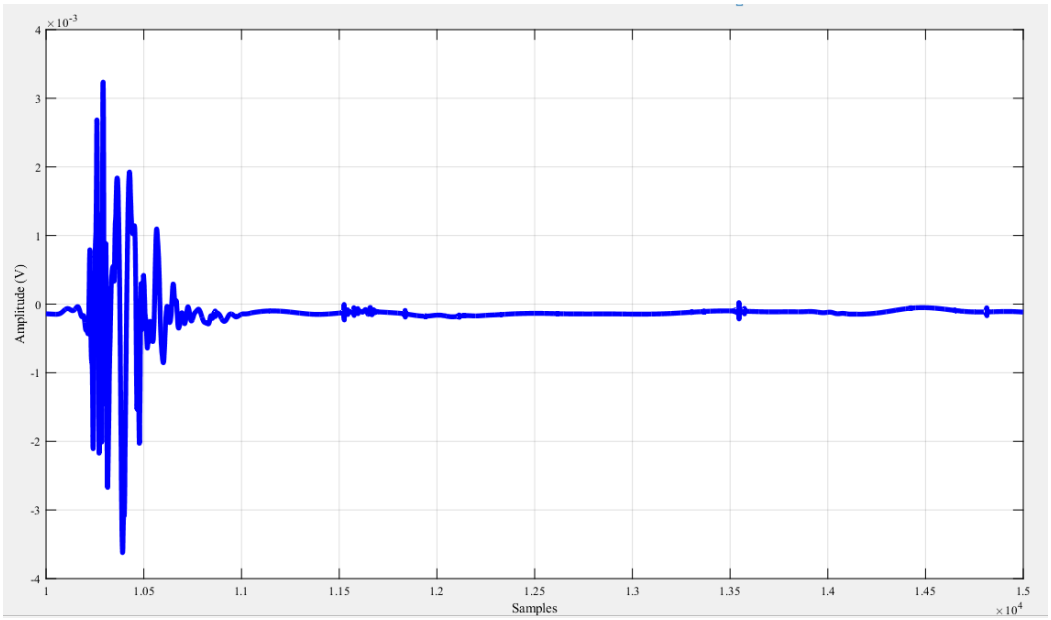
(a) Pulse 1



(b) Pulse 2



(c) Pulse 3



(d) Pulse 4

Figure 3.6: De-noised PD signals.

The PD de-noised signal samples are shown in Figure 3.6. The de-noised signals represent truncated versions of the measured signals. This is done for better visualisation of the PD pulses of interest. By comparing the measured PD signals in Figure 3.3 with the de-noised signals in Figure 3.6, it can be concluded that the wavelet-PCA technique employed in this thesis provides effective de-noising. Although there are losses in the de-

noising process which is evident in the reduced amplitude of the reconstructed de-noised signals, it presents us with representative PD signals that will be used in subsequent analysis for PD localisation.

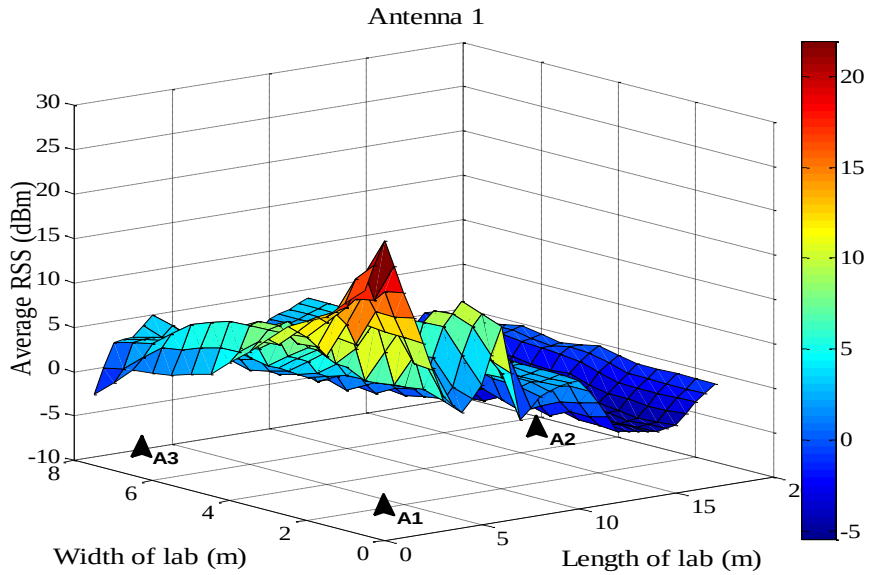
3.5 Spatial variability

Spatial variability occurs when a measured quantity at different spatial locations exhibits values that differ across the locations. In the PD localisation system proposed, it is important to assess the spatial variability of any measured quantity to be used as fingerprint to a localisation algorithm. Here, the PD dataset collected as described in section 3.3 is used to test for spatial variability of PD fingerprints. RSS is used to demonstrate PD fingerprint spatial variability. The RSS of the PD signals is calculated as the sum of squares of the samples. The contributions of the three RF sensors used during the measurement campaign is also investigated.

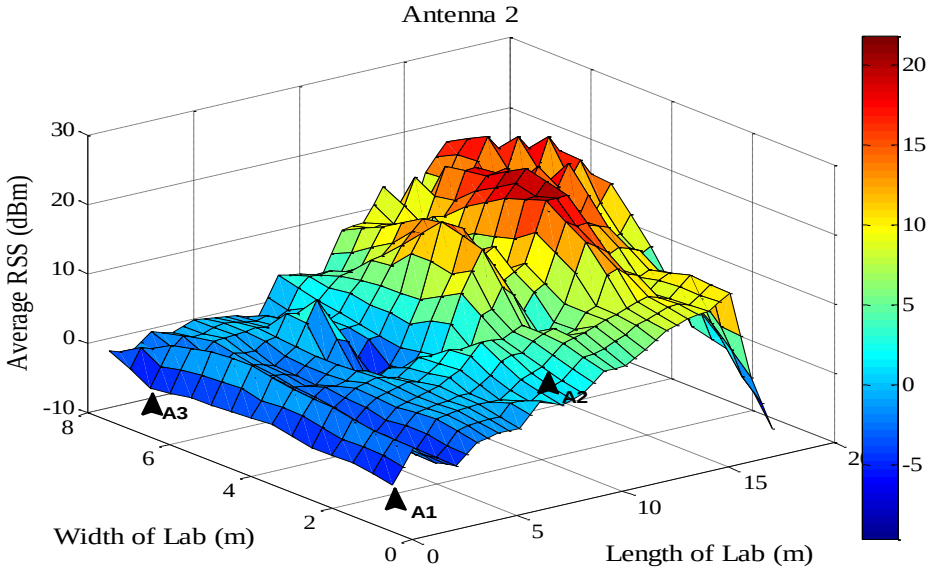
3.5.1 Spatial variation of PD location fingerprints

The performance of the proposed PD fingerprinting localisation system depends greatly on the variation of location fingerprints. Signal variation gives an indication of the degree to which PD fingerprint, such as RSS patterns differ from location to location. Theoretically, a source of emission at a given location is expected to produce similar RSS values over time. And two different locations with different distances from the same measuring device should have different RSS values. However, in a propagation environment, RSS attenuates with increasing distance between the source of emission and receiver. Obstruction, reflection and scattering from various objects within the environment leads to variations in RSS over small time scales. This causes the RSS to decay in a random fashion. The net effect of these variations is a lower SNR, which leads to high error. Therefore, the average of the 20 measurements taken at each location is computed and

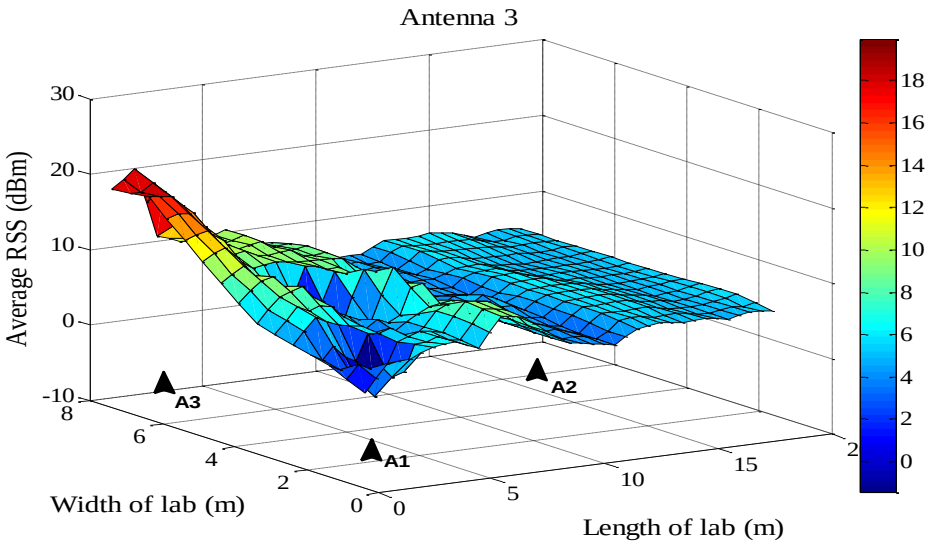
used in order to smooth out the effect of variations and increase the SNR. These values can be grouped together as patterns of RSS at a particular location. In order to investigate the spatial variability of RSS patterns at different locations, RSS measurements collected at each grid point in the experimental laboratory were considered. Figures 3.7 (a), (b) and (c) show the PD signal strength pattern in the radio environment for each of the three sensors. The figures reveal both the complexity of the propagation environment and the unique spatial signatures created by PD signal strength at different locations which facilitate the application of fingerprinting technique.



(a) Antenna 1



(b) Antenna 2



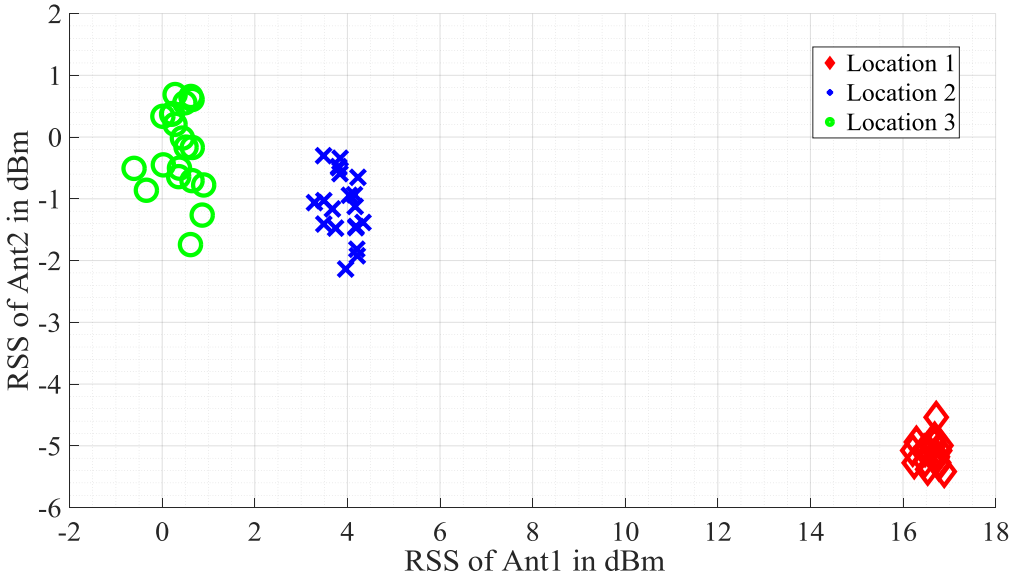
(c) Antenna 3

Figure 3.7: PD RSS spatial pattern.

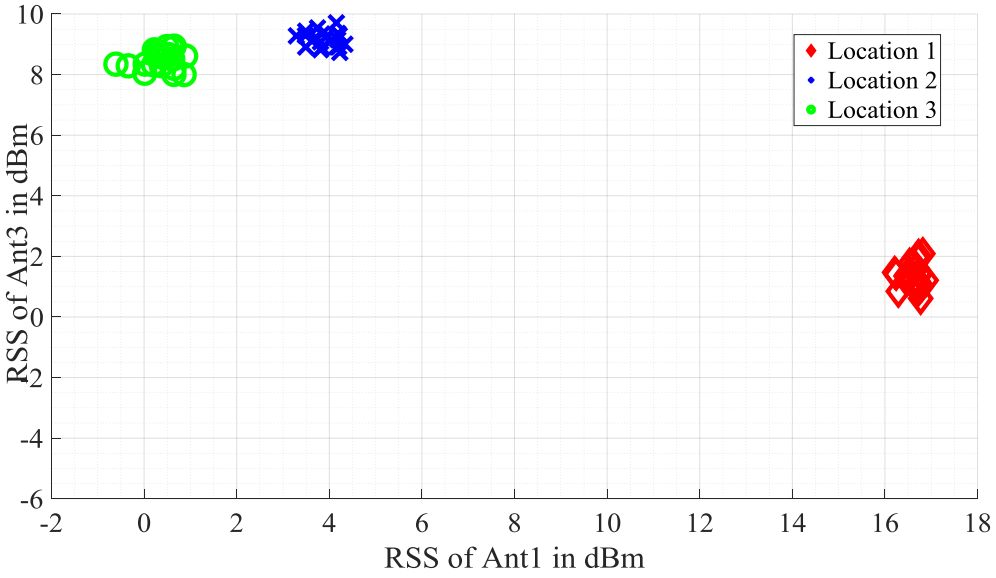
3.5.2 Effect of sensor positioning on PD fingerprint separation

To investigate the effect of each receiving sensor on RSS spatial variation at different locations, six locations in the experimental laboratory were considered. Three of these locations are from one half of the laboratory where antennas 1 and 3 are positioned and the other three locations from the other half where antenna 2 is positioned as shown on the floor grid in Figure 3.2. This forms two groups of three locations each. For clarity, we call the first group top half locations (where antennas 1 and 3 are positioned) and the second group bottom half locations (where antenna 2 is positioned). The distance between one location and another in each group is 6 meters. Figures 3.8 (a), (b) and (c) show the location separation of RSS measured from pairs of antennas at three locations at top half and Figures 3.9 (a), (b) and (c) show that for RSS measured from pairs of antennas at three locations at bottom half. The red markers represent the pattern from location 1, the blue markers represent that from location 2 and the green markers those measured

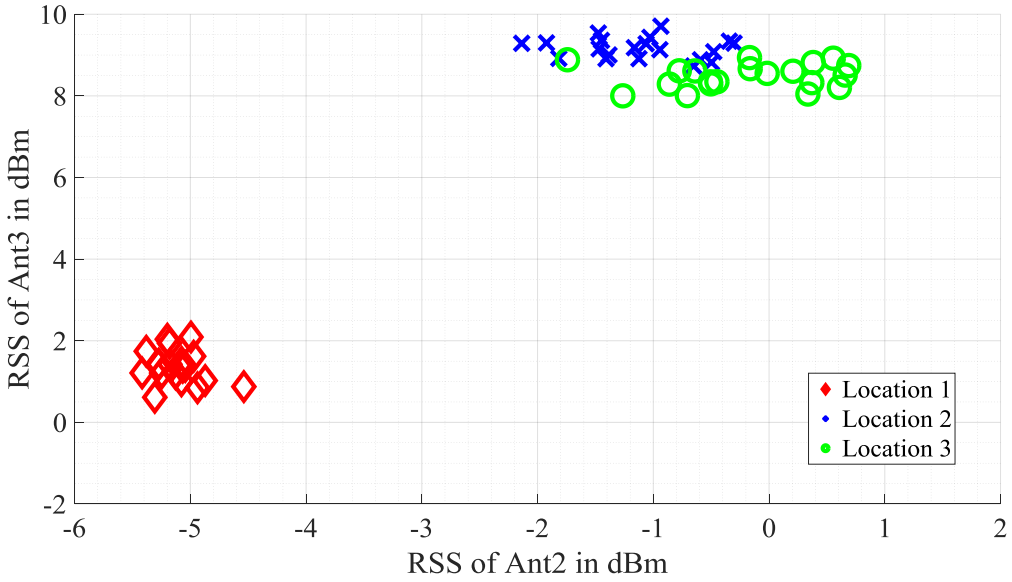
from location 3 in both groups. Patterns from each location can be grouped together as a cluster. This is because RSS has limited range and the signal often concentrates around some average value. The more distinct this separation is i.e, the further apart the marker clusters are, the easier it is for any location estimation method to make decision. For the locations in the top half of the laboratory, it can be seen that sensor 3 does not really help in separating RSS from location 2 (blue) and location 3 (green). RSS from sensor 2 on the other hand are beneficial when it comes to differentiate between location 1 (red) and the other two locations. RSS from sensor 1 are the most convenient in differentiating the three locations considered. In the case of locations in the bottom half of the lab., RSS from sensor 3 helps in separating location 3 (green) from location 1 (red) and 2 (blue). Sensor 2 is beneficial in differentiating location 2 (blue) from the other two locations. Again RSS from sensor 1 contribute more to the separation of all the three locations.



(a) Antenna 1 and 2

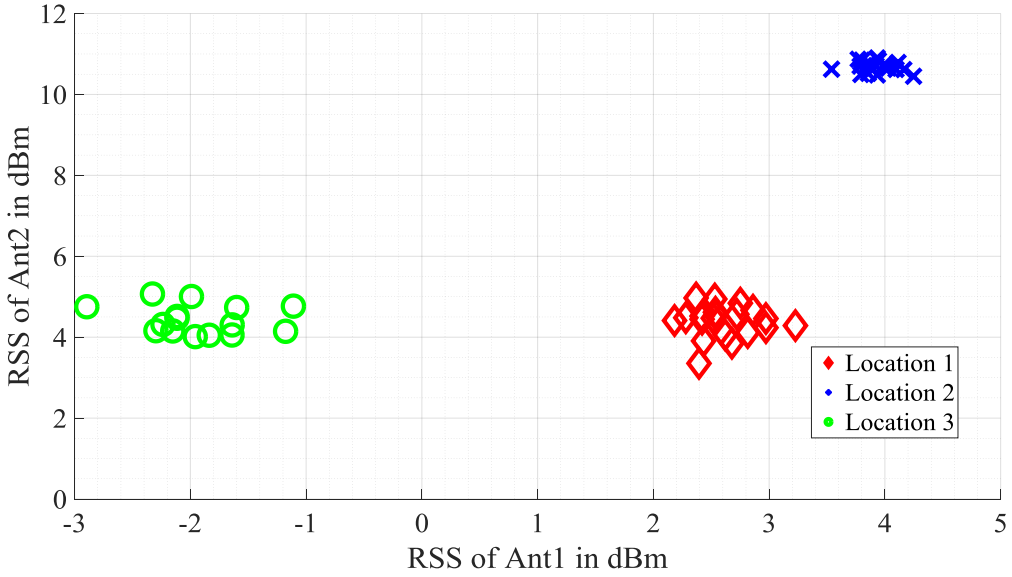


(b) Antenna 1 and 3

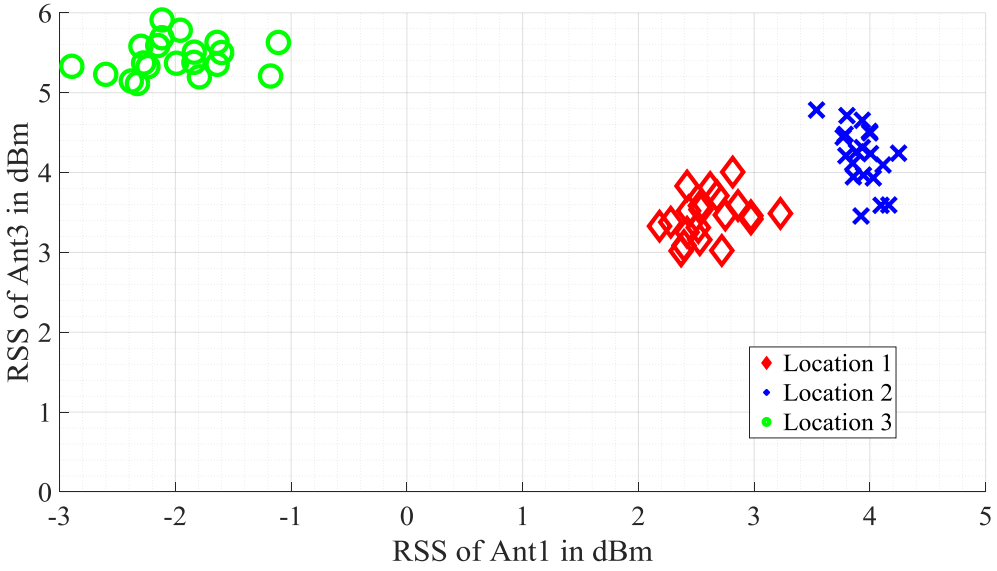


(c) Antenna 2 and 3

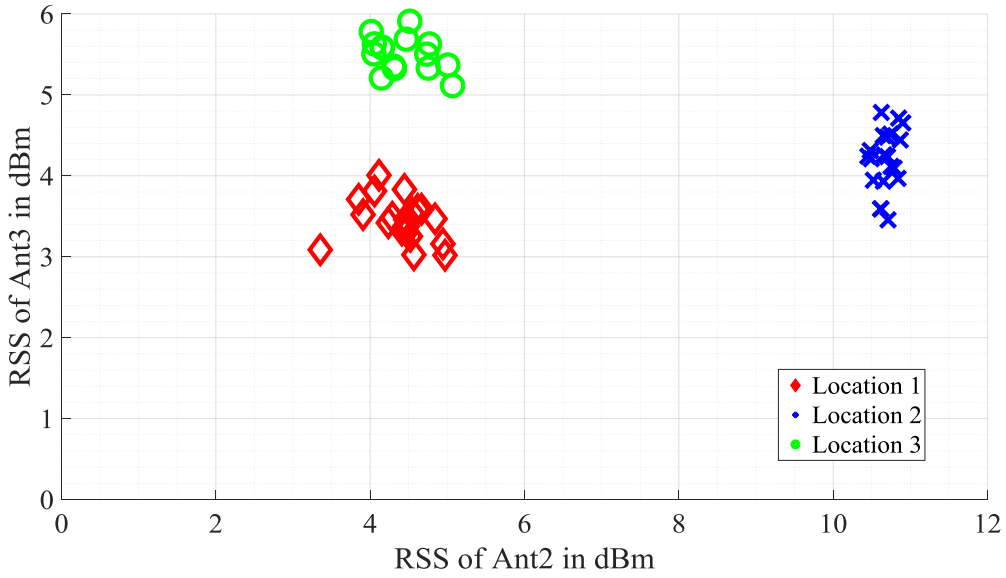
Figure 3.8: PD Figerprint Separation from three location at top end of the lab.



(a) Antenna 1 and 2



(b) Antenna 1 and 3

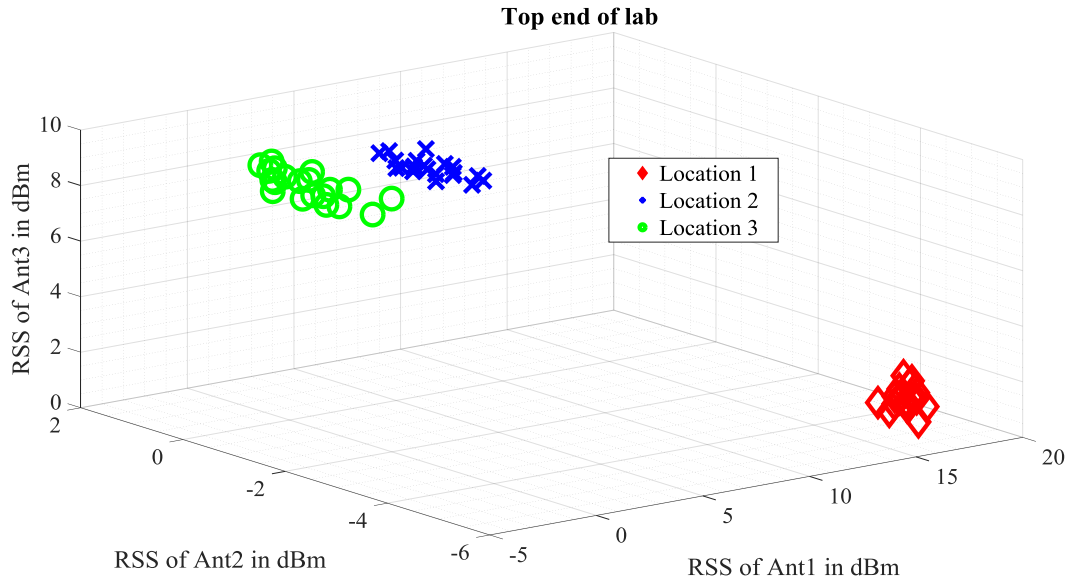


(c) Antenna 2 and 3

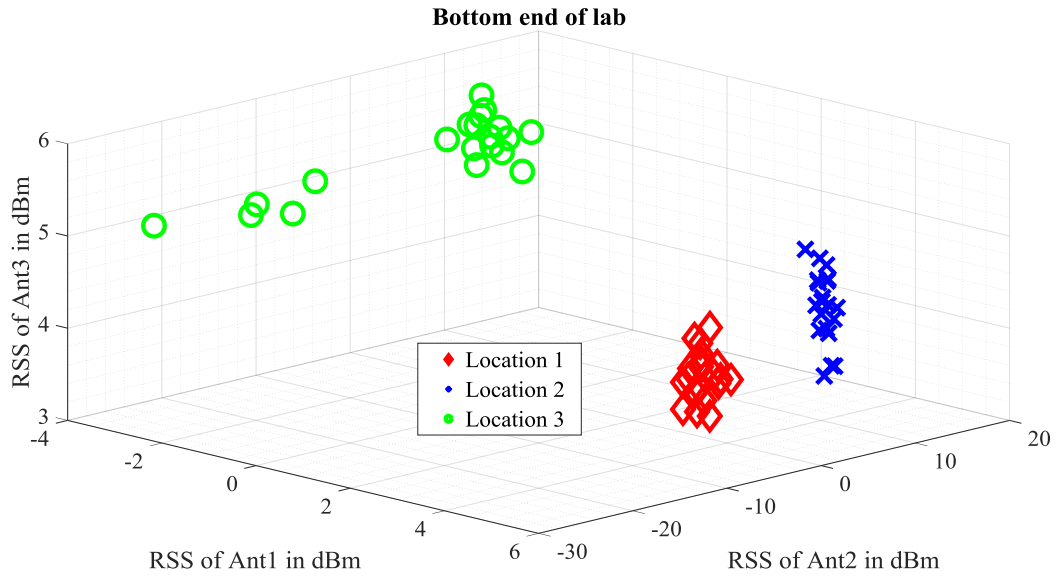
Figure 3.9: PD Figerprint Separation from three location at bottom end of the lab.

It can also be seen that measurements from each sensor reveals very different signal separation patterns at different locations. It is important to note that the sensors do not have the same level of significance when it comes to PD localisation at a particular location. This may be attributed to a sensor's propagation path. The thought in this thesis is that combining the effect of the three sensors on signal separation pattern at

each location will be beneficial to the overall performance of the PD localisation system. This is demonstrated in Figures 3.10 (a) and (b), where the RSS of PD from the three sensors are considered. It can be seen that there is an increase in the separation distance of the RSS clusters for each location when the contributions of the three sensors are taken into consideration. This is further proved in our investigation in the next chapter.



(a) Top end



(b) Bottom end

Figure 3.10: PD Figerprint Separation from three location at both ends of the lab.

3.6 Assumptions for RSS location fingerprints

For the purposes of modelling the proposed PD localisation system based on location fingerprints, each RSS element r_i in the location fingerprint $R = (r_1, r_2, ..., r_n)$ is assumed to be a random variable with the following assumptions.

- The RSS features in each location fingerprint are assumed to be mutually independent.
- The mean of the RSS is assumed to be constant and can be used as fingerprint since the RSS samples exhibit clustering.

3.7 Summary

This chapter investigates the feasibility of deploying a network of inexpensive radio sensors to monitor RF pulses emitted during PD events in order to determine the location of PD. The PD measurement campaign carried out to demonstrate the proposed substation-wide PD monitoring and localisation system is reported. PD data collection and preparation is also outlined in this chapter. The measured RF signals have been de-noised via wavelet technique to extract PD pulses. It is these de-noised signals that will form the bases for further analysis. The location-dependent property of the RF received signals have been demonstrated using RSS. The spatial variability of RSS as PD location fingerprints was established. In the next chapter, an investigation into the use of RSS in conjunction with machine learning algorithms for PD localisation would be carried out.

CHAPTER 4

Investigation into the use of Radio Fingerprinting for Localisation of Partial Discharge Sources

4.1 Introduction

In the previous chapter, the underlying assumption that PD received signals exhibit location-dependent characteristics has been established using Received Signal Strength (RSS). This implies that RSS can be used to infer PD location. Generally, RSS-based localisation methods require knowledge of the underlying radio propagation environment such that a suitable propagation model that defines the relationship between RSS and distance to receiver sensor can be built. However, in the case under study it is not appropriate to use a ready-made propagation model due to multipath problem that often characterise the real-life radio environment in which PD is experienced. Moreover, a detailed RF propagation modelling is non trivial. To this end, this chapter investigates an

alternative method based on radio fingerprinting technique for PD localisation. Firstly, the PD localisation problem is presented, followed by an empirical investigation into the use of RSS as PD location fingerprint for PD source localisation. The state of the art is outlined. Machine learning algorithms have been identified and used to model RSS-location relation and hence derive PD location. Furthermore, performance evaluation of the developed models is carried out with particular focus on accuracy.

4.2 Problem Statement

Many PD localisation methods proposed in the literature are based on estimating the distance between the measuring unit and the discharge source. Such methods are either time-based using the TDoA techniques [67][68], angle-based using the AoA techniques [82], or power-based by employing the RSS [104][105]. Distance estimates are then combined using either triangulation, or trialateration to estimate PD locations. TDoA and AoA techniques are complex and have power hungry signal processing cost. They also have stringent measurement (e.g LoS) requirement in order to performance well. The cost/complexity of these approaches prompt an investigation into methods based solely on the use of RSS measurements due to their low-power consumption and cost competitiveness. One common approach using RSS is based on variation of radio path-loss law inversion [106]. The intensity of radio waves is said to diminished with distance and this can be exploited to determine signal location. Theoretically, free space path-loss varies in strength as an inverse square law [106] and is very simple and straightforward to implement. However, in real life applications such as PD localisation considered in this thesis, the measured signal power can be significantly altered by the presence of multipath caused by reflection and diffraction, additive noise, shadowing and/or obstructions and No Line of Sight (NLOS). Due to the harsh environment, unreliable measurements of RSS result in deterioration of the localisation performance. Moreover, the radiated power of the PD RF signal is not known a priori, therefore accurate range values (distance between

the transmitter and receiver) become difficult to compute. Furthermore, the path-loss model for the considered environment is both unknown and generally different for each path travelled by PD signals from their sources to receiving sensors. Assumptions can be made regards radiated power and path-loss index. However, this can further degrade the localisation accuracy.

Localisation techniques based on fingerprinting are more reliable because they do not make any assumption about the propagation channel or require estimate of RF channel parameters. However, fingerprinting often involve a resource intensive survey stage where a new radio map (of signatures) is needed each time a change occur in the propagation environment in order to maintain the needed accuracy. In this thesis we suggest a low-cost Wireless Sensor Network (WSN) approach where the network itself builds the spatial RSS map of signatures autonomously and continually. With this approach PD monitoring system can be permanently deployed and thus monitor the substation in real-time at low cost without interruption. This chapter provides a method for PD localisation using fingerprinting technique based on PD RSS data in conjunction with K-nearest neighbor algorithm (KNN) [107][105] and Artificial Neural Network (ANN) [108][109].

4.3 Mathematical Formulation

The low-cost approach proposed in this thesis is based on the deployment of wireless sensors at arbitrary but known locations in the substation compound; these sensors can record radio emission from PD sources in real time. The locations of these sources can be inferred by sophisticated machine learning algorithms. Owing to the topography of assets in a substation, PD localisation in this setting can be regarded as a 2 dimensional (2-D) problem. Let the compound be modelled as a finite 2-D location space L. Each location $l \in L$, is taken as a reference PD source emitting RF signals where our training data set is gathered. The signal strengths of PD are collected from m receiving sensors.

We denote the m-dimensional received signal strength space as R. Each element in this space is a vector whose entries represent the signal strength values captured by different sensors. Sampling from the received signal strength space R at location l at time t, denoted by $r(l,t)$, can be viewed as a vector stochastic process. However, in this work, the PD localisation problem is defined as a stationary process, therefore the time index is dropped. For the remaining of the thesis, we refer to this process as $r(l)$ and whenever the location index l is dropped, we refer to the received signal strength vector as r. It is also assumed that the samples measured from different sensors are independent.

Therefore the PD localisation problem becomes a mapping of the received signal strength onto PD source location. This can be modelled as in Equation 4.1;

$$\mathbf{L} = f(r) + e(r) \tag{4.1}$$

where $r \in R$ is the signal strength vector from n locations, captured by m sensors. f is the function of RSS. And e accounts for the noise. The PD RSS fingerprint map can therefore be expressed as in Equation 4.2;

$$\mathbf{R} = \begin{bmatrix} r_i^1(1) & r_i^1(2) & . & . & . & r_i^1(m) \\ r_i^2(1) & r_i^2(2) & . & . & . & r_i^2(m) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ r_i^n(1) & r_i^n(2) & . & . & . & r_i^n(m) \end{bmatrix} \tag{4.2}$$

A column of the matrix is a RSS vector that contains measurement collected by a sensors from n distinct PD sources (locations).

The key challenge is to develop a model that can determine the planar location of PD sources based on PD fingerprint data at low cost. Both the receiving sensor nodes and PD sources are made stationary during measurement.

4.4 Location Fingerprinting Technique

Location Fingerprinting (LF) in the context of PD localisation is a technique that utilises radio signals to determine PD location by characterising the radio signal environment in which PD is expected. The location fingerprinting approach exploits the relationship between PD signal measurements and a specific spatial location. In other words, a PD physical location is mapped into a unique measure in radio signal space and the model created is used to infer PD location. In this chapter, the measured parameter used is the received signal strength. The strength of a signal is known to have a high correlation with distance travelled from it source to a receiver [110]. In the harsh environment of the substations, wireless signal propagation is severely affected by dense multipath, attenuation and NLOS which renders traditional parametric geolocation techniques (e.g trilaterarion) ineffective and complex for actual implementation. The location fingerprinting technique however, turns the multipath phenomenon to surprisingly good use: by combining the multipath pattern with other signal characteristics, it creates a signature unique to a given location.

Generally, the procedure for fingerprinting based localisation technique is divided into two phases: a database construction phase and localisation phase [111]. The wireless nodes are deployed in an approximately regular grid topography. The grid spacing, which is the distance between the closest locations is reported in meters. In the database construction phase, the RSS from PDs generated at different grid points are collected with associated known locations and stored in a database otherwise called a radio map. The radio map is a reflection of how the signal propagates. The vector of RSS values at a point (location)

on the grid is known as the location fingerprint of that point. This vector creates a unique RSS pattern for that location. In this thesis, the average RSS suggest a single feature per fingerprint. In the localisation phase, whenever PD occurs at any location within the area considered, the measured fingerprint is compared to fingerprints entries stored in the radio map and appropriate algorithm is used to determine the location of the PD source. The basic structure of the location fingerprinting system is as shown in Figure 4.1.

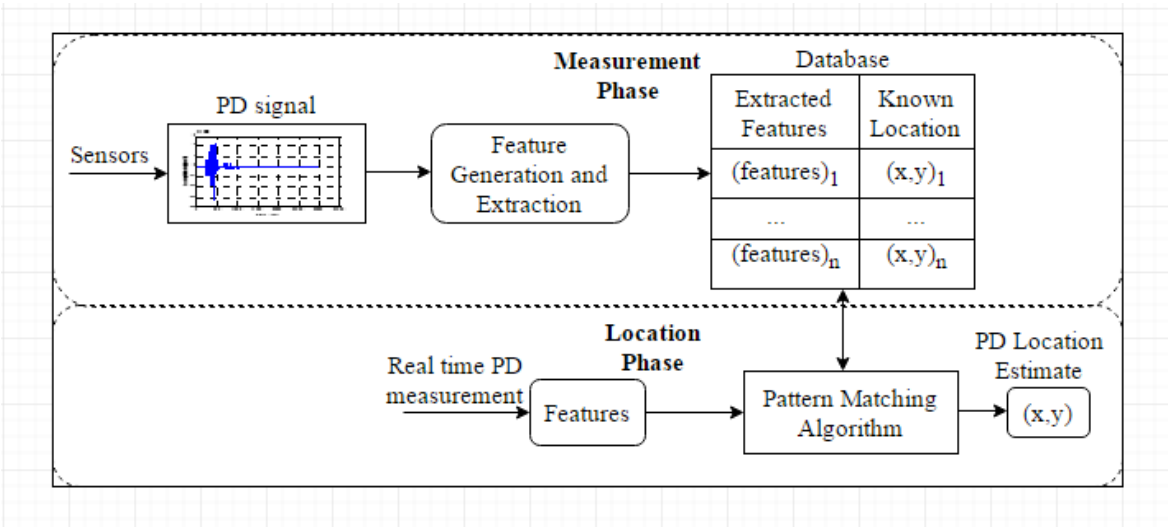


Figure 4.1: PD location fingerprinting system.

