A Framework for Developing a Prognostic Model Using Partial Discharge Data from Electrical Trees

Nur Hakimah Ab Aziz

A thesis submitted for the degree of Doctor of Philosophy to

the Department of Electronic and Electrical Engineering
University of Strathclyde

2018

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.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed : Hakimah Aziz Date : 22/06/18

Dedication

To my parents mak and abah, my husband Zamzam, and my children Balqis, Adam and Athif for all their love, support and sacrifices.

Acknowledgment

Alhamdulillah. Thank you, Allah.

My sincerest thanks to Dr Victoria Catterson, for her guidance, patience, time, and encouragement, throughout the duration of this work.

Special thanks to Prof Martin Judd, Prof Simon Rowland, Dr Sanjay Bahadoorsingh, Dr Xiao Hu and Francis Cox for their kind assistance and valuable suggestions.

I'd also like to express my gratitude to my colleagues in the Intelligent System group, who have been really kind and helpful.

A big thank you to the Malaysian Government and Universiti Teknikal Malaysia Melaka (UTeM) for giving me the opportunity to pursue my doctoral study.

I owe a special thanks to my family who supported me and helped me throughout my life and during this study. I dedicate this work to you all. Mak, Abah, I do not know how to thank you enough for providing me with the opportunity to be where I am today. I love you so much.

I also dedicate this Ph.D. thesis to my lovely children, Balqis, Adam and Athif who are the pride and joy of my life. I love you more than anything and I appreciate all your patience and support during ibu’s Ph.D. studies.

Last but not least, thank you to my dearest husband, Zam. I love you for everything, for being so understanding and for putting up with me through the toughest moments of my life. I thank Allah for enlightening my life with your presence.

Abstract

Insulation breakdown is a key failure mode of high voltage (HV) equipment, with progressive faults such as electrical treeing leading to potentially catastrophic failure. Electrical treeing proceeds from defects in solid insulation, and cables are particularly affected. Research has shown that diagnosis of the fault can be achieved based on partial discharge (PD) analysis. Nonetheless, after diagnosis of a defect, engineers need to know how long they have to take action. This requires prognosis of remaining insulation life.

The progression of a defect is far less well understood than diagnosis, making prognosis a key challenge requiring new approaches to defect modelling. The practical deployment of prognostics for cable monitoring is not currently feasible, due to the lack of understanding of degradation mechanisms and limited data relating defect inception to plant failure. However, this thesis advances the academic state of the art, with an eye towards practical deployment in the future. The expected beneficiaries of this work are therefore researchers in the field of HV condition monitoring in general, and electrical treeing within cables in particular.

This research work develops a prognostic model of insulation failure due to the electrical treeing phenomenon by utilising the associated PD data from previous experiment. Both phase-resolved and pulse sequence approaches were employed for PD features extraction. The performance of the PD features as prognostic parameters were evaluated using three metrics, *monotonicity*, *prognosability* and *trendability*. The analysis revealed that features from pulse sequence approach are better than phase-resolved approach in terms of monotonicity and prognosability. The key contributions to knowledge of this work are three-fold: the selection of the most appropriate prognostic parameter for PD in electrical trees, through thorough analysis of the behaviour of a number of candidate parameters; a prognostic modelling approach for this parameter based on curve-fitting; and a generalised framework for prognostic modelling using data-driven techniques.

Contents

List of Figures

List of Tables

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AC	Alternating current
AE	Acoustic emission
AI	Artificial intelligence
ANN	Artificial neural network
AR	Autoregressive
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
AWG	Arbitrary waveform generator
CBM	Condition-based maintenance
CEIDP	Conference on Electrical Insulation and Dielectric Phenomena
CH	Convergence horizon
CRA	Cumulative relative accuracy
CV	Cross validation
DAM	Discharge Avalanche Model
DBM	Dielectric Breakdown Model
DMM	Dissado-Mazzanti-Montanari
DC	Direct current
DGA	Dissolved gas analysis
EL	Electroluminescence
EoL	End of life
EoUP	End of useful prediction
EPR	Ethylene propylene rubber
EAA	Ethylene–acrylic acid
EVA	Ethylene–vinyl acetate

GIS	Gas-insulated substation
GOF	Goodness-of-fit
HFCT	High frequency current transformer
HPLC	High performance liquid chromatography
HV	High voltage
ICSD	International Conference on Solid Dielectrics
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IQR	Interquartile range
IPM	Inverse power model
IPTM	Inverse power threshold model
LDA	Linear discriminant analysis
LDPE	Low density polyethylene
LOOCV	Leave-one-out cross validation
MA	Moving average
MI	Measuring instrument
MSD	Multiresolution signal decomposition
MSE	Mean squared error
NAR	Nonlinear autoregressive
NARX	Nonlinear autoregressive with external input
NI	National Instruments
pC	Picocoulomb
PCA	Principal component analysis
PCI	Peripheral component interconnect
PD	Partial discharge
PDEV	Partial discharge extinction voltage
PDHD	Partial discharge height distribution
PDIV	Partial discharge inception voltage
PE	Polyethylene
PH	Prognostic horizon
POF	Probability of failure
PRPD	Phase-resolved partial discharge
PRPDA	Phase-resolved partial discharge analysis

PSA	Pulse sequence analysis
RA	Relative Accuracy
RMS	Root mean squared
RMSE	Root mean squared error
RUL	Remaining useful life
SiR	Silicone rubber
S/N	Signal-to-noise
SSE	Sum of squared errors
SSR	Sum of squared regression
SST	The total sum of squares
SVM	Support vector machine
TDNN	Time-delay neural network
TEAM	Thermal, electrical, environmental and mechanical
THD	Total harmonic distortion
TTF	Time-to-failure
UHF	Ultra-high frequency
UHVnet	Universities High Voltage Network
XLPE	Crosslinked polyethylene

Nomenclature

φ	Phase angle occurrence
φ_{inc}	Phase inception
Φ	Phase inception symmetry
σ	Standard deviation
Λ	Lacunarity
Δ	Difference between two values
λ	Specific time index
$\#\frac{d}{dx}$	Number of sequential data points
α	Weibull scale parameter
β	Weibull shape parameter
c	y-intercept
C_a	Capacitance of virtual test object
C_k	Coupling capacitor
C_m	Centre of mass as a measure of convergence for a metric M
cc	Correlation coefficient
cc_p	Pearson correlation coefficient
cc_s	Spearman correlation coefficient
cov	Covariance
D	Tree damage
d_f	Fractal dimension
du	Voltage change
du_{all}	Voltage change that comprises du_{large} and du_{small}
$du_{all(bd)}$	Voltage change at t_{bd}
$du_{fit(bd)}$	Fitting of du_{all} value at t_{bd}
$du_{fit(gnd)}$	Fitting of du_{all} value at t_{gnd}

du_{large}	Large voltage change
du_{small}	Small voltage change
du_{th}	Threshold value of du
du/dt	Voltage derivative with respect to time
dt	Change in time
$d\theta_{all}$	Phase change
$d\theta_{large}$	Large phase change
$d\theta_{small}$	Small phase change
ε	Permittivity
e	Error
E	Electric field strength
f	Frequency
F	Number of features
G	Gibbs free energy
h	Harmonic order
H_{qm}	PD maximum amplitude
H_{qn}	PD mean amplitude
H_n	PD pulse count
i	Index number
i_α	The first time index when the prediction enters the α-bounds
i_{bd}	Time index at t_{bd}
i_{CH}	Time index when the predictions begin to remain unchanged
I	Current
j	Radius
k	Empirical factors
k_B	Boltzmann constant
k_P	Planck constant
K_p	Peak parameter of harmonics
K_{rms}	RMS parameter of harmonics
K_s	Waveshape parameter of harmonics
Ku	Kurtosis
l	Length

l_b	Length of a new tree channel
L	Lifetime
μ	Mean
m	Slope
m_{fit}	Slope of the fit
M	Mass of the area under a curve
n	Voltage cycle
N	Number of occurrences or observations
N_b	Number of branches
N_p	Number of phase bucket per half cycle
N_q	Number of discharge
p	The value of prognostic variable
p_{fail}	Failure value of prognostic variables
p_{start}	Initial value of prognostic variables
P	Probability
q	Charge
Q_s	Discharge symmetry
r	Radius
rg	Ranking of variable
R	Resistance
Rc	Chemical reaction rate
R^2	Coefficient of determination
R_{adj}^2	Adjusted R^2
s_λ	A set of time indexes
$\|s_\lambda\|$	Cardinality of s_λ
Sk	Skewness
t	Time
t_{bd}	Breakdown time
$t_{bd(est)}$	Estimated breakdown time
$t_{bd(act)}$	Actual breakdown time
t_{ch}	Time taken for the formation of new tree channel
t_{cons}	Breakdown time of constant voltage

t_{crit}	Critical time in tree growth
t_d	Decay time
t_{gnd}	Touch ground time
t_{prog}	Breakdown time of progressive voltage
t_r	Rise time
t_{th}	Time at threshold point
t_w	Pulse width
T	Temperature
u	Instantaneous voltage
U_{bd}	Breakdown voltage
U_p	Peak voltage
U_{rms}	RMS voltage
U_{th}	Threshold voltage
ω	Angular frequency
w	Weight factor
ΔW	Activation energy
x_c, y_c	x and y coordinates for centre of mass, C_M
X	Variables
X_{max}	Maximum value of X variables
X_{min}	Minimum value of X variables
X_{norm}	Normalised X variables
X_R	Residual values of X
X_T	Mean trend values of X
y	Actual response
$\bar{y}$	Average of y
$\hat{y}$	Predicted response
Z	Z-score normalization
Z_m	Measuring impedance
Z_n	Noise blocking filter

Chapter 1

Introduction

1.1 Research Background

A significant proportion of high voltage equipment in power networks is reaching or exceeding its anticipated design life [1]–[3]. It is a big challenge for network providers to manage these ageing assets effectively without compromising the availability and reliability of the network. The performance of these assets is expected to deteriorate and the consequences of equipment failure can be catastrophic. The associated cost of failure is not only for repair or replacement but also the loss of customer revenue.

The need to balance asset availability and cost effectiveness is becoming increasingly important in power network industries. Providers are coming to realise the potential wastefulness of scheduled (time-based) maintenance that may not accurately reflect the usage of the asset, thus resulting in unnecessary maintenance [4]. The transition between scheduled maintenance to condition-based maintenance (CBM) is accelerating [5]. CBM provides a more cost-effective service as well as the capability to predict and prevent failures through health monitoring, diagnostics and fault prognostics; the process has been outlined in ISO 55002 [6].

Unlike scheduled maintenance, the aim of CBM is to predict the upcoming failure so maintenance can be proactively scheduled when it is needed [7]. The diagnosis and

prognosis of a failure are performed when certain indicators show signs of decreasing or anomalies are detected. An advanced prognostic capability is desired because the ability to forecast this future condition enables a higher level of condition-based maintenance for optimally managing total life cycle cost.

The major cause of failure of high voltage (HV) equipment comes mostly from the insulation system [1]. This could be the consequences of the insulation deterioration or due to insulation defects introduced during manufacturing process or maintenance routines. Over the years, a number of methods have been developed to diagnose and monitor the degradation of the insulation system. Among them, partial discharge (PD) analysis is a well-accepted indicator of the degradation of electrical insulation, permitting early detection of insulation faults [8]–[11].

The insulating capability of various insulating materials is influenced by many factors. Physical, chemical, and electrical stresses contribute to the deterioration of the insulating materials. The formation of electrical treeing is one of the main causes of insulation degradation under high electric stress [12]. Electrical trees progress by periodic PDs after initiating at a point of high divergent stress [13]. Hence, PD measurement has long been associated with electrical treeing for detection [14], determining the type of electrical tree (e.g. branch or bush) [15] and classifying the growth stage [16], [17].

The increased integration of renewable and distributed generation has challenged asset management, particularly in tolerating power quality issues due to a steady rise in the number of power electronic devices. The gradual change in working environment results in a different ageing mechanism of the insulation system [2]. Previous work at the University of Manchester conducted experimental studies to investigate the effect of harmonics (one type of power quality issue) on electrical tree in terms of tree size, time to breakdown and phase-resolved partial discharge (PRPD) pattern [18]. The corresponding results are as follows:

- No changes were detected in electrical tree growth characteristics due to variation of harmonic order, waveshape factor, K_s and total harmonic distortion (THD) in the excitation voltage at 14.4 kV constant peak.

- The variation in breakdown trends did not reveal a deterministic relationship with THD and K_s.
- The composite waveforms influenced the partial discharge pattern produced.

The results show that harmonics do not give a very significant impact to the electrical growth process. Towards an advanced CBM, the same electrical treeing data reported in [18] has been utilised and analysed in this thesis for further investigation into the prognosis of solid insulation lifetime.

1.2 Justification for Research

Electrical treeing is one of the main degradation mechanisms in high voltage cables [12], [19] that is strongly related to the existence of defects in the insulation, such as cavities, conducting particles and protrusions from the electrodes. These defects, either developed during manufacturing or assembly, promote the inception of PD that may trigger the initiation of electrical treeing and grow into hollow channels towards the opposite electrode which can lead to breakdown. As it is closely linked with insulation breakdown, many researchers have investigated the morphology of electrical trees towards breakdown and some of them have extended their work to the development of life models.

It was mentioned in [20]–[22] that the extension of electrical treeing is due to electrical discharge activity inside the tree tubules. Hence, the electrical tree growth is usually studied along with the PD activities. The available models that relate PD mechanisms with the tree morphology can be either physics-based [20]–[24] or data-driven [25] or the combination of both approaches [26]. Life models in [25] and [26] proposed the mean of PD magnitude per cycle (for every 20 seconds) as the prognostic parameter although skewness and kurtosis of PD magnitude per cycle also show a monotonic trend in [25]. The selection of the prognostic parameter was unclear, thus, a set of metrics should be proposed to characterise these features, aiding in the selection of the most appropriate feature.

Most of the developed models for electrical treeing are physics-based, and relate the tree growth with the activity of the discharges within the tree. In [26], the breakdown is expected to occur when the tip reaches the ground plate (negative electrode). Also, the simulated tree growth in [20]–[22] only considers the forward growth (the tree propagates from the inception point to the ground electrode). This assumption may be misleading due to the existence of a return tree that has been observed and discussed very briefly in [27]–[29]. This means, an immediate breakdown may not occur when the original tree traverses the insulation gap since the tree may continue to grow from the ground electrode to the start point of the tree.

Literature describing the physics of the return tree is scarce and the reason for its existence remains unclear. Therefore, a data-driven approach is considered in this thesis since not much effort has been applied to modelling the return tree. Most of the published data-driven approaches utilise Weibull analysis of the time-to-failure (TTF) data. This approach lacks information on the degradation process. Alternatively, this research work aims to investigate the degradation behaviour of the selected PD feature and develop a prognostic model based on its characteristics.

Also, this thesis incorporates the effect of harmonics on the behaviour of the selected PD feature. Little work has been reported on the effect of harmonics on electrical treeing growth. The findings in [18] did not reveal a deterministic relationship between THD and breakdown time. In contrast, a study in [30] has reported that harmonics with higher THDs could accelerate the tree growth resulting in early insulation failure. These contradictory findings may be due to the constant peak voltage used in [18]. This work therefore, aims to identify any distinct behaviour of the selected PD feature for different THD and harmonic orders.

1.3 Principal Contributions

This research provides the following contributions to knowledge:

- Selection of the most appropriate prognostic parameter from analysed PD features through the application of a set of metrics introduced in [31] which characterise the suitability of a prognostic parameter.
- Relatedly but distinctly, identification of the characteristics of the selected prognostic parameter that correspond to the electrical tree growth. Three stages have been identified and the selected prognostic parameter shows a unique characteristic in distinguishing the stages.
- Identification of no deterministic behaviour of the selected PD feature for different THD, K_s and harmonic orders.
- Proposal of a generalised prognostic framework specifically for data-driven techniques where the step-by-step procedure is described in this thesis.
- Proposal of an algorithm to improve the curve fitting approach for predicting the failure time in the scope of the studied samples.
- Confirmation of the accurate and robust performance of the proposed prognostic algorithm. This analysis is based on the availability of the input data, in which the prediction's performance is evaluated. The accuracy is based on the error of the predictions while the robustness is expressed as the convergence of the predictions, measuring how fast the predicted breakdown time converges to an approximately correct prediction.

1.4 Thesis Overview

This chapter has highlighted the maintenance issues in the current power network industry that motivates this research work, as well as giving the outline of the thesis and the contributions.

The next two chapters provide the background information for this research. Chapter 2 describes the solid insulation degradation focusing on electrical treeing in cable application. The key literature on partial discharge analysis is identified. The influence of harmonics on partial discharge activity, electrical tree growth and insulation life is given. The previous work in life prediction using electrical treeing or PD data is reviewed. Chapter 3 proceeds with the identification of techniques for feature selection, model fitting, sample validation and algorithm performance.