4.5 Location Fingerprint

A location fingerprint based on radio signal characteristics such as RSS is the basis for representing a unique location. An underlying assumption is that each location in the radio propagation environment exhibits a unique RF signature [112][105]. The environment is calibrated to capture the full dynamic of the signal characteristic at each location. Thus, the more cluttered the signal propagation is, the more unique the location fingerprint. A location fingerprint used in this chapter is a vector of RSS values measured by the receivers and corresponding location label. To create a location fingerprint, a number of measurements of RSS vectors are collected at each location. Each RSS element of a

vector can be considered as a random variable. The average RSS from each location is calculated and recorded as an element in the location fingerprint. The dimension of the vector is determine by the number of receivers used.

4.6 Machine Learning Techniques for PD Location Estimation

A PD location estimation technique is a procedure that exploits the relationship between location dependent signal parameters otherwise known as fingerprints and location information to infer PD source location. As earlier mentioned in section 4.2, RSS is used here as PD fingerprints. The key challenge is how to effectively and efficiently model RSS/location relation and hence derive PD location from RSS. The complexity of this problem motivates the use of sophisticated machine learning algorithms for PD location estimation. The RSS fingerprints from different locations and their corresponding location coordinates form a training data set that is used to create estimator models that relate PD fingerprints and location coordinates.

4.6.1 K-Nearest Neighbour Regression Model

The k nearest neighbour algorithm is one of the simplest and well known machine learning algorithms widely used for classification and regression tasks (function approximation which interpolates between known samples). It is an intuitive fit with the problem of inferring the propagation environment between arrays of radio sensor nodes. For PD location, k nearest neighbour algorithm is considered as a regression problem which consists in mapping PD signal input features (fingerprints) onto dual output corresponding to locations in 2-dimensional space. The idea of k nearest neighbour technique is based on the assumption of locality in feature space. KNN regression for PD localisation consists of

database construction and the location estimation phases. Database construction comprises the collection and storage of known examples. Each example consists of a data point having PD fingerprint (i.e RSS from a source) labeled with its physical location coordinate. In the location estimation phase, KNN finds k examples in the database whose RSS values are closest (most similar) to the new PD observation in feature space (of fingerprint). In the nearest-neighbour calculations each fingerprint is represented as a vector of average RSS with entries for each receiver sensor. To determine PD location, is based on averaging the physical location coordinates of the selected closest examples otherwise known as k-nearest neighbours. The estimated PD location of the new observation is given by Equation 4.3, which is the arithmetic average of the location coordinates of these k closest neighbours.

$$(\hat{x}, \hat{y}) = \frac{1}{k}\Sigma_{i=1}^{k} x_i, y_i \tag{4.3}$$

where $(\hat{x}, \hat{y})$ represent the coordinates of the estimated PD location and (x_i, y_i) is the location of the i^{th} closest neighbour.

The determination of the similarity of measurements is based on distance measure in signal (feature) space. The common choice for the comparison measure is Euclidean distance [107]. Suppose there are n location fingerprints in the database denoted by $R = r_1, r_2, ..., r_n$ and each fingerprint has a one-to-one mapping to a set of locations $L = l_1, l_2..., l_n$. A sample of new fingerprint vector measurement collected is denoted as s. Assuming that these measurements were collected by m sensor nodes, then the sample of new fingerprint vector is $s = (s_1, s_2, ..., s_m)$ and each fingerprint in the database can be denoted as $r_i = (r_{i,1}, r_{i,2}, ..., r_{i,m})$. Therefore, the similarity between the new observation s and the fingerprints in the database r_i from m sensor nodes according to Euclidean distance can be expressed as in Equation 4.4;

$$D_{euc}(s, r_i) = \sqrt{(\Sigma_{j=1}^{m}(s_j - r_{i,j})^2)} \tag{4.4}$$

which is the distance in the signal (feature) space between the new observation s and the fingerprint in the database r_i. Given the euclidean distance D_{euc}, that computes the similarity in signal space, the procedure with k nearest neighbour algorithm is to select the location of the fingerprint j that has the shortest signal distance. Another similarity measure considered in this work is the Mahalanobis distance metric [113]. The Mahalanobis distance $D_{mal}(s, r_i)$ between a location fingerprint r_i and a new sample vector s, given a covariance matrix of location fingerprint Σ can be defined as in Equation 4.5;

$$D_{mal}(s, r_i) = \sqrt{(s - r_i)^T \Sigma^{-1} (s - r_i)} \tag{4.5}$$

Mahalanobis distance measure takes into account the correlations of the random fingerprint components and is independent of the scale of measurements [114]. Thus it can be used for measuring the similarity of an unknown sample set to a known one. Both the Euclidean distance and the Mahalanobis distance are used in this thesis and direct comparisons between the two are given.

The value of the parameter k has an impact on the performance of the KNN localisation model. When only one neighbour (k=1) is considered in the k nearest neighbour algorithm, the algorithm becomes the so-called nearest neighbour algorithm (NN). Here, the closest neighbour becomes the location estimate. This is demonstrated later in this chapter where optimal value of the parameter k is obtained through cross validation.

In the KNN algorithm, all the k nearest neighbours in the fingerprint space contribute uniformly to the final PD location estimate regardless of their distances from the new observation in the coordinate space. This however limits the performance of the models especially in a large space. In the next subsection, an approach to overcome this bottle neck is discussed.

4.6.2 Weighted KNN

The KNN algorithm discussed in Section 4.6.1 can be improved by introducing a weight factor. The weights account for the degree of contribution of each of the k neighbours according to their similarity distance to the new observation in the fingerprint space such that the closest neighbours contribute more to the final location estimate than neighbours farther away. This balances the variations in the fingerprints. To this end, the distance measure used to search for the nearest neighbours is transformed and used as weights to the k nearest neighbours. The commonly used weight factor is the inverse of the neighbours similarity distance from the new observation. It is important to note that there are other weighting function that can be considered. The choice of this function will be discuss later in this chapter. The improved KNN algorithm has the ability to smooth out the impact of noisy training data by taking the weighted average of the nearest neighbours. This enhances localization accuracy. The improved KNN algorithm, herein called the Weighted K-Nearest Neighbour (WKNN), replaces Equation 4.3 with Equation 4.6;

$$(\hat{x}, \hat{y}) = \frac{\Sigma_{i=1}^{k} w_i * (x_i, y_i)}{\Sigma_{i=1}^{k} w_i} \tag{4.6}$$

where

$$w_i = \frac{1}{D(s, r_i)} \tag{4.7}$$

Equation 4.7 is the weight of each neighbour and D is the fingerprint similarity distance between the new observation and the samples in the database.

Despite the advantages of the KNN algorithm such as ease of implementation and the minimal training or tuning requirement, KNN still has some limitations. The KNN algorithm discussed is based on a smoothing approximation. In reality the locations in the signal or fingerprint space do not perfectly correspond to locations in the physical space. Also since the estimates are a function of the grid point's position, the accuracy

of the KNN algorithm will be dependent upon how much location fingerprints can be separated in signal space. As expected, the location error decreases when the density of fingerprint increases, but the decreasing rate slows down during the thickening of the radio map. Unfortunately, the PD signal measured by the sensor nodes represents a complex, non-linear surface which cannot be captured by the KNN algorithm. In addition, as the number of fingerprints in the database increases, the computational complexity of KNN also increases and may be prohibitive for deployment in a large substation. Conversely, an Artificial Neural Network (ANN) presents itself as a viable alternative for PD location fingerprinting due to its ability to accommodate highly non-linear relationships.

4.6.3 Artificial Neural Network Model

Artificial neural networks (ANNs) are the most commonly used computational learning techniques for solving function approximation problems [115][116]. They essentially imitate the learning process of biological neural networks through the use of interconnected nodes, often called neurons. Each node is a simple processing element that responds to the weighted inputs it receives from other nodes. In its basic form, a neuron computes its output using a weighted sum of its inputs and its activation function. The activation function introduces non-linearity and robustness in the network. Each connecting link between nodes is associated with a weight that can be tuned based on experience, making the neural network capable of learning. Given sufficient neurons and a set of input-output data, the ANN can be trained to approximate any continuous function arbitrarily well.

The attractive features of ANN include its robustness to noise in the inputs due to the orthogonality among the output values of the hidden nodes, its ability to learn, and subsequently recognise complex and non-linear relationships between input and output vectors without any prior knowledge of the relationship. In this thesis, ANN is employed

as a tool to infer PD source locations from measured PD features (RSS) based on prior knowledge obtained during the training of the neural network. Two variants of ANN; the Multilayer Perceptron (MLP) [117], and Generalized Regression Neural Network (GRNN) [118] models have been developed to determine PD locations given the received signal strength of the PD traces.

4.6.3.1 Multilayer Perceptron Partial discharge Localisation Model

The Multilayer Perceptron (MLP) is the most popular, and widely used feedforward neural network [107] model for any arbitrary input-output mapping problem. MLP is applicable to function approximation, pattern classification, signal analysis amongst others [105][107][109][119]. In the context of PD localisation, it consists of a non-linear mapping from a set of PD input features (RSS) onto the dual output variables representing the location coordinates of the PD source. Here, the neural network learns from knowledge represented by an RSS-location data set consisting of input-output examples. The MLP network architecture consists of three layers as shown in Figure 4.2. The first layer is the input layer, where the nodes are the elements of the PD feature (RSS) vector. The second layer is the hidden layer where the computation takes place. It is so called since it does not contain input or output units. The third layer is the output layer which represent the output of the localisation process. A linear activation function is used in the input and output layers. A sigmoidal activation function shown in Equation 4.8 which normalises all values between the range of 0 and 1 is used in the hidden layer to provide robustness against outliers or extreme values. Other activation functions could be used. The performance of the MLP model is reported later in this chapter.

$$f(x) = \frac{1}{(1 + e^{-x})} \tag{4.8}$$

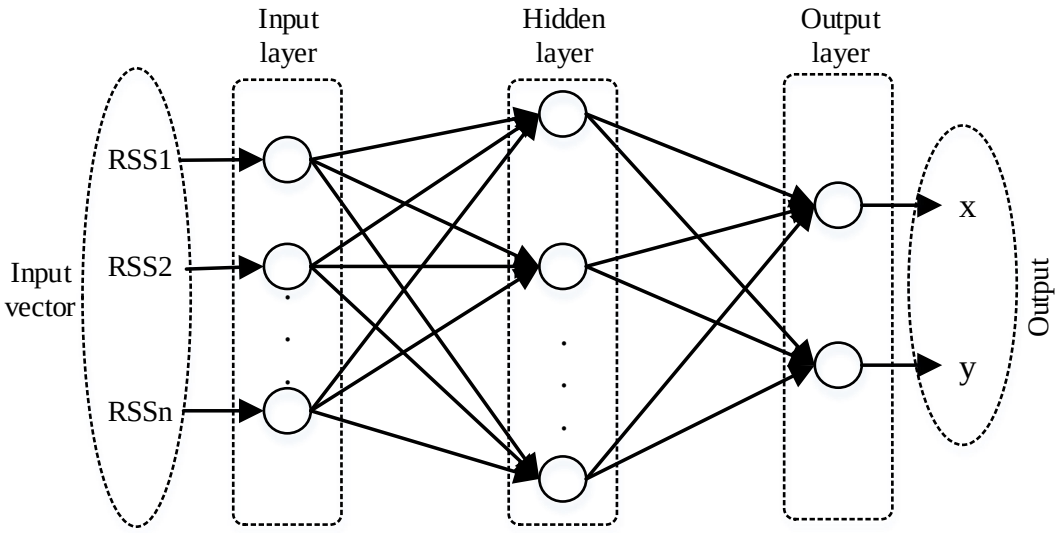


Figure 4.2: MLP network architecture

In general, a backpropagation algorithm [120][121][122] trains a feedforward network. In the training process, the network is trained to build a model of a set of PD features (RSS) as a function of PD locations so that the network can generalize and predict outputs from inputs that it has not seen before. The backpropagation training algorithm works as follows: first, an input pattern (PD fingerprints) is presented to the network via the input layer and these input signals are passed to the next layer in a feed-forward fashion. The network output is then compared to the desired output and the error computed. This error is then back-propagated through the network and the neuron weights are adjusted to minimize the error. This process of feeding forward signals and back propagating the error is repeated iteratively until the error of the network is minimized. It is through the successive modification of the adaptive weights that the neural network is able to learn the underlying non-linear relation between the PD features and PD location. A fully connected MLP PD localisation model is as shown in Figure 4.3.

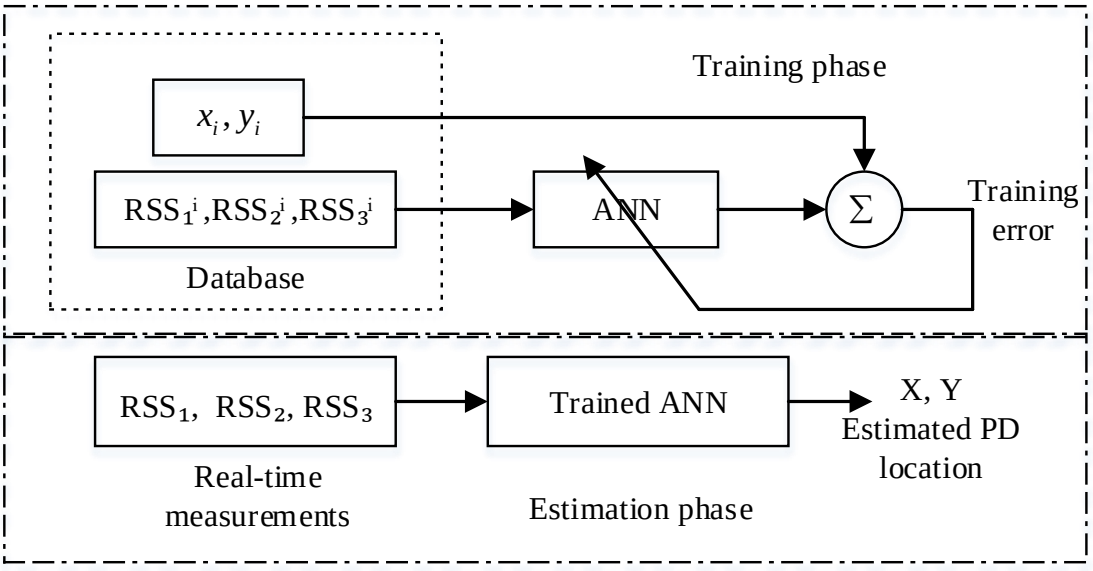


Figure 4.3: MLP model for PD localisation

The choice of an optimum network structure is important in modelling MLP neural network. This includes, the determination of input and output nodes, number of hidden layers and number of hidden neurons. Determination of the required number of input and output nodes is a relatively easy task because it is a problem-oriented task. For PD localisation problem, the number of input nodes is directly determined by the number of PD features; in the case under study, we have a network of three input nodes since we have three PD features (RSS from 3 antennas). The number of output nodes is determined by the dimension of the location coordinate. For a 2-dimensional location problem, the number of output nodes is two. Also for function approximation problems with limited data, only one hidden layer is sufficient to approximate the function. In addition, the number of hidden layers increase feature extraction capability of the network at the cost of significant extended training and operational time of the network. Therefore, one hidden layer is adopted in this thesis. The number of hidden neurons is set to approximately half the product of the number of input and output plus one. Therefore a 3-4-2 network model with three inputs nodes, four hidden neurons and two outputs nodes is adopted in

this work.

In order to improve the generalisation of the MLP model (that is models ability to do well on unseen data rather than just training set) and avoid overfitting, Bayesian Regularisation (BR) [123][124] is used to train the network. In a BR network, the objective function (such as mean sum of squares of the network error) to be minimised is modified by adding a penalty term (sum of squares of the network weights). This penalises (set near zero) large weights that may be introduced in order to obtain smoother mapping. Hence BR network trains on a number of effective network parameters or weights. The network uses the knowledge acquired during training to provide interpolated values for the location coordinates of the unseen test data.

4.6.3.2 Generalised Regression Neural Network Partial discharge Localisation Model

A generalized regression neural network (GRNN) is proposed as an alternative neural network method for PD localisation. This approach presents multiple advantages over MLP networks. The main advantage of GRNN is in its learning speed. In the GRNN network optimisation process only one free parameter (smoothness parameter) has to be adjusted in "one-pass" (forward) through the training data. The single-pass learning means no back-propagation [116][125] is required. GRNN learns in a fraction of time it takes other training algorithms such as BP. Unlike MLP, it does not require an iterative procedure for training. It is consistent and its estimate always converges to a global minimum. Hence it produces high accuracy in the estimation. The algorithm provides smooth approximation of target function even with sparse data in a multidimensional space. Whereas BP needs a large size of training data. In this thesis, GRNN carries out a non-linear mapping from PD features onto PD source location, drawing the function directly from the training data. The GRNN network has a fixed structure with an input

layer, a hidden (pattern) layer, a summation layer and an output layer that are fully connected as shown in Figure 4.4.

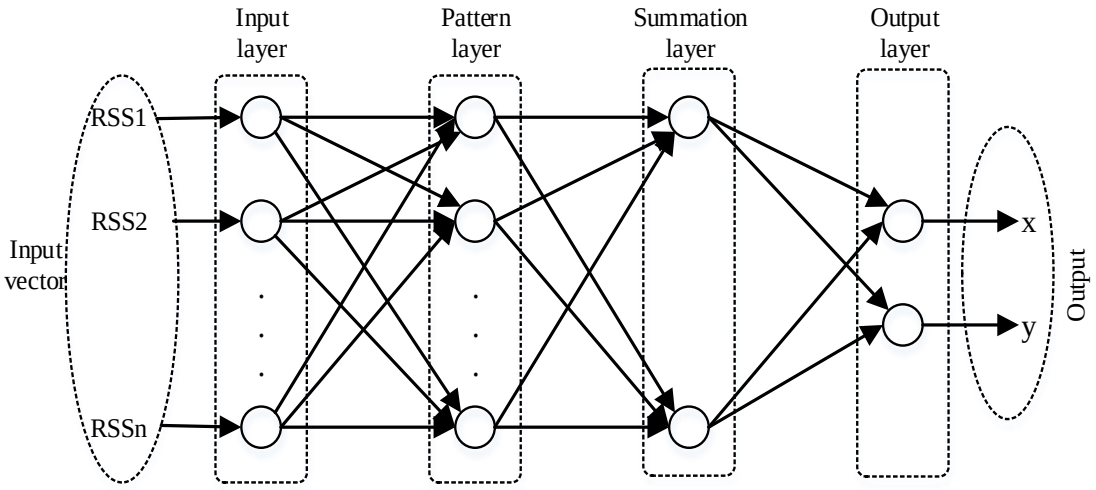


Figure 4.4: GRNN model for PD localisation

During training, the relationship between the fingerprints and the PD location coordinates is memorised and stored. The training data contain input-output mapping. Once the inputs are presented to the GRNN the weight parameters are determine instantly. The weights between the input and each pattern neuron are set to unity. In each pattern unit, a radial (Gaussian) basis function kernel is used as an activation function to calculate the output value. The Gaussian kernel is given by Equation 4.9;

$$\theta(r, r_i) = exp(-\frac{d\theta}{2\sigma^2}), \qquad d\theta = (r - r_i)^T(r - r_i) \tag{4.9}$$

where $d\theta$ is the squared euclidean distance between the training samples r_i and the new observation r. σ is the smoothing (spread) factor of the kernel. The summation layer is subdivided into S-summation neuron (summation units) and D-summation neuron (a single division unit). The S-summation neuron determines the sum of the weighted out-

puts of the pattern layer. The D-summation neuron determines the unweighted outputs from the pattern layer such that the estimated output to an unknown input vector is computed using Equation 4.10.

$$(\hat{x}, \hat{y}) = \frac{\Sigma_{i=1}^{n}(w_i) * \theta(r, r_i)}{\Sigma_{i=1}^{n}\theta(r, r_i)} \tag{4.10}$$

where $(\hat{x}, \hat{y})$ represent predicted values of x and y coordinates of PD location associated with the observed fingerprint r, w_i is the activation weight for the pattern layer neuron at i and n is the number of training samples. θ is the activation function defined in (4.9). The weights of the pattern neuron to D-summation neuron connections are set to unity, while weights of the pattern neurons to S-summation neuron connections are set to the output value of the training samples.

The GRNN network has only one unknown parameter, spread factor, σ. The appropriate value of the spread factor can be determined using exhaustive search method. This process of optimisation will be carried out in the next section.

4.7 Performance Evaluation

This section provides an empirical evaluation of the performance of the RSS-based PD localisation models developed in section 4.6. The PD data collected during the experiment described in chapter 3 is used for this evaluation. The first data set collected from the 144 grid points is treated as training data points and their location coordinates are assumed to be known. The second data set collected from 32 distinct locations within the same environment is treated as test data with only the RSS values used for location estimation. Their location information are assumed to be accurate and subsequently used as reference data for validation. All the estimated locations using the developed models are compared with the reference locations to calculate distance error.

4.7.1 Performance Evaluation Metrics

Performance metrics provide the basis for evaluating and comparing partial discharge localisation models. Therefore, the performance of the models described in Section 5.5 is evaluated and compared based on the following metrics: accuracy, percentile, precision, Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Cumulative Distribution Function (CDF).

4.7.1.1 Accuracy

Accuracy is the most used metric to evaluate the performance of a localisation model. It represents the difference between the actual location and the estimated location. This metric is generally measured as the Euclidean distance between the estimated location and the actual location, as defined by 4.11

$$Acc = \sqrt{(x_{est} - x_{actual})^2 + (y_{est} - y_{actual})^2} \tag{4.11}$$

where (x_{est}, y_{est}) are the x and y coordinates estimated by the localisation model, and (x_{actual}, y_{actual}) are the actual x and y coordinates. This metric is often represented by the mean distance error.

4.7.1.2 Mean Squared Error

Mean Squared Error takes the average of the square of the difference between the actual values and the predicted values.

$$MSE = \sqrt{(x_{est} - x_{median})^2 + (y_{est} - y_{median})^2} \tag{4.12}$$

4.7.1.3 Root Mean Square Error

RMSE of a model measures the square root of the average squared difference between the estimated location and the actual location. RMSE metric allows us to compute the localisation error for both X and Y coordinates. The RMSE value per each coordinate can be computed by 4.13 and 4.14

$$RMSE_x = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(x_{est} - x_{actual})^2} \tag{4.13}$$

$$RMSE_y = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_{est} - y_{actual})^2} \tag{4.14}$$

Here n is the number of estimates. The $RMSE_x$ and $RMSE_y$ values can be combined to compute the overall RMSE, which is the net error of the localization algorithm. The ovreall RMSE can be computed as in 4.15:

$$RMSE_{overall} = \sqrt{(RMSE_x^2 + RMSE_y^2)} \tag{4.15}$$

4.7.1.4 Precision

Precision is defined as a measure of the consistency in the performance of a location algorithm at a certain level of accuracy. In other words it measures the reproducibility of successive location estimates. It is used to assess the robustness of a localisation model as it reveals the performance variation of location estimates over the entire considered environment. This metric can be measured as the standard deviation in the location error. However, in this thesis precision is measured as the distribution of distance error between the estimated location and the actual location. The distribution function described in section 4.7.1.6 is used for measuring the precision of each localisation model.

4.7.1.5 Percentile

A percentile is a measure used to indicate the value below which a given percentage of observations in a group of observations falls. To determine percentiles in this thesis, the rank (R) of the percentile is first computed. This is done using the following formula in Equation

$$R = (P_{ct}/100) * (N + 1). \tag{4.16}$$

where P_{ct} is the desired percentile and N is the number of observations. If the rank is an integer, the percentile is the observation with rank R. However, when R is not an integer, the percentile is computed by interpolation as follows;

1. Separate the integer portion of R defined as IR and the fractional portion defined as FR.
2. Find the observations with rank IR and with rank $IR + 1$.
3. Compute the difference between the observations found in 2.
4. Finally, compute the percentile by multiplying the difference in 3 by FR and add the result to the lower observation.

4.7.1.6 Cumulative Distribution Function

The cumulative distribution function is used to describe the distribution of random variables. In this thesis, CDF indicates the time fraction or probability that the location error is less than or equal to a certain value. Mathematically, CDF of a random variable (localisation error) E is expressed as in Equation 4.17;

$$F_E(e) = P[E \leq e] = \int_{\infty}^{e} F_E(u)du, \quad for \quad all \quad e \in R. \tag{4.17}$$

This implies that, for a given localisation error e, $F_E(e)$ is the probability of having localisation error less than or equal to e. $F_E(e)$ goes to 0 as e gets smaller ($-\infty$). Conversely, $F_E(e)$ goes to 1 as e gets larger (∞). In this thesis, CDF is described by the percentile format. The location model whose CDF graph reaches high probability values faster is preferred because its distance error is concentrated in small values. In the next section, the critical parameters that affect the performance of the machine learning algorithms used in the PD localisation problem under investigation are optimised via exhaustive search method.

4.7.2 Optimal k-parameter in KNN

One important evaluation of the KNN location algorithm is to study how its performance varies with the number of neighbours (k) chosen for interpolation in the localisation phase. We start with $k = 1$, which is otherwise known as the nearest neighbour (NN) method. Figure 4.5 represents the variations in the localisation Mean Square Error (mse) with 10 different values of k. It can be observed that as k increases, the mse decrease to an absolute minima and then increase again. When the value of k increases initially, the underlying variations of signal strength at a single neighbouring point is averaged out with increasing k, resulting in decrease of error distance. The errors in signal strength vectors at different neighbours are assumed to follow different patterns. Hence averaging over more neighbours may yield more accurate results. However, for larger values of k, the errors again increase as many grid points which are closer in signal space but much farther in physical space are at this point included in the computation of the final result thereby leading to greater deviations from the actual location. The optimal value of k (the value of k with the lowest mse) is given as 5.

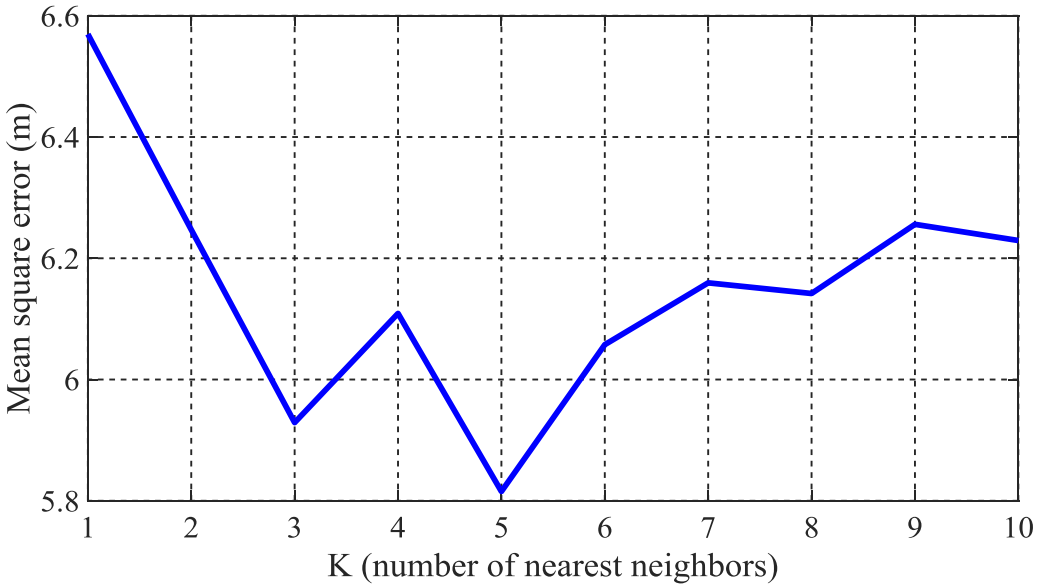


Figure 4.5: Optimal value of k for KNN model.

4.7.3 Optimal spread factor in GRNN

The hyper-parameter in the GRNN model; spread factor affects the performance of the model and hence needs to be carefully chosen in order to optimise the GRNN model. In this study, an exhaustive search method is used on a range of spread factors to determine the optimum spread factors for the model. The optimum spread factor for the GRNN model was determined according to the mean squared error (mse) criterion using the PD training data. Figure 4.6 shows the variation of mse vs spread factor. It appears from the figure that the optimal value of the spread factor is approximately 0.31 with the lowest mse value. Therefore, for optimum performance, 0.31 is used as the value of spread factor in GRNN model.

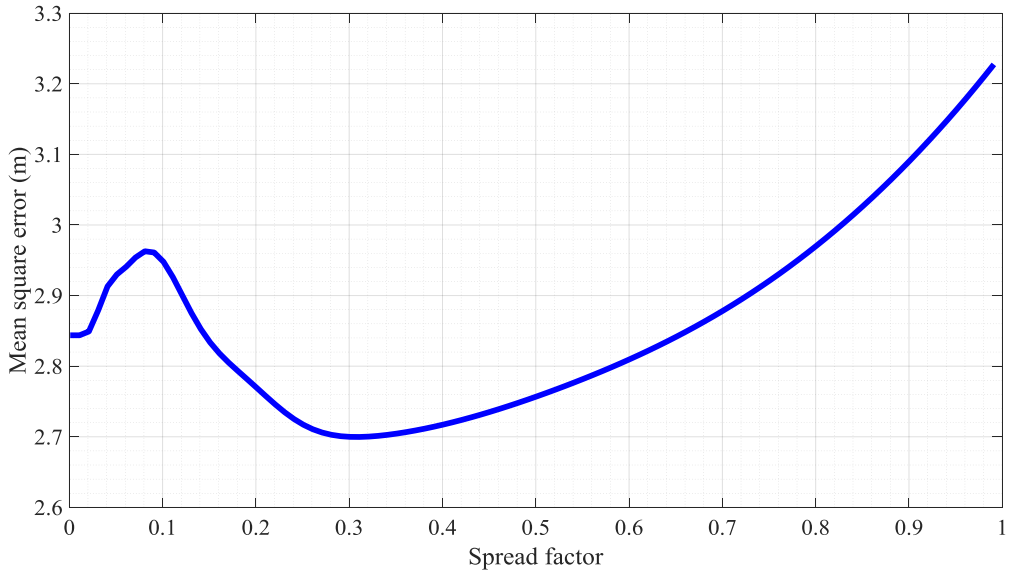


Figure 4.6: Optimal value of spread factor for GRNN model.

4.7.4 Effect of weight factor on performance of WKNN

In the WKNN algorithm, the weight factor is important in contributing to the localisation accuracy. Nearest neighbours are expected to have higher weights, in order to contribute more to the location coordinates compared to farther neighbours which should have lower weights. We investigate the impact of different weight functions on the location accuracy of WKNN model. The weight functions considered in this thesis are listed in Table 4.1 and assigned as types $A - E$.

Table 4.1: Weighting function.

Type	Expression of w_k
A	$\frac{1}{D}$
B	$\frac{1}{D^2}$
C	$\frac{1}{D^3}$
D	$\frac{1}{exp(D)}$
E	$\frac{1}{D*exp(D)}$.

where $D(s, r_i) = (\Sigma_{m=1}^{q}(s_m - r_{i,m})^2)^{\frac{1}{2}}$. Table 4.2 represent the variation in mean, 50^{th},

75^{th} and 90^{th} of the error distances with 5 different weight function used in evaluating the WKNN model. From the result in Table 4.2, the type A weight function which is the simplest function among the ones considered outperforms the other ones and will be used in the rest of the thesis for analysis.

Table 4.2: WKNN performance using different weight function.

Weight function	mean(m)	50th percentile(m)	75th percentile(m)	90th percentile(m)
$\frac{1}{D}$	1.74	1.15	2.56	3.91
$\frac{1}{D^2}$	1.76	1.19	2.86	3.89
$\frac{1}{D^3}$	1.83	1.26	2.91	4.24
$\frac{1}{exp(D)}$	1.77	1.52	2.66	3.90
$\frac{1}{D*exp(D)}$	1.80	1.23	2.82	3.97.

4.7.5 Impact of distance metric on performance of KNN models

It is important to consider the issue of correlation between signal strength; a common phenomenal in a real environment which can affect the PD localisation accuracy. For KNN, it is believed that using the Mahalanobis distance metric to compute the similarity between locations in RSS signal space can correct the high correlation between signal strength from different sources taking into account the covariances between the RSS. Table 4.3 compares the statistics of the error distance for KNN methods using Mahalanobis and Euclidean distance functions. It can be observed that KNN/WKNN-Euclidean outperforms the models based on Mahalanobis. More specifically, the mean of the distance error for KNN-Euclidean is $1.83m$, whereas for that based on Mahalanobis is $2.07m$. This result indicates that the Euclidean distance is preferable here and will be used in the rest of the thesis for KNN PD location determination. This may be as a result of low correlation in our data.

Table 4.3: Performance of different distance metric.

	KNN		WKNN	
	Euclidean	Mahalanobis	Euclidean	Mahalanobis
mean(m)	1.83	2.07	1.74	1.93
50th percentile(m)	1.36	1.86	1.15	1.54
75th percentile(m)	2.39	2.63	2.56	2.76
90th percentile(m)	4.03	3.49	3.92	3.86

4.7.6 Effect of grid spacing

Two scenarios were created to investigate the effect of grid spacing on the performance of PD localisation models considered in this work. It is worth noting that for the PD location models considered, only the training grid is of concern. The grid spacing for the test set do not affect the performance of the models. In the first scenario, the grid describe in chapter 3 for PD measurement, a 1×1 grid spacing with 144 training locations is used for training our models and the test data (from 32 test locations) is used to benchmark the models. The second scenario is created by taking the test locations on a 2.5×2.5 grid spacing as our training locations. In this scenario, the initial training data set from 144 training locations is used as test set to evaluate the performance of the models. For the purpose of clarity, the training grid in the first scenario is tagged Grid B (denser) and the training grid in the second scenario is tagged Grid A (sparse). The result shows that, as the grid spacing increases, the accuracy decreases for each model. Figures 4.7 and 4.8 show the result for KNN model and GRNN model. As expected, the denser the training grid, the better the results for the models considered. However, the improvement between the sparse and denser grid for the KNN model is larger than the improvement for the GRNN model. The mean distance error increase by 33.3% for KNN and 21.8% for GRNN. The KNN models benefit the most from refinement of the grid spacing.

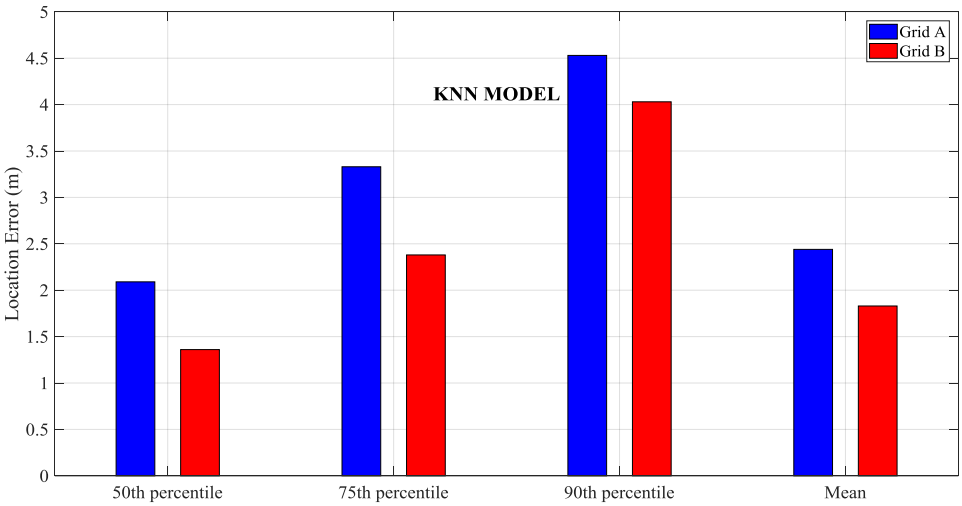


Figure 4.7: KNN performance for different grid spacing.

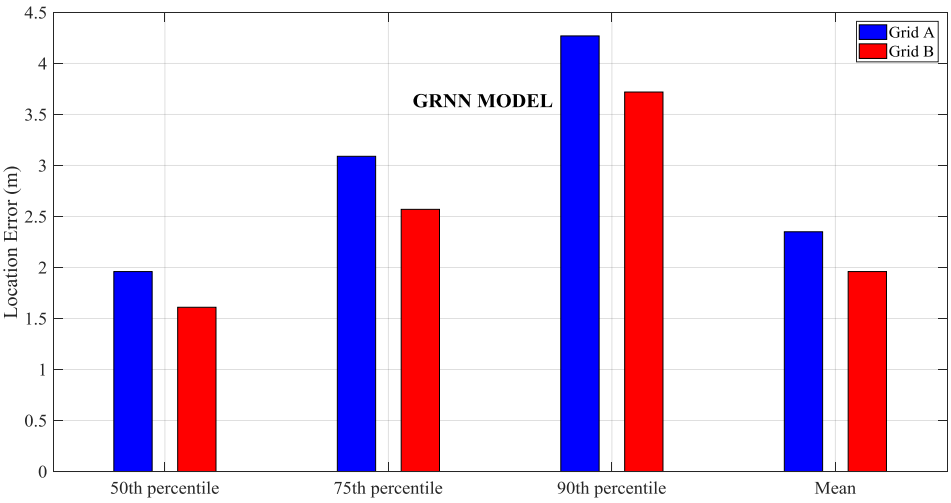


Figure 4.8: GRNN performance for different grid spacing.

4.7.7 Impact of localisation algorithm

The result of PD localisation using the four models (KNN, WKNN, MLP and GRNN) is given in Figure 4.9. The results show that the performance of the models are comparable with KNN models performing better in terms of RMSE. MLP and GRNN however, have higher location precision. The localisation error is less than $3.5m$ with probability of 90% for MLP and less than $3.75m$ for GRNN. This shows the robustness of the ANN models. A more detail analysis of the performance of the models with respect to the number of sensors is presented in the next subsection.