The prognostic model is developed in the next two chapters, which are the main contributions for this work. Chapter 4 details the process of identifying the prognostic indicators using phase-resolved partial discharge analysis (PRPDA) and pulse sequence analysis (PSA). The developed prognostic model is described in Chapter 5 with the aim of predicting the failure time of the treeing samples. The model is validated using the holdout and 4-fold cross validation technique while the performance of the model is evaluated in the aspects of accuracy and convergence.

Finally, Chapter 6 concludes and summarises the main points of this thesis, as well as identifying areas of future work, which could benefit and advance the developed model described in this thesis.

1.5 Publications

The research detailed in this thesis resulted in the following publications:

1. N. H. Aziz, V. M. Catterson, S. M. Rowland and S. Bahadoorsingh, "Analysis of Partial Discharge Features as Prognostic Indicators of Electrical treeing", in *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 1, 2017.

2. N. H. Aziz, V. M. Catterson, S. M. Rowland and S. Bahadoorsingh, "Effect of harmonics on Pulse Sequence Analysis Plots from Electrical Trees", *2014 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, Des Moines, IA, 2014, pp. 377-380.

3. N. H. Aziz, V. M. Catterson, M. D. Judd, S. M. Rowland and S. Bahadoorsingh, "Prognostic Modelling for Electrical Treeing in Solid Insulation using Pulse Sequence Analysis", *2014 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, Des Moines, IA, 2014, pp. 377-380.

4. N. H. Aziz, M. D. Judd and V. M. Catterson, "Identifying prognostic indicators for electrical treeing in solid insulation through PD Analysis", *2013 IEEE International Conference on Solid Dielectrics (ICSD)*, Bologna, 2013, pp. 152-155.

5. N. H. Aziz, M. D. Judd and V. M. Catterson, "Identifying prognostic indicators for electrical treeing in solid insulation through Pulse sequence analysis", *7th Universities High Voltage Network (UHVnet) Colloquium*, 15th – 16th January 2014, University of Surrey, United Kingdom.

6. N. H. Aziz, M. D. Judd, and V. M. Catterson, "Generating PD data from electrical treeing in silicone rubber for insulation lifetime modelling", *6th Universities High Voltage Network (UHVnet) Colloquium 2013*, 16th – 17th January 2013, University of Strathclyde, United Kingdom.

Chapter 2

Solid Insulation Degradation

2.1 Solid Polymeric Materials

Materials are commonly classified as metals, ceramics, and polymers [32], [33]. Ceramic insulators were first used in power transmission applications around 1880 before natural and later synthetic polymers took the place due to their lightweight, hydrophobic, ease of handling, reduced cost and improved contamination performance [34], [35]. The evolution of solid polymeric insulation in high voltage applications has been discussed comprehensively in [36], [37].

Generally, polymers can be divided into three types: thermoplastics, thermosets and elastomers. Thermoplastics are polymers that will melt when heat is applied and reform (harden) when cooled while thermosets on the other hand, have no melting point and are formed through an irreversible chemical reaction often referred to as *polymerisation* or curing [32]. Polyethylene (PE) was the first (1960) thermoplastic used in HV cables, replacing oil-impregnated paper insulated cables, before the thermoset version of PE, crosslinked polyethylene (XLPE), was introduced in 1963. XLPE brought improvements in mechanical and thermal properties and also an increased resistance to treeing [37]. However, the high melting point of XLPE raises environmental concerns around XLPE recycling technologies [38]. Lastly, elastomers are soft and compliant polymers that are able to experience large and reversible deformations. Since elastomers have lower cross-link density than

thermoset materials, they can be made to function as elastomers above their glass transition temperatures e.g. natural rubber, ethylene propylene rubber (EPR) and silicone rubber (SiR) [33].

Improvement on the performance of dielectric materials can be achieved through material composition, i.e. a mixture of two or more dielectric materials. Although the use of composite dielectrics have been practiced before the introduction of polymers e.g. oil-impregnated paper cable, nowadays, the technology has evolved to the nanocomposites with superior thermal, electrical and mechanical properties that elevate commodity plastics to engineering plastics [39], [40].

This research aims to study the electrical treeing characteristic in epoxy resin material and correlate it with the corresponding partial discharge measurements for the development of a degradation model. Epoxy resin is a thermoset polymer thus requiring a hardener as the curing agent [39]. Epoxy resins have excellent electrical insulation properties as well as high thermal resistance, hence they are mostly used for casting and moulding in high voltage applications, for example in bushings, cable accessories, instrument transformers and gas-insulated substation (GIS) spacers [41], [42]. The transparency of the epoxy resin permits a non-destructive approach for tree growth monitoring, aiding the investigation of this research work.

2.2 Ageing Phenomenon

Solid insulating materials hold an important role in electrical equipment although sometimes with a combination of liquid (paper-oil) and gaseous insulations (gas insulated) especially when operating at very high voltage [43]. During their service life, the insulation systems are subjected to ageing mechanisms thus suffering from degradation and deterioration. Many papers [44]–[48] have defined insulation ageing in accordance with the earlier International Electrotechnical Commission (IEC) [49] and Institute of Electrical and Electronics Engineers (IEEE) [50] standards, i.e. “an irreversible deleterious change to the serviceability of insulation systems”. In addition, a more specific definition can be found in [51] which defines ageing as the

“reflection of changes of a material’s or a system’s electrical and physical properties with time while exposed to a variety of stresses and/or environment”.

The ageing factors can be either individual or a combination of stress factors. Thermal, electrical, environmental (or ambient) and mechanical (TEAM) stresses are classified as intrinsic ageing in [12], [50]. The extrinsic ageing on the other hand, refers to physical of the material itself i.e. contaminants, defects, protrusions and voids, [50] that are unintentionally introduced during material processing, transportation, installation, or in service [52].

In practice, the multifactor ageing is more realistic than the individual stress. The synergy effect when two or more stresses are present or applied can be either direct or indirect interaction [44], [45], [53]. The former results in a different effect when the stresses are applied simultaneously compared with sequentially applied stresses, while in the latter the order makes no difference. Temperature and voltage stresses are regarded to happen simultaneously resulting in direct interaction that causes a faster degradation than individual applied stress [54]–[56]. Adding more stress factors e.g. vibration (mechanical) stress, as expected, reduces the insulation’s lifetime more, however, the ageing rate is the concern of research in [57], [58].

All of the previous examples are considered as bulk degradation processes or macroscopic approaches which can aid the study of long-term life prediction [59]. In contrast, the microscopic approach concentrates on a local area of a whole insulation system, e.g. partial discharge and electrical treeing.

This research work aims to correlate the growth of electrical treeing with the corresponding PD data. The extrinsic ageing factors usually result in localised modifications in material structure. Interaction between physical and electrical ageing factors yields four ageing mechanisms i.e. space charge, partial discharge, water treeing and electrical treeing. Electrical treeing is often regarded as the final stage in electrical ageing resulting from either partial discharge or water treeing [60], [61]. Space charge could influence partial discharge activity through alteration of the local electric field thus affecting the electrical treeing indirectly.

2.3 Space Charge

Space charge can be defined as excessive charges that accumulate in the bulk of the dielectrics, at the interfaces between conductors and dielectrics, or at the interfaces between different dielectric materials [62], [63]. Space charge occurs when the rate of charge accumulation differs from the rate of charge leaving the insulation. This is mostly associated with mobile and trapped charges, which can be explained in terms of electrons, holes and ions, depending upon the mechanism of charge transfer [64].

During the application of an electric field, charges are injected into the insulating material by the electrodes [65]. However, due to the existence of trapping sites in polymer, most of the charges may be localised and not contribute to the current flow [66]. Most of them are trapped in shallow traps compared to deep traps [65], where the former are due to physical defects (e.g. void) while the latter are due to chemical defects or/and impurities [67]. These trapped charges are called space charges.

In alternating current (AC) applications, detrapping and recombination between mobile injected charges with trapped charges of opposite polarity may occur in the shallow traps during the reverse half-cycle [65]. Over a period of time, the residual of the trapped charges or space charges may move progressively to deeper traps and modify the localised electric field around the trapping sites [62]. Such field distortion can be significant. The enhanced electric field that is beyond the design electric field strength can increase the local conductivity and potentially accelerate the degradation of the dielectric material, and even lead to breakdown [62], [68].

In comparison to direct current (DC), space charge tends to be less of an issue under AC excitation, and hence attracts less research attention [69]. Findings in [70], [71] show that only a small amount of space charge is accumulated under AC electric fields, and the amount was significantly reduced under high frequency conditions. At high frequencies, the very short duration of polarity change limits the accumulated amount of space charges [72]. Nevertheless, research in [72], [73] found contradictory results. It is suggested in [72] that different material of electrode and insulator could be a factor contributing to the inconsistency in the literature. Hence,

the role of space charge under AC fields cannot be completely ignored. The effect of space charge in PD initiation is vital since PD can lead to electrical treeing and eventually breakdown of the insulation.

The presence of space charge has also been associated with the occurrence of electroluminescence (EL) prior to partial discharge inception. EL is a phenomenon that occur when an insulating material emits light in response to recombination of charge carriers of both polarities [74], while the remaining trapped charges becoming space charges. The same process is repeated for every cycle of the AC voltage resulting in the light emission of EL and a polarity reversal of the space charge. EL does not contribute to damage formation but it can be an indicator whereby charge injection may lead to deterioration of the polymeric structure [74]. Hence, it is suggested in [67] that the EL technique is a valuable tool to evaluate the dielectric properties of novel insulating materials, such as nanodielectrics.

2.4 Partial Discharge

IEC 60270 defines PD as "a localized electrical discharge that only partially bridges the insulation between conductors and which may or may not occur adjacent to a conductor" [75]. PDs that normally happen at defect sites, such as voids, cavities, contaminants, and cracks, are referred to as internal discharges while PDs in between the edges of a conductor and the surface of insulation are called surface discharges [76]. Both discharges can cause progressive degradation to the insulation through electrical treeing and tracking phenomenon respectively.

In the case of epoxy resins, the curing process may form gas-filled cavities within the insulation which can be either due to air leaking into the epoxy mould, or due to insufficient pressure on the epoxy liquid [77]. Early investigations revealed that PD characteristics are affected by the shape, size and location of defects and by the thickness and type of insulation [76]. The defects can affect insulation performance in the long term but not an immediate breakdown. Hence, PD is an indicator of the presence of defects in an insulating system.

Partial discharge and space charge are closely related. The trapped charges within a material may consist of space charges and charges deposited by earlier PD events which serve as initiatory electrons. Therefore, the PD itself also acts as a source of charge injection into the dielectric. The presence of these charges gives rise to the local electric field enhancement, thus lowering the PD inception voltage (PDIV) that is the lowest voltage which must be applied to initiate PD in insulation [76]. When voltage is reduced, the voltage at which PD ceases is called the PD extinction voltage (PDEV) [76]. During the discharge process, a portion of the discharge is trapped in the cavity surface and some migrates deeper into the dielectric. It was mentioned in [78] that a later stage of the degradation process deposits a PD by-product i.e. crystals where the ignition of PD occurs at the crystal tips.

A general approach for an automated PD classification is given in [79] as shown in Figure 2-1. The first stage of the system is the measurement of PD including sensors, data acquisition and preprocessing, mainly for detecting the PD signals. In the latter stage of PD measurement, the captured signal is digitised and purified through denoising. The next stage is the representation of PDs either in the form of phase-resolved, time-resolved or pulse sequence. At this stage, classification of PD defects is possible but with low accuracy (i.e. the pattern might be classified to more than one potential defect). A more accurate classification can be achieved through feature extraction by identifying possible descriptors for different aspects of the discharge pattern. The defect classification is made by comparing the extracted features with a defect database that is mainly generated from the knowledge of the experts.

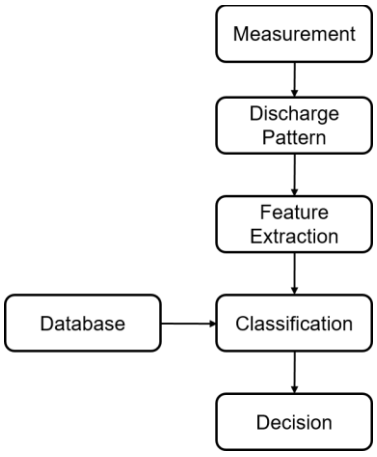


Figure 2-1. A general approach for PD classification [79]

2.4.1 PD Detection

PD monitoring allows the observation of trends in PD activity for failure diagnosis and prognosis of high voltage plant. The presence of PD may indicate a need for maintenance and as such is important information for the asset manager. PDs are often accompanied by emissions of sound, light, heat and chemical reactions thus there are various methods to detect the presence of PD [75]. This includes electrical, acoustic, thermal, chemical and ultra-high frequency (UHF) monitoring.

Guidelines for electrical detection for PD measurement are available in the IEC 60270 standard, depicting the maturity of the technique [75]. The original PD current pulses are characterised by a very short duration (nanoseconds) thus it is not easy to capture the shape of such pulses. Therefore, the current-time integral (charge of the captured PD pulses) is measured instead of the peak value of the PD current pulses, giving the apparent charge levels in picocoulombs (pC), which is obtained from the transient voltage drop across the test object terminals [80].

IEC 60270 recommends three basic measuring circuits which differ by the arrangement of the measuring impedance, Z_m [75]. Figure 2-2 shows the most common circuit employed in practice where Z_m and the coupling capacitor, C_k, are connected in parallel with the test object, C_a. During PD occurrence, the voltage across C_a decreases momentarily due to the voltage drop across the HV source impedance, Z_n. As a result, a transient current flows through Z_m, allowing voltage to be measured across it. A measuring instrument (MI) is then used to identify the apparent charge from the voltage change. The scale factor of these two parameters is determined through a calibration procedure by repetitively injecting a short duration current pulse of known charge magnitude into the terminals of the test object.

This conventional method requires a controlled environment hence, is not suitable for on-site PD measurement. The electromagnetic interference and electrical noise produced during the operation of HV equipment may leads to false detection of PD [81]. In cable systems, distortion of PD pulses occurs as they propagate away from the PD source [82]. It is a challenge to avoid the attenuation and dispersion of the signal that affect the measurement of PD.

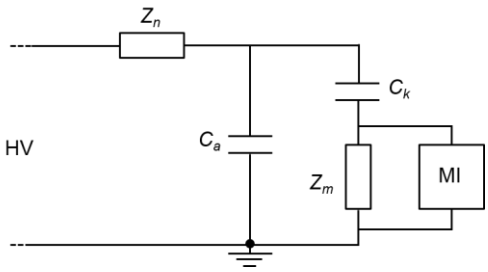


Figure 2-2. Most common PD measuring circuit recommended in IEC 60270 [83]

The UHF method is an alternative with an improved signal-to-noise (S/N) ratio. This method was first introduced for gas-insulated switchgear and then was applied to power transformers [84] and cable terminations (on-line) [85] for identification of defect type and location. The main advantage of this method is its ability to identify the location of the PD by using multiple UHF sensors and the "time-of-flight" technique (measuring the time difference for the PD signal to reach each of the sensors) [84]. These sensors are placed to surround the PD source, and PD can be detected by measuring electromagnetic emissions originating from transient currents of PD in the 500-1500 MHz range. The generated UHF signal is detected by the coupler that produces an output in the form of an oscillatory voltage signal. In most circumstances, the magnitude of the UHF signal was found to be dependent on the current pulse magnitude thus can be represented as the equivalent PD magnitude [86].

In comparison to UHF sensors, acoustic emission (AE) sensors are cost effective, simple, easy to install and are insusceptible to external electrical and electromagnetic interference [87]. The basic principle of this method is the detection of the mechanical energy wave that propagates from the discharge site through the insulation. However, the application of this method is limited due to the complicated nature of the acoustic propagation pathways. The attenuation of the acoustic wave is high thus requires an appropriate level of sensitivity [88]. The guidelines for both UHF and AE methods are available in the IEC 62478 standard [89].

Finally, chemical detection is the exploitation of the changes in chemical composition due to PD activities, and is mostly used in GIS and power transformers rather than power cables [88]. Dissolve gas analysis (DGA) and high performance

liquid chromatography (HPLC) are the common methods for this approach. DGA quantifies the composition levels of different gases in insulation due to decomposition of the material when subjected to thermal and electrical stresses [90]. HPLC on the other hand, measures the by-products (e.g. glucose) instead of the dissolved gasses. A common drawback related to both methods is the lack of information on the nature, intensity, and location of PD [88].

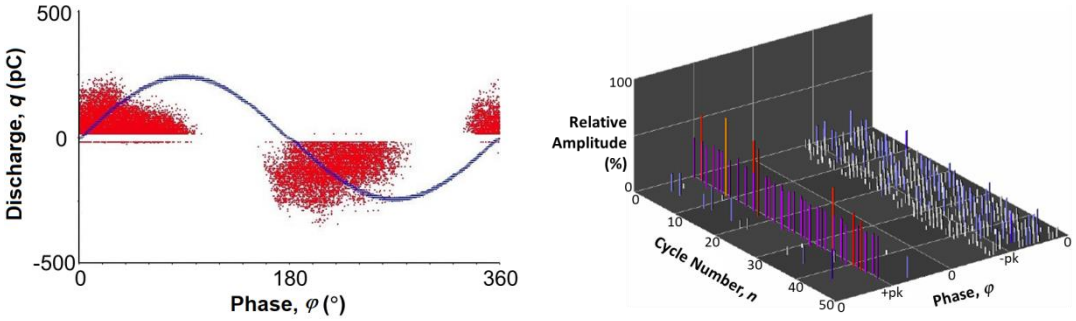
2.4.2 PD Data Representation

The PD data can be represented in either phase-resolved or time-resolved format. The most common approach has traditionally been through the PRPD representation in which the variation of PD pulses (either in magnitude or repetition rate) is often represented by statistical quantities [91]–[94]. Later in 1990's, a more meaningful interpretation of PD phenomena was introduced in PSA approach [95]. PSA examines the relationship between two consecutive PD pulses, which relates to the physical processes occurring within the localised degradation region.

2.4.2.1 Phase-Resolved

Phase-resolved PD data are acquired based on the phase angle of the AC test voltage waveform. For electrical detection, three basic quantities of the PD pulse are quantified at the measurement stage over a predetermined time duration: phase angle occurrence, φ, charge, q and voltage cycle occurrence, n. The presentation of this data is commonly known as φ–q–n or PRPD patterns. The UHF detection method on the other hand, quantifies the output voltage from the UHF coupler instead of discharge magnitude.

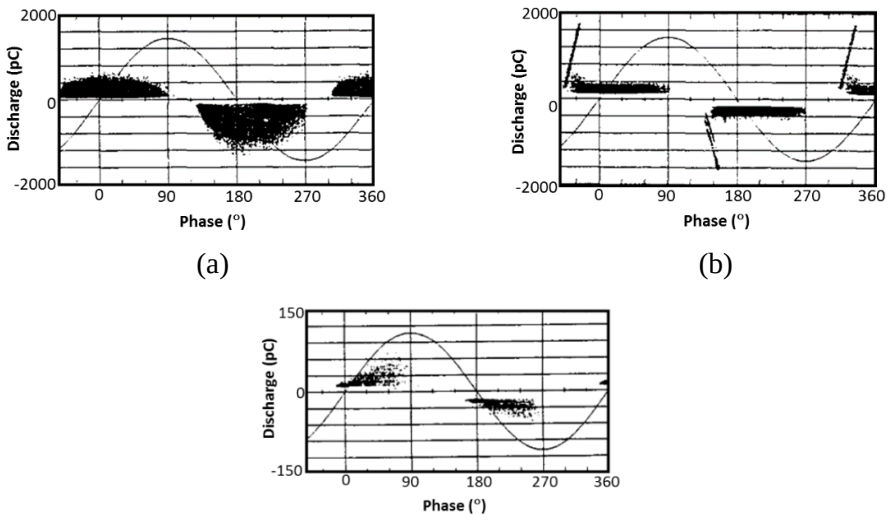
Figure 2-3a and Figure 2-3b illustrate the PRPD pattern from electrical and UHF detection respectively for different PD sources. The scatter plot in Figure 2-3a shows the distribution of apparent charge, q, on φ, unrevealing the information of n. Alternatively, the three-dimensional (3D) plot in Figure 2-3b allows the observation of relative amplitude of PD pulse (obtained from the pulse captured by the UHF sensor) on φ and n. This plot displays the PD activity in 50 cycle bursts which is the total data in 1 second and is used for data interpretation [96].



(a) (b)

Figure 2-3. Example of PRPD pattern in (a) two-axes and (b) three-axes from different PD sources.

Special terms have been used to describe the PRPD pattern and relate them to the nature of PD. Figure 2-4 shows the typical PRPD pattern of internal discharges. Observation on the PRPD pattern of a test sample with a void reveals a transition of a "turtle-like" pattern (described by the flat top shape in Figure 2-4a) into a "rabbit-ear" like pattern (higher discharge magnitude at the earlier phase as shown in Figure 2-4b) [97]. This can be discriminated from electrical treeing that has a right-angled triangle shape or "wing-like" pattern as shown in Figure 2-4c [98], [99]. These descriptions are considered in the feature extraction stage for more accurate and detailed diagnosis of the PD defects.



(a) (b)

(c)

Figure 2-4. Description of PRPD pattern [97]: (a) "Turtle" like pattern, (b) "Rabbit-ear" like pattern, and (c) "Wing" like pattern

Before extracting possible features from φ–q–n pattern, three different two-dimensional (2D) phase distribution graphs are constructed in [100] as shown in Figure 2-5. This requires a predetermined phase window which depends on the intensity of the PD data. Dividing one full cycle (360°) into phase windows gives a number of phase buckets. All the PD pulses within each phase bucket are then represented by three quantities, the maximum charge, H_{qm}, the average charge, H_{qn}, and the number of charge, H_n. Plotting these quantities onto phase values allows various statistical features to be extracted in order to characterise the defect type.

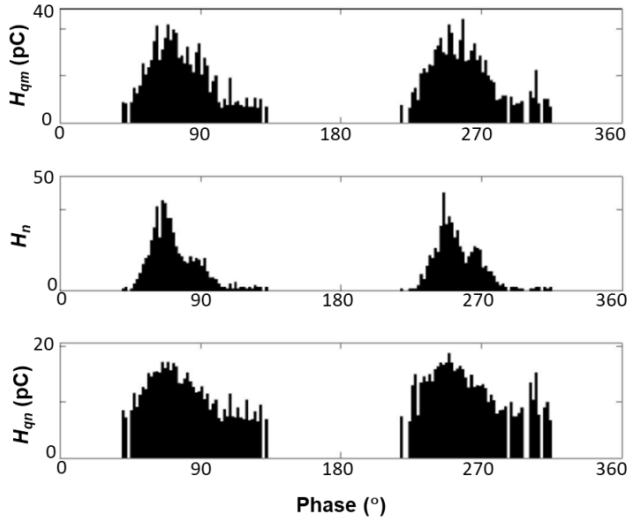


Figure 2-5. 2D examples of PD maximum pulse height, H_{qm}, PD pulse count, H_n and PD mean pulse height, H_{qn} [101]

2.4.2.2 Time-Resolved

Recently, the high frequency current transformer (HFCT) has been widely used as an online PD detector in power cables for PD identification and discriminating interference [102], [103]. The HFCT detects the original PD pulse with the pulse shape as represented in time-resolved format as shown in Figure 2-6 with the following parameters [104]:

- Pulse rise time, t_r: time required to rise from 10% to 90% levels of the peak value.
- Pulse decay time, t_d: time required to decay from 90% to 10% levels of the peak value.

- Pulse width, t_w: time interval between 50% levels on both sides of the peak value.

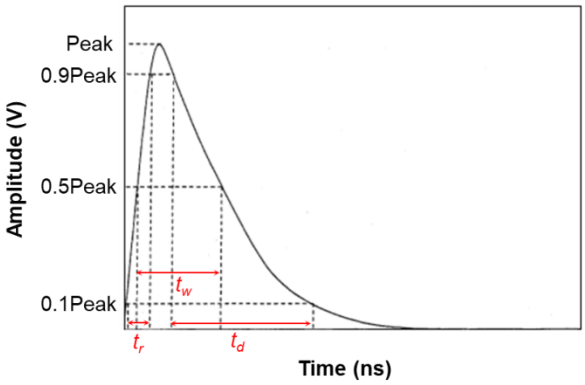


Figure 2-6. Typical parameters describing the shape of an ideal PD pulse [104]

Previous research has revealed the direct relationship between the physics of the PD defect and the shape of the signal [103], [105]. This is due to the difference in PD pulse signal when generated by different PD sources. Not only that, this approach requires a less expensive measurement system compared to phase-resolved measurements. However, a distorted signal maybe expected especially for a long cable that increases the travel path between the PD site and the terminals of the test object [80].

2.4.2.3 Pulse Sequence

Pulse sequence analysis (PSA) treats PD pulses as events within a sequence. The rationale behind this approach is that the history and condition of a sample, including recent discharge events, influence the ignition and nature of the next discharge pulse [96]. In particular, key governing parameters of each discharge are the local electric field and its change from the last pulse, which are both dependent on the voltage difference between consecutive pulses. The voltage differences do not occur at random but in specific sequences characterising the discharge processes in the defect, at least in part due to the build-up of space charges [96].

Figure 2-7 shows the basic principle of the PSA approach, where the solid circles represent three PD pulses within the reference cycle numbered as 1, 2 and 3. Three parameters were introduced in [96] for representing the changes in consecutive PD

pulses: the instantaneous voltage, *u*, the voltage difference, *du* and the voltage derivative with respect to time, *du/dt*. Considering consecutive PD pulses of 1 and 2, *du* and *dt* of pulse 1 can be determined using equations 2-1 and 2-2 where *n* = 1. Both equations can then be used to calculate *du/dt*.

$$du_n = u_{n+1} - u_n \tag{2-1}$$

$$dt_n = t_{n+1} - t_n \tag{2-2}$$

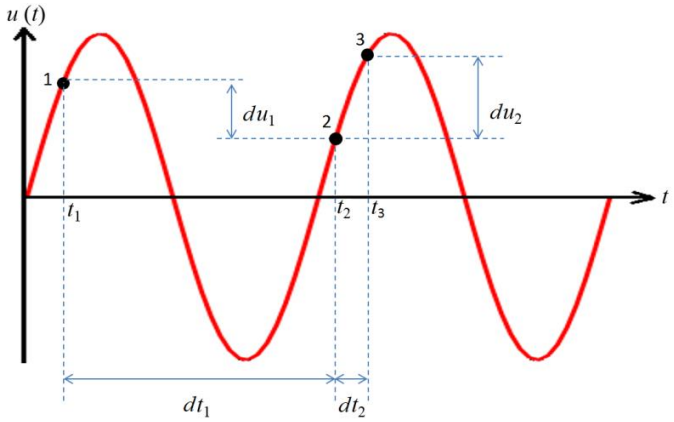


Figure 2-7. Basic principle of PSA [95]

The three parameters are commonly presented in a scatter graph with consideration of the previous (*x*-axis) and current (*y*-axis) PD pulses. Several papers have used these plots to identify the nature of the PD [96], [106], [107]. For example, consecutive PDs from surface discharge form six clusters in the *du* plot of Figure 2-8a while PDs from a void have an extra two clusters giving eight clusters all together in Figure 2-8b [106].

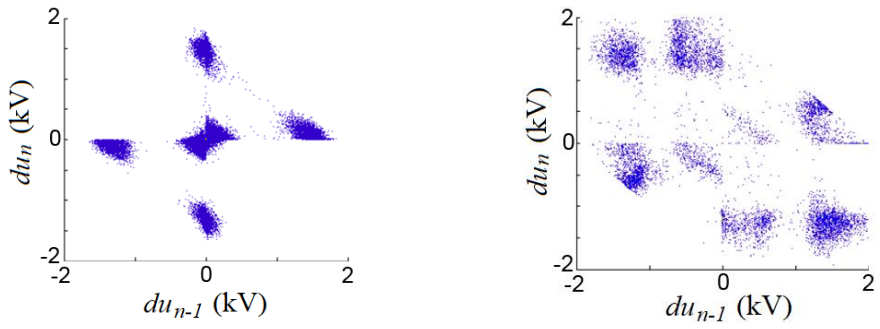
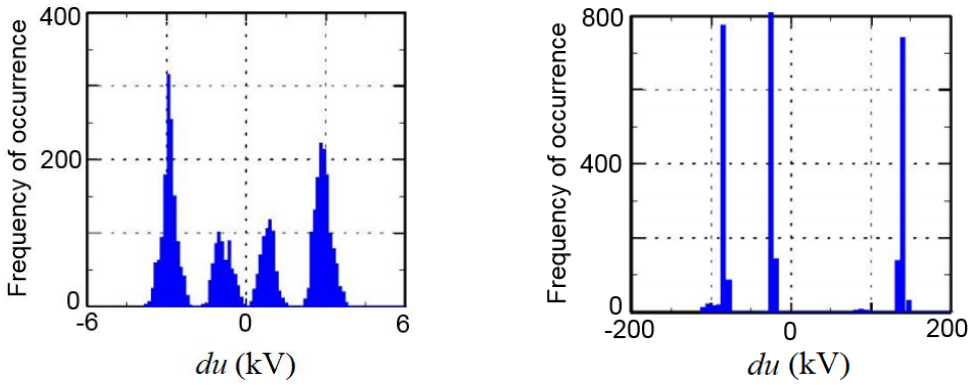


Figure 2-8. Example of *du* plots considering consecutive PD pulses resulting from (a) surface discharge and (b) void [106]

Alternatively, the PSA parameters are presented as a histogram of the number of occurrence. Investigation in [108] found that occurrence of *du* in electrical treeing is concentrated in four characteristic values as can be seen in Figure 2-9a which can be considered as a systematic shift. As a comparison, surface discharge yields only three values of *du* as shown in Figure 2-9b. This non-symmetric behaviour is due to different PD magnitudes in the positive and negative half cycles.



(a) (b)

Figure 2-9. Example of accumulated *du* plots resulting from (a) electrical treeing and (b) surface discharge [108]

2.4.3 PD Feature Extraction

Different PD sources result in their own PD pattern due to the geometry, location in insulation, dielectric properties and applied electric field [109]. The unique PD characteristics are commonly described in terms of informative features by experts, that is, what they believe are the important variables to build a model. In addition to feature extraction, the *curse of dimensionality* can be a problem due to the large number of variables [110]. As the number of features increases, the amount of data needed to support the result grows exponentially. Hence, feature selection or feature reduction is required to discard the irrelevant or non-informative features. The techniques are mostly generic and not limited to PD data, thus will be discussed in the next chapter. In this research work, the available PD data is in phase-resolved format, therefore, only features of the phase-resolved pattern will be discussed here.

2.4.3.1 Statistical Parameters

The effort to improve the interpretation of PD data for diagnosis was initiated by Gulski with the implementation of statistical parameters for correlating PRPD pattern with the type of defect [91]. The statistical variation can be observed either in magnitude or in the phase of both positive and negative half cycles of the 2D phase distributions mentioned previously in Section 2.4.2.1. Thus, the three 2D distributions as shown in Figure 2-5 can now be expanded and expressed as $H_{qm}^{+}(\varphi)$, $H_{qm}^{-}(\varphi)$, $H_{qn}^{+}(\varphi)$, $H_{qn}^{-}(\varphi)$, $H_{n}^{+}(\varphi)$ and $H_{n}^{-}(\varphi)$ where the positive cycle ranges from 0° to 180° while the negative cycle ranges from 180° to 360°. The correlation between positive and negative half cycles was studied in [91] using the features in equations 2-3 to 2-6. As a final feature, the correlation factor in equation 2-6 is modified as in equation 2-7 to include the symmetrical measurement of discharge.

- Discharge symmetry, Q_s

$$Q_s = \frac{\sum_{i=1}^{N_q^-} q_i^- \Big/ N_q^-}{\sum_{i=1}^{N_q^+} q_i^+ \Big/ N_q^+} \tag{2-3}$$

  where q_i^- and q_i^+ are the discharges in the negative and positive half cycle respectively while N_q^- and N_q^+ are the number of discharges in the negative and positive half cycle respectively. Discharge symmetry examines the magnitude variation over the two phases and checks which of the two half cycles is experiencing the greater magnitude.

- Phase inception symmetry, Φ

$$\Phi = \frac{i_{\varphi_{inc}^-}}{i_{\varphi_{inc}^+}} \tag{2-4}$$

  where φ_{inc}^- and φ_{inc}^+ are the phase inception in the negative and positive half cycle respectively and *i* is the index number of the phase inception. The initial pulse of PD activity in each half cycle infers the phase inception.

Phase inception symmetry compares the position of the initial PD pulses between half cycles. For the positive half cycle, the value of $i_{\varphi_{inc}^{+}}$ is the same as φ_{inc}^{+} thus ranges from 1 to 180°. This range also applies to $i_{\varphi_{inc}^{-}}$ in the negative half cycle which can be determined using equation 2-5.

$$i_{\varphi_{inc}^{-}} = \varphi_{inc}^{-} - 180°, \qquad 180° < \varphi_{inc}^{-} \leq 360° \tag{2-5}$$

The phase inception is symmetry when $i_{\varphi_{inc}^{-}}$ equals to $i_{\varphi_{inc}^{+}}$ giving Φ = 1. Hence, the nearer Φ is to 1, the more symmetrical the phase inception.