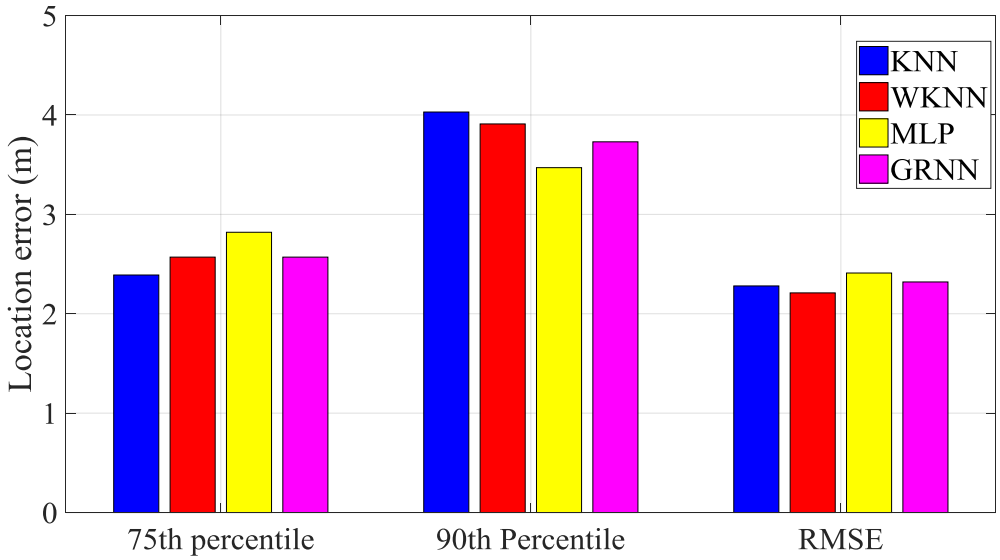
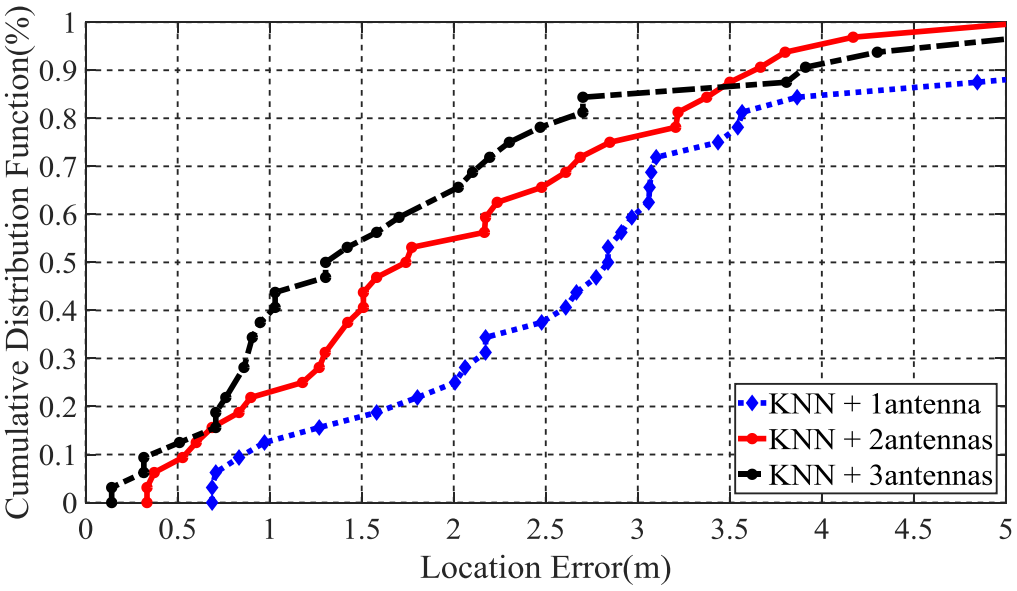


Figure 4.9: Performance of PD localisation models (KNN, WKNN, MLP and GRNN).

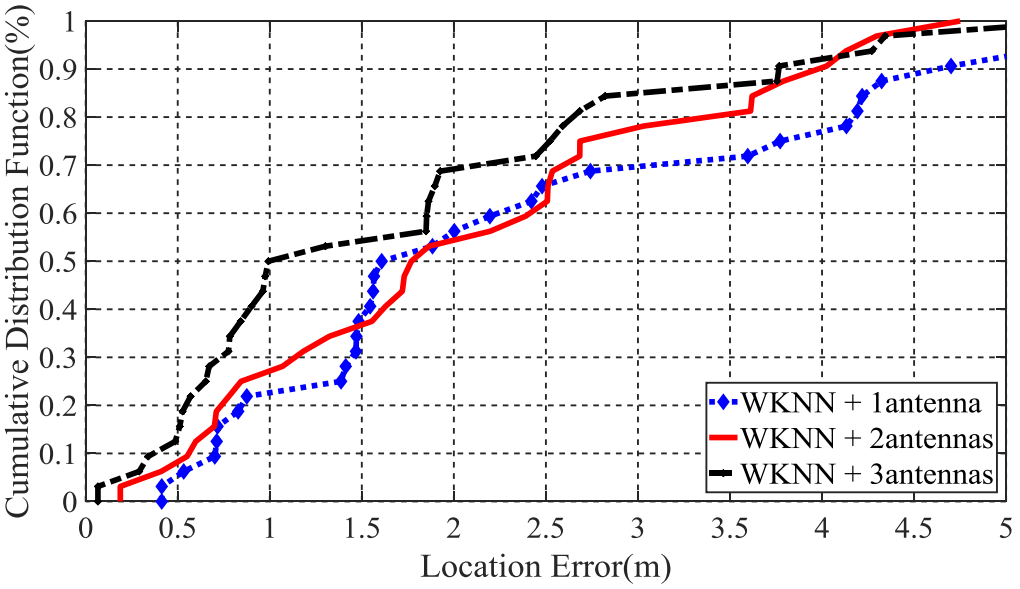
4.7.8 Impact of number of sensors

Radio fingerprinting technique relies on the unique RF signature created at each location in the propagation space. The number of possible unique signatures in turn depends on the number of receiving antennas. In order to evaluate the impact of number of antennas on the localisation accuracy, the data collected as describe in chapter 3 is used. First, RSS values from one antenna are used to infer PD locations using each model developed. We

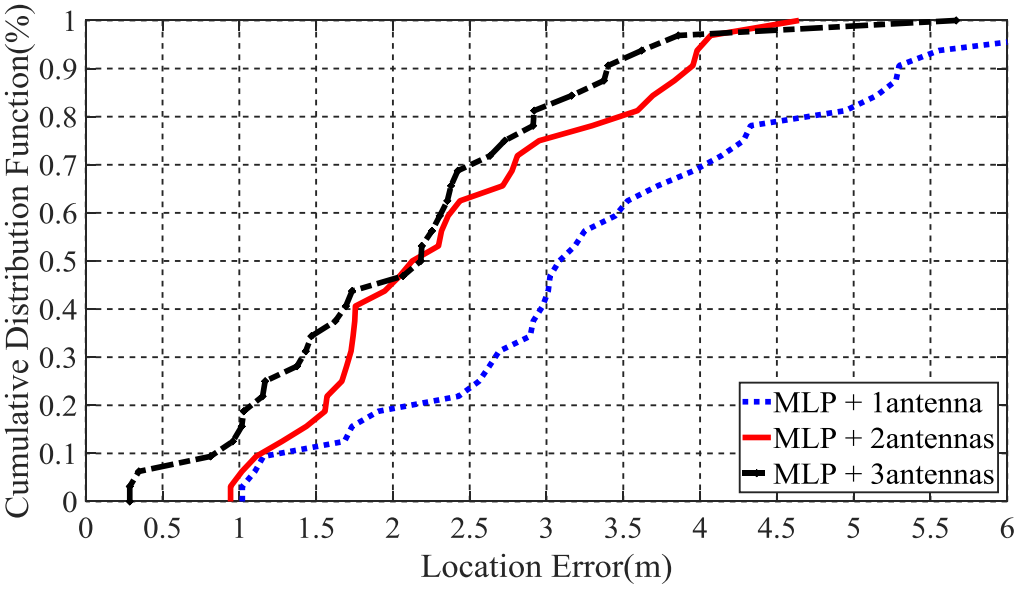
then used RSS values from two antennas and finally RSS values from the three antennas. Figures 4.10 (a), (b), (c) and (d) show the impact of the number of receiving antennas on PD localisation accuracy for each of the PD localisation models. The results indicate that there is a steady improvement in localisation accuracy as the number of antennas increases from 1 through 3. In GRNN model the median and the 75^{th} percentile localisation error were reduced by 42% and 34% respectively when 3 antennas were used instead of 2. This indicates that in practical reality where a concentration of sensors would be deployed in electrical substations for PD localisation there would be a corresponding increase in localisation accuracy.



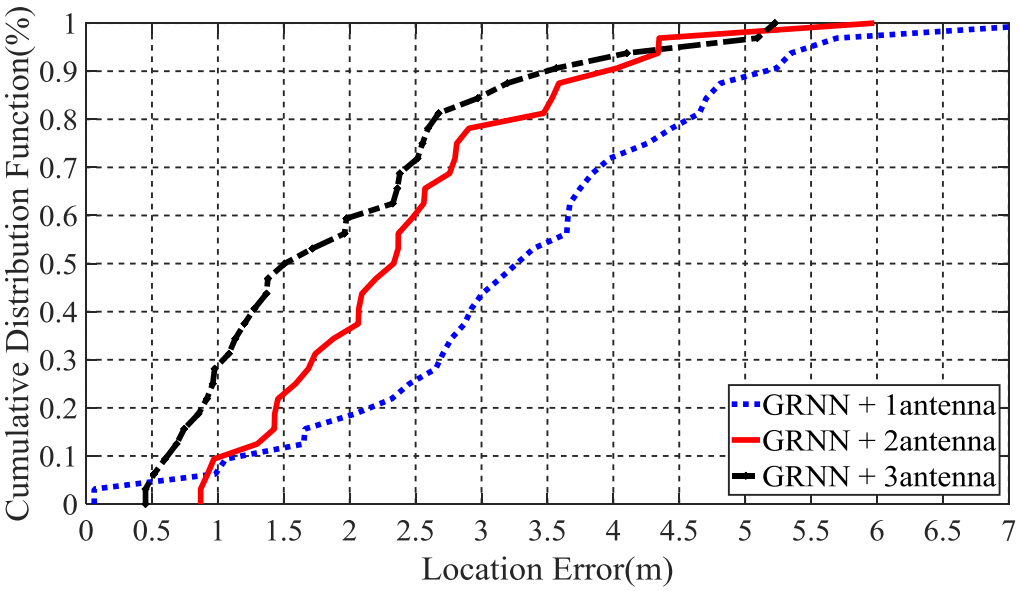
(a) KNN


Cumulative Distribution Function(%)
Location Error(m)
WKNN + 1antenna
WKNN + 2antennas
WKNN + 3antennas

(b) WKNN


Cumulative Distribution Function(%)
Location Error(m)
MLP + 1antenna
MLP + 2antennas
MLP + 3antennas

(c) MLP



(d) GRNN

Figure 4.10: CDF of localisation error with varying number of antennas for each model.

4.8 Summary

PD source localisation is an important aspect of continuous monitoring of entire substation to facilitate maintenance. This chapter introduces a substation-wide RSS-based PD localisation system. Four different algorithms have been used to model the RSS-location relation for PD localisation. Two variants of artificial neural network; MLP and GRNN are presented in this work along with two variant of nearest neighbour algorithm; KNN and WKNN. RSS values of PD signal collected during the measurement campaign described in chapter 3 are used as PD fingerprint and input to the developed models. Performance evaluation of the four models is carried out. The results indicate that RSS-based substation-wide PD localisation is feasible.

The PD events in this work were generated with a pulse generator to ensure that the emulated PD traces are identical for each location and repetitive. Though we obtained good results with our models using RSS, in practical reality where different sources of

PD generate signals with different shapes and amplitude, the absolute values of RSS will differ significantly across PD types, which will affect the performance of our models. The nature of the inverse problem of PD localisation presented in this work, the impact of propagation on received signals and sparseness of PD data motivate an investigation into sophisticated machine learning technique. In the next chapter, we will explore kernel based machine learning techniques for PD localisation. Furthermore, robust PD location features can be reconstructed based on power ratios of received signals on each sensor pair to counterbalance the variation in received signals from different PDs.

CHAPTER 5

Support Vector Regression Approach for Partial Discharge Localisation using Signal Strength Ratios

5.1 Introduction

In the previous chapter, the absolute received signal strength (RSS) patterns of partial discharge (PD) radio measurements were used as PD fingerprints to infer PD location. Although we obtained good results, the complexity of PD occurrence, RSS patterns and the inverse problem posed by PD localisation motivate an investigation into the use of more robust PD features and sophisticated kernelised methods. This method combines radio-location fingerprinting and kernel based machine learning technique for PD localisation. Kernel-based algorithms effectively model the complex relationship between PD features and location, along with correlation in the PD features at proximate locations, in a kernel. They take the measured PD features such as RSS as inputs and give a mea-

sure of location as output. Specifically, two kernelised models based on Support Vector Machines (SVMs) are developed: Support Vector Regression (SVR) and Least-Squares Support Vector Regression (LSSVR). The performance of the approaches in terms of location accuracy is examined. Furthermore, instead of absolute RSS measurements, a derivative of RSS patterns otherwise known as received signal strength ratio (SSR) is used as PD features. We assert that the relative RSS derived in this chapter is robust and can provide improved performance in terms of PD localisation accuracy.

5.2 Inverse Problem Formulation

The measured PD RSS r at every observation point are closely related to the location of discharge. This relation can be expressed mathematically as shown in Equation 5.1;

$$r = f(l) \tag{5.1}$$

where l is the location of PD and f represent the relationship between l and r. The aim of this formulation is to find the unknown location of a PD source from PD features such as RSS or its derivative. This reduces to determination of the following relation as shown in Equation 5.2;

$$l = f^{-1}(r) \tag{5.2}$$

This problem can be reformulated as a regression problem, where the unknown function f^{-1} must be approximated from the knowledge of a number of known pairs of vector (r_i, l_i) with $i = 1, 2, ..., N$. The known pairs of vectors (r_i, l_i) serve as inputs into any regression model. Once f^{-1} is known, the unknown location of PD can be immediately inferred from measured PD features such as RSS, and r. In this chapter, Support Vector Regression (SVR) and its least square version are used to approximate f^{-1} and hence locate PD.

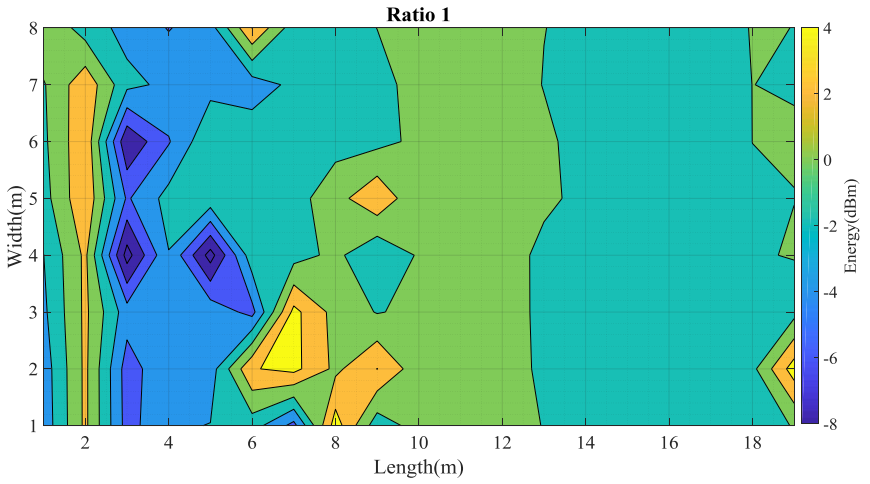
5.3 PD location fingerprint

The absolute RSS measurements used in chapter 4 are often used as fingerprints for radio beacons that are characterized by constant transmission powers [126]. Given that the energy emitted by each PD event may be different due to the progressive nature of PD severity as deterioration continues, absolute RSS is not well suited to the PD location problem under investigation. A more robust fingerprint is the ratio of RSS components received by multiple receivers. In this chapter, instead of the absolute RSS values, the Signal Strength Ratios (SSR) between pairs of sensor nodes are used as the location fingerprint. Lets Assume $P = p_1, p_2, ...p_m$ is a set of receiving nodes deployed in the compound for monitoring PD and $L = l_1, l_2, ...l_n$ represents the finite location space. For each location space, l_i, an observation can then be represented by a pair of nodes $p \epsilon P$ and a measured signal strength value $r \epsilon R$ where $R = r_1, r_2, ...r_p$. Here, the signal strength ratio is defined at each location for a unique node pair $p_i \times p_j \epsilon P \times P$ with the constraint $i < j$ for uniqueness. Therefore, the signal strength ratio for the receiving nodes p_i and p_j can be computed for PD measurement taken at location $l = [(p_i, r_i), (p_j, r_j)]$ as in Equation 5.3;

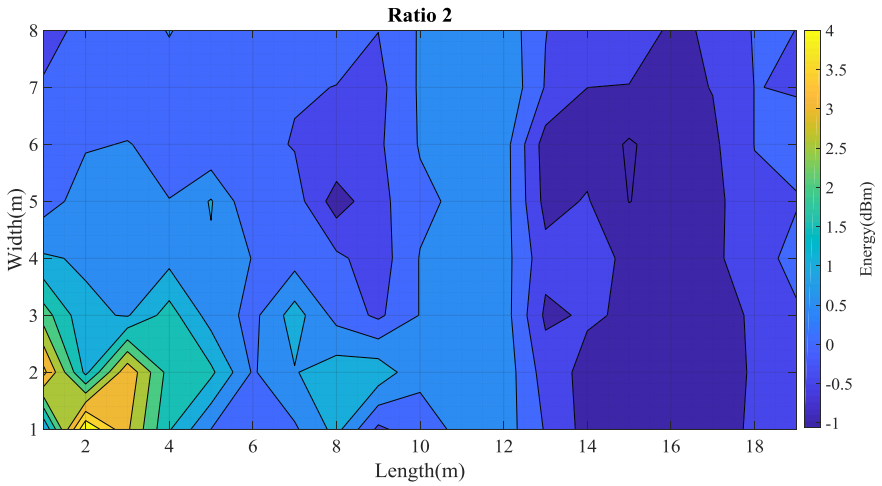
$$SSR(p_i, p_j) = \frac{r_i}{r_j} \tag{5.3}$$

By way of demonstration, Figures 5.1 (a), (b) and (c) show the spatial variation of the computed SSR in the experimentation laboratory. Here, each of the signal strength maps is not exclusive to one receiving sensor but a ratio of a pair of receiving sensors. For clarity, the ratio of signal strength between sensor 1 and 2 is called Ratio 1 (Figure 5.1 (a)), the SSR between sensor 1 and 3 is Ratio 2 (Figure 5.1 (b)), and SSR between sensor 2 and 3 is Ratio 3 (Figure 5.1 (c)). These represent the radio map of the laboratory with respect to PD SSR features. The colour changes indicate the variation in signal strength ratio with locations. This shows the complexity of the radio environment that is characterise by multipath propagation phenomenon. The yellow colour indicate high reception of

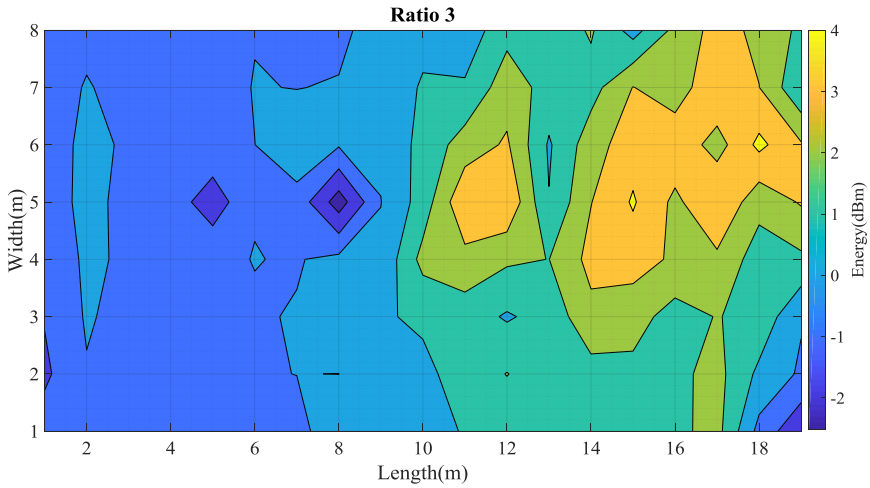
emulate PD pulse from locations close to a particular receiver. As the PD source moves away from a particular receiver, there is more complexity in propagation as signals from relatively distant sources seem to be captured better compared to sources that are nearest the receiver. The blue colour is indicative of low reception of PD signals from locations quite far away from that particular receiver. These spatially correlated measurements (regions with similar wireless propagation characteristics) will be exploited to create a database which will be used to infer PD location using machine learning technique. We will demonstrate the viability of SSRs for PD localisation in the next section using two of the models developed in chapter 4.



(a) Ratio of antennas 1 and 2



(b) Ratio of antennas 1 and 3



(c) Ratio of antennas 2 and 3

Figure 5.1: PD signal strength ratio spatial pattern.

5.4 Verification of RSS ratio performance

In order to demonstrate the viability of using signal strength ratios for PD localisation, two of the machine learning algorithms used in chapter 4; KNN and GRNN are used to model the SSR-location relation and then estimate PD location. These algorithms are chosen for their simplicity and ease of implementation. The relative PD RSS (ratios) constructed from PD training data collected during the measurement campaign described in chapter 3 and their reference locations are used as input to the models. The training of the models is done as explained in chapter 4. The models are then tested using SSR from the 32 distinct test locations as inputs. The outputs of the models are the estimated x and y coordinates of the 32 test locations. Because we are using signal strength ratios as PD input features in this chapter, there is need to optimise again both KNN and GRNN for values of k and σ respectively. An exhaustive search is used to find the optimal values of the parameters. The results of the exhaustive search for KNN and GRNN are shown in Figures 5.2 and 5.3. For KNN, Figure 5.2 shows that the optimal value of k is 4, which is

different from $k = 5$ that was used in the previous chapter where absolute values of RSS were used. Figure 5.3 shows that the optimal σ is 0.13. This differs from 0.31 that was the optimal value of σ in chapter 4 analysis where absolute values of RSS were used.

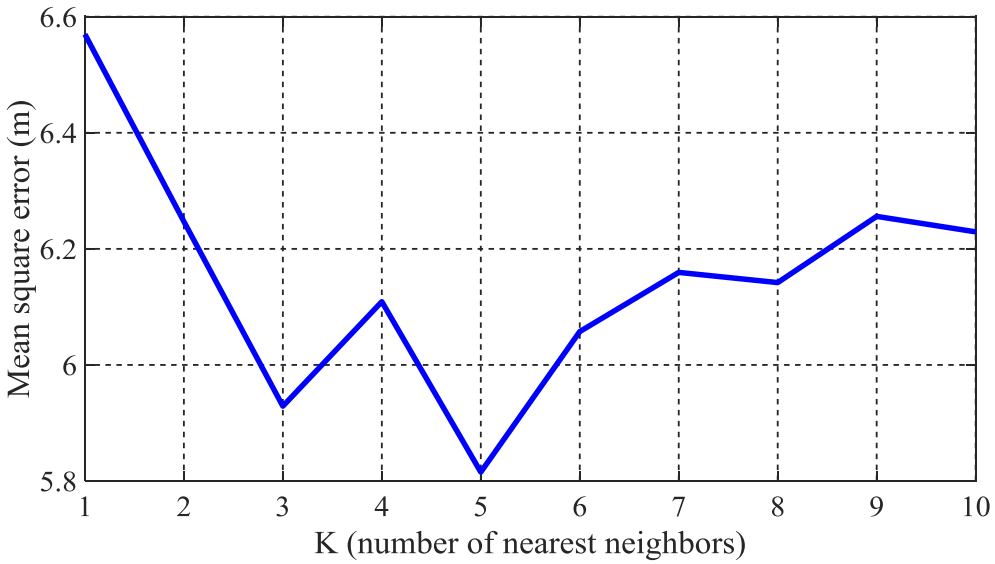


Figure 5.2: Effect of K on performance of KNN.

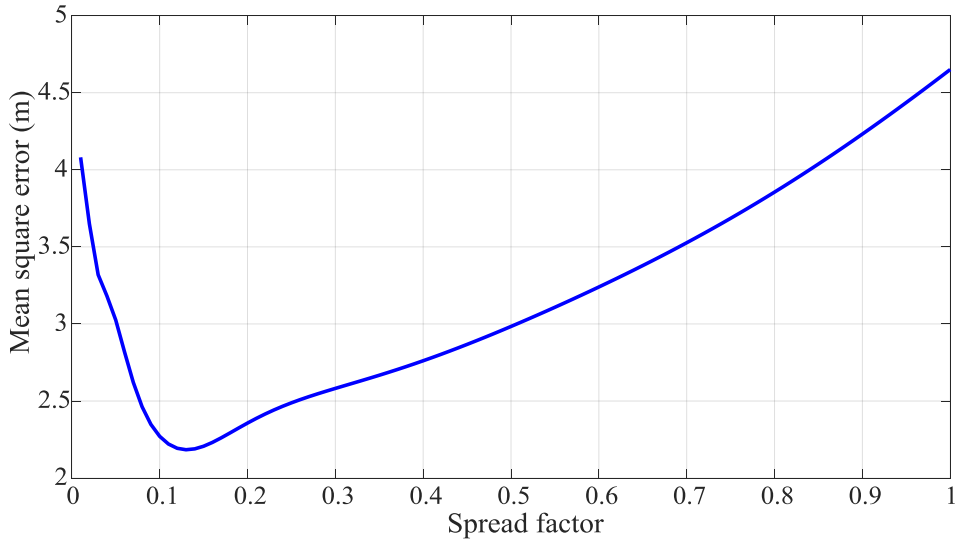


Figure 5.3: Effect of spread factor on performance of GRNN.

With these optimal parameters, KNN and GRNN PD localisation models are developed

using signal strength ratios. The performance of KNN and GRNN models based on signal strength ratios (SSR)is quite encouraging with both models having mean location error of $1.81m$. It is interesting to note here that with SSR, KNN is able to locate PD with a minimum location error of 0 (exact location of PD), while GRNN has a minimum location error of 0.31. In comparison, the localisation result of both models has improved using SSR compare to using RSS as PD fingerprints. The summary of the result is shown in Table 5.1. This result does not only demonstrate that signal strength ratios are robust features for PD localisation but can also improve localisation accuracy.

Table 5.1: KNN and GRNN performance using signal strength ratio.

Parameter(m)	KNN	GRNN
Mean	1.81	1.81
75%	2.60	2.38
Max	5.00	4.41
Min	0.00	0.31

5.5 Support Vector Machine Models

Support Vector Machines (SVM) [127][128][129] are kernel-based learning techniques applicable to both classification and regression problems. SVMs are based on the idea of mapping the original input vectors into a higher dimensional feature space through a kernel function where a separating hyper-plane can be found. Without the need to compute the non-linear mapping explicitly, dot-products can be computed efficiently in higher dimensional space. The dominant feature which makes SVM very attractive is that problems which are non-linearly separable in the original space can be linearly separated in the higher dimensional feature space. SVMs have shown tremendous success in applications such as data classification, time series prediction, identification systems and data clustering [129][130][131][132][133]. In the context of PD localisation, we formulate SVM as a regression task, which consist of training a model that defines the non-linear mapping

function between the PD features and their spatial location in high dimensional feature space, leading to Support Vector Regression (SVR) [134][135][136][137].

5.5.1 Support Vector Regression

5.5.1.1 Theory of SVR

The support vector regression technique is a learning procedure based on statistical learning theory which employs structural risk minimisation principles [134]. SVR uses training data to build its prediction model. This method can be applied to both linear and nonlinear regression problems. In the case where the relation between a multidimensional input vector and the output is unknown and very likely to be non-linear, SVR maps the input vectors into a high-dimensional feature space by a nonlinear mapping function, where samples become linearly separable and an optimal regression surface is constructed [138]. SVR aims at finding a linear hyperplane, which fits the multidimensional input vectors to output values. The constructed model is then used to predict output values for data that is not previously seen during training. Given our training dataset of signal strength ratios (SSR) values and locations $T_d = (r_i, l_i)$, $(r_i \in R, l_i \in L)$ where $i = 1, ...n$ with r_i the input vector of data point i, l_i the actual locations and n the number of data points, the goal of SVR is to find the mapping $f : R \rightarrow L$ and make $f(r_i) \approx l_i$. For a nonlinear problem, the training patterns are pre-processed by a map into some feature space before SVR is applied. SVR finds the best or optimal regression surface $f(r)$ within a deviation ϵ as the prediction model, leading to ϵ-SVR [139]. For theoretical purposes, lets begin with the linear case, suppose we have a set of training data where r_n is a multivariate set of N observations with corresponding locations l_n, the linear function to be determine can be expressed as in Equation 5.4;

$$f(r) = \omega^T r + b \tag{5.4}$$

where ω is the weight vector and b is the bias. The training of the SVR is to find $f(r)$ that minimises the norm value ($\omega^T\omega$). This is formulated as a convex optimisation problem to minimise Equation 5.5;

$$\begin{aligned} min_\omega J = \frac{1}{2}||\omega||^2 \\ \text{s.t.} \quad & l_i - < \omega, r_i > -b \leq \epsilon \\ & < \omega, r_i > +b - l_i \leq \epsilon \end{aligned} \tag{5.5}$$

The problem in Equation 5.5 might be restrictive by bounding the range of errors of the training data within ϵ. Thus, to deal with the otherwise infeasible constraints, we introduce slack variables ξ_i, ξ_i^* for each point. This approach is similar to the "soft margin" concept in SVM classification. The slack variables allow regression errors to exist up to the value of ξ_i and ξ_i^* without degrading performance. With these slack variables the minimisation problem in Equation 5.5 becomes Equation 5.6 [140];

$$\begin{aligned} min_{\omega,b,\xi} J = \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ \text{s.t.} \quad & l_i - < \omega, r_i > -b \leq \epsilon + \xi_i \\ & < \omega, r_i > +b - l_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{5.6}$$

Where C is the box constraint, a positive numeric value that controls the penalty imposed on data points that lie outside the ϵ margin and helps to prevent overfitting (regularisation). To solve the problem in Equation 5.6, a standard dualisation method with Lagrange multipliers α_i, α_i^* can be used [139]. To obtain the dual formula, we construct a Langrangian function from the primal function by introducing the nonnegative multipliers

$\alpha_i, and \quad \alpha_i^*$. This leads to the dual formula in Equation 5.7:

$$\begin{aligned} minL(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)r_i^T r_j + \\ \epsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} l_i(\alpha_i^*) - \alpha_i \\ \text{s.t.} \quad \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ 0 \leq_i \leq C \\ 0 \leq_i^* \leq C \end{aligned} \tag{5.7}$$

By solving the dual problem, ω can be described as a linear combination of the training points using Equation 5.8;

$$\omega = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)r_i \tag{5.8}$$

Where $\alpha_i \geq 0$ and $\alpha_i^* \geq 0$. The function used to estimate locations depends only on the support vectors and can be expressed as Equation 5.9.

$$f(r) = \sum_{i=1}^{SV}(\alpha_i - \alpha_i^*)(r_i^T, r) + b \tag{5.9}$$

To obtain a nonlinear SVR model, we replace the dot product $r_1^T r_2$ with a nonlinear kernel function $K(r_1, r_2) =< \phi(r_1), \phi(r_2) >$, where $\phi(r)$ is a transformation that maps r to a high-dimensional space. With this, the linear model in Equation 5.9 becomes a nonlinear SVR model expressed as in Equation 5.10;

$$f(r) = \sum_{i=1}^{SV}(\alpha_i - \alpha_i^*)k(r_i^T, r) + b \tag{5.10}$$

where α_i are the Lagrange multipliers which satisfy $0 < \alpha_i^* < C$, r_i are the support vectors whose α_i is not zero, and SV is the number of support vectors. Equation 5.10 shows that the decision function depends on support vectors. This means an optimal

regression surface is constructed by these support vectors. The idea of support vectors form a sparse subset of the training data that can be used and is particularly useful for resource constraint applications such as the one under investigation.

5.5.1.2 PD Localisation based on SVR

In our approach, we propose to deploy a number of sensors at arbitrary but known locations in a two-dimensional substation compound; these sensors record radio emission from PD sources in real-time and estimate the location of the PD. During the PD measurement campaign described in chapter 3, the compound was divided into a 1×1 squared grid. Each grid point represented by an x, y coordinate was considered a PD location (source). Therefore, to compute PD location, two SVR models are required; one for each x and y coordinates. The PD feature used to develop the SVR models is the received signal strength ratio (SSR). Firstly, RSS from known locations are gathered by the three sensors deployed in the compound. The ratios of these RSS are calculated as described in section 5.3. These SSR values and their corresponding location coordinates form a training database for the models. Suppose the true coordinates of PD location l_i are (x_i, y_i) and the corresponding values of SSR for the PD from this locations are (r_i^1, r_i^2, r_i^3). The vector $R_i = [r_i^1, r_i^2, r_i^3]$ with their correct locations are taken as the SVR input feature vector for training and the resulting model is used to infer the PD location $l_i(x_i, y_i)$ of new observed PD measurement. All the feature vectors from known locations and their corresponding location coordinates constitute the SVR training sample set. Using this training set, the SVR is trained to build a PD localisation model that would subsequently be used to predict or estimate PD location given an SSR vector that is not previously seen. During training, the input fingerprints (SSR/location) from known locations are transformed to a new feature space with N features, one for each support vector. That is to say that, they are represented only in terms of their dot products with support vectors (special data points chosen by the SVR optimisation algorithm). A Kernel function [141] is em-

ployed in the transformation to provide a nonlinear mapping from the input space to the new feature space. The kernel functions that were used in this thesis are listed in Table 5.2. In the localisation phase, for a given SSR vector the kernel finds the similarity or a distance measure between the vector and the support vectors stored after training. The corresponding coordinates of the support vectors closest to the SSR vector are used to compute the PD location for the given SSR vector. The SVR PD location model is as shown in Figure 5.4.

Table 5.2: Kernel functions.

Kernel Name	Kernel Function
Linear	$K(r_i, r) = r_i^T r$
Polynomial	$K(r_i, r) = (1 + r_i^T r)^q, q = 2, 3, ...$
Gaussian (RBF)	$K(r_i, r) = exp(-\gamma \|\|r_i - r\|\|^2)$

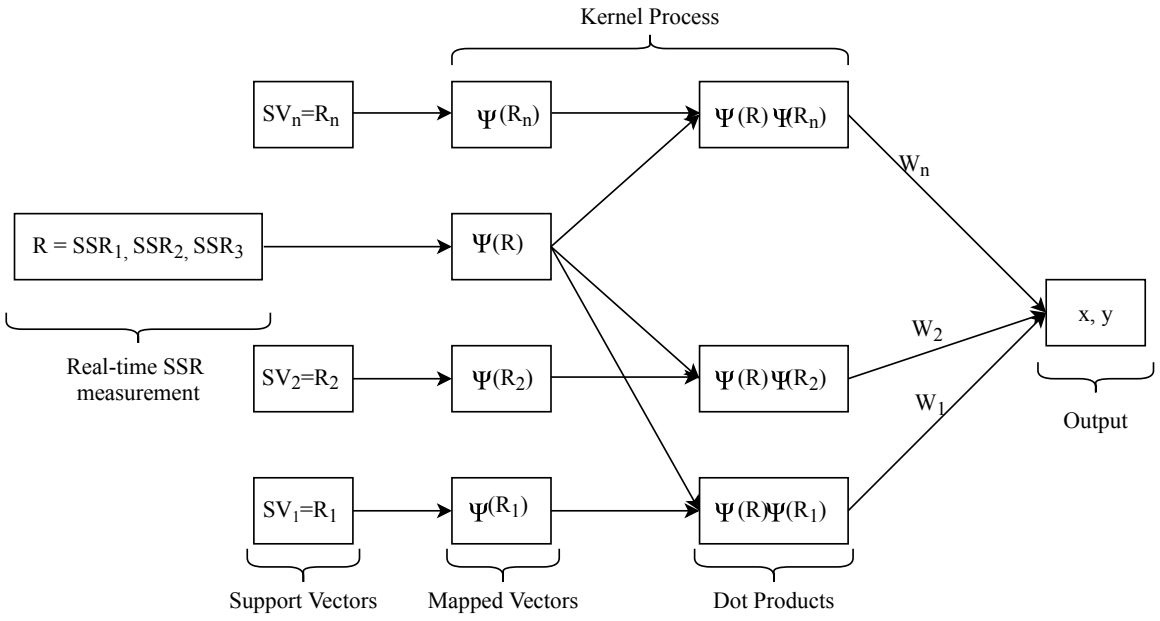


Figure 5.4: SVR PD Localisation Model.