- Cross-correlation factor, *cc*

$$cc = \frac{N_p \sum xy - \sum x \sum y}{\sqrt{\left[N_p \sum x^2 - \left(\sum x\right)^2\right]\left[N_p \sum y^2 - \left(\sum y\right)^2\right]}} \tag{2-6}$$

where x is the discharge in the positive half cycle, y is the discharge in the negative half cycle and N_p is the number of phase buckets per half cycle. The discharge can be either H_{qm}, H_{qn}, or H_n. This measurand is used to evaluate the difference in shape of the positive and negative half cycle.

Finally, the *cc* is modified to include the discharge and phase asymmetry factors, *Q* and Φ, as represented in equation 2-7.

- Modified cross correlation factor, *mcc*

$$mcc = \Phi.Q.cc \tag{2-7}$$

The evaluation of the 2D PRPD shape with respect to a Normal distribution is performed using skewness and kurtosis in equations 2-10 and 2-11 respectively, with the mean and standard deviation first computed using equations 2-8 and 2-9 where x_i is a discrete value and N is the number of the discrete values.

- Mean, μ

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2-8}$$

- Standard deviation, σ

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2} \tag{2-9}$$

- Skewness, *Sk*

$$Sk = \frac{1}{N\sigma^3}\sum_{i=1}^{N}\left(x_i - \mu\right)^3 \tag{2-10}$$

  The skewness describes the symmetry of the 2D distributions. Positive skewness indicates that the distribution is asymmetric to the left while negative skewness indicates asymmetric to the right. Zero skewness represents a symmetric distribution.

- Kurtosis, *Ku*

$$Ku = \left[\frac{1}{N\sigma^4}\sum_{i=1}^{N}\left(x_i - \mu\right)^4\right] - 3 \tag{2-11}$$

  The kurtosis describes the sharpness of the distributions. Zero kurtosis indicates a Normal distribution, positive for sharp distribution and negative for flat distribution.

By taking the skewness as an example, the PD from a cavity with and without electrical treeing yields a positive and negative skewness respectively in [91] thus could be a reference feature for PD diagnosis.

2.4.3.2 Waveshape Descriptors

Strachan and Rudd et al. [92], [93] have improved on Gulski's approach by adding more features of the PRPD pattern, not limited to the statistical features. These key features highlight the underlying physics occurring at the site of the discharge and

could only be identified by experts based on their experience. Therefore, research in [92], [93] has interviewed experts in order to construct the expertise model.

The hierarchical model of the identified features for PD diagnosis is shown in Figure 2-10. This model includes all the possible features that have been highlighted by the experts as important features of the PRPD pattern, along with their subcategories. These features are called descriptors, where the asterisk (*) indicates that derived features were originally from Gulski's list [111].

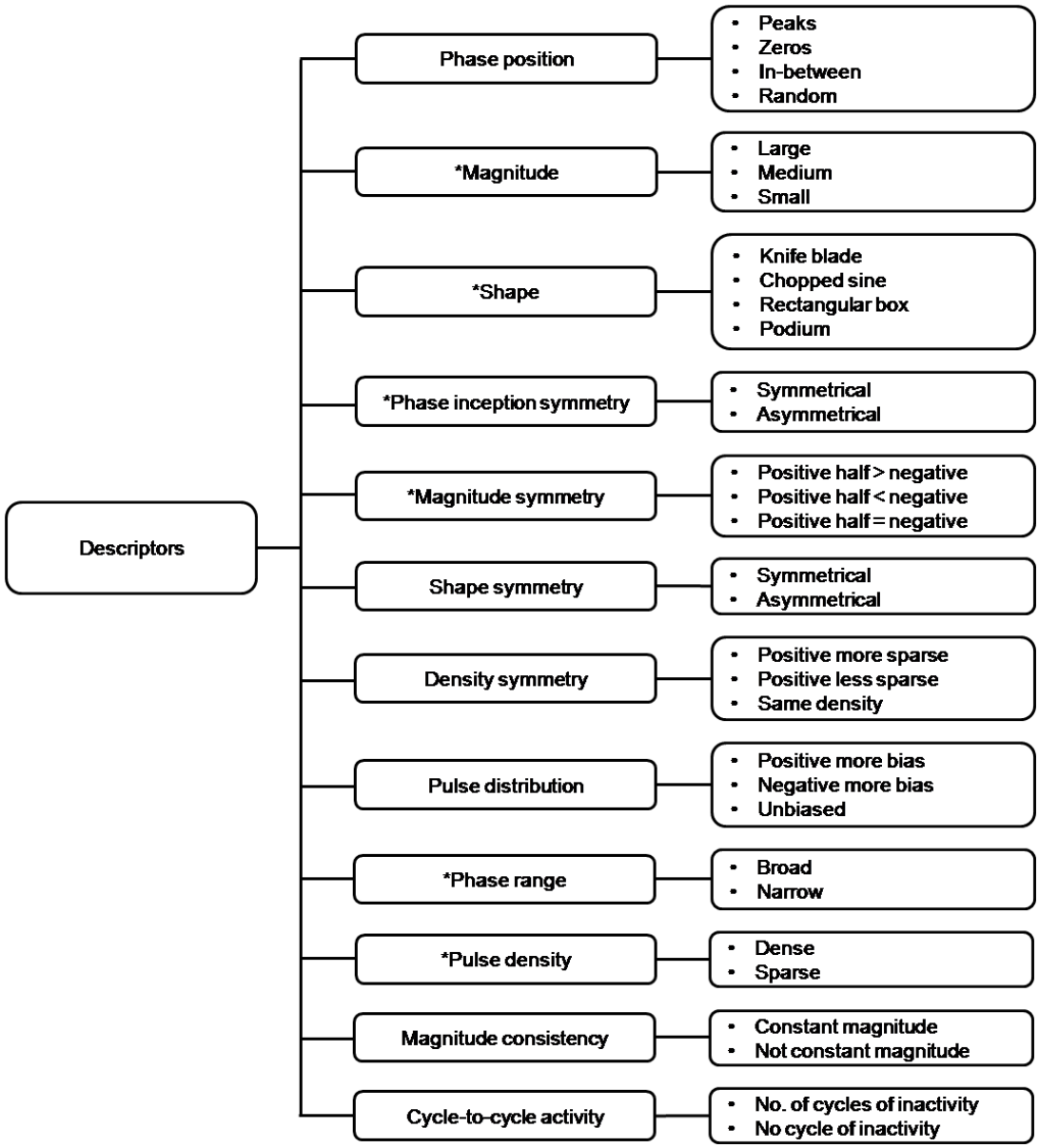


Figure 2-10. Descriptor hierarchy for PRPD pattern [112]

Not all the twelve descriptors are used to classify each of the defects. Different PD defects may be described with the same descriptors but at least one descriptor will differentiate the PD defect characteristics. For example, rolling and bouncing particle are described by phase position, shape, phase range and density. Both defects fall in the same subcategories for the first three descriptors. Thus, only the feature of the density descriptor distinguishes the two defects.

2.4.3.3 Weibull Parameters

Weibull analysis was originally applied to sample failure data to derive a mathematical model for the lifetime distribution of the sample [113]. Equation 2-12 indicates the two-parameter Weibull distribution that has been applied for the insulation life model in [114]. The two parameters are α and β which determine the scale and the shape of the distribution respectively.

Alternatively, a number of research publications [115]–[118] have applied the Weibull distribution to PD data by replacing the failure time, t in equation 2-12, with the discharge pulse height (discharge magnitude) q, yielding the probability distribution of PD pulse rate, $F(q)$ instead of probability of failure $F(t)$ as can be seen in equation 2-13. It is shown in [117] that the different discharge sources can be identified through the β and the identification holds even when two sources are applied simultaneously [115], [117]. For the latter case, a five-parameter Weibull function has been found to permit the separation of PD sources [115], [117] but will not be discussed further here. The finding in [118] however, shows that α gives better identification than β using the summation and the ratio of α in both positive and negative half cycles i.e. $\alpha^+ + \alpha^-$ and α^+/α^-.

$$F(t) = 1 - \exp\left[-\frac{t}{\alpha}\right]^{\beta} \tag{2-12}$$

$$F(q) = 1 - \exp\left[-\frac{q}{\alpha}\right]^{\beta} \tag{2-13}$$

2.4.3.4 Image Processing Tools

The statistical analysis discussed previously was mainly applied to the 2D PRPD pattern while the Weibull analysis only needs a single parameter. To fully utilise the three φ–q–n parameters of PRPD pattern, the 3D plot is regarded as an image on which image processing algorithms can be used to extract the distinguishing features. A literature survey in [104] has given three image processing approaches that have been applied to the PRPD data. The first two have been used to extract features from the 3D pattern, and the last approach is an image decomposition technique to separate the individual PD pattern from a multi-defect PD pattern.

The first approach is the texture analysis algorithm that primarily investigates the grey level variation in images. The basic principle is to divide the image into M x N resolution pixels. Originally, the description of an image is based on the grey level values of the pixels. In order to apply the texture analysis algorithm for PD feature extraction, the grey level values were replaced by the pulse magnitudes, q whereas the M and N are represented by the phase divisions and AC voltage cycles respectively. Examples of potential PRPD image features reported in [119] are homogeneity, heterogeneity, local variation, local similarity and average value of image grey level. For PD source identification, each image feature is extracted from each of the tested PD defects at different levels of the voltage cycle. Features with high and constant variance for each tested voltage cycle would be selected for PD classification.

The complex nature of the 3D PRPD pattern can be treated as a fractal surface, thus has encouraged researchers in [120] to introduce the second approach of fractal features for interpreting the pattern. The authors defined fractal as "any shape where the parts, when magnified, reveal as much detail as the whole". By assuming a stick with length l is used to measure the length of (let us say) a coastline, where $N(l)$ denotes the actual length of the coastline, it could be expressed (as a power law) by a parameter d_f, called the fractal dimension as in equation 2-14 where K is a constant.

$$N(l) = Kl^{-d_f} \tag{2-14}$$

For a curving surface such as a 3D PRPD pattern, square boxes (l x l) are used instead of the stick. The area where the shape resides is divided into the square boxes, and the number of boxes that contain part of the shape are counted. The value of d_f can be estimated from the slope of a line fit to the –log ($N(l)$) versus log (l), for different values of l. In addition to d_f as the fractal feature, the lacunarity, Λ is also extracted to quantify the gaps or lacunae present in a given surface. These two fractal features represent the 3D pattern, where d_f is used to quantify the surface roughness while Λ gives a measure of the denseness of the fractal surface. This approach has shown its capabilities in discriminating patterns from different PD defects [120]–[124] with an advantage of less extracted features (two fractal features) compared to the texture analysis. Figure 2-11 shows an example of pattern discrimination reported in [120].As can be seen, patterns from the same class lie close to each other and separated from other classes.

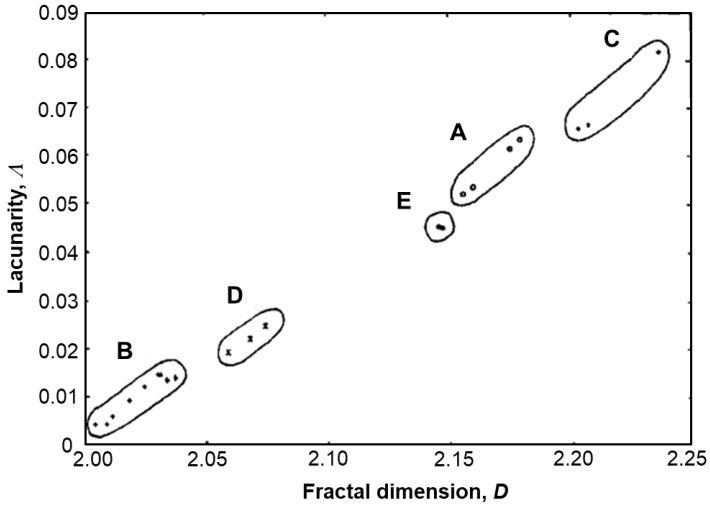


Figure 2-11. Cluster representation on the feature plane showing pattern discrimination capabilities [120]

The final approach is the wavelet-based image decomposition technique that can be used to identify individual PD sources present in multi-defect PD patterns. It was mentioned in [125] that the resultant PD pattern from more than one defect is an overlapping of the individual defect patterns. The effectiveness of the multi-defect PD recognition thus depends on the degree of the overlap. The multiresolution signal decomposition (MSD) technique of the wavelet transform has been successfully used for image decomposition.

In the application of the 2D PRPD pattern, for example, the $H_n(\phi)$ distribution, the number of PD pulses (vertical) and the phases (horizontal) are treated separately for decomposition. The outputs of the decomposition are then reconstructed yielding four subimages i.e. one approximate image and three detailed images of vertical, horizontal and diagonal. Investigation in [125] reveals that only the vertical and horizontal images represent salient features of the individual sources in a separable form hence could aid the identification of the individual defect from a multi-defect pattern.

2.4.4 PD Classification

The classification of a PD pattern to type of the defect requires either the knowledge from experts based on their experience or a database of historical PD patterns corresponding to particular defects. In the former case, with consideration of waveshape descriptors in Figure 2-10, the formulation of the descriptors are based on the PD behaviour determined from the experts. Taking one of the descriptors, *phase position* for example, the issue of space charge will result in PD at *zero* and *in-between* phase positions unless, if only minimal space charge is present and no memory effect extends beyond the half cycle, PDs will take place at the *peaks* [112]. PD pulses could also result at *random* phase positions, which is one of the characteristics of a rolling particle. The classification of a rolling particle defect using the knowledge-based approach is depicted in Figure 2-12. Four descriptors are identified to be affected by the defect i.e. *phase position*, *phase range*, *shape* and *pulse density*.

In the case of historical data, either individual or hybrids of classical and artificial intelligence (AI) approaches are used to learn the data for PD classification. The distance classifier [126]–[130], artificial neural network (ANN) [118], [128], [131]–[134], fuzzy logic [128], [135], [136], support vector machine (SVM) [128], [129], [135], [136] and decision trees [137], [138] are examples of these approaches. Since classification is not the focus of this research work, further discussion on the AI techniques will be given in Chapter 3 that concerns its employment in prognostic studies.

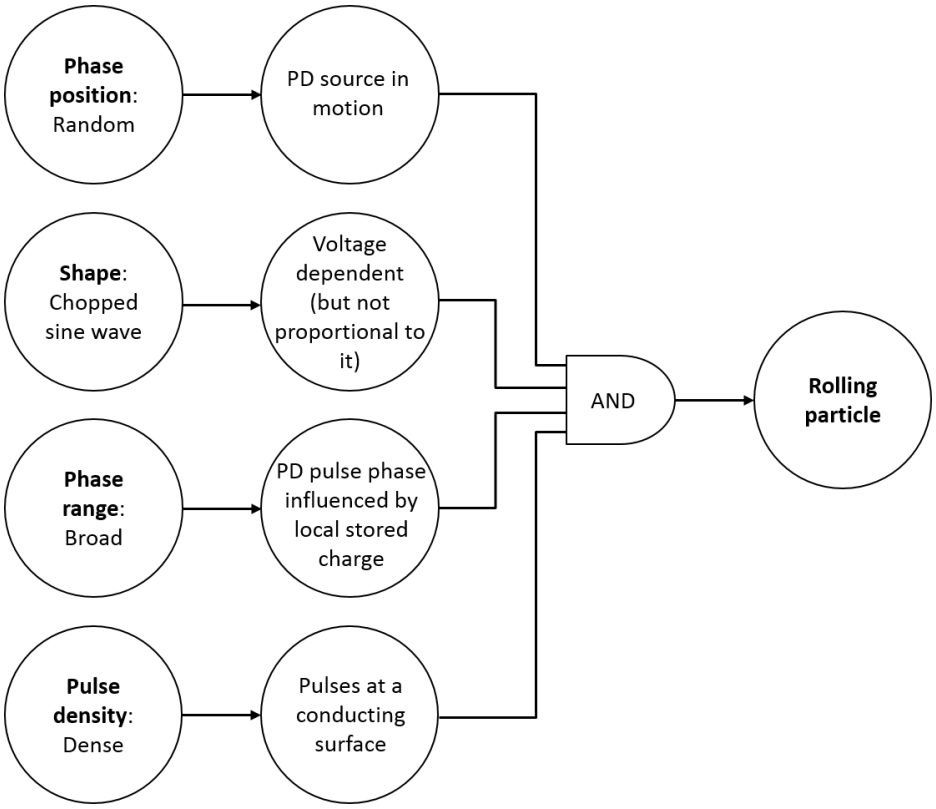


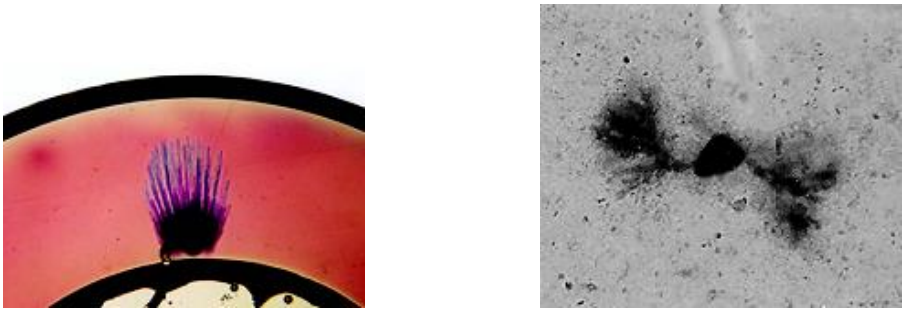
Figure 2-12. Semantic network model of knowledge flow for rolling particle defect [112]

2.5 Water Treeing

Treeing is a type of damage which progresses through a dielectric section by resembling a tree-like path. The effect of moisture has proved so important that there are two major classes of treeing, i.e. water treeing and electrical treeing. Water trees, also known as electrochemical trees, are generally observed as a dendritic pattern of water-filled microcavities in the polymer [139]. The initiation of a water tree not only requires water and electric field, but also contaminant. Under laboratory conditions, it is found to be very difficult to grow water trees in pure water [140].