5.5.2 Least Squared Support Vector Regression

The Least Squares Support Vector Regression (LSSVR) algorithm [133][142] is a reformulation of the standard SVR algorithm described in Section 5.5.1.1 by adding a least squares term in the cost function, which leads to solving a system of linear equations. The idea of linear equations makes LSSVR more appealing and computationally more economical compare to solving the convex Quadratic Programming (QP) for standard SVR. LSSVR for PD localisation consists of two phases: training and localisation. During the training phase, the optimal parameters of LSSVR algorithm are tuned using PD signal strength ratios (SSR) at known locations (training points) and PD localisation model developed. In the localisation phase, PD measurements taken at unknown locations are analysed, and their locations obtained using the developed model. Given a training data set $(r_1, l_1), (r_2, l_2), ..., (r_n, l_n)$ of n points, with PD input (SSR) data $r_i \in R^m$, with m-dimension (in this case number of receiving sensors), and output (PD location coordinate) data $l_i(x_i, y_i) \in R^2$ in 2-dimension, the reformulated PD regression problem can be expressed as the following optimisation problem in Equation 5.11

$$\begin{aligned} min_{\omega,b,e} J = \frac{1}{2}||\omega||^2 + \frac{1}{2}\gamma \sum_{i=1}^{N}(e_i^2) \\ \text{s.t.} \quad l_i =< \omega, \psi(r_i) > + b + e_i, \qquad i = 1, ..., N \end{aligned} \tag{5.11}$$

with $\psi(.) : R^n \rightarrow R^{n_h}$ being a non-linear function which maps the input data into the so-called higher dimensional feature space, weight vector $\omega \in R^{n_h}$ in primal weight space, bias term b and error variable e_i. The error term here represents the true deviation between actual values and estimated values, rather than a slack variable which is needed to ensure feasibility (as in SVR case). $\gamma \geq 0$ is a regularisation constant. The standard SVR is modified in two ways; equality constraints are used instead of inequality constraints, and a squared loss function is taken to represent the error variable. These modifications are expected to simplify the problem. To solve the optimisation problem in the dual space,

one defines the Lagrangian in Equation 5.12;

$$L(\omega, b, e, \alpha) = J(\omega, b, e) - \sum_{i=1}^{N} \alpha_i \omega^T \Psi(r_i) + b + e_i - l_i \tag{5.12}$$

with Lagrange multipliers $\alpha_i \rightarrow R$. Differentiating Equation 5.12 with respect to ω, b, e and α gives us the conditions for optimality as shown in Equation 5.13;

$$\begin{cases} \frac{\partial L_{LSSVR}}{\partial \omega} = 0 \rightarrow \omega = \sum_i^n = 1\alpha_i \Psi(r_i) \\ \frac{\partial L_{LSSVR}}{\partial b} = 0 \rightarrow \sum_i^n = 1\alpha_i = 0 \\ \frac{\partial L_{LSSVR}}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L_{LSSVR}}{\partial \alpha_i} = 0 \rightarrow < \omega, \Psi(r_i) > + b + e_i - l_i = 0. \end{cases} \tag{5.13}$$

After eliminating ω and e, the solution yields the following linear equations as shown in Equation 5.14;

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \frac{1}{\gamma} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ L \end{bmatrix} \tag{5.14}$$

Here, γ is a constant value which balances the generalisation capability and the accuracy of LSSVR; $1_v = (1, ..., 1)^T$ is an N-dimensional column vector, and I is a $N \times N$ identity matrix; $l = (l_1, ..., l_n)^T$, and $\alpha = (\alpha_1, ..., \alpha_n)^T$ the Lagrange multipliers. $\Omega_{ik} = k(r_i, r_k)$ is the kernel function. Both the kernel parameters and γ are determined during the training process.

The resulting LSSVR model for PD location estimation is as shown in Equation 5.15;

$$\hat{l}(r_i) = \sum_{i=1}^{SV} \alpha_i k(r_i, r) + b \tag{5.15}$$

where α, b are obtained during the training phase by solving Equation 5.14.

5.6 Performance Analysis

5.6.1 Performance Evaluation Metrics

The following performance metrics; cumulative Distribution Function (CDF), accuracy, percentile, precision, and Root Mean Square Error (RMSE) as defined in chapter 4 are used to evaluate and compare the partial discharge localisation models described in Section 5.5.

5.6.2 Impact of kernel function on SVR performance

Kernel functions are used in Support Vector Machines to map the non-linear inseparable data into a higher dimensional feature space where the computational power of the linear learning machine is increased [10], [16]. In this thesis, we investigate the effect of three commonly used kernel functions namely; linear, polynomial and Gaussian RBF on the performance of SVR algorithm for PD localisation. For the polynomial kernel, an order of 2 is chosen for analysis. Result obtained with different kernel functions is as shown in Table 5.3. It can be seen that the Gaussian RBF kernel function has the highest

Table 5.3: Performance of SVR model using different kernels.

Kernel Function	mean (m)	25%	50%	75%
Linear	2.73	2.02	2.65	3.38
Polynomial	2.34	1.49	2.49	3.15
Gaussian RBF	1.97	0.89	1.54	2.73

localisation accuracy with median location error of 1.54m. This implies that 50% of the time SVR-RBF locates PD with error less than 1.54m. Meanwhile, the kernel that has the lowest localisation accuracy is the linear kernel function. With linear kernel, our model has a median location error of 2.65m. Therefore, Gaussian RBF kernel is adopted for use in subsequent analysis in this thesis.

5.6.3 Impact of localisation algorithm

In order to evaluate the performance of the PD localisation models developed, the data collected from the measurement campaign described in Chapter 3 was used. The received signal strength ratios (SSR) derived as described in section 5.3 were used as PD location features. To train the PD location models, features generated from the radio signals collected as training data set ($144 \times 3 = 432$, a combination of locations and sensors) and their corresponding ground truth locations were used as training inputs into each model. Each (feature, location) pair constitutes training data point. The PD SSR from the test data at 32 distinct locations without their true locations were then presented to the models to estimate the PD source locations. The corresponding point accuracy is visualised in the form of heat-maps for each of the models in Figures 5.5, and 5.6.

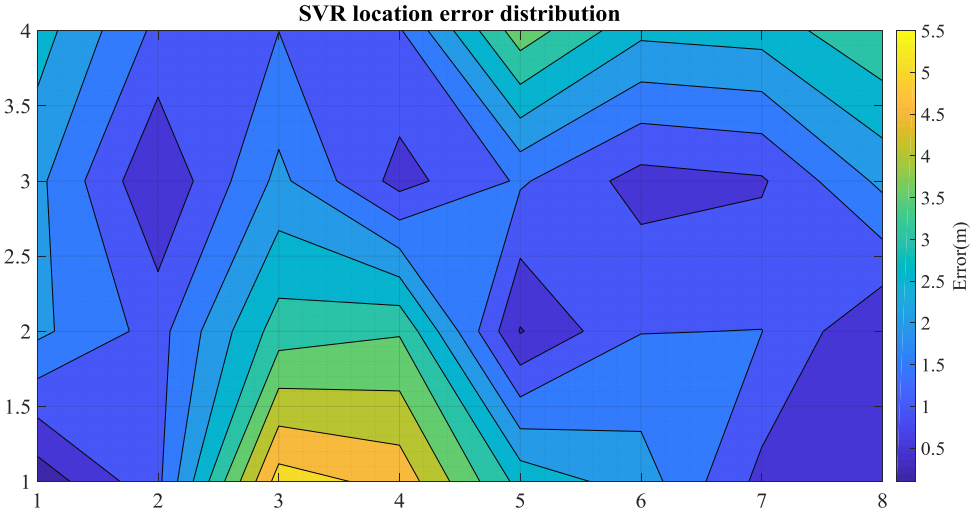


Figure 5.5: SVR location error distribution.

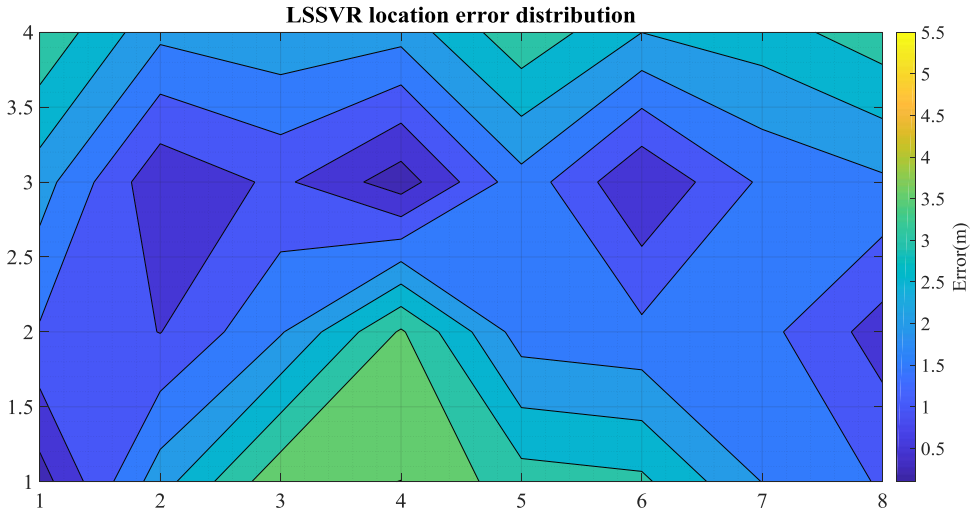


Figure 5.6: LSSVR location error distribution.

Blue areas refer to good accuracy results (point accuracy of $2m$ or less), whereas the accuracy worsens when the colour changes to green, orange and finally yellow. A yellow zone corresponds to a distance error around $5m$. Based on Figure 5.6, it is clear that LSSVR model achieves the most accurate result in general. The SVR method is less accurate, achieving an average point accuracy around $1.97m$. In addition, results show that the accuracy on one part of the environment is significantly higher than the accuracy in the other part. This may be as a result of uniqueness of the signatures created by multipath effect. It is observed that the corresponding point accuracy near the walls is significantly lower. This can be explain by the fact that a uniform surface like a wall will produce some kind of similarity in signature at different locations. A more detail overview of location errors is given in Figure 5.7.

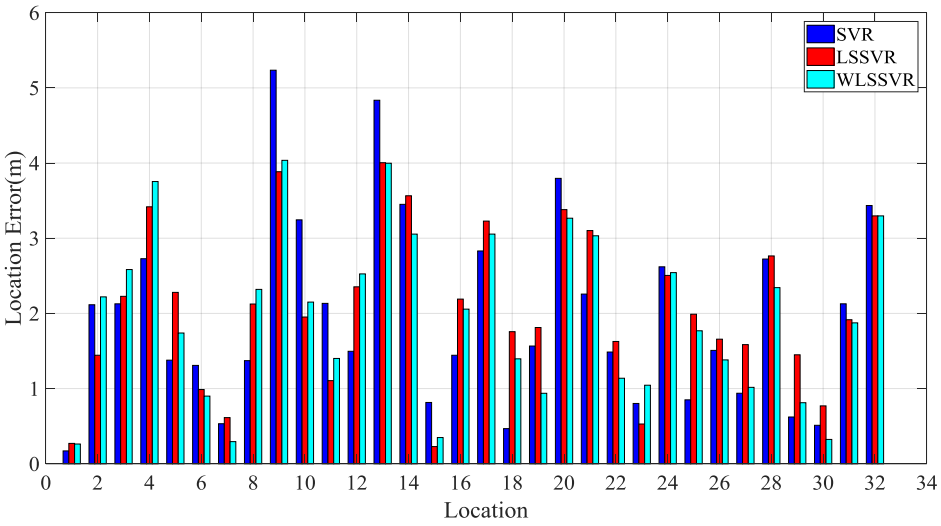


Figure 5.7: Model location error.

This shows the location error in terms of Euclidean distance for each test location. The maximum location error for SVR and LSSVR models is found to be $5.24m$, and $4.00m$ respectively. At four different locations LSSVR is able to provide PD location estimation with location errors less than $0.5m$. It is rather interesting to see SVR model producing location estimate with the lowest location error of $0.17m$. Figures 5.8 and 5.9 further show the errors in each coordinate (x and y) of the test locations for the three models.

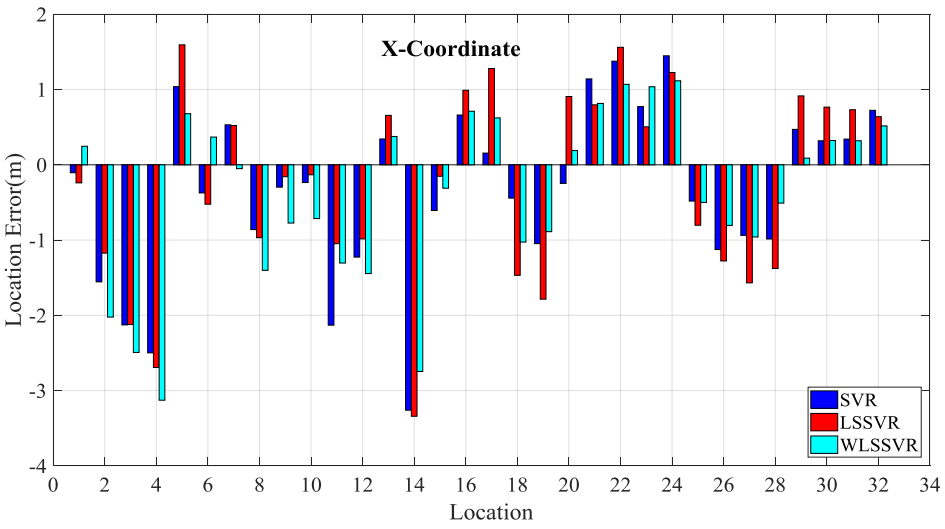


Figure 5.8: Errors in X-component.

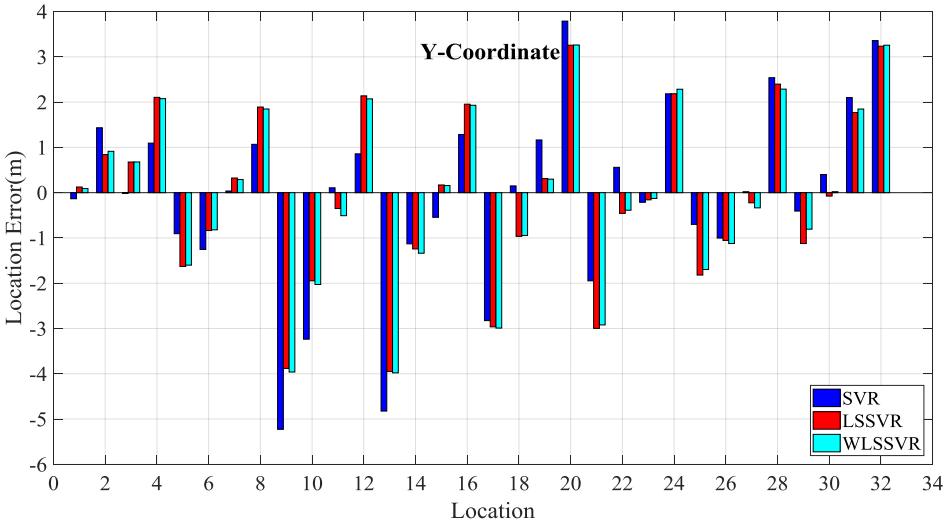


Figure 5.9: Errors in Y-component.

This result shows that the errors in x and y vary from $-3.26m$ to $1.45m$ and $-5.23m$ to $3.79m$ respectively for the SVR model. For LSSVR model, the errors in x and y vary

from $-3.34m$ to $1.59m$ and $-3.95m$ to $3.26m$ respectively. It is observed that the overall accuracy of each of the models is mostly affected by the y-coordinate predictions which may be due to the nature of NLOS (soft-uniform) in this direction. Regarding precision metric, Figure 5.10 shows the cumulative density function (CDF) of localisation error for the models. Given that the spacing between test locations is $2.5m$, the proportion of test data with localisation error below this value is also evaluated. The SVR model has a precision of 68.75% within 2.5m. This means that 68.75% of test locations are estimated with localisation error below 2.5m. Whereas, the LSSVR model has a precision of 71.90% within 2.5m. Which imply that in the case of LSSVR, 71.90% of test points are estimated with error less than 2.5m.

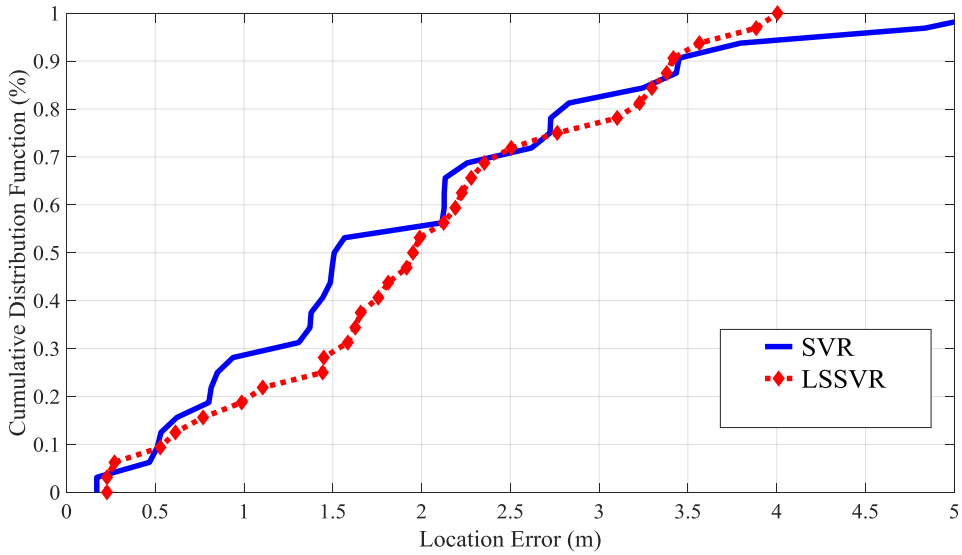


Figure 5.10: Probability of location error for each model.

In terms of RMSE metric, the LSSVR model outperforms the SVR model with overall RMSE of 2.30m which shows that 95% of the estimates are within two RMSE. Again, for both models, the Y component of the RMSE is the predominant source of error in the overall RMSE. This increase in the RMSE of the Y component can be explained by the soft-NLOS (uniform) scenarios created in this dimension, which are predominant in the Y axis. The complete result of the performance of SVM-based models using SSR in

terms of RMSE metric is summarised in Table 5.4.

Table 5.4: Performance of SVM-based models in terms of RMSE.

Model	$\text{RMSE}_x(m)$	$\text{RMSE}_y(m)$	$\text{RMSE}_{overall}(m)$	$\text{Min}_{error}(m)$	$\text{Max}_{error}(m)$
SVR	1.18	2.00	2.33	0.17	5.24
LSSVR	1.29	1.91	2.30	0.22	4.00

5.7 Summary

A robust PD localisation system based on signal strength ratios (SSR) have been considered. The principle and computational realisation of the methods based on support vector machine have been described. Specifically, two variants of support vector machine: support vector regression and least squares support vector regression have been used to model the relationship between SSR and PD source location. In this chapter, SSR and reference locations are used as location fingerprints and inputs to the developed models. The performance of the developed models have been evaluated in terms of statistical operators of localisation error. The results indicate that PD localisation based on SSR and LSSVR are robust and represent practical alternatives for PD source location.

In the next chapter, we additionally exploit the knowledge that multipath effects will influence the delay spread of a signal, and this in turn can provide information on the location of the source. Multipath is highly frequency-selective and so specific frequency components will yield useful localise information. Furthermore, due to the high dimensionality of the features that may be extracted in this work, feature reduction technique is employed to select the best possible set of features that will be used in the localisation process to reduce computational cost whilst maintaining accuracy.

CHAPTER 6

Improving PD Source Localisation

6.1 Introduction

This chapter investigates techniques to improve the proposed PD localisation method. The improvement is targeted towards accuracy and robustness. To this end, instead of using the RSS or relative RSS of the RF signals to construct a database of fingerprints, other signal parameters in both the time and frequency domain have been investigated and used for PD localisation. We propose that a combination of multiple PD features may enhance performance. This chapter is divided into two parts. The first part is based on extracting time domain statistical features from the received RF signals to improve PD localisation. With high dimensional features space created, a feature selection algorithm is employed to select the best possible features which are then used in the localisation process to reduce computational cost whilst maintaining a reasonable level of accuracy. In the second part, wavelet packet transform technique have been investigated for PD

localisation. Extracting PD location features is the essential criterion for this analysis. The received RF signals are first transformed into the time-frequency domain using a time-invariant version of wavelet packet transform; Maximal Overlap Discrete Wavelet Packet Transform (MODWPT). PD features are extracted from the transformed signals. The wavelet based PD features are then used to infer PD localisation. Finally, ensemble learning methods are employed to model the feature-location relation and used to improve accuracy and precision of PD localisation algorithms.

6.2 PD Statistical Feature Characterisation

In any complex radio environment, such as an electricity substation, multipath propagation will occur as a result of reflection and diffraction; this gives rise to many delayed and attenuated versions of the transmitted radio pulse. The superposition of these received pulses produces a temporal signature (energy versus time). Multipath effects are also known to be frequency-selective in that the shape of the signature will change based on the frequency band of the pulse. In order to obtain additional dimensions from which to produce a set of unique fingerprints, we exploit the frequency-selective nature of multipath. It has been found that specific frequency components will yield useful information for PD localisation. The use of relative frequency measures will account for the different frequency components inherent in different types of PD [143]. Signatures extracted from the frequency components represent PD features. The features generated in this section for PD location are as shown in Table 6.1. These features are computed from PD pulse waveforms. Intuitively, the derivation and use of these features for PD location is based on the fact that a discharge event produces an RF signal that will propagate as a travelling wave through the environment. As a result, the radiated signal amplitude and shape are modified by the propagation path between the PD source and the receiving sensor due to path loss attenuation, signal shadowing and multipath effects as mention earlier. These effects may vary quite markedly within an electricity substation leading to the uniqueness

of the received signal to the location of discharge. This suggests that the signatures of the received pulses can be exploited to provide a rich database for inferring PD location. The nature of the received PD pulse is not only a function of the location but also of the measuring device. With the response of the measuring device known, more characteristics of the signals at defined intervals may be explored and used for localisation.

Table 6.1: PD time-domain features.

S/No	Features
1	Variance (VAR)
2	Root Mean Squared (RMS)
3	Peak to Peak Value (PPV)
4	Impulse Factor (IF)
5	Area under the waveform (A)

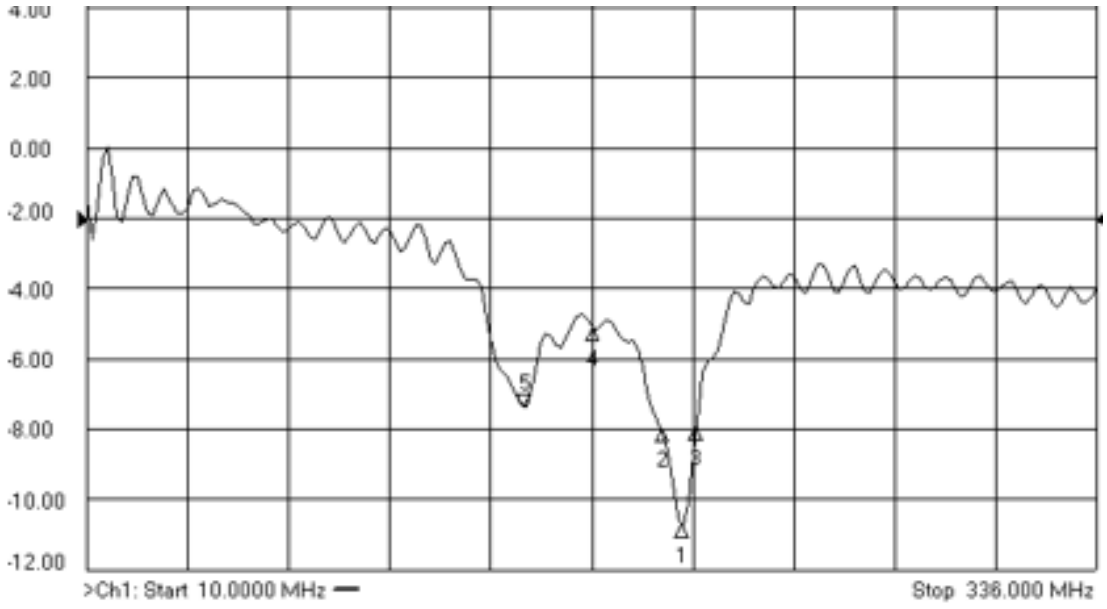


Figure 6.1: Antenna frequency response

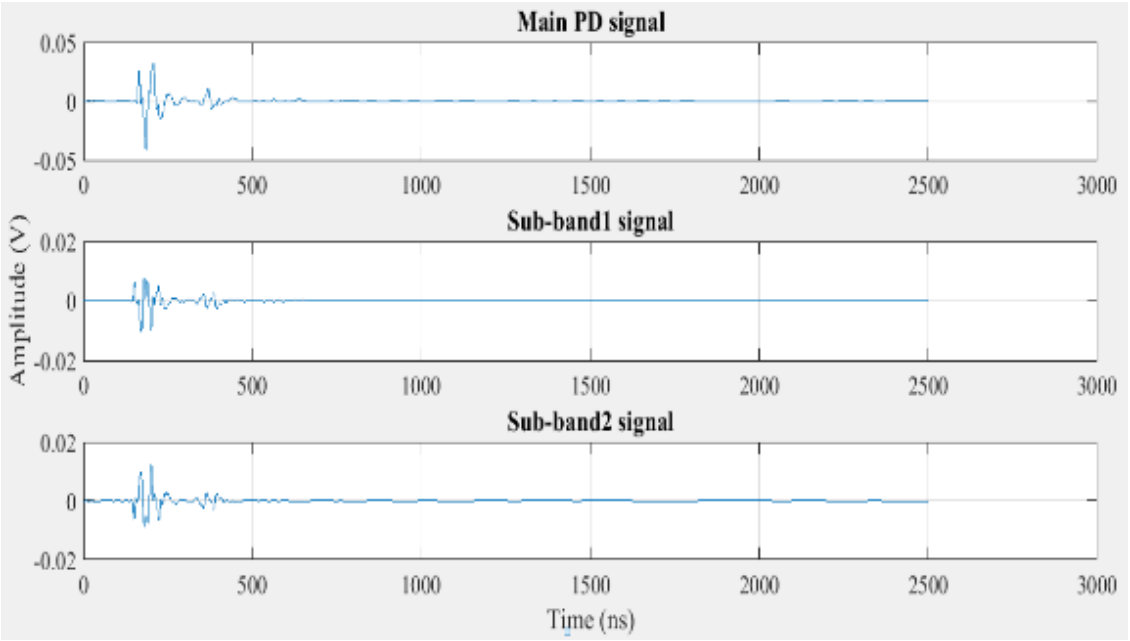
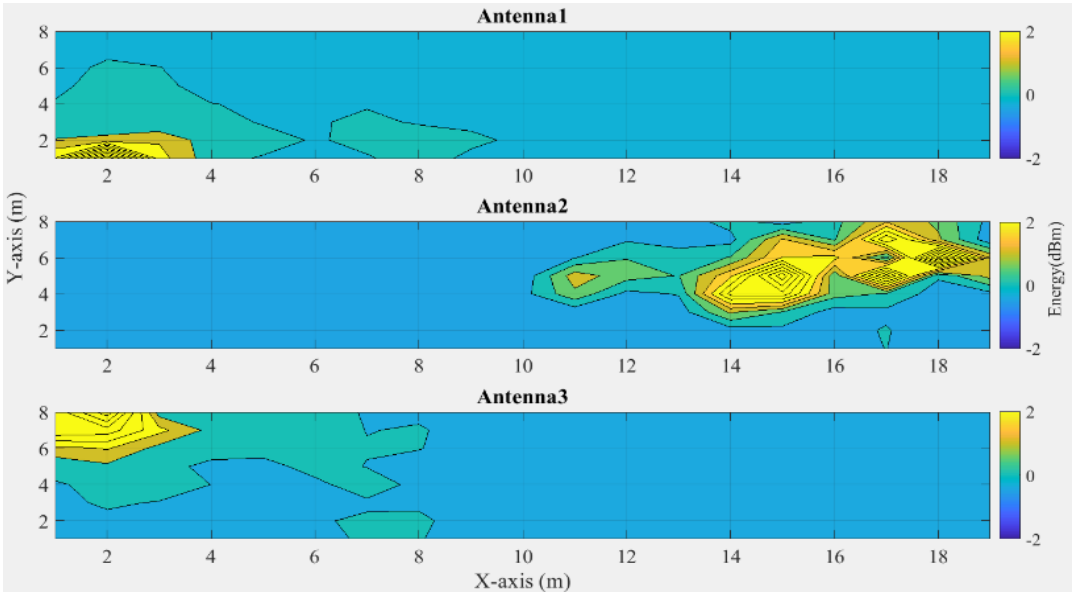
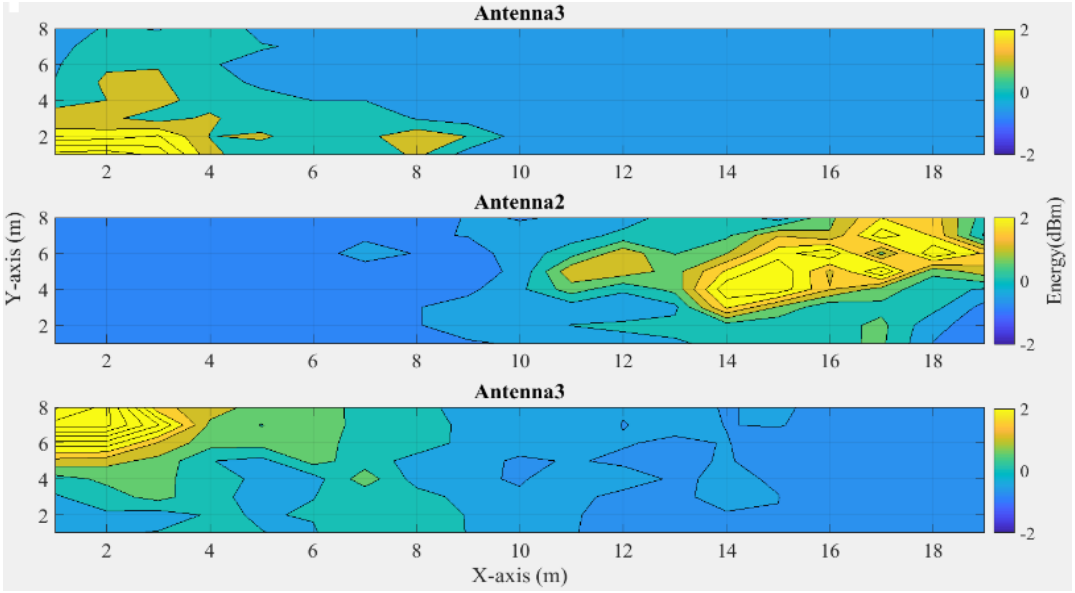


Figure 6.2: PD broadband and sub-band signals

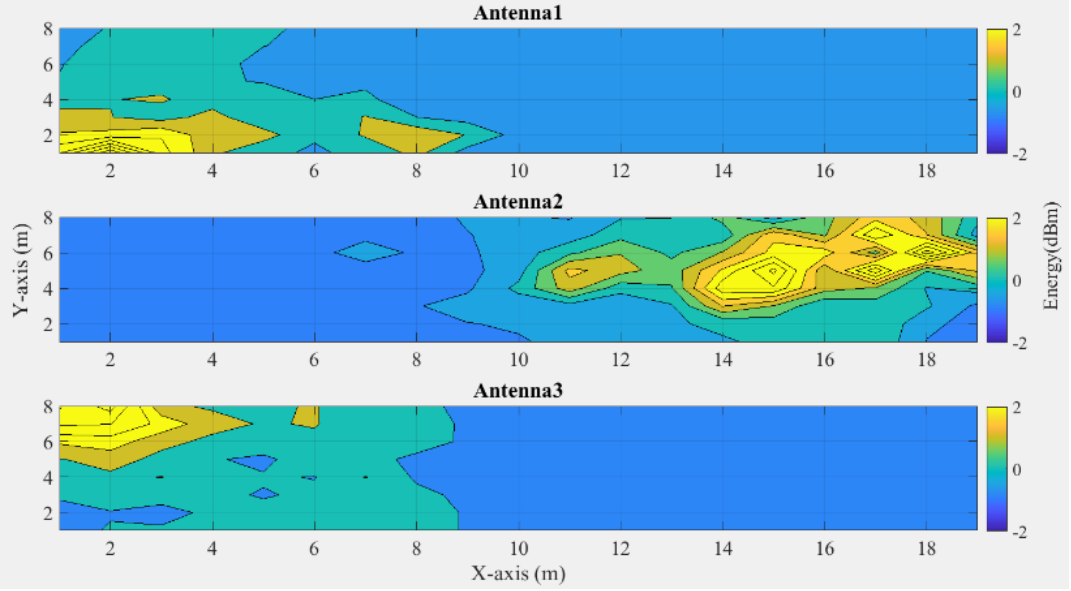
The frequency response of the antennas used during the measurement campaign reported in chapter 3 produce two peaks at 151 MHz and 200.7MHz respectively as shown in Figure 6.1. We believe that these peaks indicate the regions where most of the information about the discharge are captured. Based on the information from the frequency response, two bandpass filters are used to filter the PD signals in these regions for more information. In this work, we take the centre frequencies of the filters to be the peaks observed from the frequency response. The upper and the lower cut-off frequencies are 145.5 MHz and 155.5 MHz for band 1 and 195.35 MHz and 205.35MHz for band 2; these correspond with the 3dB point of each peak. This provides us with two distinct PD sub-signals in addition to the PD broadband signal (main PD signal) as shown in Figure 6.2 for analysis. PD features in Table 6.1 are then extracted from the main PD signal and the two sub-band signals. These features are used as input into PD localisation models. This work is based on the premise that the derived PD features provide meaningful information that can be used to infer PD location. To demonstrate that this is a reasonable premise, Figure 6.3(a-e) shows the maps of how the PD features (variance, root mean square, peak to



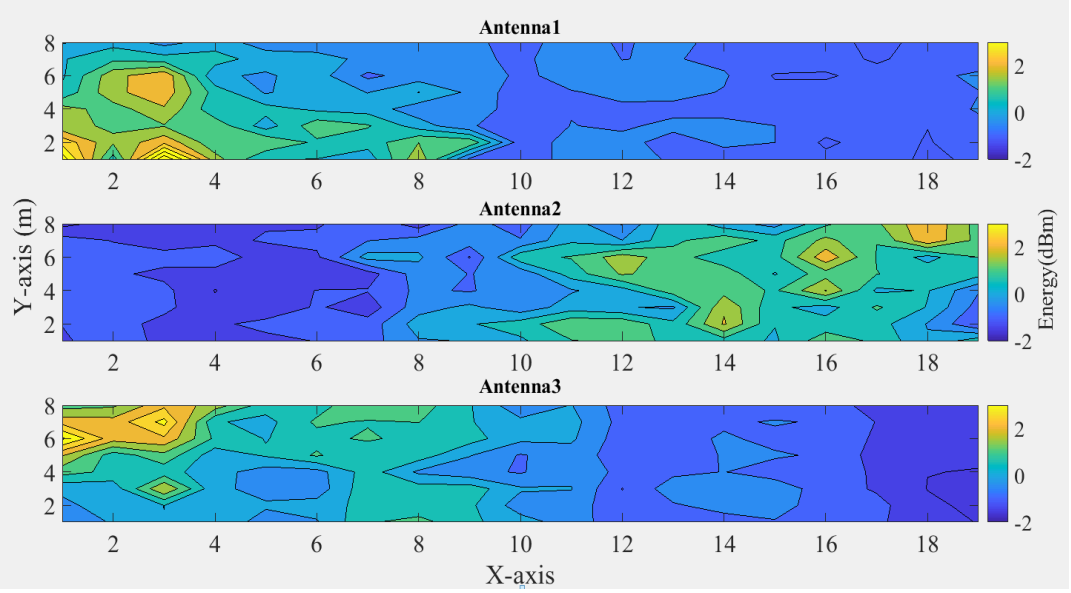
(a) variance



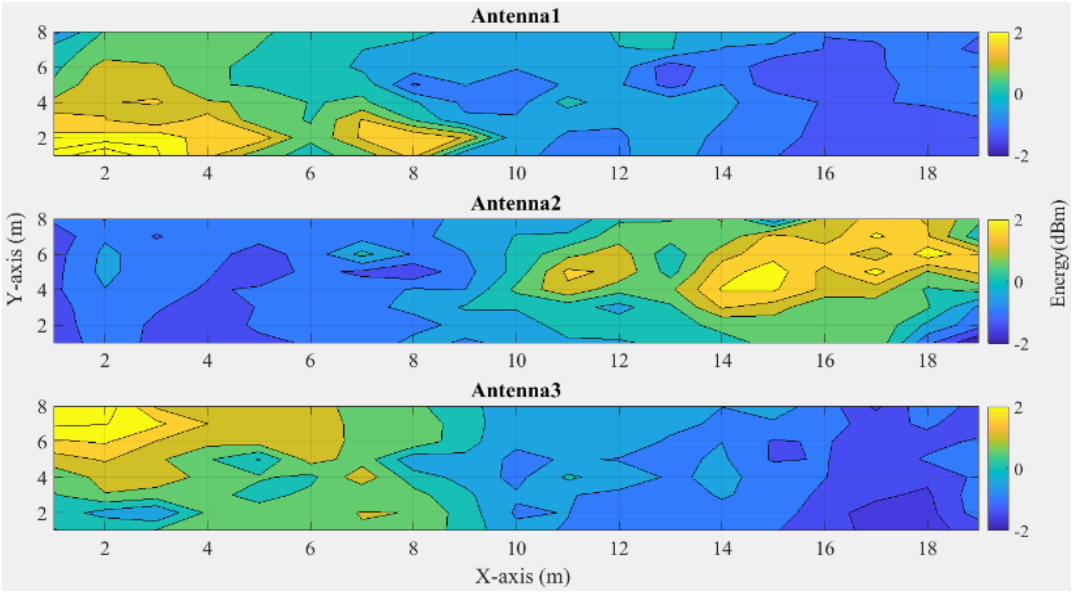
(b) peak to peak values



(c) RMS



(d) impulse factor



(e) area under waveform

Figure 6.3: Spatial map of PD statistical features

peak value, impulse factor, and area under waveform) measured at each of the 3 receiving sensors vary with location. Unsurprisingly, the values measured closer to a PD source are stronger than the ones received from a distance farther away. The effect of multipath and signal distortions add to the unique signatures created at different locations. The spatial map shows a clear correlation between the features and distinctive PD locations. This unique spatial pattern is an indication that the derived features are informative and can be used to infer PD location by matching the patterns. Figure 6.4 represent the architecture of the pattern matching technique described for PD localisation.