The water may be present on the interface between two materials or within the insulating material. The former results in vented trees (Figure 2-13a) that have a direct contact with a reservoir of aqueous electrolyte [13] and can grow completely through a dielectric section to bridge the electrodes. The latter on the other hand, results in bow-tie trees (Figure 2-13b) which progress symmetrically from either a contaminant, boundary surface or water filled void within the insulation, where there

is limited access to an aqueous reservoir [13]. The growth of a bow-tie tree is not as significant as a vented tree since a bow-tie tree rarely grows large enough to cause electrical breakdown [141].



(a) Vented tree (b) Bow-tie tree

Figure 2-13. Examples of water tree: (a) Vented water trees growing from the conductor screen [37], (b) Bow-tie tree in clay filled EPR cable insulation initiating from a contaminant [139]

An increase in the concentration of the water molecules at the tips of water trees, as well as the applied stress to the insulation, will increase the degradation rate, hence, reducing the breakdown strength of polymeric insulation [142]. However, when a water tree fully bridged an insulator, breakdown does not necessarily occur immediately, although the breakdown strength is reduced.

The progress of water treeing is often followed by the initiation of an electrical tree [28], [142] . In this case, the transport of charges at the water tree tip creates a local field that high enough to initiate the electrical tree. Also, there is evidence that a water tree may propagate without the presence of partial discharge activity unless an electrical tree was present [12]. The electrical stress required for electrical treeing formation (< 100 kV/mm) is higher than water tree (< 10 kV/mm) [143]. Investigation in [144] reveals that after the transition of water tree to electrical tree, the PDIV dropped with respectively 50% and 44% of the original PDIV value. This highlights the severity of the electrical treeing compared to water treeing in terms of ageing.

2.6 Electrical Treeing

Electrical treeing is known to be one of the routes to cause failure in solid insulation [145]. Current developments of DC power transmission has brought concerns about electrical treeing not only in AC fields but also in DC fields, in the presence of voltage ramps, short circuit, polarity reversal, impulses or constant DC voltages [13]. Before the term of treeing was introduced, it was described as a growing pit from a void, which can be either carbonised or uncarbonised depending on the type of electrode, insulating material and stage of treeing [146]. Discharge activity in the voids erodes their surfaces, creating nonconducting pits protruding into the solid material, which later become conducting in the presence of incident discharges, raising the stress at the tip to intrinsic strength levels and creating localised breakdown.

Electrical trees may start to grow from water trees, sharp conducting particles or gaseous cavities [74]. The initiation and growth of an electrical tree are accompanied by PD activities within the developing tree-shaped hollow channels of micrometre (μm) diameter and length. Figure 2-14 outlines three distinct stages of electrical tree growth. The inception stage is characterised by a finite initiation time. Under continual AC field application, the electrical tree propagates across the insulation with a decelerating growth rate which then accelerates leading into the runaway stage before breakdown.

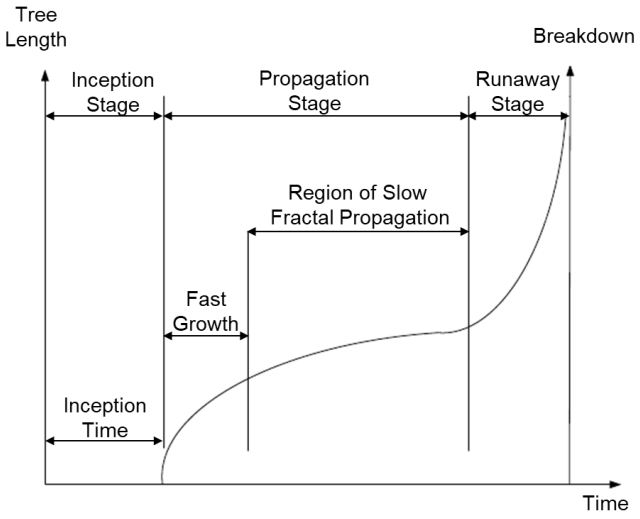


Figure 2-14. Schematic representation of electrical tree growth [13]

2.6.1 Inception Stage

Generally, the common defects in polymeric dielectrics that initiate electrical trees can be categorised into two classes: defects producing gas discharges and defects producing strong local field enhancement [147]. The former are mainly dedicated to pit formation through partial discharge activities in gas-filled voids. The voids are produced either from defects formed during production, installation and operation, from cracks and crazes due to electromechanical stress [148], [149] or dielectric heating [150], or at the interface between the electrode and the dielectric due to electrostrictive force [151]. In addition, micro cracks produced in epoxy resins during casting and cooling processes would also be filled with gas.

For a tree to be initiated, the PD activity in the cavity needs to be concentrated in a certain area [78]. This is automatically fulfilled in a prolate shaped cavity. In the case of oblate and flat cavities, the PD activity is localised by the crystal growth (solid by-product due to PD in cavity) on the insulation surface. However, the tree initiation may be intrinsic if the tip of the pit becomes highly electrically stressed. Material experiencing thermal ageing could also initiate a tree intrinsically through the oxidation mechanism [148].

The latter category of defects results from sharp point electrode protrusion either from material defects or from the tip of water tree. Tree initiation takes place by a process of electron injection and extraction from field enhancement tips, where charge carriers move back and forth repeatedly between the electrodes and the stressed dielectric [53], [151]. Trapping, detrapping and recombination mechanisms of space charge (explained in Section 2.3) contribute to the formation of small pits and electroluminescence respectively and are followed by the tree initiation. During this processes, the electrons will gain energy that enables them to ionize gas atoms when they collide and lead to electron avalanches. This intrinsic process may not result in a measurable PD, thus could not be detected through PD monitoring. Tanaka [151] divided this process into three mechanisms: field distortion charge accumulation, joule heating and oxidation.

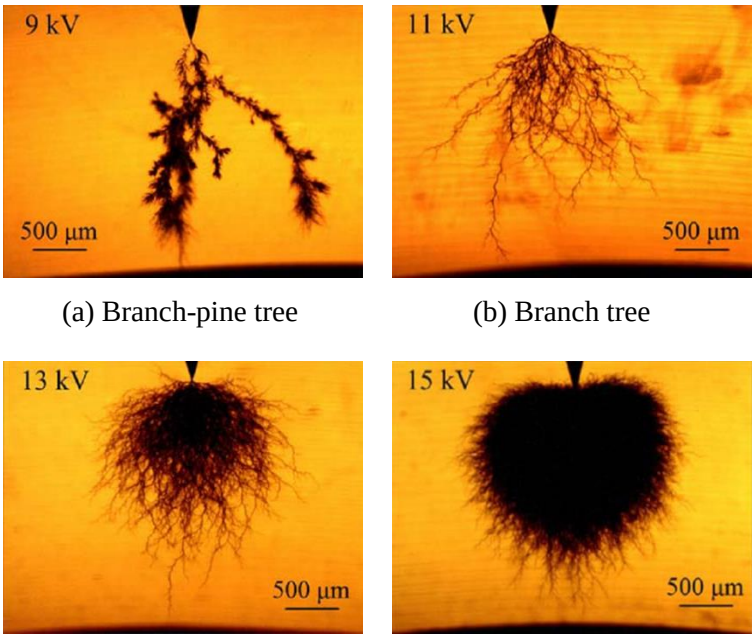
Both defect categories explained above occur during an incubation period, defined as the time required for tree initiation from the time the voltage is applied [148]. This time period is also known as inception time, as shown in Figure 2-14. Treeing inception is considered to be the time required for generating an observable tree (usually about 10 μm in length) [13]. Since the electron currents are small, no externally detectable signals occur during the initiation period [152]. Once a tree has been initiated after the incubation period, PD can be detected for both defect categories.

2.6.2 Propagation Stage

After the tree inception, electrical tree growth is driven by PD activity in the existing tree tubules. These discharges erode the insulation material and create tree-like branches. Unlike the tree initiation, the propagation process produces detectable current and visible light that can be the means for treeing detection [152].

The resultant tree depends upon various factors, e.g. applied voltage [149], [153]–[155], temperature [156], frequency, needle tip radius and sample preparation [27]. For instance, the increase of frequency or applied voltage results in the transition of branch tree to bush tree. A branch tree has multiple branched structures; the discharge activities are restricted to a few branches at a time and thereby the tree extends only from those branches [155]. Discharges in bush trees on the other hand, are spread throughout the body of the bush; new tree tubules are generated and packed together in the bush form [155]. Higher applied voltage may result in larger PD and lead to higher gas pressure that promotes bush formation [157]. Nevertheless, the higher damage density causes slower propagation in bush trees compared to branch trees. As this transition was mainly witnessed at ambient temperature, studies in higher temperature (70° for XLPE [156] and 80° for PE [149]) exposed the increased in gas diffusion reduces the gas pressure, yielding the formation of a branch tree rather than a bush tree. Both tree types are normally characterised through fractal dimension, d_f i.e. branch trees have $d_f < 2$ whereas for bush trees $2 < d_f \leq 3$ [158].

In addition, a number of researchers have introduced terms “branch-pine” or “monkey-puzzle” to describe the type of electrical tree that requires voltage that is lower than the inception of a branch tree [154], [155], [159]. Investigation on tree growth in XLPE cable insulation at different voltage levels discovered the transition of branch-pine to branch and finally to bush [155], [159] as shown in Figure 2-15. The difference between branch and branch-pine trees is the presence of many short side branches on each side of the long branches in the branch-pine tree. This type of tree was also observed in epoxy resin [154].



(a) Branch-pine tree (b) Branch tree

(c) Bush tree (d) Bush tree

Figure 2-15. Electrical trees obtained from XLPE samples at 9, 11, 13 and 15 kV [160]

As can be seen in Figure 2-14, the propagation stage can be separated into two phases i.e. fast and slow growth. At the beginning of the propagation stage, only a single branch structure will grow from the tip electrode and the growth is mainly in the vertical direction towards the ground electrode. The tree then either continues to grow from the same branch or enters a slow propagation phase where more channels will grow out from the initial branching structure and results in a bush tree [159]. However, as the tree length increases, the potential and the field at the tree tip reduces thus decelerating the growth rate especially for a long distance gap between needle tip and ground plate or for low applied voltage [157].

2.6.3 Runaway Stage

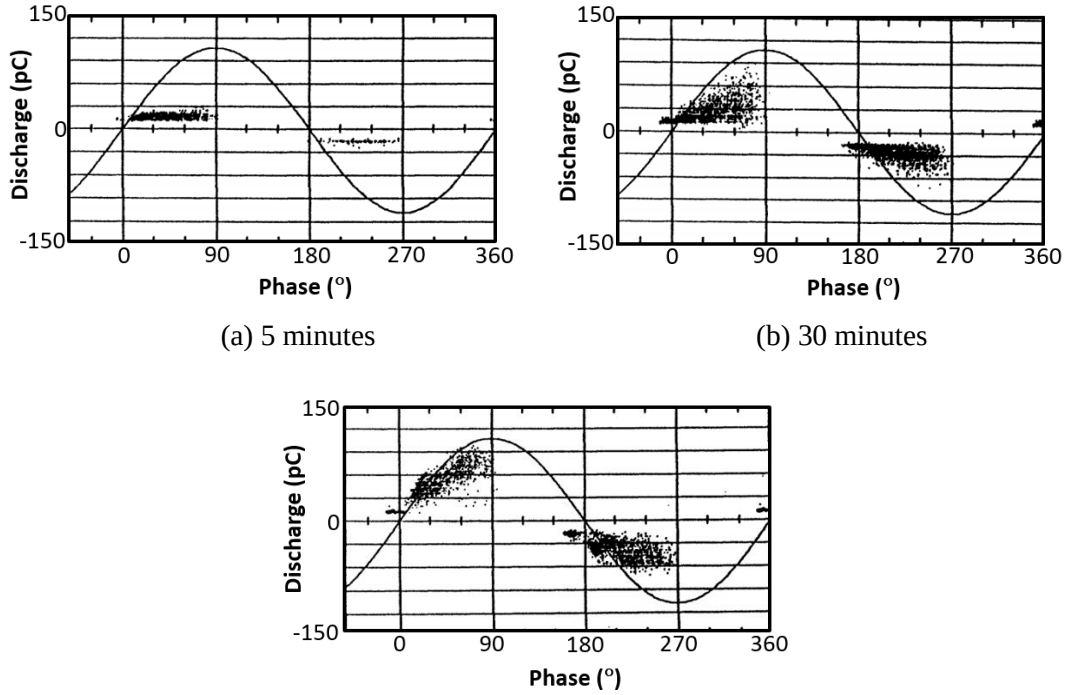
The runaway growth may predominantly occur at the leading branches, where one of the branches continues to grow towards the plane electrode. Once a leading branch comes close to the plane electrode, the electrical field enhancement due to the proximity of the ground plane enables runaway tree growth to occur. However, this may not causes an immediate breakdown since a return tree might occur after the original tree traverses the insulation gap. A return tree is the tree that grows from the ground electrode to the start point of the treeing. Results from electrical treeing experiments using a double needle electrode system molded in PE [28], showed a return tree was grown after a branch tree traversed the insulation at applied root mean squared (RMS) voltage of 10 kV. At a higher voltage of 20 kV_{rms}, a bush tree was observed to grow and a complete breakdown occurred when the tree reached the ground electrode. A return tree was also observed in [27], [29] which the resultant current in [29] was much larger than the forward tree.

2.6.4 Corresponding Partial Discharge Analysis

As the initiation and the growth of electrical trees are accompanied by PD activities, many research studies have incorporated the morphology of electrical treeing with the analysis of PD data. The presence of electrical treeing can be detected through PD pattern analysis as different PD sources result in unique PD patterns. Since PD data available for this research work is in the form of phase-resolved patterns, thus, PD analysis is limited to these parameters: phase occurrence, magnitude of PD, number of PD and time of PD occurrence.

Explanation of the PRPD representation approach in Section 2.4.2.1 mentioned some of the terms used for describing the PRPD pattern, namely “turtle”, “rabbit-ear” and “wing” like patterns. Early efforts in finding the correlation between electrical tree growth and the corresponding φ–q–n plot have used these terms to distinguish the different stages in tree growth. Investigation in [98] has highlighted three unique PRPD patterns for describing the inception, early growth and late growth stages as illustrated in Figure 2-16. Since the initiation of an electrical tree is related to void

formation, the *φ–q–n* plot during this stage depicts a “turtle” like pattern, which is a typical pattern for a void. A branch tree grew from the low density polyethylene (LDPE) sample after half an hour of the experiment and yielded a “wing” or “triangle” like pattern. The branch tree then changed to a bush tree where fewer small PD appeared when the instantaneous voltage reached the peak. These patterns are showed to be reproducible for a triangle wave applied voltage and different insulation materials i.e. ethylene–vinyl acetate (EVA) and ethylene–acrylic acid (EAA) copolymers.