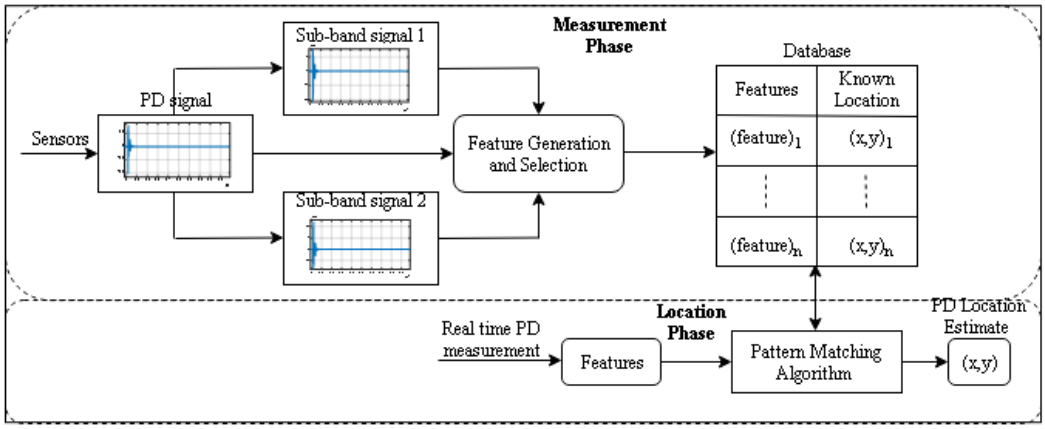


Figure 6.4: Block diagram of PD localisation.

6.3 Correlation based feature selection

Following the method introduced in this chapter for generating PD features using different frequency components, each of the observed PD radio signal captured by a sensor produces 15 features (that is 5 features for each of 3 bands) making it a total of 45 features for the three sensors used in the measurement campaign described in chapter 3. There is a trade-off between the number of features used as inputs to a machine learning algorithm and performance. A delicate balance must be achieved between too few features (insufficient information leading to under fitting) and too many features (curse of dimensionality). Moreover, different features may correlate: two features may be highly correlated with the output, but also with each other, hence providing little or negligible information. Accordingly, we seek to transform our original feature set to a potentially smaller set of decorrelated features using the Correlation based Feature Selection (CFS) algorithm [144][145] [146]. CFS is based on correlation which is a similarity/interdependence measure between two random variables. Two variables are said to be linearly dependent when their correlation coefficient is 1, and uncorrelated when the correlation coefficient is 0. Generally, in CFS, the mutual correlations of all feature pairs are evaluated and the

feature with the highest average absolute mutual correlation is removed at each iteration step of the algorithm [145]. However, in this thesis, we develop a new feature selection criterion for the CFS algorithm. Instead of removing features with highest average correlation, at each iteration step of the selection algorithm, we first pick features with higher correlation and then among them discard the feature with the highest average mutual correlation. For example, if two features have correlation of 1 between them, we discard the feature with the highest average correlation and keep the other. When a feature is removed from the feature set, it is also discarded from the remaining average correlation. This continue until average absolute mutual correlation of all remaining feature is less than 0.4. The resulting feature matrix from the CFS algorithm is the optimal lower dimensional subset from the original feature set. This method is applied to the PD feature data for feature selection. The result brings the total number of features from the three antennas used for localisation from 45 to 9. The 9 best features selected by the algorithm are the area under curve of sub-band 1, sub-band 2 and the broadband signal on each of three antennas (i.e, each antenna produces 3 features in the 2 sub-band and the main signal for a particular signal).

6.4 Performance Evaluation

In order to verify the use of time domain statistical features to infer PD location and the effectiveness of CFS in reducing PD feature dimensionality whilst maintaining an acceptable level of accuracy, all the PD features extracted from each of the 144*3 =432 combinations of PD location and sensors as described in chapter 3 are used to train a machine learning algorithm in order to build PD localisation model. During the training phase, PD data from reference locations (known locations) are used and the algorithm is fed two inputs; the mean values of the PD features and their corresponding location. Each (feature, location) pair constitutes a training data point. After applying the CFS technique, the selected features are used in the input. PD data collected from the 32 distinct

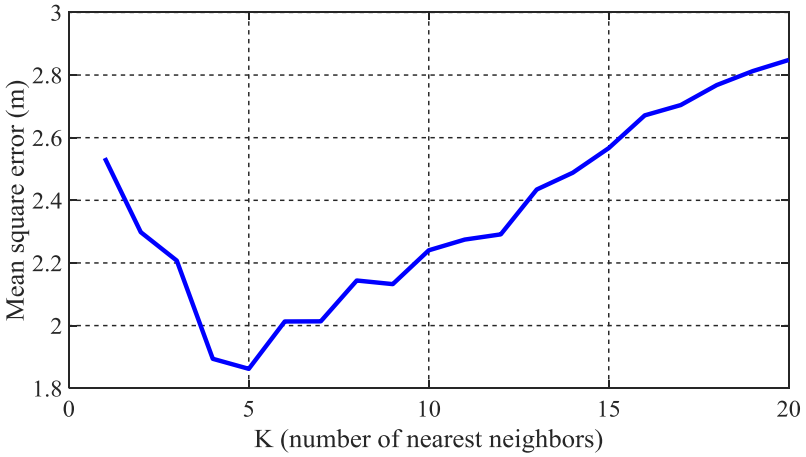
locations otherwise known as test locations is used for validation. In this section, we revisit two of the developed PD localisation models discussed in the chapter 4 namely; weighted k-nearest neighbour (WKNN) and generalised regression neural network (GRNN) for evaluation. Any other pattern matching algorithm could be used; however, WKNN and GRNN are chosen for their performance and ease of implementation. These algorithms (WKNN and GRNN) can be optimised before applying the generated features. This is achieved by tuning the hyper-parameters.

6.4.1 Parameter tuning

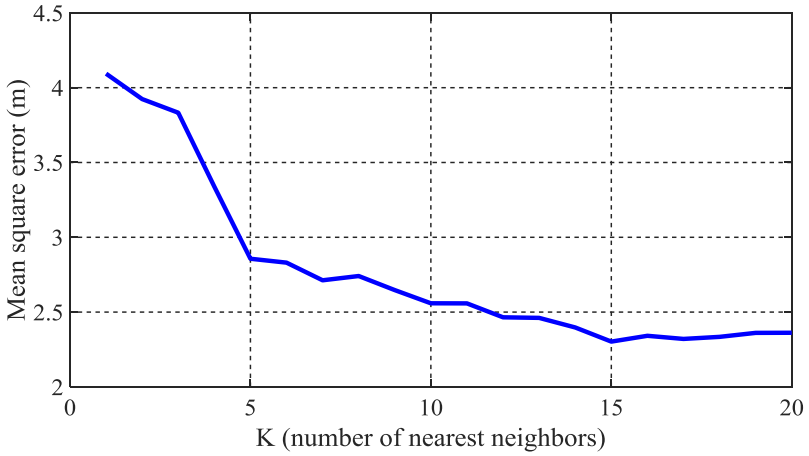
6.4.1.1 The effect and choice of WKNN model parameters

The performance of WKNN algorithm is usually affected by the value of the parameter k and the weight function used. The choice of k is very critical. It indicates how many neighbors influence the localisation process. However, it is known that a small value of k places limit to the area covered in a given prediction and provides a model with low bias and high variance. On the other hand, a large value of k covers more area in each prediction thereby producing a model with lower variance but increased bias. In some instances larger k increases the computational cost without corresponding increase in accuracy. There is a trade-off between bias and variance of any given KNN model. Therefore, for optimal performance, we need to choose the best value of k. In this work, an exhaustive search method is employed. Figures 6.5 (a) and (b) represent the variations in the mean error distances with 20 different values of k for PD data before and after CFS respectively. It can be observed that as k increases, the mean error distances decrease to an absolute minima and then increase again. When the value of k increases initially, the underlying variations of signal parameters at a single neighbouring point is averaged out with increasing k, resulting in decrease of error distance. The errors in signal parameter vectors at different neighbours are assumed to follow different patterns. Hence averaging

over more neighbours may yield more accurate results. However, for larger values of k, the errors again increase as many grid points which are closer in signal space but much farther in physical space are at this point included in the computation of the final result thereby leading to greater deviations from the actual location. Optimal values of k (the value of k with the lowest average error) are given as 5 and 15 before and after CFS respectively. These values are used in subsequent analysis. The weight function has discussed in chapter 4 is used here (that is inverse of the euclidean distance in signal space).



(a) WKNN-before CFS

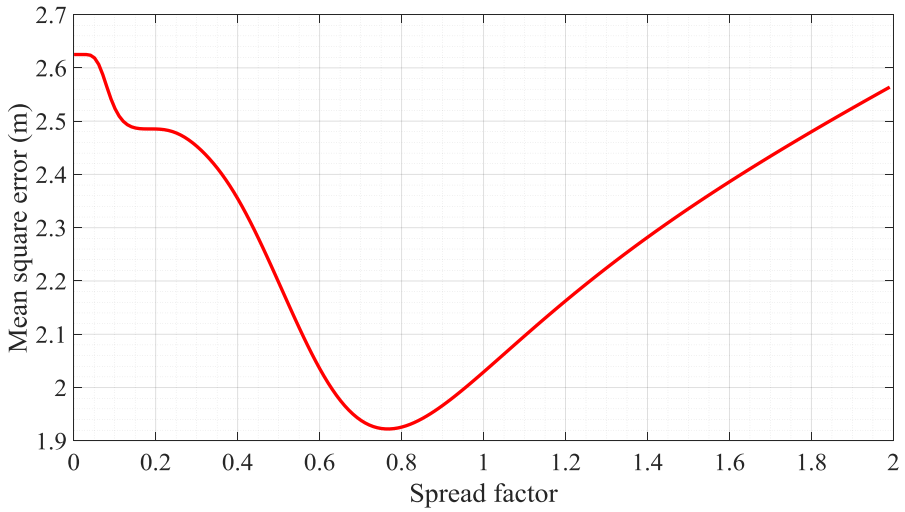


(b) WKNN-after CFS

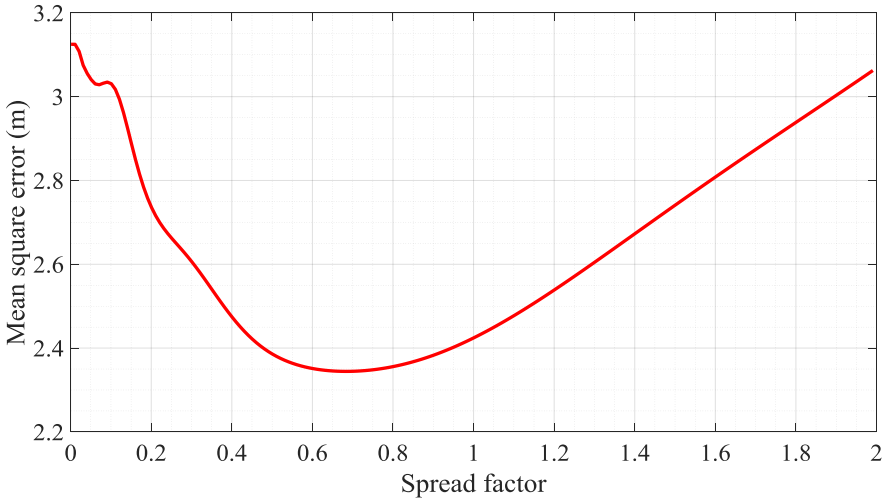
Figure 6.5: The effect of k on WKNN performance before and after CfS.

6.4.1.2 The effect of spread constant on GRNN performance

In GRNN model, the choice of an appropriate spread factors σ to be applied to each of the input data sets before and after the application of CFS is critical. The spread factor affects the degree of generalisation of the network and hence the performance of GRNN model. In this study, an exhaustive search method is used on a range of spread factors to determine the optimum spread factors in the GRNN models. For each input combination, the optimum spread for the GRNN model was determined according to the mean squared error (mse) criterion. Figures 6.6 (a) and (b) show the variation of mse statistic vs spread factor for the input before CFS and after CFS respectively. This implies that GRNN before CFS where we have 45 features has a minimum mse with the spread value of approximately 0.77 and GRNN after CFS has a minimum mse with spread factor of approximately 0.69 . Therefore, the optimum spread factors determined for PD input data before and after CFS are 0.77 and 0.69 respectively. These values are used to evaluate the performance of the GRNN models for each data set (before and after CFS).



(a) GRNN-before CFS



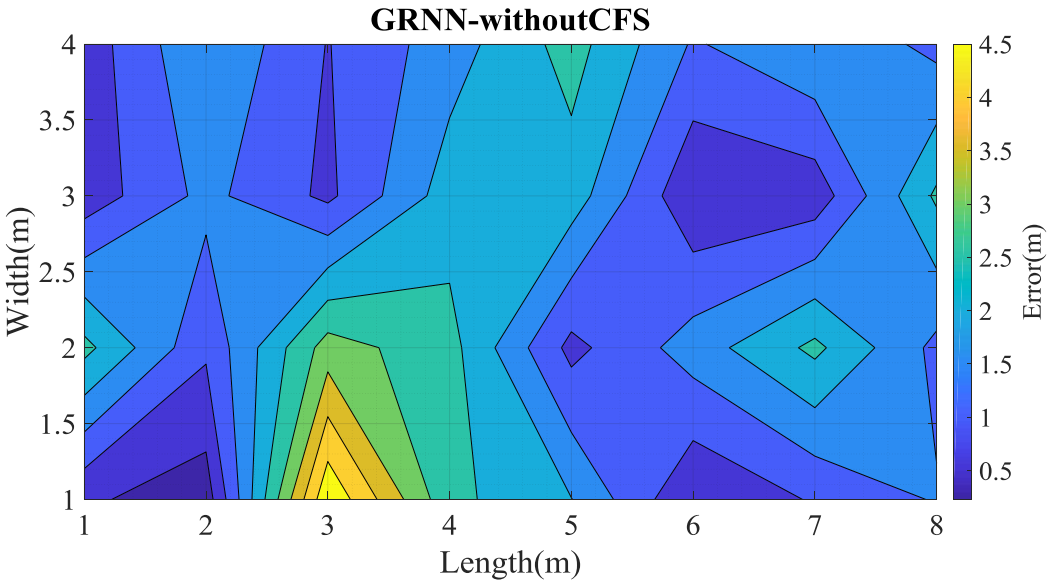
(b) GRNN-after CFS

Figure 6.6: The effect of spread factor on GRNN performance before and after CFS.

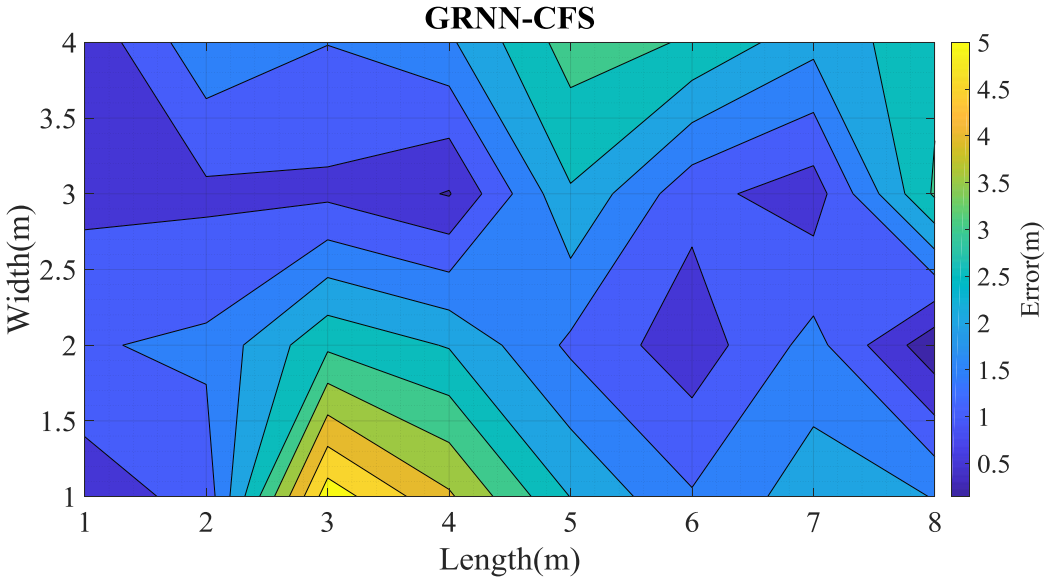
6.4.2 Impact of the PD features on accuracy

The performance of each algorithm is evaluated before and after the application of CFS to the PD features. We have 45 PD features before the application of CFS and 9 features after applying CFS. To train the PD location algorithms, PD features from training data

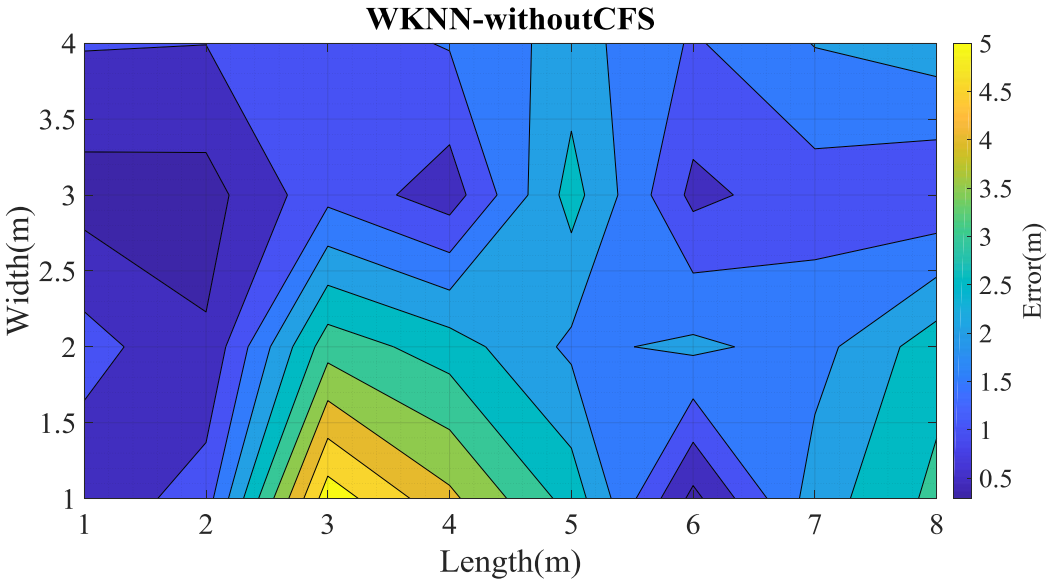
set and their corresponding ground truth locations are given as inputs to each algorithm. For validation, the PD data (i.e PD features) from the 32 test locations is presented to the developed models (GRNN and WKNN) to predict each test location. The performance is evaluated in terms of distance error defined as the Euclidean distance between the estimated location and the true PD location.



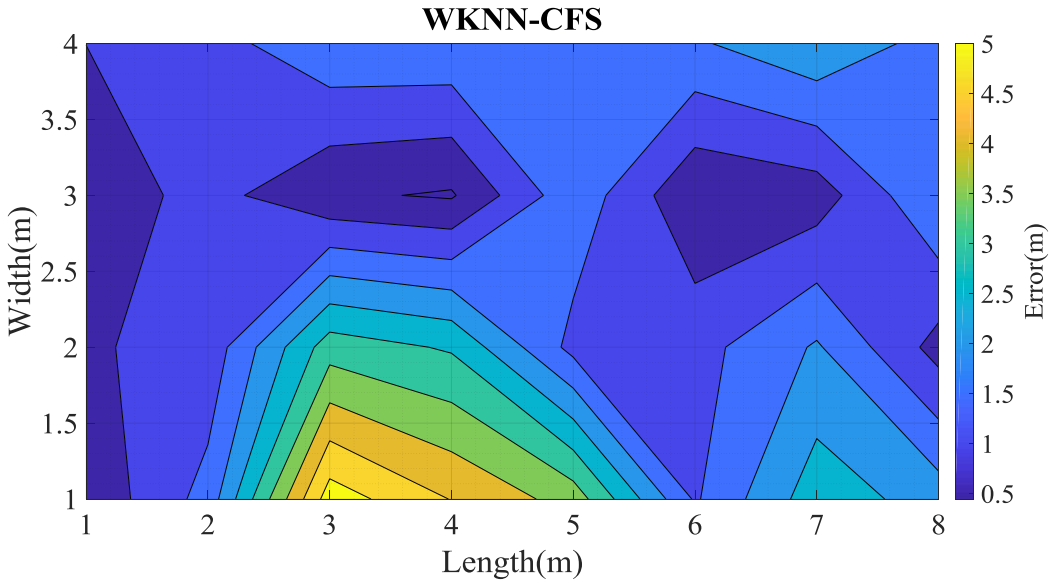
(a) GRNN-before CFS



(b) GRNN-after CFS



(c) WKNN-before CFS



(d) WKNN-after CFS

Figure 6.7: Location error map for GRNN and WKNN models before and after CFS

Figure 6.7 shows the heat-maps of point accuracy for each of the models. Here, the areas covered in blue refer to areas where PD is located with good accuracy (location errors of $2m$ or less), whereas areas covered in green, orange and finally yellow show corresponding decrease in accuracy. A yellow zone corresponds to location error of $4m$ and above. Based on Figure 6.7, it is clear that the localisation results before CFS (45

features) and after CFS (9 features) are comparable irrespective of the localisation model used. Again, results show that the accuracy on one part of the laboratory is significantly higher than the accuracy in other parts, and is consistent regardless of the technique used.

6.4.3 Evaluating localisation accuracy

Figure 6.8 shows the localisation performance of the algorithm before and after CFS. The first thing to note is that there is an overall improvement in the location accuracy of each algorithm compared to the results obtained in chapters 4 and 5 where only received signal strength was used as input features to the algorithms. More specifically, the mean error for WKNN and GRNN before CFS is 1.80m and 1.70m respectively. Also the application of CFS for dimensionality reduction brings about a reduction in computational cost whilst maintaining reasonable level of accuracy with the mean error of 1.82m and 1.86m for WKNN and GRNN after CFS respectively.

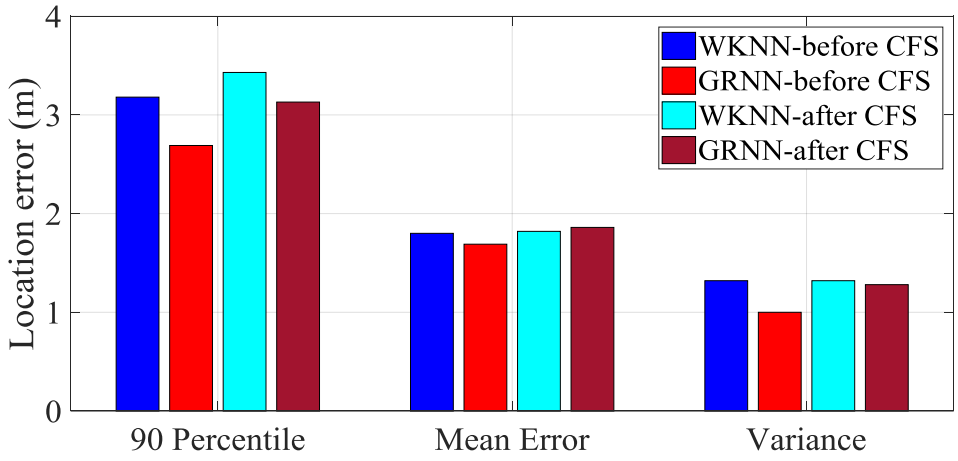


Figure 6.8: Model localisation accuracy.

The summary of results in terms of 0.25, 0.50, 0.75, and 0.90 overall cumulative error probabilities (CEP) before and after CFS is given in Table 6.2. This clearly shows that

Table 6.2: The performance statistics of the models before and after CFS.

Model	CEP=0.25(m)	CEP=0.50(m)	CEP=0.75(m)	CEP=0.90(m)
WKNN-Before CFS	1.02	1.49	2.48	3.18
GRNN-Before CFS	0.87	1.59	2.43	2.69
WKNN-After CFS	0.98	1.74	2.17	3.43
GRNN-After CFS	0.89	1.64	2.47	3.13

half of the time PD sources were located with error less than 1.75m for each algorithm before and after CFS. When WKNN is used after CFS, 75% of the PD sources were located with distance error of 2.17m or less. And 90% of the sources were located with distance error of 2.69m or less when GRNN is used before CFS. The results indicate that the use of numerous PD statistical features (leveraging on the frequency selective nature of PD received signals) other than RSS improves the localisation accuracy of our models. However, features extracted directly from measured time domain PD signals may not be sufficient in situations where PDs of different type occur. In the next section, we will explore time-frequency domain technique for feature extraction.

6.5 Wavelet Analysis of Partial discharge signals

Another method considered in engineering PD features for source localisation is to analyse these signals in the frequency domain. In the frequency domain, PDs of different types can be effectively characterised for localization. Most of the frequency domain methods used in PD signal analysis and localization are based on the well-known Fast Fourier transform (FFT) techniques [147][148]. However, PD signals are stochastic and often demonstrate a non-stationary and transient nature, carrying small yet informative components embedded in larger repetitive signals [149]. This limits the application of FFT based techniques for PD localization. To this end, we propose to use Wavelet Transform [84] to decompose RF PD signals into different frequency bands and extract localised PD location dependent parameters for robust PD radio-location.

The application of wavelet transformation in PD signal analysis is a well-known technique which overcomes the problems of other signal processing techniques such as the classical Fourier Transform which can describe the frequency components contained within a complex signal, but cannot indicate where in time those frequencies reside. Although the Short-Time Fourier Transform (STFT) [150][151] which divides a signal into short segments of equal length and computes the FT on each segment gives an indication of time of impact from a signal, it has a fixed resolution. This implies that for a wide window, STFT gives better frequency resolution but poor time resolution and for a narrow window, it gives good time resolution but poor frequency resolution. A multiresolution wavelet-based approach which gives good time resolution for high-frequency and good frequency resolution for low-frequency permits efficient isolation and extraction of localised signal features.

6.5.1 Wavelet transform theory

Wavelets are mathematical functions that are used to transform data into time-frequency components. Wavelets break up data and analyse them in different scales or resolutions. With a wide-window, a wavelet examines prominent features and with a small window, a wavelet brings out the small local features. In wavelet analysis, a wavelet function known as mother wavelet is usually adopted. A contracted, high frequency mother wavelet is used for temporal analysis, while a dilated, low-frequency mother wavelet is used for frequency analysis. In other words, the wavelet transform splits a signal into shifted (translated) and scaled versions of a mother wavelet function. The wavelet function $\psi_{a,b}(t)$ is a family of long-duration low-frequency and short-duration high frequency functions defined in Equation 6.1;

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}),\ \ a > 0,\ \ b \in Re \tag{6.1}$$

where a is the scale or dilation parameter, and b the shift or translation operator. The scale parameter is related to the reciprocal of frequency and the translation parame-

ter stands for time. Consequently, there is usually a non-uniform contraction/expansion over the time-frequency plane. The value of the scale parameter controls this contraction/expansion process. Depending on its value, the wavelet function expands or contracts in the time domain triggering a corresponding expansion or contraction in the frequency domain. When a is large ($a > 1$), the basis function becomes a stretched version of the mother wavelet ($a = 1$) and demonstrates a low-frequency characteristic. When a is small ($a < 1$), this basis function is a contracted version of the mother wavelet function and demonstrates a high-frequency characteristic. The continuous wavelet transform (CWT) of a time domain signal $s(t)$ is given in Equation 6.2;

$$CWTa, b = \frac{1}{\sqrt{a}} \int_{-\inf}^{+\inf} \psi(\frac{t-b}{a}).s(t).dt \tag{6.2}$$

The result of this integral is called wavelet coefficients. Each coefficient represents the similarity between the portion of the signal examined and the scaled and translated mother wavelet.

The wavelet transform described, takes care of the deficiencies inherent in Fourier analysis. The low-frequency component of the signal is analysed using a long-duration function while the high frequency component is analysed with a short-duration function. Thus, through the wavelet transform, any finite-energy signal can be mapped from the time-domain to the wavelet-domain, which in this case is a two-dimensional distribution over time and frequency. The wavelet coefficients $CWT(a, b)$ are calculated over a range of a and b, that is the wavelet transform breaking the signal down into component parts in the wavelet-domain. However, in practice, the choice of continuous a and b for calculating wavelet coefficients in CWT causes computational complexity and generates a lot of redundant data. In order to overcome this drawback, an efficient solution is to discretise the parameters a and b, leading to discrete wavelet transform.

6.5.2 Discrete Wavelet Transform

Discrete Wavelet transform (DWT), which is one of the popular methods in the wavelet transformation family has been widely used in PD signal processing, classification and de-noising [84] [152] [153][154][155]. The DWT is derived from the CWT by using the discrete forms of the scale parameter a and the translation parameter b, which gives only a subset of the scale and time shifts. Generally, a is replaced by 2^j and b is replaced by $k2^j$, where k and j are integer values, which means that a and b are based on powers of two, consequently reducing computational cost and eliminating redundancy with sufficient accuracy. Therefore, the modified wavelet function is as given in Equation 6.3;

$$\psi_{j,k}(n) = 2^{\frac{j}{2}}\phi(2^j n - k), j, k \in Z \tag{6.3}$$

which is an orthonormal basis. Z is a set of integers. Although it is called discrete wavelet, it normally is a (piecewise) continuous function. The effect of discretising the wavelet is that the time-scale space is now sampled at discrete intervals. In addition to the wavelet function, DWT introduces a new function, known as the scaling function. The wavelet function determines the detail coefficients while the scaling function determines the approximation coefficients. The scaling function $\phi(t)$ is used to construct orthogonal wavelets $\psi(t)$ using the dilation equations as shown in Equations 6.4 and 6.5;

$$\phi(\frac{n}{2}) = \sqrt{2}\sum_k h(k)\phi(n-k) \tag{6.4}$$

$$\psi(\frac{n}{2}) = \sqrt{2}\sum_k g(k)\phi(n-k) \tag{6.5}$$

These orthogonal wavelets could be thought of as a system of cascaded filters that decompose the signal. In general, a signal is passed through a filter bank consisting of a high (the wavelet function) and a low-pass filter (scaling function). The $h(k)$ and the $g(k)$ represent the coefficient of the low-pass filter and the high-pass filter respectively [156].

This filtering procedure is equivalent to the two-channel subband coding that uses Quadrature Mirror Filters (QMFs) [157]. This implies that by multiresolution, all we need to do to transform a signal into wavelets, is to pass the signal through cascaded system of filters. Rather than taking the scalar product of the scaling function or the wavelet with the signal, we convolve the signal with these filters. Figure 6.9 shows the filtering process, which demonstrates the discrete wavelet decomposition procedure.

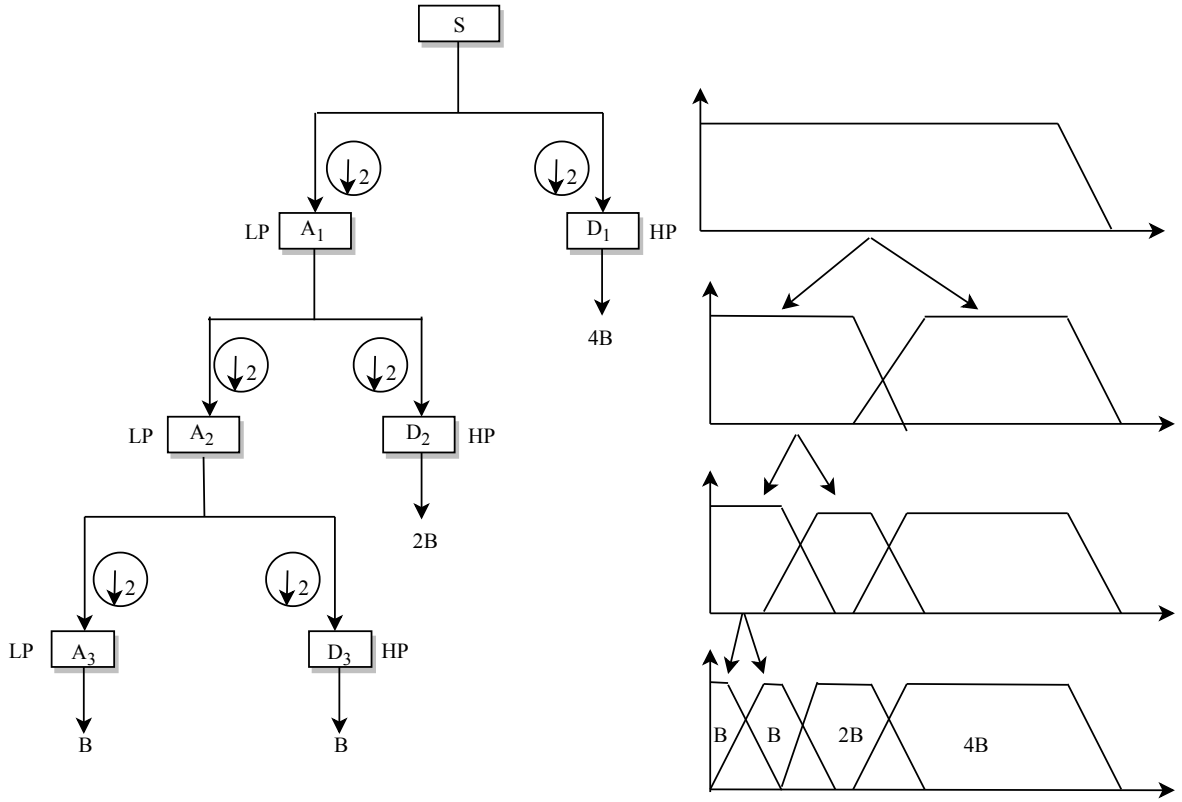


Figure 6.9: DWT filter decomposition.

DWT is therefore a linear operator that decomposes the PD signal via a cascaded system of low and high pass filters into two components: approximation coefficients (As), which are the low frequency, high scale information of the PD signal and detail coefficients (Ds), which capture the high frequency, low scale information in the signal. While the Ds remain unchanged, the As are further decomposed into new Ds and As. This process continues until the final decomposition level. Owing to its time and frequency localization ability, DWT can reveal the local characteristics of the input PD signals. In addition, the

multi-level decomposition of PD signal into different scales by DWT generates multi-scale features, each of which represents particular characteristics of PD signal. However, since DWT decomposes only As in each level, it is difficult to extract distinctive information from Ds. The DWT is also time-variant and often divides the frequency spectrum into non-uniform bands. In the next section, wavelet technique that overcome these drawbacks is considered.

6.5.3 Wavelet Packet Transform

One possible drawback of DWT is that it often fails to sufficiently capture high frequency information in signals and this can be a challenge in PD analysis where important information is located in high frequency component. This leads us to the wavelet packet transform [158] [159] [160] [161], a generalisation of the DWT form by linear combination of wavelet functions providing more possibilities for signal processing. WPT can be said to be a tree of subspaces, providing level by level transformation of a signal from the time domain to the frequency domain, with the root node (top or first level) of the tree $\omega_{0,0}$ representing of the original time domain PD signal. The tree is typically a binary tree, where node $\omega_{j,k}$, with j representing the scale and k representing the subband index within the scale, is decomposed into an approximation (left sub-node) space $\omega_{j,k} \rightarrow \omega_{j+1,2k}$ [162] and a detail (right sub-node) $\omega_{j,k} \rightarrow \omega_{j+1,2k+1}$. This is done by splitting the orthogonal basis $\phi_j(n-2^j k)_{k\in Z}$ of $\omega_{j,k}$ into two new orthogonal bases $\phi_j(n-2^{j+1}k)_{k\in Z}$ of $\omega_{j+1,2k}$ and $\psi_j(n-2^{j+1}k)_{k\in Z}$ of $\omega_{j+1,2k+1}$ [163], where $\phi_{j,k}(n)$ and $\psi_{j,k}(n)$ are the scaling and wavelet functions respectively that are given in Equations 6.6 and 6.7 [163];

$$\phi_{j,k}(n) = \frac{1}{\sqrt{2^j}}\phi(\frac{n-2^j k}{2^j}) \tag{6.6}$$

$$\psi_{j,k}(n) = \frac{1}{\sqrt{2^j}}\psi(\frac{n-2^j k}{2^j}) \tag{6.7}$$

where 2^j, the dilation or scaling parameter determines the level of contraction or scaling and $2^j k$, the position or translation parameter determines the time location of the wavelet. It is important to note that at every subsequent level of decomposition, there is a decrease in temporal resolution and a corresponding increase in frequency resolution. In other words, unlike DWT, WPT decomposes both the ACs and the DCs simultaneously at every level. WPT represents high frequency information in a better manner as compared to DWT. WPT has the same frequency bandwidths in each resolution. This property enables WPT to preserve the information in the 'original' PD signals, resulting in robust features. For n levels of decomposition the wavelet packet decomposition produces 2^n different sets of coefficients (or nodes). For example, a three level WPT decomposition tree shown in Figure 6.10 produces a total of 8 sub bands with each sub band covering one-eighth of the signal frequency spectrum. The top or first level of the wavelet packet transform is the time representation of the signal where as bottom level has better frequency resolution which also provides more features about a signal as compared to DWT. WPT has been used in [164] to decompose PD signals into multiple scales, and extracted PD features used for PD classification. Though, the WPT decomposes both the ACs and DCs at specific decomposition level dividing the frequency spectrum into uniform bands, it is still time-variant. To overcome this problem of time-variant, a new version of WPT otherwise known as maximal overlap discrete wavelet packet transform (MODWPT)[165][166][167] is introduced. Here, a signal is decomposed into both ACs and DCs at all levels considered through filter banks, presenting uniform frequency bands without down-sampling making it a time-invariant transform. That is to say that at each level of decomposition, the output has the same number of samples of the input signal. As with the WPT, the MODWPT is also an energy preserving transform. It partitions the energy among the wavelet packets at each level such that the sum of energy over all the packets equals the total energy of the input signal. MODWPT is therefore adopted in this thesis and will be use in subsequent analysis to extract localised PD features for locating PD sources.