(a) 5 minutes (b) 30 minutes

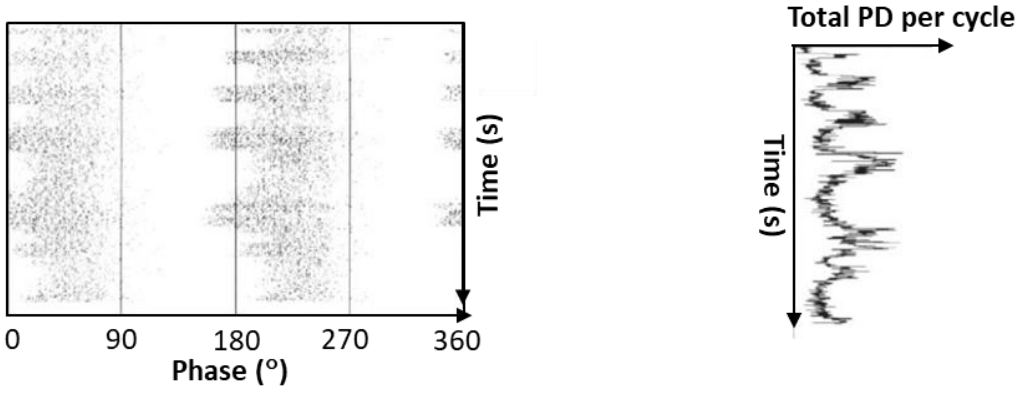
(c) 180 minutes

Figure 2-16. The *φ-q-n* plots from electrical treeing in a LDPE sample under 8kV 60Hz sinusoidal applied voltage; (a) tree initiation (5 min), (b) branch tree (30 min), and (c) bush tree (180 min) [98]

From Figure 2-16, it can be seen that the PD pulses mainly occur in the first and third quadrant of the voltage waveform. The reason for such occurrence has been discussed comprehensively in [161]. At these quadrants, the magnitude of the instantaneous voltage is increasing which results in an increase to the total field distribution [99], thus triggering the next partial discharge [96]. The transition from “turtle” like to “wing” like pattern emphasises the increase in discharge magnitude

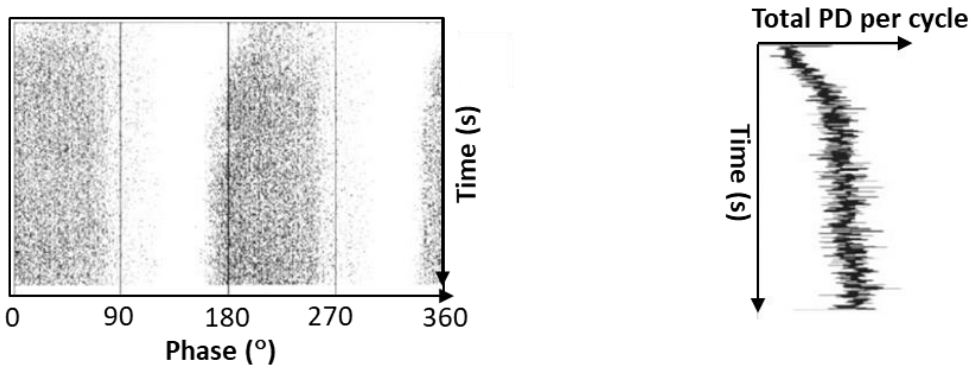
with the phase angle and the instantaneous voltage during the treeing process. The PD yields the maximum magnitude when instantaneous voltage is nearly at its peak, and then becomes inactive when reaching the peaks where the derivative of voltage with respect to time, *du/dt* is equals to zero. The magnitude of instantaneous voltage in the second and fourth quadrants is decreasing thus lowering the total field distribution, resulting in less PD occurrence. This suggests that the voltage derivative plays a role in the occurrence of PDs [98].

The difference between branch and bush types of treeing could also be identified by observing the PRPD pattern throughout the tree growth. For a branch tree, Champion and Dodd [15] found that a temporary phase shift occurred at each PD burst as shown in Figure 2-17a. These short interval phase shifts in the partial discharge activity were associated with a sudden increase in the PD rates as depicted in Figure 2-17b. In contrast, a gradual phase shift occurred during the growth of a bush tree in Figure 2-18, where the phase distribution widens from only the first quadrant to include the fourth quadrant, and from only the third quadrant to include the second quadrant. The PD rates in Figure 2-18b are less chaotic than those of the branch tree. The same characteristic of the PD rates was also found in [20], [153], [162].



(a) Phase-time plots (b) Total PD per cycle

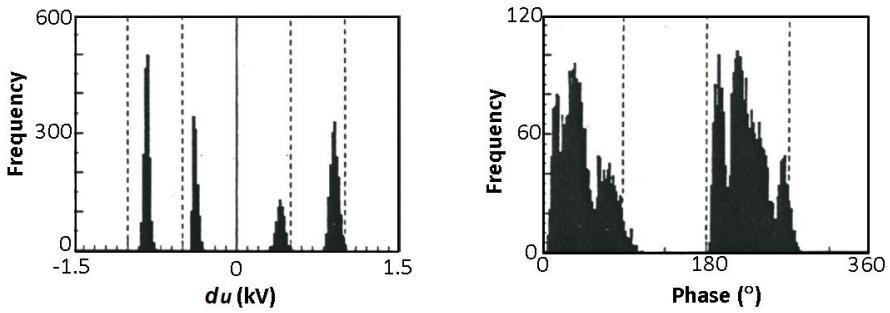
Figure 2-17. PD characteristics for branch tree in (a) phase–time plots, and (b) the corresponding total PD rate as a function of time [15]



(a) Phase-time plots (b) Total PD per cycle

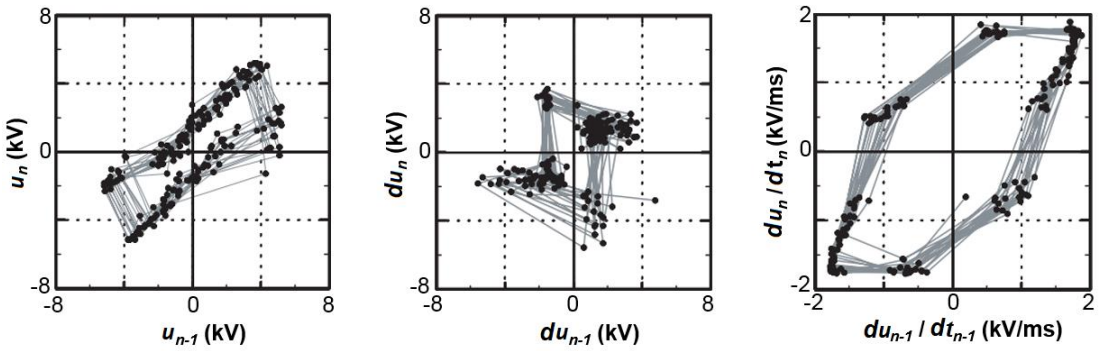
Figure 2-18. PD characteristics for bush tree in (a) phase–time plots, and (b) the corresponding total PD rate as a function of time [15]

The analysis of the PD sequences reveals that consecutive discharges occur in specific sequence [96]. For the case of electrical treeing, the specific sequence can be verified from the histogram of voltage difference shown in Figure 2-19a. It shows that there is a systematic periodic shift of the external voltages due to the space charge built up by the discharge process. In contrast, no specific sequence is provided by the phase occurrence due to the broad distribution in the Figure 2-19b. Since the voltage of PD occurrence brings significant impact, in PSA, three plots of consecutive discharges based on the instantaneous voltage are proposed in [96] i.e. the instantaneous voltage, *u*, the voltage difference, *du* and the ratio of the voltage difference to the time difference, *du/dt*. Figure 2-20 shows the consecutive plots of electrical treeing in PE (represented by the lines which connect two consecutive discharges) that can be a reference for PD pattern recognition.



(a) (b)

Figure 2-19. Frequency distributions of (a) *du* and (b) phase angles for electrical treeing in polyethylene [96]



(a) (b) (c)

Figure 2-20. Pulse sequence analyses of PD signals after tree initiation in polyethylene on the basis of scatter plots of (a) *u*, (b) *du*, and (c) *du*/*dt* [96]

2.7 Impact of Harmonics on Electrical Ageing

Harmonic pollution in the distribution network is a great concern due to the increased use of power electronic converters in both residential and industrial areas. The growing interest in distributed generation makes it even worse. These converters generate harmonic components, which propagate towards the network supply side, and nonsinusoidal voltages at the load side. Consequently, there is a gradual change in the working environment [2]. Hence at the plant level, insulation systems will age differently influencing electrical ageing mechanisms such as partial discharges and electrical treeing.

Harmonics cause overheating to the conductor as well as the insulation material that can accelerate thermal ageing of the insulation [163], [164]. In power cables, the additional heat is primarily due to copper loss, P_{copper}, expressed as follows:

$$P_{copper} = I^2 R \tag{2-15}$$

This loss is dependent on two electrical parameters: the current that flows through the cable, *I*, and the cable resistance, *R*. The nonsinusoidal components increase the net RMS load current thus increases the copper loss. The resistance on the other hand, increases due to the skin effect and proximity effect. Both phenomena vary as a function of frequency as well as conductor size and spacing.

Harmonics are components of a periodic wave having a frequency that is an integer multiple of the fundamental power line frequency. Harmonics, therefore, are a multiple of the fundamental frequency and are usually defined in the Fourier series as a periodic steady state distortion waveform shown in equation 2-13 [165].

$$u(t) = \sum_{h=1}^{N} U_{hp} \sin\left(h\omega_1 t + \varphi_h\right) \tag{2-16}$$

where N is the number of harmonic components contained in the voltage waveform, h represents the harmonic order, while U_{hp}, ω_h and φ_h are the peak voltage, angular frequency and phase shift for the h^{th} order harmonic respectively. In power engineering, the fundamental frequency is 50 or 60 Hz depending on the network. Figure 2-21 shows the resultant waveform when 3rd, 5th and 7th harmonics at 30% of amplitude pollute the fundamental at 0 phase shift.

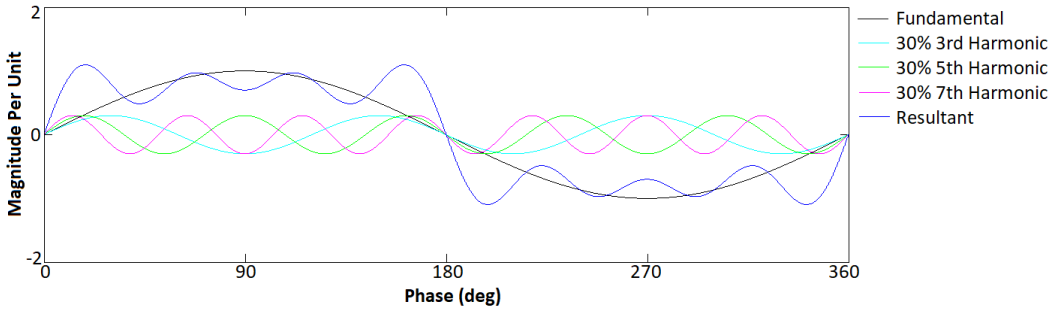


Figure 2-21. The resultant of the fundamental and 3rd, 5th and 7th harmonics when $\boldsymbol{\varphi_h} = 0$

The measure of harmonic content is commonly expressed as total harmonic distortion (THD),

$$THD = \sqrt{\sum_{h=2}^{N} \left(\frac{U_{hrms}}{U_{1rms}}\right)^2} \tag{2-17}$$

where U_{hrms} and U_{1rms} is the RMS voltages of the h^{th} order harmonic and fundamental respectively. In addition, Montanari and Fabiani [165] have formulated three other measures to describe the resultant amplitude and waveshape, i.e. peak parameter, K_p, RMS parameter K_{rms}, and waveshape parameter, K_s as in equations 2-16 to 2-18 respectively. In the case of a pure sinusoid, $K_p = K_{rms} = K_s = 1$, while in distorted regimes they are probably greater than unity.

$$K_p = \frac{U_p}{U_{1p}} \tag{2-18}$$

where K_p is the ratio of the peak voltage of the resultant waveform, U_p to the peak fundamental, U_{1p}.

$$K_{rms} = \frac{U_{rms}}{U_{1rms}} \tag{2-19}$$

where K_{rms} is the ratio of the RMS voltage of the resultant waveform, U_{rms} to the RMS fundamental, U_{1rms}.

$$K_s = \sqrt{\sum_{h=1}^{N} h^2 \left(\frac{U_{hrms}}{U_{1rms}}\right)^2} \tag{2-20}$$

where K_s is proportional to the RMS derivative of the waveform and thus is related to its steepness.

Life tests on a self-healing capacitor in [164]–[167] show that the effect of peak parameter, K_p, is the most critical among the three parameters since even a small increase in K_p can cause a failure time reduction of ten times or more. For this reason, only K_p is considered for life modelling in [168], [169]. The increase of the peak amplitude can be noted from equation 2-16 to depend on the phase shift, $\boldsymbol{\varphi_h}$ of the h^{th} order harmonic. As can be seen in Figure 2-21, the peak of resultant polluted waveform is almost similar to the peak of the fundamental when no phase shift is applied. However, when the 3rd and 7th harmonics are shifted by 180°, the resultant waveform in Figure 2-22 yields a factor of 1.9 increase in the voltage peak.

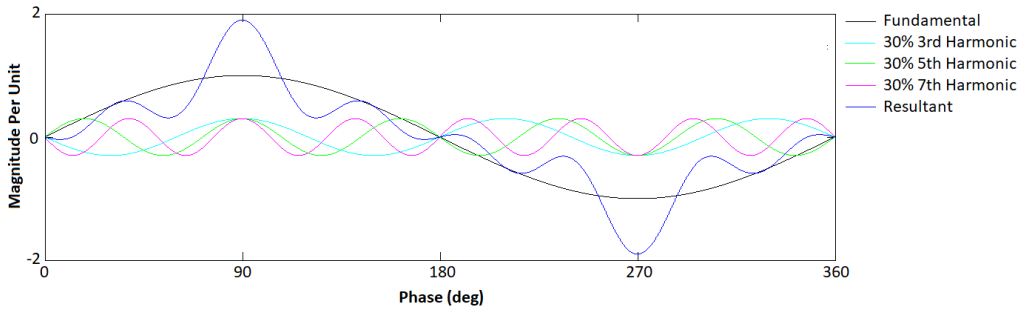


Figure 2-22. The resultant of the fundamental and 3rd, 5th and 7th harmonics when $\varphi_1 = \varphi_5 = 0°$ and $\varphi_1 = \varphi_5 = 180°$

Further investigation on the impact of the three parameters discovered the significance of waveshape factor, K_s , in life reduction [170]. As can be seen in Figure 2-21 and Figure 2-22, not only the peak could change remarkably but also the slope, *du/dt*. The increase in *du/dt* involves bursts of fast voltage rise that promotes PD activity [170]. This can be observed in Figure 2-23 where the higher *du/dt* of 11th harmonic results in higher magnitude of PD. The effect of K_{rms} however, is not very significant and may be considered negligible [170].

As the amplitude of instantaneous voltage and the derivative may influence the occurrence of partial discharges (explained in Section 2.6.4), the distortion on the voltage waveform in the electrical network thus could change the normal PRPD pattern. The changes in working conditions and exploitation stresses by harmonics have a crucial influence on the PD arising in insulation systems compared to normal conditions. The influence of harmonics on the partial discharge activity can be observed through *φ*–*q*–*n* plots in Figure 2-23. The following changes can be observed by comparing PRPD patterns at different harmonic orders and THD [171]:

- Change in the maximum discharge magnitude
- Change in the phase location of discharge activity
- Change in the number and density of the discharge activity
- Change of the symmetry of the PD images and phase distributions
- Existence of regions with no discharge activity, known as ‘dead’ zones

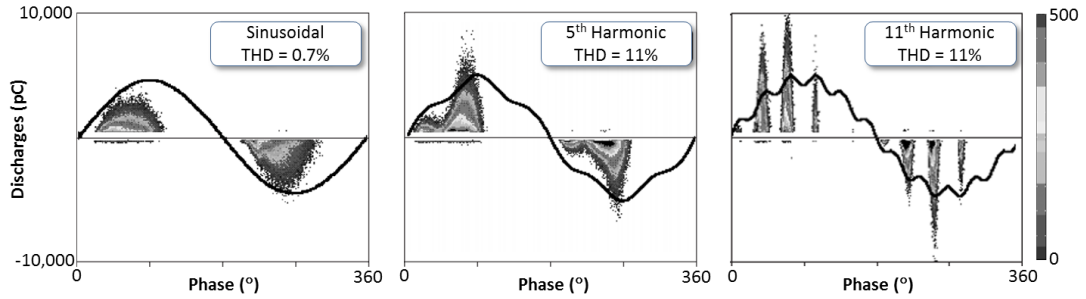


Figure 2-23. Influence of harmonic content on ***φ***–*q*–*n* plot, (a) “Pure” test voltage with THD = 0.7%, (b) 5th harmonic with THD = 11%, and (c) 11th harmonic with THD = 11% [171]

IEC 61000-2-4 [172] however, has a limit on the compatibility level for harmonic voltage on power systems in the range 1 kV – 69 kV to a THD of 5%. This means, the THD at various points in a grid system shall not exceed this level, otherwise it may cause disturbance to the power system at plant and consumer level. At consumer level, it is normally required by the utility to limit the THD by installing passive or active filters. Nevertheless, published work on the impact of harmonics on PD patterns [171] and electrical tree growth [30], [173] purposely exceed the maximum THD allowance since no significant effect to insulation failures was found for THD lower than 5%. Investigation on higher THD therefore, highlights the voltage distortion effect which may occur due to failure of harmonics filters.

Very few papers have reported the impact of harmonics on electrical tree growth. Research in [30], [174] observed bush type trees at lower harmonic voltages (even and odd) that are independent of THD. However, at higher harmonic orders, lower THD generated bush-branch trees while higher THD generated fibrillar trees i.e. a tree that spreads with thin fine light coloured branches. This fibrillar tree was also noticed at the very low frequency of 0.1 Hz where fewer branches are generated than when the frequency is increased to 1 Hz. Also, it was found that tree inception is faster at higher harmonic orders and THD level.

Bahadoorsingh and Rowland [173] utilised seven composite waveforms (including one fundamental) at different THD and K_s but constant peak values to study the impact of harmonics on tree growth and breakdown times. Analysis of partial discharge height distribution (PDHD) using the Weibull distribution of equation 2-13 pointed out that the 7^{th} harmonic produced a larger scale parameter, α than the 5^{th} harmonic for waveforms with similar THD levels. Larger α values are associated with higher PD magnitudes, hence this suggests the 7^{th} harmonic is more influential on electrical tree growth than the 5^{th} harmonic [175]. Using different electrical treeing samples, analysis of breakdown times using the Weibull distribution of equation 2-12 at increasing THD level yielded reduced shape parameters, β indicating a shorter lifetime. The growth rate of electrical trees considering both vertical and horizontal dimensions of electrical tree growth appears to be reproducibly independent of power quality variation.

As electrical treeing is associated with the PD mechanism, hence charge injection and extraction from the electrode to dielectric material could explain some of the tree behaviour due to harmonics. Harmonics that are greater than supply frequency will boost the charge injection and extraction process because of shorter time per half cycle. More charges being trapped will enhance the localised electric field, thereby accelerating the initiation of the electrical tree. The tree size and shape will then greatly depends on partial discharges in the tree channel [30]. Increased PD activity particularly due to the higher voltage peak of a distorted waveform can cause life reduction.

2.8 Review of Polymeric Ageing Models

When the insulation is subjected to either single or multiple of TEAM factors, there will be irreversible changes of the material properties hence reducing its endurance to the stress. The ageing process will eventually cause insulation breakdown when the material can no longer withstand the applied stress. The development of ageing models not only can relate the dependence of breakdown time to the applied stress but also could estimate the remaining useful lifetime (RUL) of the insulation. This information is critical for CBM in making decisions for the next maintenance plan with the possibility of prolonging the insulation's lifetime.

The emergence of polymeric ageing models has been reviewed by Montanari [48], [176]–[179] showing the transition from empirical to physical models. Very recently, Mazzanti updated the review and focusing on the life and reliability estimation of HVDC cables with extruded insulation [180]. The majority of the developed models are focused on thermal and electrical effects hence are the focus in this section. The models were initially derived for single stress ageing until the multistress ageing, particularly electrothermal stress (a combination of electrical and thermal effects), started to gain more attention in 1980's.

The empirical ageing model is often described as a phenomenological model in the literature. Experimental studies are carried out in more severe conditions than normal operation in order to induce early failures. The stress endurance from accelerated life

testing mostly obeys the exponential (equation 2-21) or inverse-power law (equation 2-22) depicting the decrease in ageing rate as insulation life increases [178].

$$y = k_1 \exp\left[-k_2 x\right] \tag{2-21}$$

$$y = k_1 x^{-k_2} \tag{2-22}$$

where k_1 and k_2 are the empirical factors. The linear representation of both models in either semilog or log-log scale respectively, allows interpolation for design stress information and extrapolation for life estimation. Furthermore, the adaption of Arrhenius and Eyring models into these empirical studies has initiated interest in physics-based modelling, giving the interpretation of the ageing mechanism.

2.8.1 Thermal Models

Crosslinking, chain scission and oxidation are examples of mechanisms that cause thermal aging [49]. Experimental studies into thermal effect were conducted at various temperatures and some of the material properties are expected to be reduced as the temperature increases. The well-known exponential relationship between temperature and failure time has been reported since 1930 in the application of oil-paper insulation in power transformers [181]. Montsinger demonstrated the relationship by plotting the tensile strength of thermally degraded paper with the lifetime, L, at three different temperatures, T. The relationship between L in hours and T in °C was expressed by rewriting equation 2-18 as follows:

$$L = k_1 \exp\left[-k_2 T\right] \tag{2-23}$$

It was shown in the paper that a linear relationship exists when plotting ln L vs T for a constant tensile strength.

The exponential relationship between temperature and failure time became more convincing when Dakin found a similarity to the well-known Arrhenius model, indicating chemical reaction rates vary with the temperature [182], [183] i.e.

$$Rc = k_1 \exp\left[-\frac{\Delta W}{k_B T}\right] \tag{2-24}$$

where Rc is the rate constant of a chemical reaction, k_1 is an empirical factor, ΔW is the activation energy for the reaction in Joules, k_B is the Boltzmann constant and T is the absolute temperature in Kelvin. In thermal ageing, Rc is considered to be the degradation rate which is inversely proportional to the lifetime, L, giving a life model:

$$L = \frac{1}{k_1} \exp\left[\frac{\Delta W}{k_B T}\right] \tag{2-25}$$

However, the plot of ln L vs $1/T$ rarely yields a straight line, hence, Eyring's model is used as an alternative. This was initially applied for reaction rate under constant thermal stress, and gives the empirical factor k_1 as a function of k_P, k_B and T. The thermal life model derived from Eyring's model is rewritten as [48], [184]:

$$L = \frac{k_P}{k_B T} \exp\left[\frac{\Delta G}{k_B T}\right] \tag{2-26}$$

where k_P is the Planck constant and G is the Gibbs free energy corresponding to the height of the energy barrier to be overcome for development of degradation reactions. A linear life line that is equivalent to the Arrhenius model can be obtained from the plot of ln LT vs $1/T$, which is mostly used in multistress ageing models [48].

2.8.2 Electrical Models

Research on electrical ageing is primarily focused on voltage endurance. Accelerated life testing is either conducted at much higher voltages [152], [185] or frequencies [186] than the operating conditions, while other variables are kept constant. The information on less severe conditions is then estimated by extrapolation. The life test data are fitted with either exponential [187], [188] or inverse power model (IPM) [152], [186], [189], [190] as represented in equations 2-27 and 2-28 respectively.

$$L = k_1 \exp\left[-k_2 U\right] \tag{2-27}$$

$$L = k_1 U^{-k_2} \tag{2-28}$$

where U can be either the applied voltage [152], [187], [188], the electric field strength [189] or the voltage gradient [190]. Thus, a linear plot could be obtained in either semilog or log-log with the slope of $-1/k_2$ or $-1/k_2$ when plotting U vs ln L or $\log_{10} U$ vs $\log_{10} L$ respectively.

Failure times were initially determined by applying constant voltage, until the continuously increasing voltage was proposed in 1960's [185]. The latter approach is preferred because no preselect of test voltage is required [185] and the scatter of breakdown times can be reduced [191]. Nonetheless, the testing time for constant voltage is considerably shorter than the progressive stress at the same breakdown voltage as shown in Figure 2-24 and proven mathematically in [185], [192]. The relationship between the breakdown time of constant voltage, t_{cons} and progressive voltage, t_{prog} can be found by fitting the IPM model to the life data and is given as follows:

$$t_{prog} = t_{cons}(k_2 + 1) \tag{2-29}$$

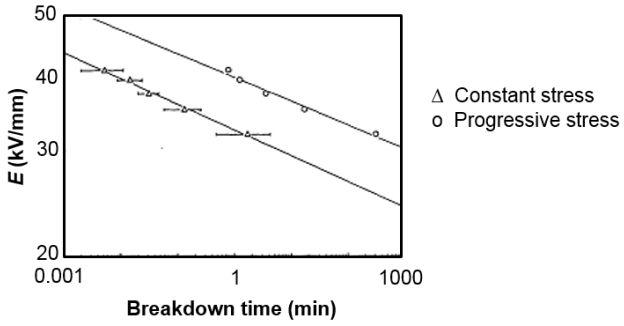


Figure 2-24. Inverse-power model for the EPR specimens subjected to constant electrical stress and progressive-stress test [191]

At low stresses, the life line tends to become horizontal which the failure times are much longer than may be expected from linear extrapolation of inverse power or exponential models [177]. Hence, it is more accurate to treat the lower stress as a threshold which has been applied in both models. As regards the exponential model, two threshold models were proposed in the literature as in equations 2-30 and 2-31.

For the case of breakdown due to surface discharge, Dakin and Studniarz [188] proposed PDIV as the threshold voltage, U_{th}, using equation 2-30.

$$L = \frac{k_1}{U - U_{th}} \exp\left[-k_2 (U - U_{th})\right] \tag{2-30}$$

$$L = \frac{k_1}{U - U_{th}} \exp\left[-k_2 U\right] \tag{2-31}$$

The inverse power model on the other hand, is modified as follows and known as the inverse power threshold model (IPTM) [193], [194]:

$$L = k_1 \left(U - U_{th}\right)^{-k_2} \tag{2-32}$$

or

$$L = \frac{k_1 U^{-k_2}}{U - U_{th}} \tag{2-33}$$

Among these four threshold models, the experimental results are found to better fit the model in equation 2-31 [194]. Plots of the exponential model (equation 2-27) and exponential threshold model (equation 2-31) in Figure 2-25 show how the exponential model could give a large discrepancy when the electrical ageing can be neglected at voltages lower than the threshold.

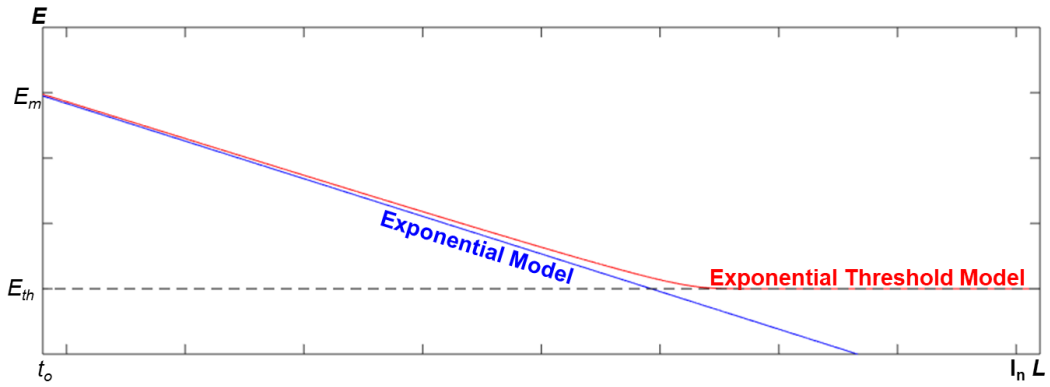


Figure 2-25. Examples of electrical endurance curves obtained by the exponential and exponential threshold models [61]

In addition to the threshold characteristic, the typical voltage versus time curve for polyethylene and epoxies has a “flat-z” characteristic which can be divided into three regions as illustrated in Figure 2-26 [193], [195], [196]. In Region I, the breakdown time typically ranges up to one hour. The change in the life curve in Region II is said to follow Eyring’s law, i.e. the exponential relationship in which the electrical ageing mechanism may be influenced by the thermal stress [196]. Finally, the life curve flattens in Region III at E_{th}. The operating voltage should be in this region where electrical ageing can be neglected, giving a very long lifetime. Applying IPM to the three regions separately would result in the voltage endurance coefficient, k_2 (also known as life exponent) as in equation 2-34, where k_2 is inversely proportional to the life line slope.

$$k_{2II} < k_{2I} < k_{2III} \tag{2-34}$$

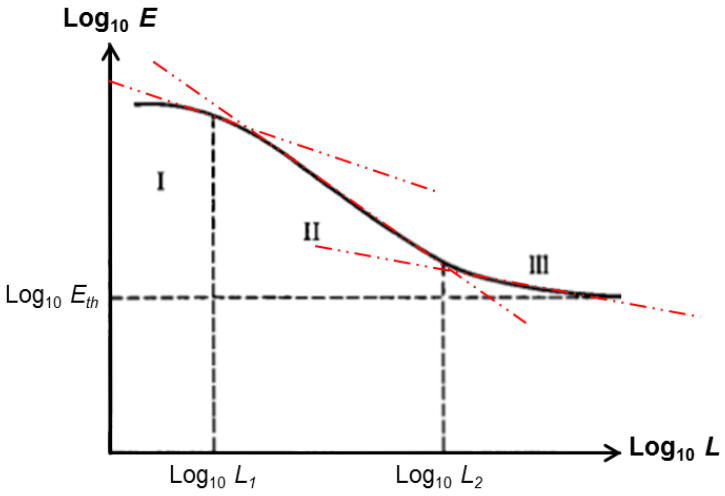


Figure 2-26. General characteristics of life curve of polyethylene and epoxy resin insulations [193], [195], [196]

2.8.3 Model Based on Electrical Treeing

Electrical treeing is often regarded as a pre-breakdown phenomenon in polymeric insulation [60], [61]. Several mechanisms have been suggested to explain the initiation of electrical treeing as discussed in Section 2.6.1. Based on the electron emission theory, Tanaka and Greenwood [148] proposed a model that relates the electrical tree inception time with the local electric field. A more recent work in

[197], has developed a model based on correlation between the electroluminescence (produced from recombination process) behaviour with the observed filamentary damage. Although both models are limited to the early stage of tree growth with no breakdown information, prediction of the inception time is worthwhile as it governs the formation of the electrical tree.

The extension of electrical tree channels depends on the activities of charged particles within the insulation material. At very high local field enhancement, the space charges that were previously deposited in the cavities are accelerated then collide with neutral gas particles and with the channel walls. The scission of molecular chains of the channel walls increases the diameter of the channels that initiate the formation of an electrical tree. According to Bahder et al. [23], the charges flow mostly in the tree stem thus resulting in an increase of stem diameter, which is called a crater. The crater depth was found to be proportional to the charge flowing forward and backward through the channel, q and can be expressed as

$$q(t,E) = k_1 ft\left(\exp\left[k_2\left(U - U_{th}\right)\right] - 1\right) \tag{2-35}$$

where f the voltage frequency, t the time of voltage application, U the applied voltage stress, U_{th} the threshold voltage and k_1 and k_2 the constants. Also, the crater depth was found to be inversely proportional to the maximum breakdown voltage stress. Both relationships are combined for determining the breakdown time, t_{bd} considering the critical depth of the crater at known breakdown voltage, U_{bd} once the electric field exceeds a threshold value, U_{th}. This gives

$$t_{bd} = \frac{1}{fk_1\left(\exp\left[k_2\left(U - U_{th}\right)\right] - 1\right)\exp\left[k_3 U_{bd} + k_4\right]} \tag{2-36}$$

where k_1 to k_4 are parameters which depend upon the material, temperature and geometry.

The early models concerning the growth of the electrical tree are mostly based on fractal dimension. Niemeyer et al. [198], [199] proposed that tree branching has a stochastic nature that is not only governed by the maximum electric field around the tree tip but also at the point which has the highest probability of growing. Since the

electrical tree in nature is 3D, the total length of all branches, $l(j)$, is represented as the radius of a circle, j,

$$l(j) \cong j^{d_f} \tag{2-37}$$

where d_f is the fractal dimension. The number of branches $N_b(j)$ is then given by

$$N_b(j) \cong \frac{dl(j)}{dj} \cong j^{(d_f - 1)} \tag{2-38}$$

This means the fractal dimension, d_f can be determined by counting the number of branches at various distances. They also found that d_f is dependent to the exponent k, from the following relation

$$P_i \propto E_i^k \tag{2-39}$$

where the probability, P_i to add segment i to the tree is related to the local electric field, E_i. Trees with low fractal dimension (branch type) were found to grow faster than trees of high fractal dimension (bush type), however, higher fractal dimension caused a greater amount of damage [24]. This model has been modified in [200] by introducing a breakdown criterion, i.e. a threshold field for the tree growth.

Based on Niemeyer et al.'s stochastic model (also known as the *Dielectric Breakdown Model* (DBM)), Dissado et al. proposed a deterministic model giving a relationship between the tree damage, D and growth time, t in terms of tree length, l [201]:

$$D(t) = \left[\frac{l(t)}{l_b} \right]^{d_t} = \frac{t}{t_{ch}} \tag{2-40}$$

where l_b is the average length of a newly created channel and t_{ch} is a characteristic time for channel formation. The discharge avalanche mechanism proposed by Bahder et al. in equation 2-35 has been adopted in the *Discharge Avalanche Model* (DAM) for determining t_{ch} i.e. the time to form a channel [158], [202]. From equation 2-35, the relationship between space charges and the depth of charge penetration is assumed to be exponential.

Montanari then modified Bahder et al.'s model by adopting Dissado's model of equation 2-40 to represent the depth of charge penetration. The model is given as in equation 2-41 with consideration to the power relationship between the tree length and applied voltage.

$$q(t,U) = k_1 \left\{ \exp\left[k_2 (U - U_{th})^{k_3} t^{1/d_f} \right] - 1 \right\} \tag{2-41}$$

It was shown in [26] that this model fits satisfactorily experimental data for both EVA and XLPE. This model may be employed to estimate the failure time of insulation once a limiting value for the amount of the chosen charge height quantile has been selected. A generalised probabilistic aging model was then derived from the Weibull distribution.