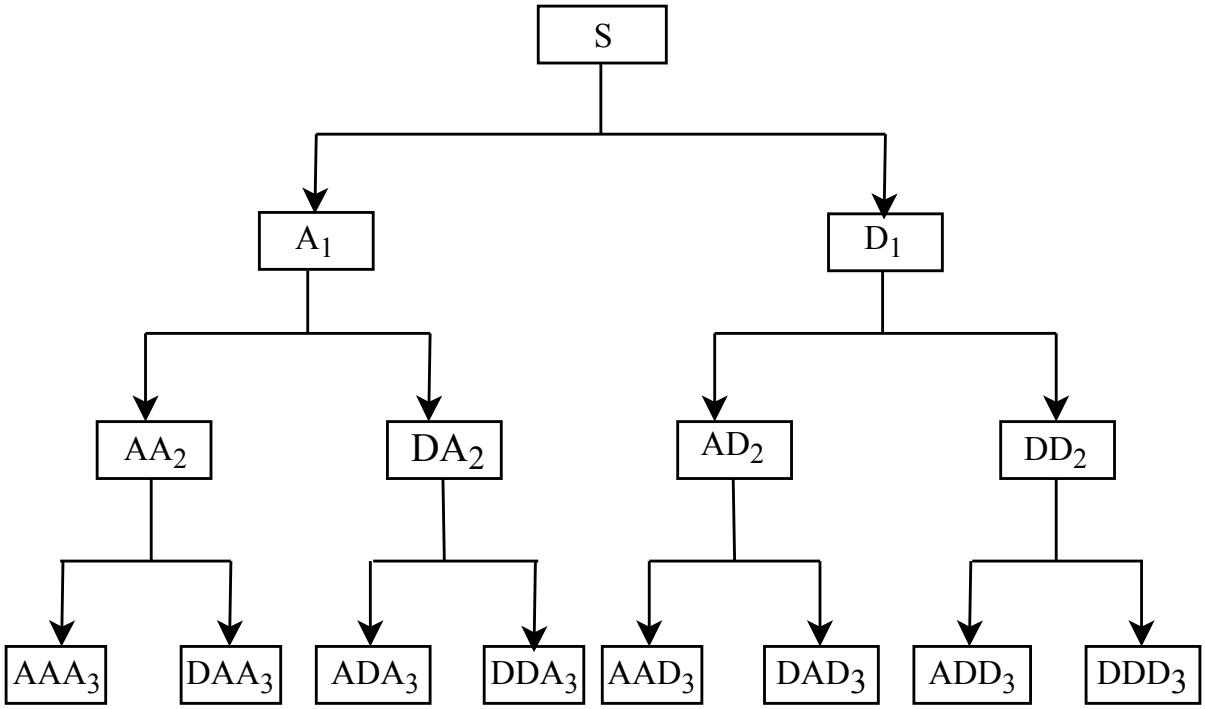


Figure 6.10: WPT decomposition Tree.

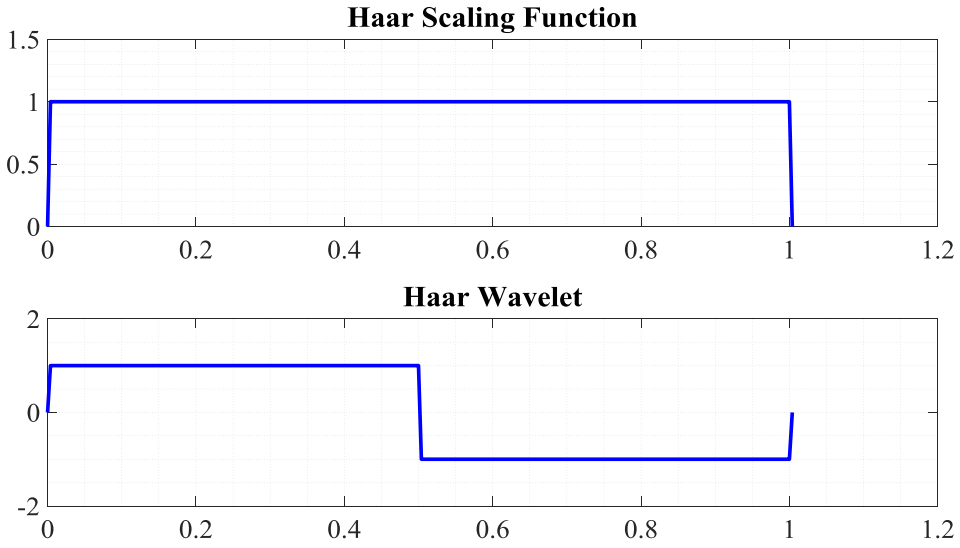
6.6 PD features extraction using MODWPT

The non-decimated version of WPT known as Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) is introduced as an alternative method of extracting the desired location dependent features from PD signals. First, the PD data is decomposed via the MODWPT to extract the time-frequency-dependent information. Thus sequenced-ordered MODWPT coefficients for the MODWPT nodes at each level of decomposition are obtained. These MODWPT coefficients are used to construct features that best represent the PD signal locations. The standard choice for depth of decomposition level j is specified as a positive integer: 4 or floor (log2 (length (signal))), whichever is smaller [168]. In our case, the transform level is set to 4 since the signal is long. The 4-level WPT decomposition generates a tree with 16 terminal nodes, corresponding to the frequency sub-bands. For analysis, it is also important to select the mother wavelet appropriate for the type of signal to be transformed since the type of wavelet chosen is crucial to the

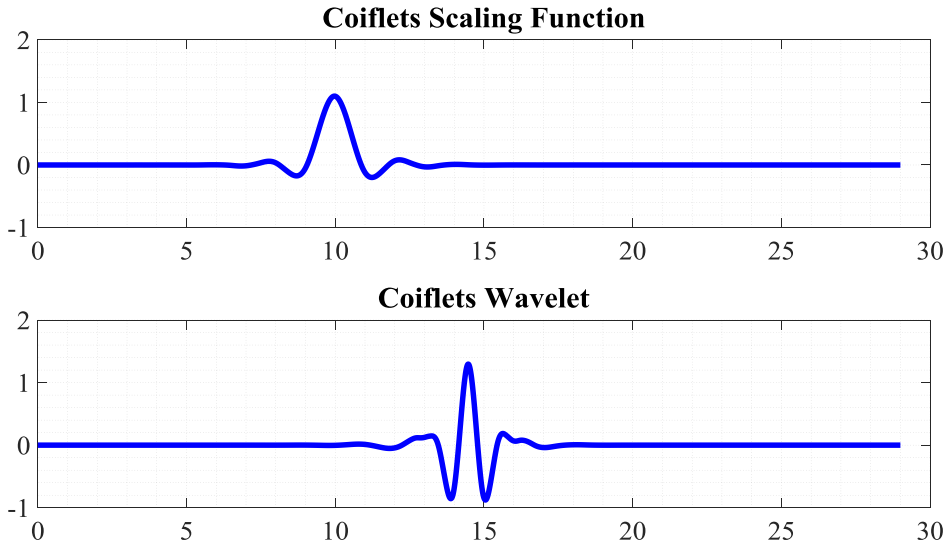
characterisation of the signals. Since we are interested in preserving the signal energy after transformation, our choice of wavelet is limited to orthogonal wavelets that are known for preserving energy. The orthogonal wavelets considered are shown in Table6.3.

Table 6.3: Energy preserving wavelets.

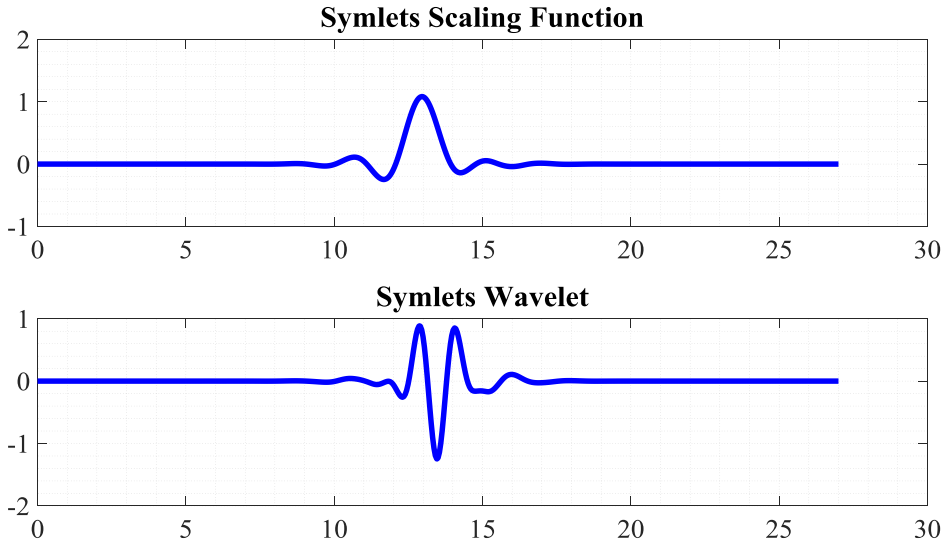
Orthogonal Wavelet	Features
Haar	Symmetric; special case of Daubechies; useful for edge detection
Coiflet	Scaling function and wavelets have the same number of vanishing moments
Symlet	Least asymmetric; nearly linear phase.
Daubechies	Nonlinear phase; energy concentrated near the start of their support
Fejer-Korovkin	Filters constructed to minimise the difference between a valid scaling filter and the ideal sinc lowpass filter; are especially useful in discrete wavelet packet transforms.



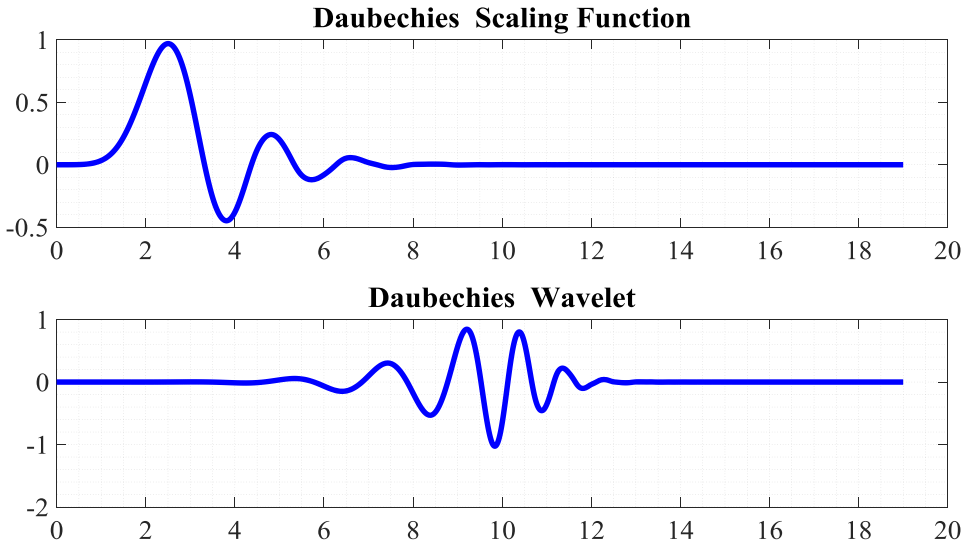
(a) Haar



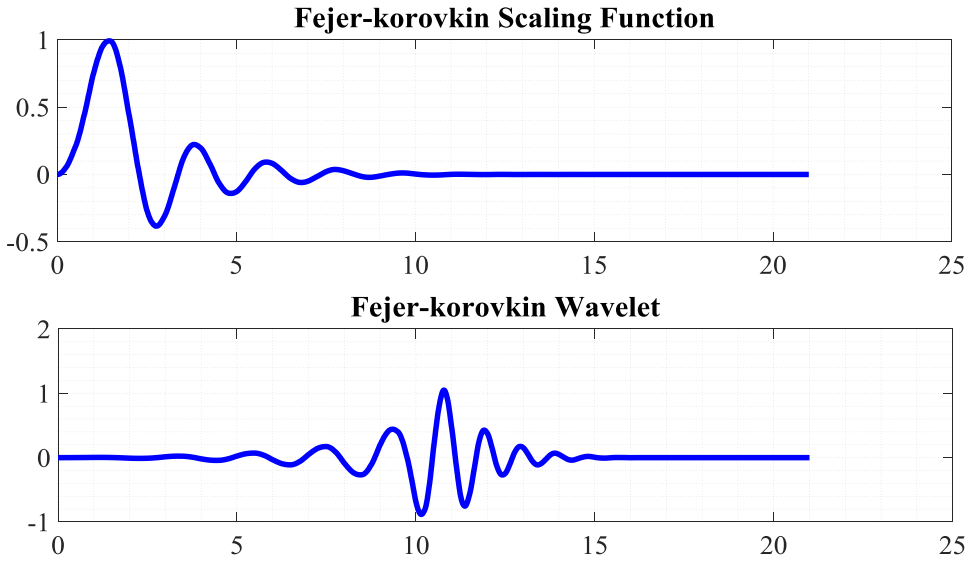
(b) Coiflets



(c) Symlets



(d) Daubechies



(e) Fejer-Korovin

Figure 6.11: Different orthogonal wavelets and their scaling functions

PD signals are characterised as decaying, oscillatory and of relatively short duration. Moreover, the wavelet analysis is a measure of similarity between the basis functions (wavelets) and the signal itself. Consequently, both a similar shape in the time-domain and similarity in the sense of frequency content are required. Figure 6.11 shows the different shapes of the orthogonal mother wavelets. Among these, frequency-localised filter; Fejer-Korovkin wavelet has been shown to be a good approximation of a decaying, oscillatory and short duration signal. In [169], the energy parameters extracted from impulsive signals using Fejer-Korovin filter outperformed other wavelet filters such as Daubechies, Symlet and Coiflet in terms of classsification accuracy even though they all are Finite Impulse Response (FIR) filters with nonlinear phase and a frequency response tending to be ideal as the support increases. The salient point is that the Fejer-Korovin wavelet is more symmetric than comparable wavelet filters [170]. Therefore, the Fejer-Korovin wavelet is adopted for use in this thesis.

The filtering operations in MODWPT, are applied to both the approximation and detail coefficients. The result is that wavelet packets provide a sub-band filtering of the input

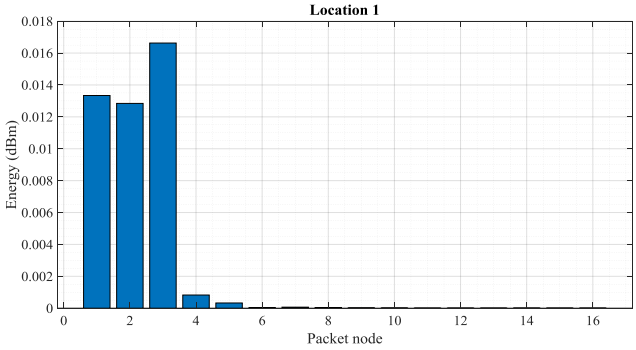
signal into progressively finer equal-width intervals. At each level j, the frequency axis is divided into 2^j sub-bands. The sub-bands in Hertz at level j are approximated as given in Equation 6.8;

$$SB_n = [\frac{nf_s}{2^{j+1}}, \frac{(n+1)f_s}{2^{j+1}}) \tag{6.8}$$

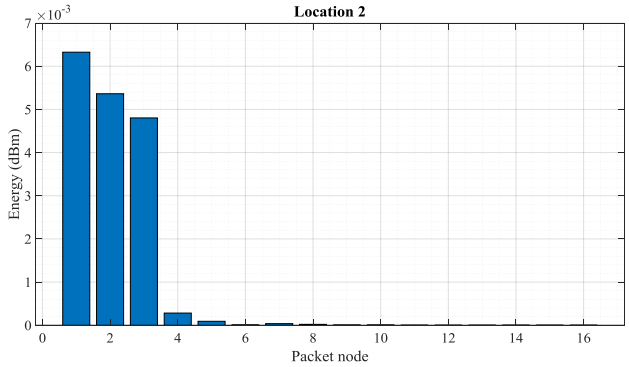
where SB_n is sub-band, $n = 0, 1, 2, ..., 2^j - 1$ and f_s is the sampling frequency. The wavelet packets have the added benefit of being an orthogonal transform, which means the energy in the signal is preserved and partitioned among the sub-bands as mentioned earlier. Therefore, the wavelet packet node energy which is the sum of square of wavelet packet transform coefficients, can represent the characteristics of PD signals. The wavelet energy of coefficients can be defined as in Equation 6.9;

$$e_{j,n} = \Sigma_k \omega_{j,k}^2 \tag{6.9}$$

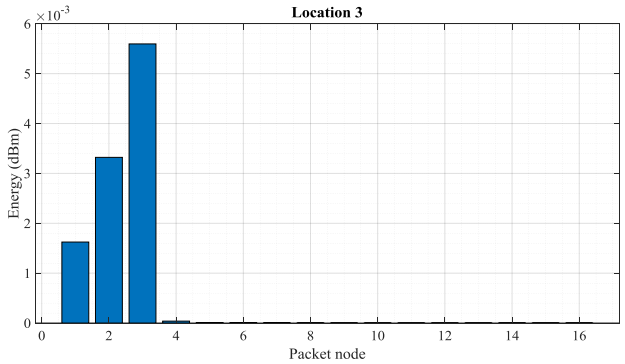
This measures the PD signal energy contained in each sub-band indexed by parameters j and n. Here, we refer to each (j, n) as a MODWPT node. The sum of the energy over all the packets nodes equals the total energy of the signal. Our interest is in the sub-band energy. As an illustration, Figure 6.12 depicts the energy profile computed based on all coefficients in each wavelet packet node for 8 PD signals captured from 8 different locations. Specifically, Figure 6.12a depicts the energy profile of PD pulse transmitted from location 1 ($x = 1$, $y = 1$) and Figure 6.12b shows the energy profile of PD pulse transmitted from another location 2 ($x = 1$, $y = 2$). The remaining figures are also energy profiles of PD pulses from different locations. By examining the packet nodes 1 to 3 (first three bars) in the energy profiles, it is evident that these are distinct signals from different locations. Thus, the wavelet packet node energy distribution represents an excellent set of features for PD localization because it is (i) time invariant, and (ii) location dependent. It is important to identify the frequency bands that contain most of the PD pulse energy.



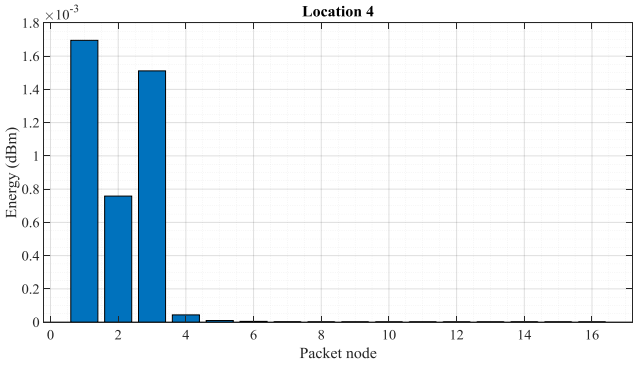
(a) PD signal energy profile from Location 1



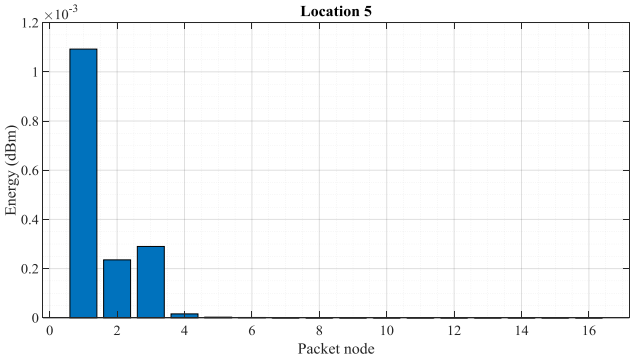
(b) PD signal energy profile from Location 2



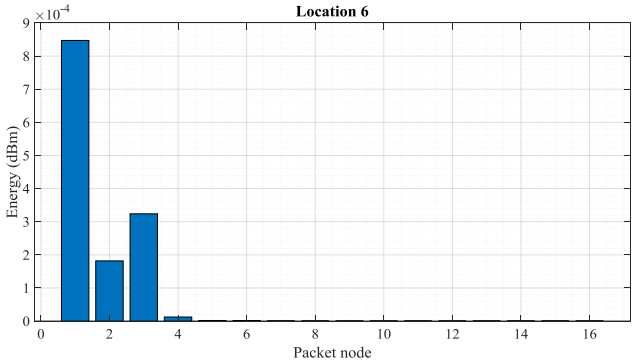
(c) PD signal energy profile from Location 3



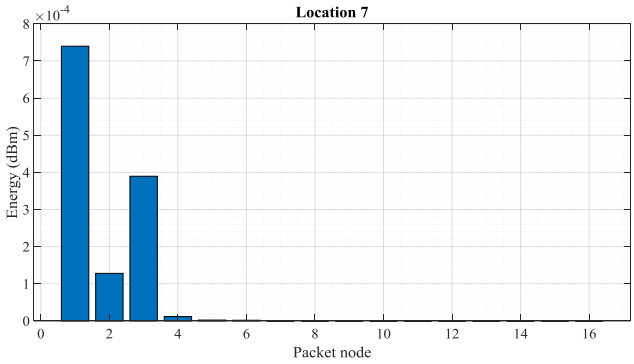
(d) PD signal energy profile from Location 4



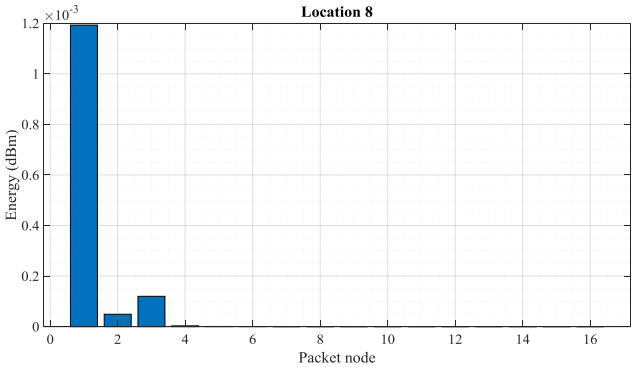
(e) PD signal energy profile from Location 5



(f) PD signal energy profile from Location 6



(g) PD signal energy profile from Location 7



(h) PD signal energy profile from Location 8

Figure 6.12: PD signals Energy distribution from different locations

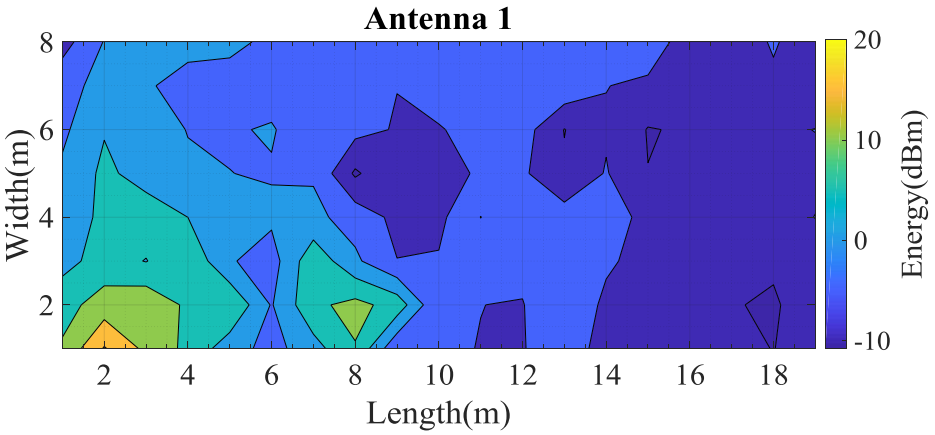
PD signals are known to have a very short duration (nanosecond), which implies a high frequency content. However, there are always dominant frequency bands that contain greater pulse energy in the entire spectrum. The energy profile by node shown in Figure 6.12 reveals the nodes (first 3 nodes) with higher percentage of pulse energy, indicating 3 frequency passbands of interest. The resulting passbands emphasize frequencies between 62.5 and 250 MHz, which contains large portions of the PD signal energy. The lower and upper frequencies of each wavelet packet frequency band are presented in Table 6.4. This demonstrates the robustness of the proposed technique. In this work, each node energy equivalent to energy in a frequency passband, represents an individual PD feature component and is used in the localization of PD.

Table 6.4: WPT-based extracted PD frequency bands

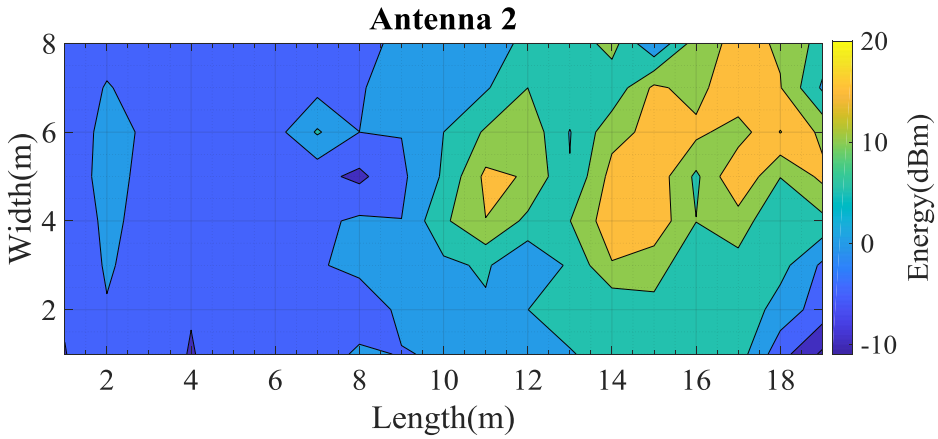
Frequency band	Lower frequency (MHz)	Upper frequency (MHz)
Band 1	62.5	125
Band 2	125	187.5
Band 3	187.5	250

This work is based on the assumption that all PD features extracted using the MODWPT provide meaningful information for inferring the PD location. To demonstrate that this is a reasonable assumption, each of the MODWPT-based PD energy maps for the three antennas in Figure 6.13(a-c) shows a distinctive variation in the energy content of the signals with respect to location in the measurement environment. They also reveal the complexity of the radio environment, which does not fit any well-known propagation model. The colour variation indicate the unpredictable decrease in energy as the source of PD moves away from the receiver in a cluttered environment. This helps in the creation of a training database for machine learning based PD localisation. The effect of multipath and signal distortions add to the unique signature created at different locations. This unique spatial pattern is an indication that the derived features are informative and can

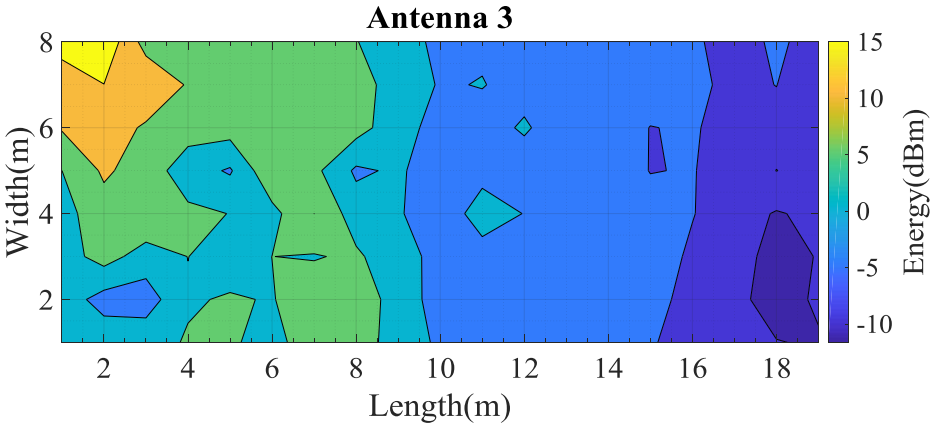
be used to infer PD location.



(a)



(b)



(c)

Figure 6.13: Spatial pattern of WPT-based PD energy

6.7 Machine Learning-Based PD Localisation Methods

6.7.1 Regression Tree-Based PD Localisation Method

The Regression Tree is one of the most simple and efficient machine learning algorithms used for predictions [171][172][173]. It is considered a variant of a binary decision tree model composed of linear regression functions at the leaf nodes. In this work, it is designed to learn and approximate the nonlinear function (PD locations/wavelet-based PD features) by piecewise linear functions at the leaf nodes. In other words, the nonlinear mapping of PD features to PD location coordinates is handled by splitting the problem into a set of smaller problems addressed with simple linear predictors. Figure 6.14 shows the architecture of a regression tree. Each node in the tree is designed to split the data so as to form clusters where accurate predictions can be performed with simple models. During training, a PD regression tree model is built top-down from the root node through binary recursive partitioning, which is an iterative process that splits the PD training data (PD features/location) into subsets that contain features with similar values. The parameters of the tree such as number of splits are optimised during training. A sum of squares reduction criterion is used in partitioning the data. The algorithm selects the split at each node that minimises the function shown in Equation 6.10;

$$SS = \Sigma_R((x,y)_i - (x,y)_R) + \Sigma_L((x,y)_i - (x,y)_L)^2 \tag{6.10}$$

Where $(x,y)_R$ and $(x,y)_L$ are the estimated values of PD location coordinates for the right and left nodes respectively. This process continues until each node reaches the terminal (leaf) node. The terminal nodes of the tree which represents a cell of the partition, store the models that approximate the best desired output. Suppose the training points at the

terminal node are $(r_i, (x, y)_i), (r_2, (x, y)_2), ...(r_n, (x, y)_n)$ where r_i represent PD features then the local model for terminal node is given as in Equation 6.11;

$$(\hat{x}, \hat{y}) = \frac{1}{n}\Sigma_{(i=1)}^{n}(x, y)_i \tag{6.11}$$

In the localization phase, the location coordinate of any PD sample (features) can be estimated by following through the branches of the obtained tree model.

Despite the simplicity that comes with the implementation of regression trees, the issue of overfitting affects its performance. Models developed using decision trees require very simple heuristics. They are composed of a set of if-else statements done in a specific order. When fully grown, it may lose some generalisation capability (with new data set that might have some bias or difference in spread of underlying features compared to training set). This leads us to explore ensemble methods that create variance amongst the data by sampling and replacing data while it test multiple models.

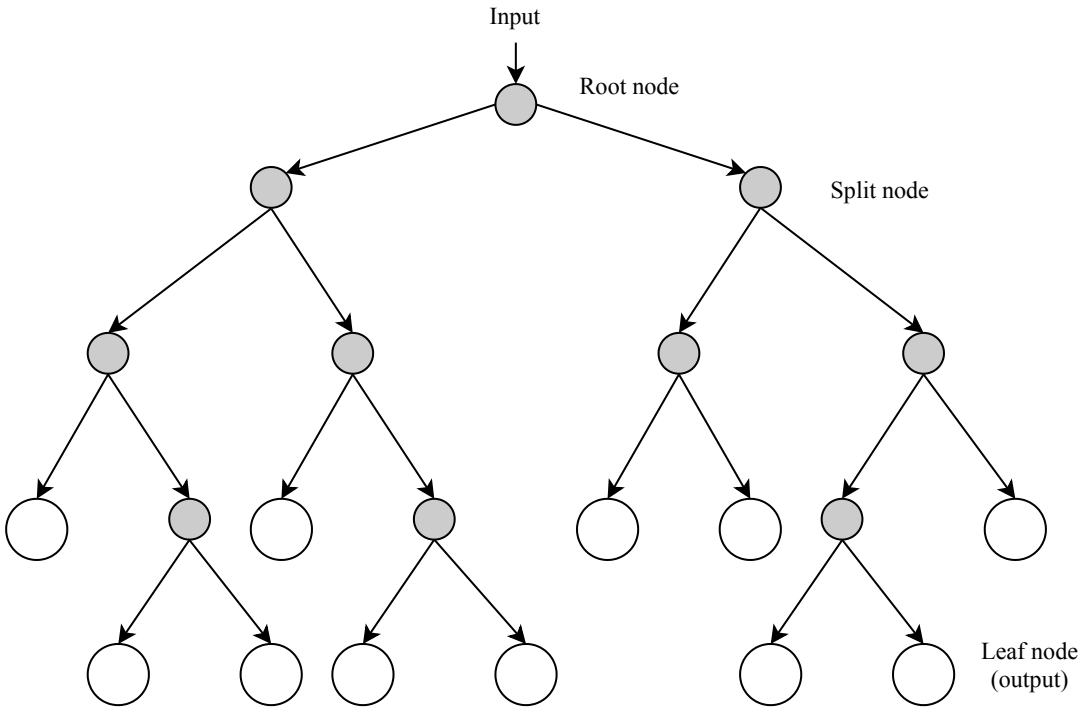


Figure 6.14: Regression tree procedure.

6.7.2 Bagging

Bagging [174][175][176], also known as boostrap aggregator is a machine learning ensemble method that combines a multitude of decision trees in order to improve performance. For PD localisation, bagging is used to model the non-linear relationship between WPT-PD features and PD location [177]. Instead of growing a single tree from the complete data set during training, bagging grows many trees using bootstrap (equiprobable sampling with replacement) samples of the PD data. Each sample is different from the original data set, yet resembles it in distribution and variability as a result of random sampling. Bootstrapping is particularly useful for small size data sets that have tendency to overfit. Different tree models are grown for each bootstrap sample. Mathematically, given a training input set $R = r_1, ..., r_n$ and corresponding training output set $(X, Y) = (x, y)_i, ..., (x, y)_n$, bagging repeatedly ($M$ times) selects subsets (randomly sample with replacement) of these training set and builds trees with each subset. For $m = 1, ..., M$: bagging builds T_m models. After training, the location estimate for unseen sample r' can be taken as the average estimates of all the predictions by individual trees on r' given in Equation 6.12;

$$\hat{T} = \frac{1}{M}\Sigma_{m=1}^{M} T_m(r') \tag{6.12}$$

The performance of bagging is affected by the particular bootstrap sample size used. Therefore, Bayesian optimisation is used to tune the model parameters. PD location is then determined by taking the average of the outputs from all the trees. Random sampling helps overcome the problem of overfitting in regression trees and improves predictions. However, there exist high correlations in prediction among some of the PD localisation subtrees, limiting the performance of the bagging regression trees. This motivates an investigation into the use of random forest for PD localisation.

6.7.3 Random Regression Forest PD Localisation Method

Random Regression Forest (RRF) is an ensemble of different regression trees widely used in prediction [178][179]. The main idea of RRF is to grow many regression trees based on some randomly selected features (subspacing) from randomly selected samples with a bootstrap strategy. Figure 6.15 shows the general procedure followed by random forest technique.

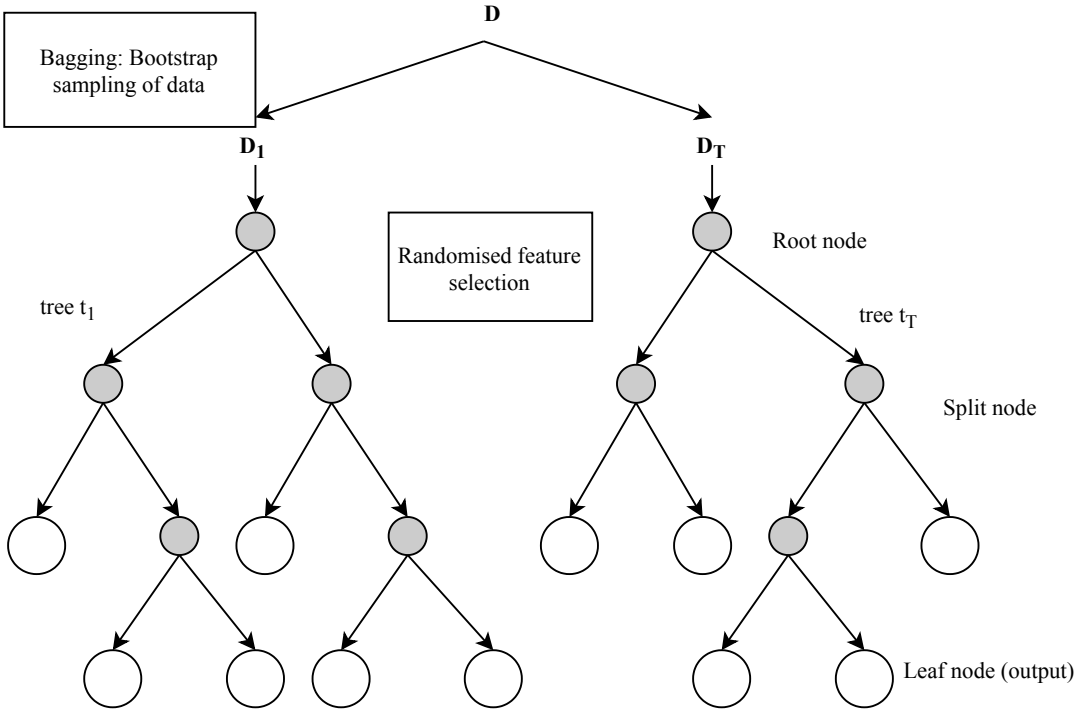


Figure 6.15: Random forest procedure.

In the context of PD localization, each tree is regarded as a function approximation problem consisting of a non-linear mapping of the PD input features onto the (x, y) coordinates representing PD location. The nonlinearity is achieved by dividing up the original PD localisation problem into smaller ones, solvable with simple models. A multivariate RRF model is developed to locate PD source to its x and y coordinate. Instead of building two models each for x and y output, multivariate RRF builds a single model which is more efficient. The RRF-PD localisation technique consists of two phases: a training phase and

a location estimation phase. In the training phase, RF grows multiple regression trees from bootstrap samples of the training data. Each sample subset comprises PD features and associated ground-truth locations. At each split or node, only a random subset of the PD features is considered. The size of the forest is fixed and all the trees are trained in parallel. During training, the parameters of the model are optimised. In the location estimation phase, each previously unseen PD features sample is given as an input to each tree in the forest starting at the root and the corresponding sequence of tests applied. Each tree gives an estimate of the location coordinate and the final prediction is taken as the average of the trees predictions. The flowchart of the RF framework is as shown in Figure 6.16. The selection of random samples of features at each split produces uncorrelated subtree predictions. Combining multiple decorrelated PD-tree models increases robustness to variance and reduces overall sensitivity to noise. This is demonstrated by the results obtained.

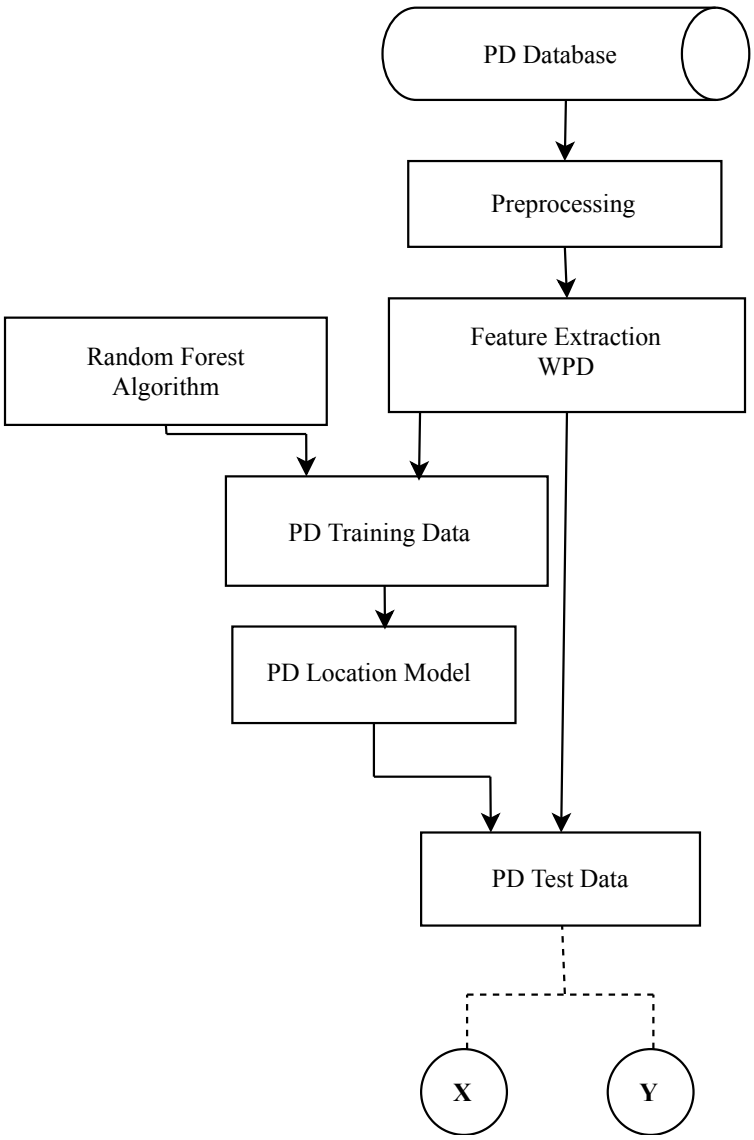


Figure 6.16: Flow chart for Random Forest PD localisation.

6.8 Performance Analysis

6.8.1 Performance Evaluation Metrics

The performance metrics defined in chapter 4; cumulative distribution function, accuracy, percentile, precision, and Root Mean Square Error provide the basis for evaluating and comparing the partial discharge localisation models described in Section 6.7. In addition,

confidence ellipse is used to further analyse the performance of the models.

6.8.1.1 Confidence Ellipse

A confidence ellipse for location estimation model represent a confidence region where a predicted location is likely to lie given a new observation in the environment considered. In other words, it approximates a region with specified percentage of probability, contains the center of the estimates. Therefore, the 95% confidence ellipse used in this thesis is the smallest region that covers the center of estimates for a given location with 95% probability.

6.8.2 Evaluating model accuracy

In order to evaluate the performance of the different machine learning algorithms used to estimate the PD location, the data collected from the measurement campaign described in section 3 is used. The MODWPT technique is applied to the PD data to extract the wavelet PD location features. PD features generated from the radio signals collected as training data set ($2880 \times 3 \times 4 = 1296$ combination of locations, sensors and features) and their corresponding ground truth locations are used to train the algorithms, as a results PD localisation models (Regression tree, bagging and regression random forest models) are built. Each (feature, location) pair constitutes training data point. The wavelet PD features from the test data at 32 distinct locations without their true locations were then presented to the models to estimate the source locations.

Localisation errors for each PD location model developed are shown in Table 6.5. The x and y coordinate error columns represent the errors in the x and y direction respectively, however, these columns may not correspond to errors in the same physical location. For example, for Reg. Tree-best the x-error is 0 for one location but has a non-zero y value. Similarly, for a different location y is 0 but has a non-zero error in the x-direction.

Table 6.5: Performance statistics of the wavelet-based developed models.

Model	X-error (m)	Y-error (m)	Overall error (m)
Reg. Tree (best)	0	0	0.167
Reg. Tree (worst)	11.50	6.50	11.51
Reg. Tree (mean)	1.18	1.65	2.29
Reg. Tree (75 percentile)	1.5	2.5	2.92
Bagging (best)	0.0007	0.0014	0.05
Bagging (worst)	6.68	6.19	6.92
Bagging (mean)	0.76	1.43	1.79
Bagging (75 percentile)	1.06	2.04	2.51
Random Forest (best)	0.0033	0.0059	0.04
Random Forest (worst)	6.15	6.08	6.26
Random Forest (mean)	0.77	1.52	1.87
Random Forest (75 percentile)	1.05	2.12	2.49

Errors in x and y coordinates are defined as absolute differences between predicted and true locations. Column 3 represents the overall localisation error. The mean and the best localisation errors are highlighted in Table 6.5. With data variability and the area covered, a mean error of 1.79m obtained is sufficient for PD location. The maximum number of PD features used at each split is 1/3 of the total number of features. In this work, 500 trees are grown for the ensemble methods.

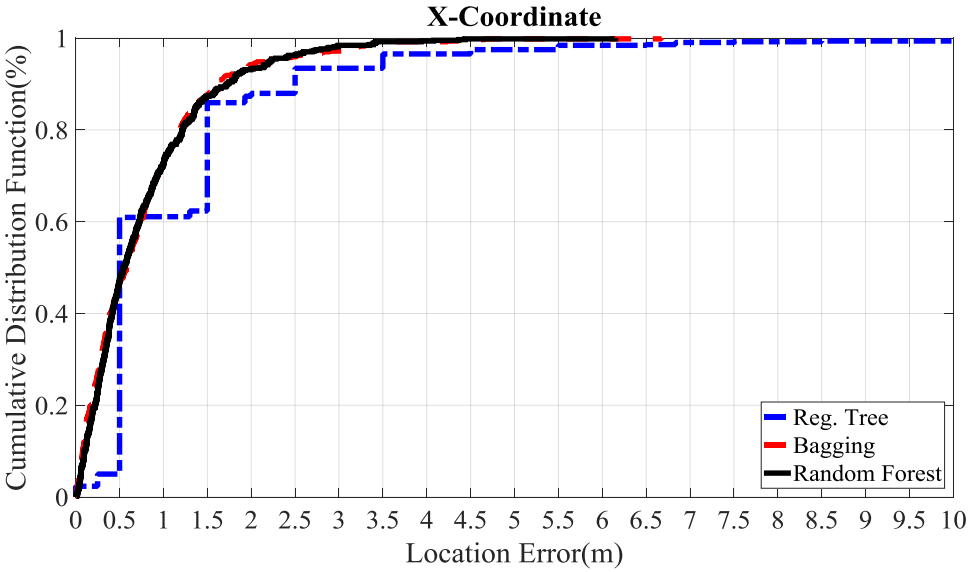


Figure 6.17: CDF of location error in x-coordinate.

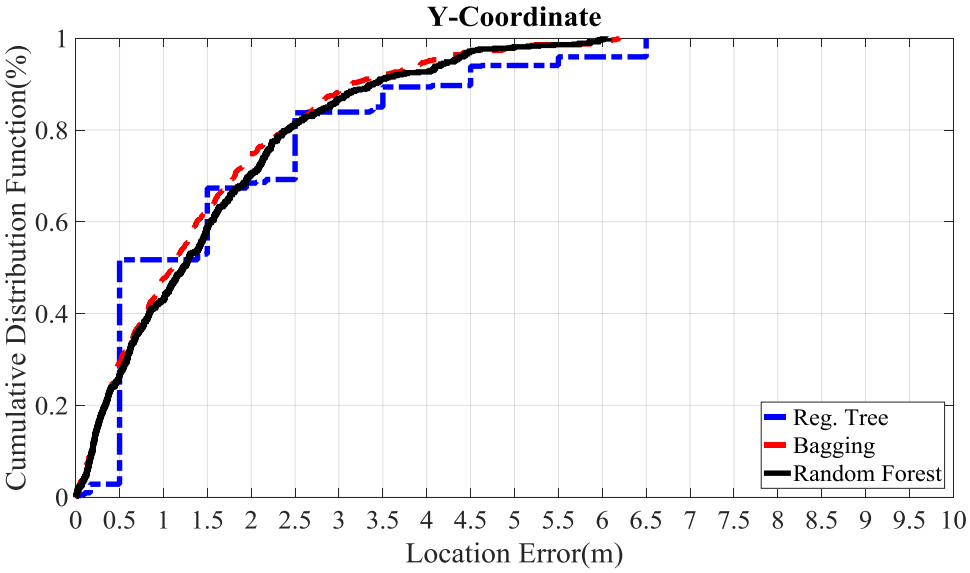


Figure 6.18: CDF of location error in y-coordinate.

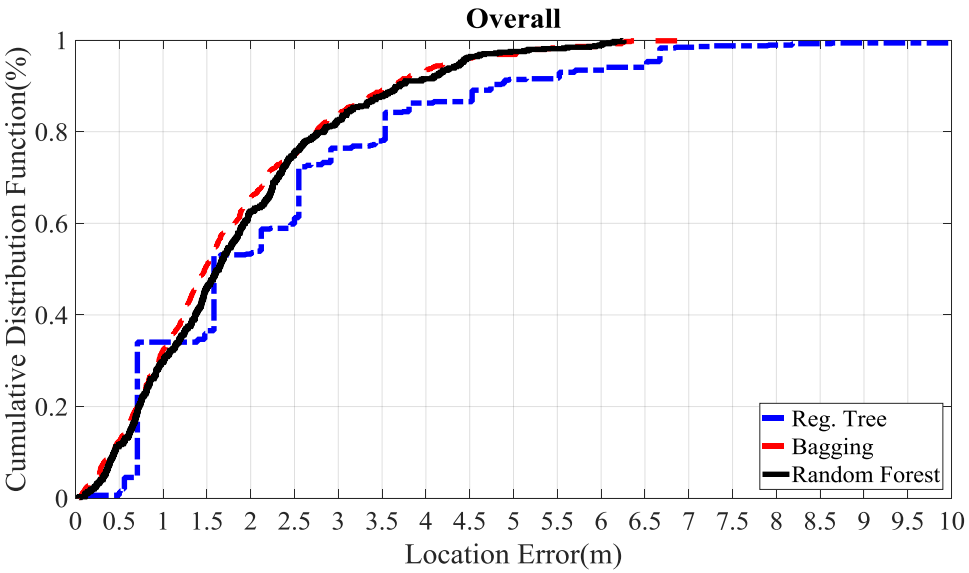
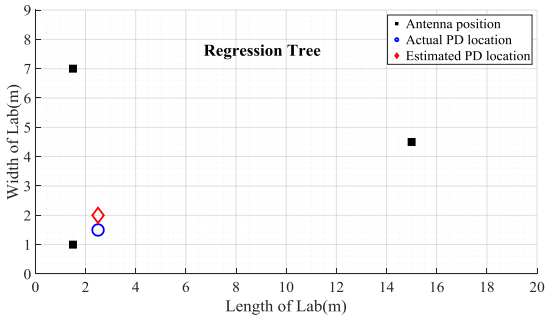
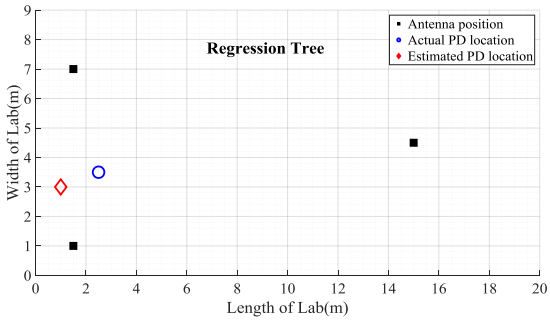


Figure 6.19: CDF of overall distance location error.

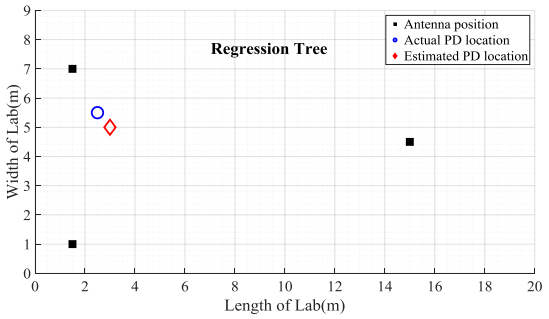
Figures 6.17, 6.18 and 6.19 further illustrate the difference in accuracy among the three models by comparing the cumulative density functions (CDFs) of location errors. It can be observed that the ensemble methods increased the overall confidence probability of errors within $3m$ to 83% which is significantly larger than the probability of 76.4% achieved by regression tree. This is however unsurprising as the regression tree is a weak learner. Bagging and random Forest ensemble methods have reduced the mean location error by 34% and 30.6% respectively compared to regression tree. Given that the spacing between test locations is $2.5m$, the proportion of test data with a localisation error below this value is also evaluated. For the regression tree method, 60% of test points were below $2.5m$. For the bagging algorithm, 74.8% are below, and for the random forest 75.4% are below.



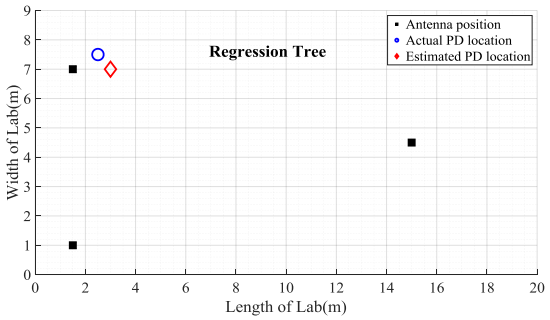
(a) Position 1



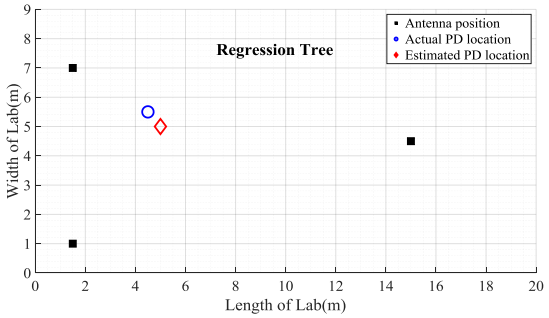
(b) Position 2



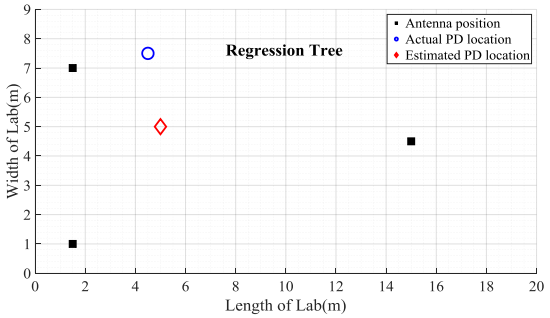
(c) Position 3



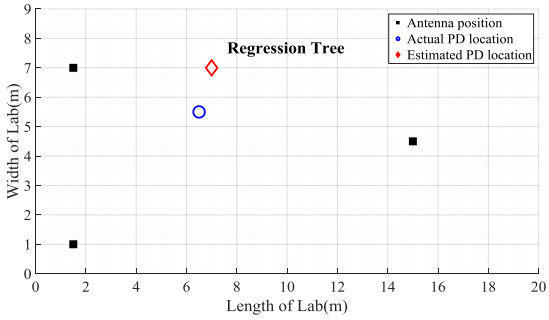
(d) Position 4



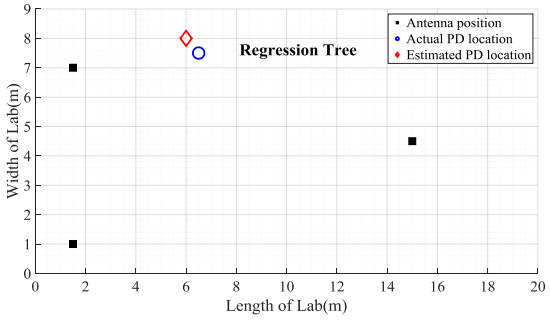
(e) Position 5



(f) Position 6

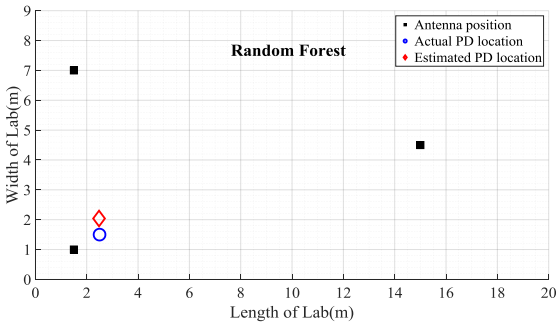


(g) Position 7

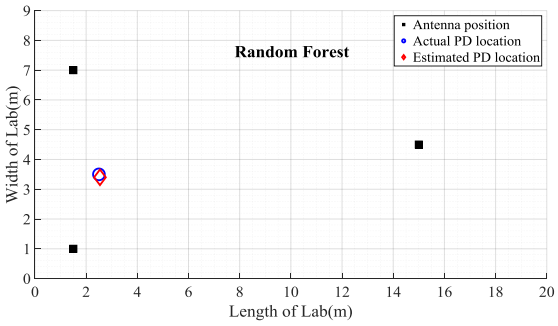


(h) Position 8

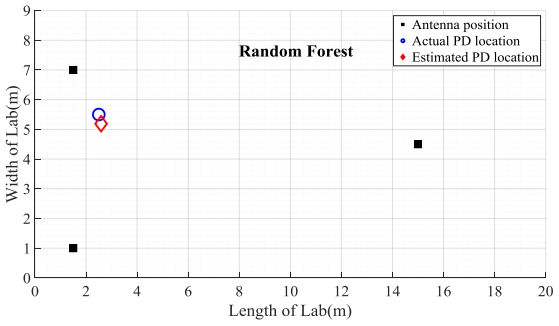
Figure 6.20: Regression Tree PD location estimates



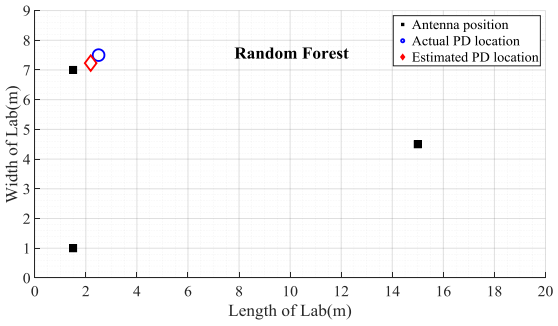
(a) Position 1



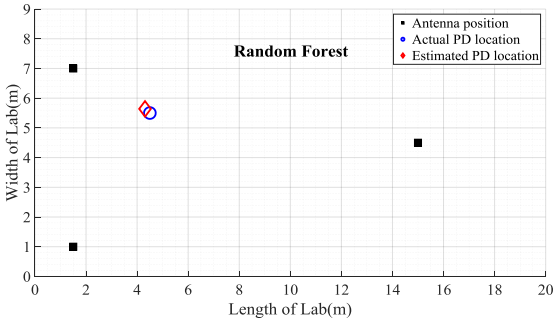
(b) Position 2



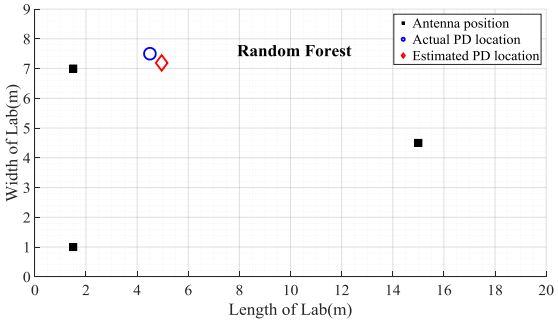
(c) Position 3



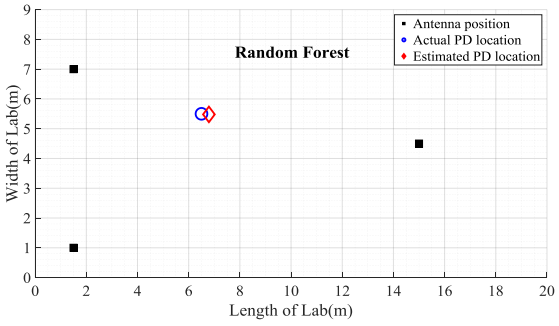
(d) Position 4



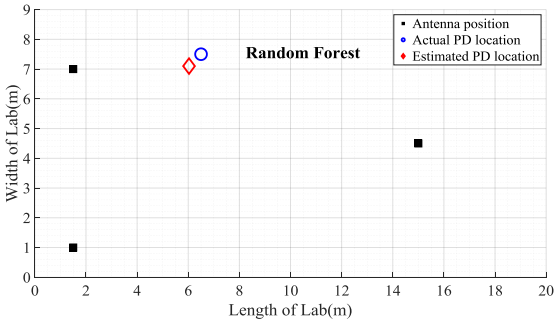
(e) Position 5



(f) Position 6

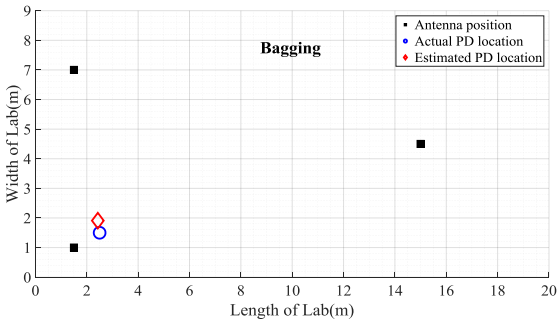


(g) Position 7

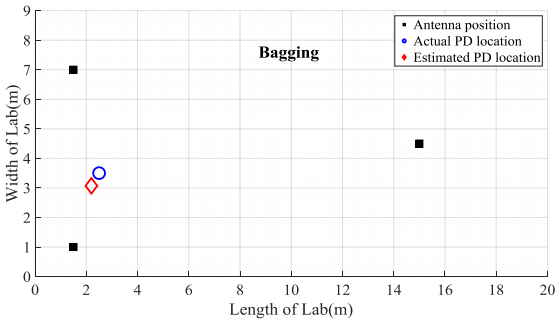


(h) Position 8

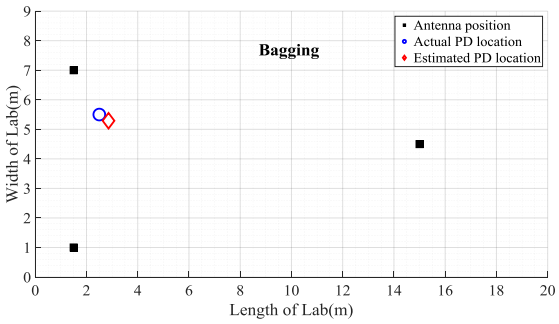
Figure 6.21: Random Forest PD location estimates



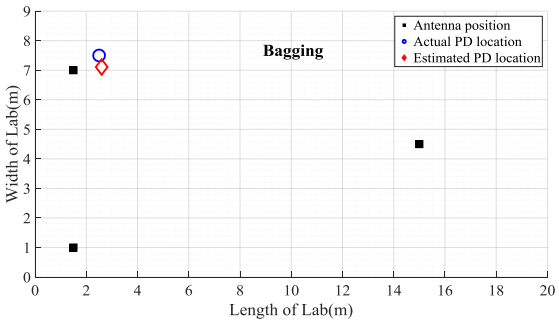
(a) Position 1



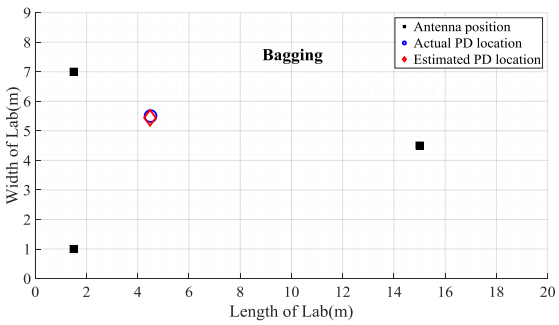
(b) Position 2



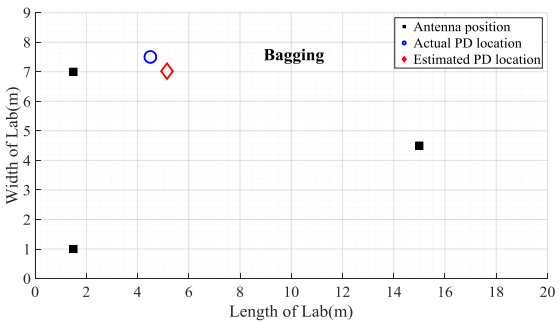
(c) Position 3



(d) Position 4

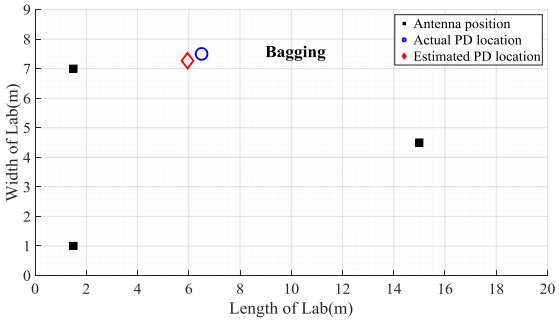


(e) Position 5



(f) Position 6

(g) Position 7



(h) Position 8

Figure 6.22: Bagging PD location estimates

The point accuracy for each test location used in the validation of this technique is of significant important. As a way of illustration, Figures 6.20, 6.21 and 6.22 shows the estimated PD source locations (red diamond) of 8 randomly selected locations for each model using wavelet-based features. The blue circle represents the actual PD source position. The black squares represent emulated sensor node positions. The results show that location error is minimal when ensemble methods are applied on the wavelet-based features. Regression tree can locate the exact x or y coordinates with zero error in some instances as shown in Table 6.5 but fails to provide a robust model, with worst location error as high as 11.5m.

6.8.3 Evaluating Model Robustness

Another key aspect of the proposed PD location scheme is its robustness. By comparing the location accuracy in x and y coordinates and the overall precision, the ensemble methods provide more robust models for PD localisation.

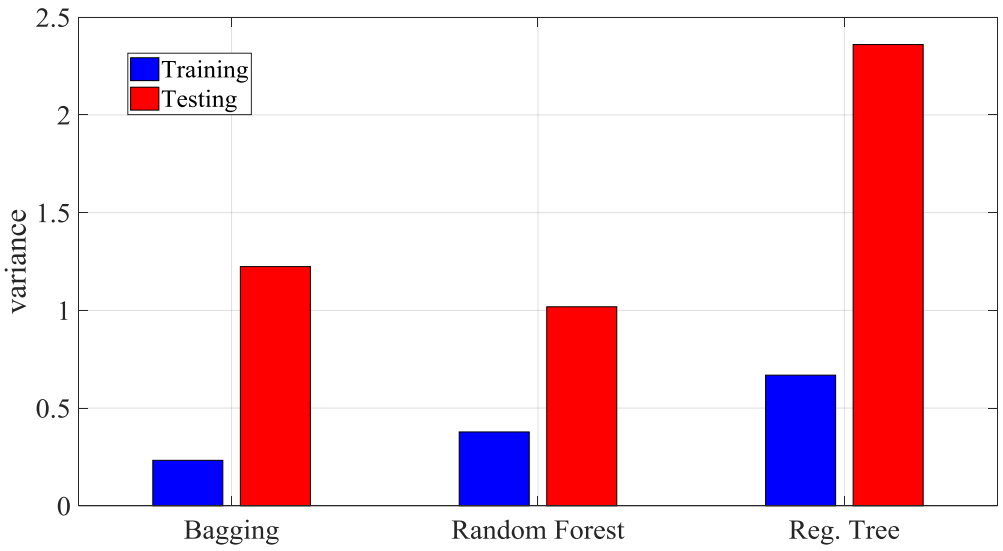


Figure 6.23: Variance of location error.

Furthermore, Figure 6.23 shows the variance in location error of the models using training

and testing datasets. Random forest, in particular, showed a much lower variance between training and testing predictions, making it more robust compared to regression tree and bagging models. Figure 6.24 shows PD localisation results from 3 random locations. Each discrete point corresponds to a single location estimate of a radio measurement. Also shown for the three PD localisation models is the confidence ellipse that contains 95% of the estimates for each location. For point $(2.5m, 1.5m)$ the regression tree estimates fall on a straight line. This means that the covariance in the y-direction is zero. Therefore, the width of the ellipse is zero in that direction. The size of the ellipse is an indication of the precision of the estimate. Our ensemble methods show more precise estimate with a smaller confidence ellipse. The large ellipses of Regression Tree indicates a less precise location estimate. This further demonstrates the robustness of the proposed wavelet-based ensemble PD localisation scheme in the presence of noise.

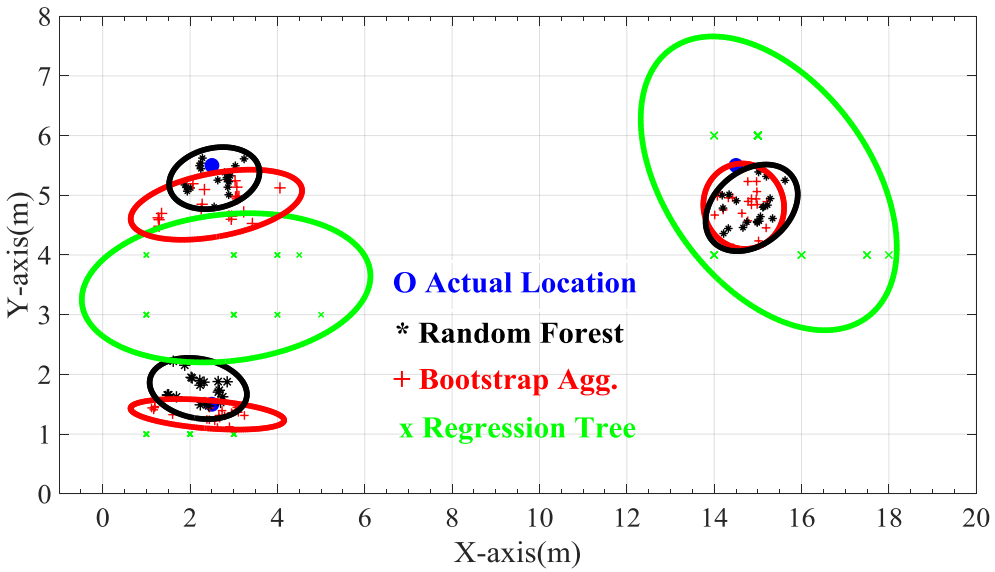


Figure 6.24: 95 confidence ellipse of three random locations for each model.

6.9 Summary

An investigation into technique that can improve the performance of the PD localisation system proposed in this thesis has been carried out. Specifically, we focus on the generation of location dependent features from PD signals and the performance of machine learning algorithms to improve the performance of the PD localisation system.

First, by examining the frequency response of the antennas used for capturing PD pulses, 2 distinct frequency bands which hold different PD information (based on the knowledge that multipath propagation is frequency selective) were exploited for PD features and integrated for fingerprinting. This provides a rich database for PD localization via pattern matching algorithm. The proposed Correlation Based Feature Selection (CFS) technique finds an effective way to select the most informative and decorrelated features from generated PD features based on correlation theory such that the dimensionality of the feature space is reduced. The GRNN and WKNN models developed using the statistical features before and after the application of CFS demonstrate high accuracy.

Secondly, a robust technique based on wavelet packet transform and ensemble learning method has been evaluated. In this method, the measured PD signals are first decomposed using MODWPT decomposition to extract PD location dependent features. The MODWPT selects the frequency bands with equal bandwidth where the energy of the PD signal is concentrated through a transformation such that the retained signal information is maximized in order to ensure high accuracy. The wavelet-based PD features are then used to build the ensemble models. PD location is obtained via ensemble machine learning technique; bagging and random forest algorithm which provides a more robust and accurate approach compared to a single regression tree model in terms of the mean localisation error. Bagging and random forest methods provide mean localisation accuracy of 1.79m and 1.87m, which enhances the precision of PD localisation by 34% and

30.6% over regression tree respectively.

CHAPTER 7

Conclusion and Future Work

7.1 Conclusion

A review of partial discharge detection and localisation systems applicable for determining the location of PD was conducted. The results of the review indicate potential for developing an enhanced PD localisation system that can be permanently deployed to monitor an entire substation.

The principle of using Received Signal Strength (RSS) as PD features in fingerprinting technique to infer PD location was successfully demonstrated. It has been found that RSS is an important parameter to consider in PD localisation as it decreases with distance from the source of discharge, creating unique signature for each location. Moreover, variations in signal strength as PD progress with passage of time and differences in PD types motivated the use of RSS derivatives (ratios) as PD features to infer PD location. It has

been shown that the use of RSS ratios increases the performance of the developed PD localisation system.

Machine learning algorithms underpin the localisation process in the developed PD location fingerprinting method. The use of machine learning algorithm obviates the need for estimating radio channel parameters or a propagation model which always falls short of capturing the underlying characteristics of any environment, degrading localisation performance when used with RSS. It has been found that this approach also uses multipath to its advantage to provide robust PD localisation. The developed machine learning models learn the unique spatial signatures created by PD signal strength to estimate PD locations. The examples used for learning are measurements taken within the environment. In this way, the underlying characteristics of the environment are well captured and this increases the performance of our method.

The combination of relative RSS and machine learning presents a novelty in the field of PD condition monitoring; an enhanced continuous solution for PD localisation. Our results dispel the impression than RSS is not a good location parameter for PD localisation. It has been established that the developed PD localisation system can monitor and locate discharges from several items of plant concurrently making it suitable for future smart grid.

A number of improvements to the developed substation-wide PD localisation system have been considered. Additional PD features beside RSS in the form of statistical characterisation of received signals were exploited and used to infer PD location. The frequency-selective nature of multipath provided useful information about PD signal in specific frequency components. The use of relative frequency measures account for the different frequency components inherent in different types of PD. Moreover, due to the high dimension of data generated, Correlation based Feature Selection (CFS) algorithm was

employed to select a set of decorrelated features hence reducing the dimensionality of the data. It has been found that the use of CFS reduces the computational complexity with no significant impact on overall precision performance. Empirical results demonstrate improvement in PD location accuracy compared to using RSS.

Another method considered for improving PD localisation was the use of wavelet transform in conjunction with machine learning ensemble technique. The time-invariant Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) was used to generate features from PD signals in different frequency bands. These features were then used as input to ensemble models: bagging and random forest. It has been shown that the ensemble-based approaches achieved higher location accuracy and robustness to noise by aggregating results from different sub-models compared with a single regression tree model.

The principles of the technique described in this thesis are flexible and can be adapted to any system that detects PD by measuring radiated radio waves. The measurement band does not need to be VHF or UHF and can be applied to any kind of high voltage installation.

7.2 Limitations

One of the limitations of the work reported in this thesis is that it addresses only a single type of PD event. In a real-world scenario, measurement of signals from more than one type of discharge can be common. Another limitation is that the developed system was not deployed in real substations. The experiment was conducted in a laboratory. However, it is expected that the localisation system will exploit the rich multipath phenomenon in real substations to achieve higher accuracy. Finally, the work did not investigate the impact of the distance between receivers.

7.3 Future Work

Most studies in PD monitoring and localisation are based on expensive and energy hungry technique; TDoA/AoA of the received RF signals and are often contact based, dedicated to a specific item of plant. The outcome of the work reported in this thesis provides a base for future study of low-cost wireless substation-wide PD monitoring and localisation systems.

Representative sensors used in the measurement campaign are off-the-shelf commercial antennas. Further work should include desire to develop customise or bespoke PD sensors that will produce all the needed data and result in one go.

Many sensors could be deployed in order to increase accuracy because the signal parameters or signatures from more sensors translate to more quality information for PD localisation. However, this comes at an additional cost of increased dimensionality and computational complexity. Further work should focus on how to select an optimal minimum number of sensors (taking into consideration possible failures of sensors) for each estimation without significant impact on the localisation accuracy.

Our low cost/complexity solution deploys a WSN based on off-the-shelf commercial radio nodes to monitor PD radio emissions. The PD events were generated with a pulse generator to ensure that the emulated PD traces are identical for each location and repetitive. A possible PD monitoring and localisation research topic based on generation of different PD emulated traces at defined locations to imitate the variability of PD in real substation should be explored in the future.

In the application of the developed PD localisation method in real substation where the area to be covered is larger than the laboratory used in this thesis, a clustering

technique would be needed to reduced computation burden in relation to the search space and database used for location estimation and also increase localisation accuracy. Thus, future research should explore and implement such technique.

The use of different antennas could be explored in the future to ascertain the effect of the type of antenna used for radio location of PD sources.

It would also be interesting to employ other signal processing techniques such as a combination of wavelet and Empirical Mode Decomposition (EMD) to extract more robust characteristics feature for enhanced PD localisation. This should be investigated further.

Bibliography

[1] P. L. Joskow, "Incentive regulation and its application to electricity networks," *Review of Network Economics*, vol. 7, no. 4, 2008.

[2] Ofgem. (). Quality of service incentives, [Online]. Available: `https://www.ofgem.gov.uk/electricity/distribution-networks/network-price-controls/quality-service/quality-service-incentives`.

[3] I. E. Portugues, P. J. Moore, I. A. Glover, C. Johnstone, R. H. McKosky, M. B. Goff, and L. van der Zel, "Rf-based partial discharge early warning system for air-insulated substations," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 20–29, 2009.

[4] F. P. Mohamed, W. H. Siew, J. J. Soraghan, S. M. Strachan, and J. McWilliam, "Remote monitoring of partial discharge data from insulated power cables," *IET Science, Measurement & Technology*, vol. 8, no. 5, pp. 319–326, 2014.

[5] H. Illias, M. Tunio, A. Bakar, H. Mokhlis, and G. Chen, "Partial discharge phenomena within an artificial void in cable insulation geometry: Experimental validation

and simulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 1, pp. 451–459, 2016.

[6] P. J. Moore, I. E. Portugues, and I. A. Glover, "Partial discharge investigation of a power transformer using wireless wideband radio-frequency measurements," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 528–530, 2006.

[7] J. Nelson and J. Stein, "A field assessment of pd and emi methodology applied to large utility generators," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 17, no. 5, pp. 1411–1427, 2010.

[8] F. Peer Mohamed, W. Siew, and J. Soraghan, "Online partial discharge detection in medium voltage cables using protection/instrument current transformers," in *Second UHVNet Colloquium on High Voltage Measurement and Insulation Research*, 2009.

[9] D. Evagorou, A. Kyprianou, P. Lewin, A. Stavrou, V. Efthymiou, A. Metaxas, and G. Georghiou, "Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network," *IET Science, Measurement & Technology*, vol. 4, no. 3, pp. 177–192, 2010.

[10] P. J. Moore, I. E. Portugues, and I. A. Glover, "Radiometric location of partial discharge sources on energized high-voltage plant," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 2264–2272, 2005.

[11] E. Lemke, S. Berlijn, E. Gulski, H. M. Muhr, E. Pultrum, T. Strehl, W. Hauschild, J. Rickmann, and G. Rizzi, "Guide for electrical partial discharge measurements in compliance to iec 60270," *Electra*, vol. 241, pp. 60–68, 2008.

[12] M. Nafar, "Simulation of partial discharge mechanism using emtp," in *International Conference on Innovations in Electrical and Electronics Engineering (ICIEE'2012)*, 2012.

[13] S. Boggs and G. Stone, "Fundamental limitations in the measurement of corona and partial discharge," *IEEE Transactions on Electrical Insulation*, vol. 17, no. 2, pp. 143–150, 1982.

[14] M. Yaacob, M. A. Alsaedi, J. Rashed, A. Dakhil, and S. Atyah, "Review on partial discharge detection techniques related to high voltage power equipment using different sensors," *Photonic Sensors*, vol. 4, no. 4, pp. 325–337, 2014.

[15] I. S. IEC, "60270, high voltage test techniques-partial discharge measurements," *International Electrotechnical Commission*, 2000.

[16] G. Stone, "Partial discharge diagnostics and electrical equipment insulation condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 5, pp. 891–904, 2005.

[17] M. D. Judd, L. Yang, and I. B. Hunter, "Partial discharge monitoring of power transformers using uhf sensors. part i: Sensors and signal interpretation," *IEEE Electrical Insulation Magazine*, vol. 21, no. 2, pp. 5–14, 2005.

[18] R. Bartnikas, "Partial discharges. their mechanism, detection and measurement," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 5, pp. 763–808, 2002.

[19] R. Bartnikas and J. Novak, "Effect of overvoltage on the risetime and amplitude of pd pulses," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 4, pp. 557–566, 1995.

[20] H. H. Sinaga, "Detection identification and localization of partial discharges in power transformers using uhf techniques," *The University of New South Wales Austrailia. PhD Thesis*, 2012.

[21] F. H. Kreuger, "Partial discharge detection in high-voltage equipment," 1989.

[22] F. Kreuger, "Industrial high voltage: Electric fields, dielectrics, constructions," *Delft University Press, Delft, The Netherlands*, 1991.

[23] J. Kuffel and P. Kuffel, *High voltage engineering fundamentals*. Elsevier, 2000.

[24] F. H. Kreuger, "Discharge detection in high voltage equipment," 1964.

[25] S. A. Boggs, "Partial discharge. iii. cavity-induced pd in solid dielectrics," *IEEE Electrical Insulation Magazine*, vol. 6, no. 6, pp. 11–16, 1990.

[26] H. Suzuki, K. Aihara, and T. Okamoto, "Complex behaviour of a simple partial-discharge model," *EPL (Europhysics Letters)*, vol. 66, no. 1, p. 28, 2004.

[27] P. Morshuis, "Partial discharge mechanisms in voids related to dielectric degradation," *IEE Proceedings-Science, Measurement & Technology*, vol. 142, no. 1, pp. 62–68, 1995.

[28] D. M. H. Foo, "Online partial discharge detection and signal analysis for high voltage cables," PhD thesis, University of Strathclyde, 2005.

[29] M. S. Naidu, *High voltage engineering*. Tata McGraw-Hill Education, 2013.

[30] A. A. Jaber, "Absolute calibration of radiometric partial discharge sensors for insulation condition monitoring in electrical substations," PhD thesis, University of Huddersfield, 2017.

[31] P. Mitchinson, P. Lewin, B. Strawbridge, and P. Jarman, "Tracking and surface discharge at the oil-pressboard interface," *IEEE Electrical Insulation Magazine*, vol. 26, no. 2, 2010.

[32] T. J. Gallagher and A. Pearmain, *High voltage: Measurement, testing and design*. Wiley, 1983.

[33] R. Patsch, "Electrical and water treeing: A chairman's view," *IEEE Transactions on Electrical Insulation*, vol. 27, no. 3, pp. 532–542, 1992.

[34] L. A. Dissado, "Understanding electrical trees in solids: From experiment to theory," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 4, pp. 483–497, 2002.

[35] R. Baumgartner, B. Fruth, W. Lanz, and K. Pettersson, "Partial discharge. ix. pd in gas-insulated substations-fundamental considerations," *IEEE Electrical Insulation Magazine*, vol. 7, no. 6, pp. 5–13, 1991.

[36] E. Forster, "Research in the dynamics of electrical breakdown in liquid dielectrics," *IEEE Transactions on Electrical Insulation*, no. 3, pp. 182–185, 1980.

[37] L. Christophorou and L. Pinnaduwage, "Basic physics of gaseous dielectrics," *IEEE Transactions on Electrical Insulation*, vol. 25, no. 1, pp. 55–74, 1990.

[38] L. Niemeyer, "The physics of partial discharges," in *International Conference on Partial Discharge*, IET, 1993, pp. 1–4.

[39] G. P. Cleary, "Interpretation of uhf signals produced by partial discharges in oil-filled power transformers," PhD thesis, University of Strathclyde, 2005.

[40] A. Reid, M. Judd, G. B. Stewart, and R. Fouracre, "Comparing iec 60270 and rf partial discharge patterns," in *International Conference on Condition Monitoring and Diagnosis*, IEEE, 2008, pp. 89–92.

[41] T. Pinpart, "Techniques for analysis and interpretation of uhf partial discharge signals," PhD thesis, University of Strathclyde, 2010.

[42] M. Shafiq *et al.*, "Design and implementation of partial discharge measurement sensors for on-line condition assessment of power distribution system components," *Aalto University, Doctoral Dissertation*, 2014.

[43] M. Muhr, T. Strehl, E. Gulski, K. Feser, E. Gockenbach, W. Hauschild, and E. Lemke, "Sensors and sensing used for non-conventional pd detection," *CIGRE, Paris*, pp. 1–7, 2006.

[44] R. Rogers, "Ieee and iec codes to interpret incipient faults in transformers, using gas in oil analysis," *IEEE Transactions on Electrical Insulation*, no. 5, pp. 349–354, 1978.

[45] X. Song, "The automation and optimisation of wavelet transform techniques for pd denoising and pulse shape classification in power plant," PhD thesis, Glasgow Caledonian University, 2009.

[46] S. Karmakar, N. Roy, and P. Kumbhakar, "Partial discharge measurement of transformer with ict facilities," in *International Conference on Power Systems, ICPS'09*, IEEE, 2009, pp. 1–5.

[47] L. E. Lundgaard, "Partial discharge. xiv. acoustic partial discharge detection-practical application," *IEEE Electrical Insulation Magazine*, vol. 8, no. 5, pp. 34–43, 1992.

[48] I. Kemp, "Partial discharge plant-monitoring technology: Present and future developments," *IEE Proceedings-Science, Measurement & Technology*, vol. 142, no. 1, pp. 4–10, 1995.

[49] D. Zhu, A. McGrail, S. Swingler, D. Auckland, and B. Varlow, "Partial discharge detection in cable termination using acoustic emission techniques and adaptive signal processing," in *IEEE International Symposium on Electrical Insulation*, IEEE, 1994, pp. 74–76.

[50] Y. Lu, X. Tan, and X. Hu, "Pd detection and localisation by acoustic measurements in an oil-filled transformer," *IET Science, Measurement & Technology*, vol. 147, no. 2, pp. 81–85, 2000.

[51] S. M. Markalous, S. Tenbohlen, and K. Feser, "Detection and location of partial discharges in power transformers using acoustic and electromagnetic signals," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 6, 2008.

[52] X. Yang, Y. Ming, C. Xiaolong, Q. Changrong, and G. Chen, "Comparison between optical and electrical methods for partial discharge measurement," in *Proceedings of the 6th International Conference on Properties and Applications of Dielectric Materials*, IEEE, vol. 1, 2000, pp. 300–303.

[53] P. Moore, I. Portugues, and I. Glover, "A nonintrusive partial discharge measurement system based on rf technology," in *2003 IEEE Power Engineering Society General Meeting*, IEEE, vol. 2, 2003, pp. 628–633.

[54] N. De Kock, B. Coric, and R. Pietsch, "Uhf pd detection in gas-insulated switchgear-suitability and sensitivity of the uhf method in comparison with the iec 270 method," *IEEE Electrical Insulation Magazine*, vol. 12, no. 6, pp. 20–26, 1996.

[55] N. Fujimoto, S. Boggs, and R. Madge, "Electrical transients in gas-insulated switchgear," in *Transactions of the March, 1981 Meeting of the Canadian Electric Association*, 1981.

[56] B. Hampton and R. Meats, "Diagnostic measurements at uhf in gas insulated substations," in *IEE Proceedings C - Generation, Transmission and Distribution*, IET, vol. 135, 1988, pp. 137–144.

[57] D. Pommerenke, I. Krage, W. Kalkner, E. Lemke, and P. Schmiegel, "On-site pd measurement on high voltage cable accessories using integrated sensors," in *Proc. 9th ISH*, 1995, pp. 1–4.

[58] M. Judd, G. Cleary, and C. Bennoch, "Applying uhf partial discharge detection to power transformers," *IEEE Power Engineering Review*, vol. 22, no. 8, pp. 57–59, 2002.

[59] M. Kawada, "Fundamental study on location of a partial discharge source with a vhf-uhf radio interferometer system," *Electrical Engineering in Japan*, vol. 144, no. 1, pp. 32–41, 2003.

[60] P. Agoris, S. Meijer, and J. Smit, "Sensitivity check of an internal vhf/uhf sensor for transformer partial discharge measurements," in *Power Tech, 2007 IEEE Lausanne*, IEEE, 2007, pp. 2065–2069.

[61] W. Hauschild and E. Lemke, *High-voltage test and measuring techniques*. Springer, 2014.

[62] E. Steennis, E. Hetzel, and C. Verhoeven, "Diagnostic medium voltage cable test at 0.1 hz," in *Proceedings of the 3rd International Conference on Insulated Power Cables (JICABLE)*, 1991, pp. 408–414.

[63] D. A. Ward and J. L. T. Exon, "Using rogowski coils for transient current measurements," *Engineering Science & Education Journal*, vol. 2, no. 3, pp. 105–113, 1993.

[64] M. Shafiq, G. A. Hussain, L. Kütt, and M. Lehtonen, "Effect of geometrical parameters on high frequency performance of rogowski coil for partial discharge measurements," *Measurement*, vol. 49, pp. 126–137, 2014.

[65] M. D. Judd, "Radiometric partial discharge detection," in *International Conference on Condition Monitoring and Diagnosis, CMD*, IEEE, 2008, pp. 1025–1030.

[66] S. Tenbohlen, D. Denissov, S. M. Hoek, and S. Markalous, "Partial discharge measurement in the ultra high frequency (uhf) range," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 6, 2008.

[67] D. Mishra, B. Sarkar, C. Koley, and N. Roy, "Localization of partial discharge source in high voltage apparatus using multiple uhf sensors," in *International Conference on Energy, Power and Environment: Towards Sustainable Growth (ICEPE)*, IEEE, 2015, pp. 1–5.

[68] K. Gülnihar, S. Cekli, C. P. Uzunoğlu, and M. Uğur, "Location estimation of partial discharge-based electromagnetic source using multilateration with time difference of arrival method," *Electrical Engineering*, pp. 1–9, 2018.

[69] Y.-H. Kim, Y.-W. Youn, S.-H. Yi, D.-H. Hwang, J.-H. Sun, and J.-H. Lee, "High-resolution partial discharge location estimation in power cables," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 2, pp. 758–765, 2014.

[70] H. Sinaga, B. Phung, and T. Blackburn, "Partial discharge localization in transformers using uhf detection method," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 6, 2012.

[71] S. Boggs, "Electromagnetic techniques for fault and partial discharge location in gas-insulated cables and substations," *IEEE Transactions on Power Apparatus and Systems*, no. 7, pp. 1935–1941, 1982.

[72] J. Pearson, B. Hampton, and A. Sellars, "A continuous uhf monitor for gas-insulated substations," *IEEE Transactions on Electrical Insulation*, vol. 26, no. 3, pp. 469–478, 1991.

[73] J. Pearson, O. Farish, B. Hampton, M. Judd, D. Templeton, B. Pryor, and I. Welch, "Partial discharge diagnostics for gas insulated substations," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 893–905, 1995.

[74] M.-X. Zhu, Y.-B. Wang, Q. Liu, J.-N. Zhang, J.-B. Deng, G.-J. Zhang, X.-J. Shao, and W.-L. He, "Localization of multiple partial discharge sources in air-insulated substation using probability-based algorithm," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 1, pp. 157–166, 2017.

[75] Z. Tang, C. Li, X. Cheng, W. Wang, J. Li, and J. Li, "Partial discharge location in power transformers using wideband rf detection," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 13, no. 6, 2006.

[76] Z. Kawasaki, J.-M. Li, K. Matsuura, M. Kawasaki, and O. Sugimoto, "Localization of partial discharges using time difference of arrival of radiation fields," in *Proceedings of the 3rd International Conference on Properties and Applications of Dielectric Materials*, IEEE, 1991, pp. 910–914.

[77] H. Hou, G. Sheng, P. Miao, X. Li, Y. Hu, and X. Jiang, "Partial discharge location based on radio frequency antenna array in substation," *Gaodianya Jishu/ High Voltage Engineering*, vol. 38, no. 6, pp. 1334–1340, 2012.

[78] J. Tang and Y. Xie, "Partial discharge location based on time difference of energy accumulation curve of multiple signals," *IET Electric Power Applications*, vol. 5, no. 1, pp. 175–180, 2011.

[79] H. Hou, G. Sheng, and X. Jiang, "Robust time delay estimation method for locating uhf signals of partial discharge in substation," *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1960–1968, 2013.

[80] H. Ishimaru and M. Kawada, "Location technique for multiple partial discharge sources using independent component analysis and direction of arrival method of electromagnetic waves based on bayesian network," *IEEJ Transactions on Power and Energy*, vol. 126, pp. 1247–1254, 2006.

[81] L. Yongfen, X. Xiaohu, D. Fei, T. Xiao, and L. Yanming, "Comparison of doa algorithms applied to ultrasonic arrays for pd location in oil," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2316–2323, 2015.

[82] H. Hou, G. Sheng, and X. Jiang, "Localization algorithm for the pd source in substation based on l-shaped antenna array signal processing," *IEEE Transactions on Power Delivery*, vol. 30, no. 1, pp. 472–479, 2015.

[83] I. E. Portugues, P. J. Moore, and P. Carder, "The use of radiometric partial discharge location equipment in distribution substations," in *18th International Conference and Exhibition on Electricity Distribution, CIRED.*, IET, 2005, pp. 1–4.

[84] M. A. Rahman, P. Lewin, and P. Rapisarda, "Autonomous localization of partial discharge sources within large transformer windings," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, pp. 1088–1098, 2016.

[85] G. Robles, J. Fresno, and J. Martínez-Tarifa, "Radio-frequency localization of multiple partial discharges sources with two receivers," *Sensors*, vol. 18, no. 5, p. 1410, 2018.

[86] J. M. Fresno, G. Robles, J. M. Martínez-Tarifa, and B. G. Stewart, "A combined algorithm approach for pd location estimation using rf antennas," in *IEEE Electrical Insulation Conference (EIC)*, IEEE, 2017, pp. 384–387.

[87] A. Baug, N. R. Choudhury, R. Ghosh, S. Dalai, and B. Chatterjee, "Localisation of single and multiple partial discharge sources based on sequence of arrival and levels of peak amplitude of acoustic emissions," in *2016 National Power Systems Conference (NPSC)*, IEEE, 2016, pp. 1–6.

[88] G. Zhang, X. Zhang, J. Tang, and H. Cheng, "Study on localization of transformer partial discharge source with planar arrangement uhf sensors based on singular value elimination," *AIP Advances*, vol. 8, no. 10, p. 105 232, 2018.

[89] J. Fresno, G. Robles, J. Martínez-Tarifa, and B. G. Stewart, "Survey on the performance of source localization algorithms," *Sensors*, vol. 17, no. 11, p. 2666, 2017.

[90] R. A. Hooshmand, M. Parastegari, and M. Yazdanpanah, "Simultaneous location of two partial discharge sources in power transformers based on acoustic emission using the modified binary partial swarm optimisation algorithm," *IET Science, Measurement & Technology*, vol. 7, no. 2, pp. 119–127, 2013.

[91] C. Boya, M. Rojas-Moreno, M. Ruiz-Llata, and G. Robles, "Location of partial discharges sources by means of blind source separation of uhf signals," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 4, pp. 2302–2310, 2015.

[92] O. El Mountassir, B. G. Stewart, S. G. McMeekin, and A. Ahmadinia, "Effect of noise on the location accuracy of partial discharges using radiated rf detection techniques," in *47th International Universities Power Engineering Conference (UPEC)*, IEEE, 2012, pp. 1–6.

[93] M. Aminghafari, N. Cheze, and J.-M. Poggi, "Multivariate denoising using wavelets and principal component analysis," *Computational Statistics & Data Analysis*, vol. 50, no. 9, pp. 2381–2398, 2006.

[94] L. Satish and B. Nazneen, "Wavelet-based denoising of partial discharge signals buried in excessive noise and interference," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 10, no. 2, pp. 354–367, 2003.

[95] C. Petrarca and G. Lupo, "An improved methodological approach for denoising of partial discharge data by the wavelet transform," *Progress In Electromagnetics Research*, vol. 58, pp. 205–217, 2014.

[96] E. Mostacci, C. Truntzer, H. Cardot, and P. Ducoroy, "Multivariate denoising methods combining wavelets and principal component analysis for mass spectrometry data," *Proteomics*, vol. 10, no. 14, pp. 2564–2572, 2010.

[97] A. Kyprianou and G. Georghiou, "Wavelet packet denoising for on-line partial discharge detection in high voltage systems," in *Proceedings of the IEEE International Symposium on Intelligent Control. Mediterrean Conference on Control and Automation*, IEEE, 2005, pp. 1184–1189.

[98] J. C. P. Mng and M. Mehralizadeh, "Forecasting east asian indices futures via a novel hybrid of wavelet-pca denoising and artificial neural network models," *PloS One*, vol. 11, no. 6, e0156338, 2016.

[99] S. Mohanty, K. K. Gupta, and K. S. Raju, "Adaptive fault identification of bearing using empirical mode decomposition–principal component analysis-based average kurtosis technique," *IET Science, Measurement & Technology*, vol. 11, no. 1, pp. 30–40, 2017.

[100] A. Kyprianou, P. Lewin, V. Efthimiou, A. Stavrou, and G. Georghiou, "Wavelet packet denoising for online partial discharge detection in cables and its application to experimental field results," *Measurement Science & Technology*, vol. 17, no. 9, p. 2367, 2006.

[101] J. Shlens, "A tutorial on principal component analysis," *ArXiv preprint arXiv:1404.1100*, 2014.

[102] L. I. Smith, "A tutorial on principal components analysis," Tech. Rep., 2002.

[103] D. A. Jackson, "Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches," *Ecology*, vol. 74, no. 8, pp. 2204–2214, 1993.

[104] U. Khan, P. Lazaridis, H. Mohamed, D. Upton, K. MIstry, B. Saeed, P. Mather, M. Vieira, R. Atkinson, C. Tachtatzis, *et al.*, "Localization of partial discharge by using received signal strength," in *2nd URSI Atlantic Radio Science Meeting (AT-RASC)*, IEEE, 2018, pp. 1–4.

[105] E. Iorkyase, C. Tachtatzis, R. Atkinson, and I. Glover, "Localisation of partial discharge sources using radio fingerprinting technique," in *Loughborough Antennas & Propagation Conference (LAPC)*, IEEE, 2015, pp. 1–5.

[106] J. de Souza Neto, E. de Macedo, J. da Rocha Neto, E. Da Costa, S. Bhatti, and I. Glover, "Partial discharge location using unsynchronized radiometer network for condition monitoring in hv substations-a proposed approach," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 364, 2012, p. 012 053.

[107] G. Ding, J. Zhang, L. Zhang, and Z. Tan, "Overview of received signal strength based fingerprinting localization in indoor wireless lan environments," in *IEEE 5th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE)*, IEEE, 2013, pp. 160–164.

[108] C. Nerguizian, S. Belkhous, A. Azzouz, V. Nerguizian, and M. Saad, "Mobile robot geolocation with received signal strength (rss) fingerprinting technique and neural networks," in *IEEE International Conference on Industrial Technology, IEEE ICIT'04*, IEEE, vol. 3, 2004, pp. 1183–1185.

[109] P.-J. Chuang and Y.-J. Jiang, "Effective neural network-based node localisation scheme for wireless sensor networks," *IET Wireless Sensor Systems*, vol. 4, no. 2, pp. 97–103, 2014.

[110] N. Swangmuang, "A location fingerprint framework towards efficient wireless indoor positioning systems," PhD thesis, University of Pittsburgh, 2009.

[111] L. Gui, M. Yang, P. Fang, and S. Yang, "Rss-based indoor localisation using mdcf," *IET Wireless Sensor Systems*, vol. 7, no. 4, pp. 98–104, 2017.

[112] K. Pahlavan, X. Li, and J.-P. Makela, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, 2002.

[113] J. Kuriakose, J. V. Naveenbabu, M. Shahid, A. Shetty, *et al.*, "Analysis of maximum likelihood and mahalanobis distance for identifying cheating anchor nodes," *ArXiv preprint arXiv:1412.2857*, 2014.

[114] K. Kaemarungsi, "Efficient design of indoor positioning systems based on location fingerprinting," in *International Conference on Wireless Networks, Communications and Mobile Computing*, IEEE, vol. 1, 2005, pp. 181–186.

[115] S. Vasanth, Y. M. Yeap, and A. Ukil, "Fault location estimation for vsc-hvdc system using artificial neural network," in *2016 IEEE Region 10 Conference (TENCON)*, IEEE, 2016, pp. 501–504.

[116] P. Werle, A. Akbari, H. Borsi, and E. Gockenbach, "Partial discharge localisation on power transformers using neural networks combined with sectional winding transfer functions as knowledge base," in *Proceedings of 2001 International Symposium on Electrical Insulating Materials, (ISEIM 2001)*, IEEE, 2001, pp. 579–582.

[117] L. Hamza and C. Nerguizian, "Neural network and fingerprinting-based localization in dynamic channels," in *IEEE International Symposium on Intelligent Signal Processing, WISP 2009*, IEEE, 2009, pp. 253–258.

[118] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, 1991.

[119] S. Mohanty and S. Ghosh, "Artificial neural networks modelling of breakdown voltage of solid insulating materials in the presence of void," *IET Science, Measurement & Technology*, vol. 4, no. 5, pp. 278–288, 2010.

[120] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception*, Elsevier, 1992, pp. 65–93.

[121] J. C. Whittington and R. Bogacz, "Theories of error back-propagation in the brain," *Trends in cognitive sciences*, 2019.

[122] A. Wanto, M. Zarlis, D. Hartama, *et al.*, "Analysis of artificial neural network backpropagation using conjugate gradient fletcher reeves in the predicting process," in *Journal of Physics Conference Series*, vol. 930, 2017.

[123] C. D. Doan and S.-y. Liong, "Generalization for multilayer neural network bayesian regularization or early stopping," in *Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference*, 2004, pp. 5–8.

[124] H. Zhang, J. Cui, L. Feng, A. Yang, H. Lv, B. Lin, and H. Huang, "High-precision indoor visible light positioning using deep neural network based on the bayesian regularization with sparse training point," *IEEE Photonics Journal*, vol. 11, no. 3, pp. 1–10, 2019.

[125] E. T. Iorkyase, C. Tachtatzis, P. Lazaridis, I. A. Glover, and R. C. Atkinson, "Low-complexity wireless sensor system for partial discharge localisation," *IET Wireless Sensor Systems*, vol. 9, no. 3, pp. 158–165, 2019.

[126] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, no. 6, pp. 689–705, 2007.

[127] R. Rostaminia, M. Saniei, M. Vakilian, S. S. Mortazavi, and V. Parvin, "Accurate power transformer pd pattern recognition via its model," *IET Science, Measurement & Technology*, vol. 10, no. 7, pp. 745–753, 2016.

[128] E. Jiang, P. Zan, X. Zhu, J. Liu, and Y. Shao, "Rectal sensation function rebuilding based on optimal wavelet packet and support vector machine," *IET Science, Measurement & Technology*, vol. 7, no. 3, pp. 139–144, 2013.

[129] L. Hao and P. Lewin, "Partial discharge source discrimination using a support vector machine," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 17, no. 1, 2010.

[130] B. Ravikumar, D. Thukaram, and H. Khincha, "Application of support vector machines for fault diagnosis in power transmission system," *IET Generation, Transmission & Distribution*, vol. 2, no. 1, pp. 119–130, 2008.

[131] Y. Khan, A. Khan, F. Budiman, A. Beroual, N. Malik, and A. Al-Arainy, "Partial discharge pattern analysis using support vector machine to estimate size and position of metallic particle adhering to spacer in gis," *Electric Power Systems Research*, vol. 116, pp. 391–398, 2014.

[132] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of Machine Learning Research*, vol. 2, no. Dec, pp. 125–137, 2001.

[133] S. A. Bessedik and H. Hadi, "Prediction of flashover voltage of insulators using least squares support vector machine with particle swarm optimisation," *Electric Power Systems Research*, vol. 104, pp. 87–92, 2013.

[134] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Information Processing-Letters and Reviews*, vol. 11, no. 10, pp. 203–224, 2007.

[135] H. Zhang, X. Wang, C. Zhang, and X. Cai, "Robust identification of non-linear dynamic systems using support vector machine," *IEE Proceedings-Science, Measurement and Technology*, vol. 153, no. 3, pp. 125–129, 2006.

[136] S. M. Clarke, J. H. Griebsch, and T. W. Simpson, "Analysis of support vector regression for approximation of complex engineering analyses," *Journal of Mechanical Design*, vol. 127, no. 6, pp. 1077–1087, 2005.

[137] E. T. Iorkyase, C. Tachtatzis, P. Lazaridis, I. A. Glover, and R. C. Atkinson, "Radio location of partial discharge sources: A support vector regression approach," *IET Science, Measurement & Technology*, vol. 12, no. 2, pp. 230–236, 2017.

[138] K. Shi, Z. Ma, R. Zhang, W. Hu, and H. Chen, "Support vector regression based indoor location in ieee 802.11 environments," *Mobile Information Systems*, vol. 2015, 2015.

[139] K. P. Bennett and C. Campbell, "Support vector machines: Hype or hallelujah?" *Acm Sigkdd Explorations Newsletter*, vol. 2, no. 2, pp. 1–13, 2000.

[140] J. Ma, J. Theiler, and S. Perkins, "Accurate on-line support vector regression," *Neural computation*, vol. 15, no. 11, pp. 2683–2703, 2003.

[141] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.

[142] Z. Wu, C. Li, Z. Yang, and P. Wang, "Research on tourists' positioning technology based on lssvr," in *Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, IEEE, 2015, pp. 826–830.

[143] E. T. Iorkyase, C. Tachtatzis, I. A. Glover, and R. C. Atkinson, "Rf-based location of partial discharge sources using received signal features," *IET High Voltage*, 2018.

[144] M. Haindl, P. Somol, D. Ververidis, and C. Kotropoulos, "Feature selection based on mutual correlation," in *Iberoamerican Congress on Pattern Recognition*, Springer, 2006, pp. 569–577.

[145] Y. Cui, J. S. Jin, S. Zhang, S. Luo, and Q. Tian, "Correlation-based feature selection and regression," in *Pacific-Rim Conference on Multimedia*, Springer, 2010, pp. 25–35.

[146] L. Ladha and T. Deepa, "Feature selection methods and algorithms," *International journal on computer science and engineering*, vol. 3, no. 5, pp. 1787–1797, 2011.

[147] A. R. Mor, P. Morshuis, P. Llovera, V. Fuster, and A. Quijano, "Localization techniques of partial discharges at cable ends in off-line single-sided partial discharge cable measurements," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 1, pp. 428–434, 2016.

[148] T. Babnik, R. Aggarwal, and P. Moore, "Data mining on a transformer partial discharge data using the self-organizing map," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 14, no. 2, 2007.

[149] J. Hernández-Mejía, J. Perkel, R. Harley, M. Begovic, R. Hampton, and R. Hartlein, "Determining routes for the analysis of partial discharge signals derived from the field," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 6, pp. 1517–1525, 2008.

[150] Y. Avargel and I. Cohen, "System identification in the short-time fourier transform domain with crossband filtering," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1305–1319, 2007.

[151] M. Sifuzzaman, M. Islam, and M. Ali, "Application of wavelet transform and its advantages compared to fourier transform," 2009.

[152] U. Khayam and T. Kasnalestari, "System of wavelet transform on partial discharge signal denoising," in *International Conference of Industrial, Mechanical, Electrical, and Chemical Engineering (ICIMECE)*, IEEE, 2016, pp. 79–83.

[153] E. Macedo, D. Araújo, E. da Costa, R. Freire, W. T. Lopes, I. Torres, J. de Souza Neto, S. Bhatti, and I. Glover, "Wavelet transform processing applied to partial discharge evaluation," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 364, 2012, p. 012 054.

[154] Q. Shan, S. Bhatti, I. A. Glover, R. Atkinson, I. E. Portugues, P. J. Moore, and R. Rutherford, "Characteristics of impulsive noise in electricity substations," in *17th European Signal Processing Conference*, IEEE, 2009, pp. 2136–2140.

[155] M. Homaei, S. M. Moosavian, and H. A. Illias, "Partial discharge localization in power transformers using neuro-fuzzy technique," *IEEE Transactions on Power Delivery*, vol. 29, no. 5, pp. 2066–2076, 2014.

[156] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 2014.

[157] A. N. Akansu, P. A. Haddad, R. A. Haddad, and P. R. Haddad, *Multiresolution signal decomposition: Transforms, subbands, and wavelets*. Academic press, 2001.

[158] C. S. Chang, J. Jin, C. Chang, T. Hoshino, M. Hanai, and N. Kobayashi, "Separation of corona using wavelet packet transform and neural network for detection of partial discharge in gas-insulated substations," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 1363–1369, 2005.

[159] S. Anjum, S. Jayaram, A. El-Hag, and A. N. Jahromi, "Detection and classification of defects in ceramic insulators using rf antenna," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 1, pp. 183–190, 2017.

[160] Z. Ye, B. Wu, and A. Sadeghian, "Current signature analysis of induction motor mechanical faults by wavelet packet decomposition," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 6, pp. 1217–1228, 2003.

[161] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*. A. K. Peters, Ltd., 1994, ISBN: 1-56881-041-5.

[162] E. C. T. Macedo, D. B. Araújo, E. G. da Costa, R. C. S. Freire, W. T. A. Lopes, I. S. M. Torres, J. M. R. de Souza Neto, S. A. Bhatti, and I. A. Glover, "Wavelet transform processing applied to partial discharge evaluation," *Journal of Physics: Conference Series*, vol. 364, no. 7, pp. 012–054, 2012.

[163] M. Eissa, "A novel digital directional transformer protection technique based on wavelet packet," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 1830–1836, 2005.

[164] J. Li, T. Jiang, R. F. Harrison, and S. Grzybowski, "Recognition of ultra high frequency partial discharge signals using multi-scale features," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 4, pp. 1412–1420, 2012.

[165] A. T. Walden and A. C. Cristan, "The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events,"

Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1976, 1998.

[166] D. K. Alves, F. B. Costa, R. L. de Araujo Ribeiro, C. M. de Sousa Neto, and T. de Oliveira Alves Rocha, "Real-time power measurement using the maximal overlap discrete wavelet-packet transform," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 4, pp. 3177–3187, 2017.

[167] B. Whitcher, "Wavelet-based estimation for seasonal long-memory processes," *Technometrics*, vol. 46, no. 2, pp. 225–238, 2004.

[168] C. R. Cornish and C. S. Bretherton, "Maximal overlap wavelet statistical analysis with application to atmospheric turbulence," *Boundary-Layer Meteorology*, vol. 119, no. 2, pp. 339–374, 2006.

[169] M. Varanis and R. Pederiva, "The influence of the wavelet filter in the parameters extraction for signal classification: An experimental study.," *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, vol. 5, no. 1, 2017.

[170] M. Nielsen, "On the construction and frequency localization of finite orthogonal quadrature filters," *Journal of Approximation Theory*, vol. 108, no. 1, pp. 36–52, 2001.

[171] A. Swetapadma and A. Yadav, "A novel decision tree regression-based fault distance estimation scheme for transmission lines," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 234–245, 2017.

[172] D. Sánchez-Rodríguez, P. Hernández-Morera, and J. M. Quinteiro-González, "A low complexity system based on multiple weighted decision trees for indoor localization," in *Sensors*, 2015.

[173] H. Ahmadi and R. Bouallegue, "Rssi-based localization in wireless sensor networks using regression tree," in *International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, 2015, pp. 1548–1553.

[174] M. M. bin Othman, A. Mohamed, and A. Hussain, "Determination of transmission reliability margin using parametric bootstrap technique," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1689–1700, 2008.

[175] D. den Hertog, J. P. C. Kleijnen, and A. Y. D. Siem., "The correct kriging variance estimated by bootstrapping," *Journal of the Operational Research Society*, vol. 57, no. 4, pp. 400–409, 2006.

[176] A. K. Jain, R. C. Dubes, and C. Chen, "Bootstrap techniques for error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 628–633, 1987.

[177] E. T. Iorkyase, C. Tachtatzis, I. A. Glover, P. Lazaridis, D. Upton, B. Saeed, and R. C. Atkinson, "Improving rf-based partial discharge localization via machine learning ensemble method," *IEEE Transactions on Power Delivery*, 2019.

[178] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Medical Image Analysis*, vol. 17, no. 8, pp. 1293–1303, 2013.

[179] X. Xie, J. Xing, N. Kong, C. Li, J. Li, and S. Zhang, "Improving colorectal polyp classification based on physical examination data-an ensemble learning approach," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 434–441, 2018.