Studies in [25], [203] have also applied the Weibull function to the PDHD (resulting from the electrical tree and cavities respectively) and suggested the scale and shape parameter could indicate the level of ageing. However, both studies depict different behaviour of the Weibull parameters towards breakdown, thus further clarification is needed.

A stochastic-deterministic model of the electrical tree growth that incorporated PD within the tree has been presented in [204], [205]. The direction of the new channel growth was determined based on stochastic dependence on the local electric field. The electric field distribution depends on the charge deposition within the tree structure that is affected by PD activities. Therefore, the damage distribution along the tree structure depends on PD inside the channels. A computer simulation has been designed based on the proposed model which allows observation on many of the known characteristics of tree growth.

Another deterministic model has been proposed by Champion et al. based on partial discharge activities within non-conducting tree structures [21], [22]. The tree in which the discharges take place is represented as a spark gap with two model properties: U_{on} and U_{off}. Bursts in the partial discharge activity are suggested to be associated with a reduction from U_{on} to U_{off} which allow the partial discharges to

propagate further into the tree structure. This is simulated by adding one or more dipoles of charge onto the tree structure. Each dipole of charge represents a local electron avalanche occurring over a distance equal to the grid spacing. Local electric fields surrounding a tree point were used to weight the local damage in each near neighbour point, with the total damage proportional to the energy dissipation at the tree point following a partial discharge. If the energy dissipation is greater than zero, damage was added to each nearest neighbour. At the end of the computation, only those segments that had discharged at least once were considered to be part of the tree structure. A tree tubule was therefore only formed provided that a discharge had occurred along it at some time in the past.

The mechanical properties of the material such as its elastic modulus, tensile strength and fracture toughness have an effect on the growth of electrical trees thus are considered for the model developed in [206]. The model is originally derived from the Eyring model of equation 2-26. The lifetime, L in equation 2-26 is inversely proportional to the degradation rate, dD/dt where D is the damage caused by the electrical tree by applying equation 2-40. This model however, denotes $l(t)$ as the linear length of the tree structure and l_b as the linear length of a growing micro-void that forms the tree. The time of failure is then determined by treating $l(t)$ as the critical tree length. In addition, research work in [207] has implemented and modified this model for predicting the failure time using ANN.

2.9 Summary and Conclusion

Electrical treeing is one of the main reasons for the long-term degradation of solid insulation that may eventually lead to breakdown. Studies on electrical treeing have mostly focusing on polymeric insulation especially in cable application. Although treeing is also observed in cellulose material in transformers, tracking phenomenon seems to gain more interest in this area. In this thesis, epoxy resin was chosen as the insulating material due to its translucent property that permits observation on the tree growth. In addition, epoxy resins have excellent electrical insulation properties as well as high thermal resistance hence, and are mostly used for casting and moulding in high voltage applications, e.g. bushings and cable accessories.

Studies on solid insulation degradation due to electrical treeing have focused on the physics behind the phenomenon. Most of the developed lifetime models relate the tree growth with the PD activities inside the tubule. Nonetheless, these models only consider the forward growth although return trees can be expected to occur. This is due to the lack of understanding on the underlying process of return trees. In contrast, no prior knowledge is required for data-driven technique hence, is a better alternative to physics based model.

Among the PD representation techniques described in this chapter, PRPD has been the most popular approach which statistical operators are proposed as the potential features. These features however, derived mathematically from the PRPD pattern thus not directly related to the physics behind the PD events. On the other hand, PSA treats PD pulses as events within a sequence. This physical interpretation is an advantage of PSA over the PRPD. The occurrence of PD in the first and third quadrant of the voltage waveform suggests that the voltage derivatives, either with respect to time, du/dt, or phase, $du/d\theta$, play a role in the occurrence of PDs. This can be analysed using PSA by observing the change in instantaneous voltage of the PDs, which is claimed to not occur at random but in specific sequences. Therefore, both techniques are employed in the work reported in this thesis.

Little studies have reported the impact of harmonics on electrical treeing growth. From the studies, no significant impact on the failure time was found for THD below 5%. Hence, this research work purposely exceeds the maximum THD allowance (up to 40%) to highlight the voltage distortion effect which may occur due to failure of harmonics filters. Based on the Weibull distribution, there is a possibility for higher harmonic order or/and THD to reduce the failure time. This is due to the increase in voltage derivative of the resultant waveform, which causes an increase in the PD activities, and hence, accelerates the tree growth and reduces the failure time. However, the conclusion of failure time reduction is made based on insufficient life data.

Published work on PD diagnosis mainly uses offline PD data and might not applied to online PD monitoring due to the differences in the applied stresses. Despite the

infeasibility of offline data in cable monitoring, this thesis advances the academic state of the art, moving beyond diagnostics towards prognostics, with an eye towards practical deployment in the future.

Most of the reported insulation life models in the literature have been derived from Weibull distribution without considering the insulation degradation process. This thesis therefore proposes a different approach, considering the degradation process, by using PD data instead of failure times. From the literature review, the life data for solid insulation fits well to exponential and inverse power models. Therefore, both models are considered in this thesis for prognostics modelling and further explanation on these models is given in the next chapter.

Chapter 3

Prognostic Modelling

3.1 Building a Prognostic Model using PD Data

The International Standard (ISO 13381-1) [208] defines prognostics as "the estimation of time to failure and risk for one or more existing and future failure modes". Basically, three different measures are used in the literature for estimating failure occurrence [209]:

- Remaining Useful Life (RUL): the amount of time for a component to continue to function in accordance with its intended purpose, warranting replacement.
- Time-to-Failure (TTF): the time when a component is expected to fail and no longer meet its design requirements
- Probability of Failure (POF): the failure probability distribution of a component.

Three types of prognostic algorithm are defined in [31], [210] based on the type of information used in making the failure prediction. Type I, or reliability-based, considers historical TTF data to model the failure distribution, commonly using Weibull analysis, without considering the operating conditions. This could be the case for a new component with no track record of health information except for the past failure times of similar components used in similar conditions. Knowing that

harsh conditions could reduce the RUL, type II or stressor-based methods can be applied only if the operating conditions and the environmental stresses are measurable and correlated to the component's degradation. As the degradation becomes apparent, this information can be employed to improve the RUL prediction. Type III, or degradation-based, applies the degradation measure to characterise the lifetime of a specific unit operating in its specific environment. Prediction is performed by extrapolation to a predefined failure threshold. The appropriate degradation measures however do not have to be a directly measured parameter. Therefore, the insulation life models discussed in Sections 2.8.1 and 2.8.2 can be categorised as type I or II, depending on whether operating conditions or environmental stresses are considered.

Most of the tree growth models described in Section 2.8.3 were derived physically from the analysis of discharges in the tree channels. Nevertheless, how well the model fits the real data is rarely discussed in detail. This could be a major concern especially for a small sample size of data. Although physics-based models have a direct relation with the degradation mechanism, the understanding of the process might be misleading. The assumptions made in model development may not be fully applicable to real world systems, thus limiting the applicability of physics of failure models. As an alternative, a data-driven approach to type III prognostics could be applied to historical degradation data for the development of empirical life models.

A typical data driven approach does not associate degradation with physical knowledge of the failure mechanism, but the use of real data from the field allows derivation of real degradation relationships. Although gathering enough run-to-failure data for training the model might be an issue, accelerated testing could be an alternative approach. However, care must be taken to ensure that the failures seen during accelerated testing are analogous to real-world failures. In addition, the primary aspect of the data driven model is not only to provide good predictions, but to provide general and repeatable predictions [211]. More description on the data driven approach is presented in Section 3.5.

Research studies on insulation failure diagnosis are numerous and this can be achieved successfully using PD analysis [10], [79], [91], [96], [122], [131], [212], [213]. An abundance of PD features have been suggested in the literature and some of them are described in Section 2.4.3. These features however, are mostly proposed for classification rather than predictive modelling. Since the progression of a defect is far less well understood than diagnosis, studies of the correlation between PD features and the tree growth may give an insight for defect modelling. The quality of these features as prognostic parameters should be examined first as discussed in the next section.

Based on the literature, the overall process of developing a prognostic model can be outlined as shown in Figure 3-1. Pre-processing the data for prognostic modelling may sometimes involve the detection of outliers, and normalisation or/and standardisation of the extracted features [214]. The suitability of the features as prognostic parameters can be examined through three main qualities [31]: *monotonicity*, *prognosability* and *trendability*. In general, fewer features are desirable as it reduces the complexity of the model. This can be done through feature selection or transformation as discussed in Section 3.4. Prognostic models are then developed from the selected or transformed features either using regression analysis or another machine learning technique. The developed models are based on training data which can then be validated using a testing data set. The performance of the models is then evaluated in terms of accuracy, precision and robustness for model selection.

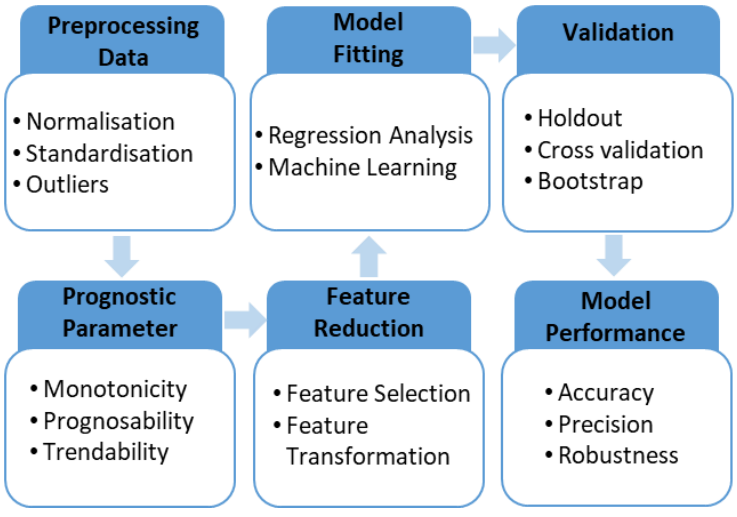


Figure 3-1. Possible steps in developing a prognostic model

3.2 Data Preparation

The potential PD features that are described in Section 2.4.3 are on different scales and/or units, thus might require data rescaling to aid the comparison between the features. This can be done either through standardisation or normalisation [214]. Normalisation scales all numeric variables in the range [0, 1] by using a linear scaling transform,

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{3-1}$$

where X_{norm} is the normalised X variables, X_{min} is the minimum value of X variables, and X_{max} is the maximum value of X variables. This approach will end up with smaller standard deviations, which can suppress the effect of outliers. Standardisation transforms each variable of the features so that they will have the properties of a standard normal distribution, i.e.

$$\begin{aligned} \mu &= 0 \\ \sigma &= 1 \end{aligned} \tag{3-2}$$

where μ is the mean and σ is the standard deviation. It is also known as Z-score normalisation in which the standard scores of the samples are determined as follows:

$$Z = \frac{X - \mu}{\sigma} \tag{3-3}$$

Finally, observations that deviate from a predetermined bound can be regarded as outliers. Looking for outliers, identifying them, and assessing their impact should be part of data preprocessing. A study in [215] has proposed a measure to characterise the tolerance of the feature to outliers in the range of [0, 1]. The measure is named as *robustness* and is determined by first decomposing the feature into its mean trend and a random part using a smoothing method:

$$X(t_i) = X_T(t_i) + X_R(t_i) \tag{3-4}$$

where $X(t_i)$ is the degradation feature value at time t_i, $X_T(t_i)$ is its trend value and $X_R(t_i)$ is the residual. A feature with good robustness is less sensitive to the parameter variations or external disturbances. The robustness of a feature is given by

$$robustness = \frac{1}{N}\sum_{i=1}^{N}\exp\left(-\left|\frac{X_R(t_i)}{X(t_i)}\right|\right) \tag{3-5}$$

where N is the total number of observations.

3.3 Prognostic Parameters

Raw data are often redundant and noisy thus not directly used for prognostics [216]. PD analysis as discussed in Section 2.4.3 is therefore used to extract meaningful information that can be linked to the degradation process. A set of metrics was introduced in [31] which characterise the suitability of a feature as a prognostic parameter. As the selection of features is vital in prognostic modelling, these metrics could aid the selection by comparing the characteristics of potential features. Three main qualities of an ideal prognostic parameter have been proposed: *monotonicity*, *prognosability*, and *trendability*.

Monotonicity is an important feature of a prognostic indicator as the degradation parameters change when the system degrades. Monotonicity characterizes the trend of time series data considering how consecutive data behaves i.e. increasing, decreasing, or static. A parameter with high monotonicity will have a strong trend in one direction, and will consequently provide good information for a prognostic model. Based on equation 3-6, the monotonicity score will be higher if most of the feature values change in one direction. The change of the feature is measured as a derivative function with respect to the allocated time interval.

$$monotonicity = \text{mean}\left(\left|\frac{\#\frac{d}{dx}>0}{N-1} - \frac{\#\frac{d}{dx}<0}{N-1}\right|\right) \tag{3-6}$$

where $\#\, d/dx > 0$ is the number of sequential data points where the second is greater than the first (positive $\#\, d/dx$), $\#\, d/dx < 0$ is the number of sequential data

points where the second is lower than the first (negative $\#\, d/dx$), and N is the number of observations.

Prognosability indicates the distribution of failure values, p_{fail}, in relation to the initial values, p_{start}, of a population of systems, i.e. the repeatability of the specific failure value between different samples. Samples that fail when a parameter reaches a particular threshold value (with a narrow distribution) will have high prognosability, and it will therefore be easier to predict the point of failure using this parameter than another with lower prognosability. According to equation 3-7, a small standard deviation of failure values and large parameter ranges will result in a greater score of prognosability.

$$prognosability = \exp\left(-\frac{\sigma\left(p_{fail} \right)}{\text{mean}\left(\left| p_{fail} - p_{start} \right| \right)} \right) \tag{3-7}$$

where σ is the standard deviation and p is the value of the prognostic feature.

Finally, trendability gives the smallest correlation coefficient, cc, between two samples in a population, which indicates the level of similarity of the shape of the failure curve between two samples. A parameter which consistently corresponds to one failure curve will be easier to model than another with low trendability. This measure is expressed as follows:

$$trendability = \min\left(\left| cc_{ij} \right| \right) \tag{3-8}$$

where i and j represents two different samples.

All the three metrics are used in this work to evaluate potential PD features as a prognostic parameter. Detailed explanation on the application of these metrics is given in Section 4.4 .

3.4 Feature Reduction

Feature reduction is an essential step before training a model to avoid *overfitting* and therefore improve the model prediction accuracy and generalisation ability. Overfitting implies that the model derived from the training data gives poor generalisation ability to unseen data, leading one to be overly optimistic about the model performance. This can be due to the size and complexity of the data. In addition to this, with less data, the training process could be simpler and faster, hence resulting in a cost-effective predictor.

The extracted features may greatly outnumber the sample size which is known as the *curse of dimensionality* as mentioned earlier in Section 2.4.3. Usually not all of these features are important or relevant and some are redundant. Consequently, without preselecting the most relevant features and effectively discarding redundant features as well as noise, the learning model has a marked risk of overfitting [110]. Therefore, feature reduction techniques [217] i.e. feature selection and feature transformation are used to remove redundant features and experimental noise by reducing high dimensionality features to a low dimensional space.

In predictive modelling, it is worth noting that the prediction of RUL or TTF employs continuous features and responses hence require regression analysis rather than classification. For this reason, the algorithms for regression analysis are the focus in this chapter.

3.4.1 Feature Selection

Feature selection approaches aid the identification of the best subset of the original features. This means, the irrelevant and redundant features will be removed from the dataset since they do not contribute to the accuracy of the prediction model. Feature selection thus reduces the number of features; as fewer features are desirable because it reduces the complexity of the model. In feature selection, the original representation of the features is not changed hence it is preferable for someone who wishes to keep the original meaning of the features. Once features have been

selected, only these features need to be calculated and analysed. In the literature, this approach is generally divided into three techniques namely: filter, wrapper and embedded methods [218]. These techniques differ in how the learning algorithm is incorporated in evaluating and selecting features.

3.4.1.1 Filter Methods

Filter methods rely on the characteristics of the features which are evaluated without utilising any learning algorithm. The aim of filter methods is to select a subset of features that optimises predetermined criteria. These criteria are usually statistical measures e.g. mean, variance, and correlation coefficients. Feature evaluation could be either univariate or multivariate. In the univariate scheme, each feature is ranked independently based on the scores, while the multivariate scheme evaluates features by batch. Therefore, the multivariate scheme is naturally capable of handling redundant features.

Most of the filter algorithms evaluate the individual features. The selection is either those that satisfy a condition or the top-ranked feature. Subsequently, no information is provided for the feature subset with the optimal modelling performance. Also, the interaction and dependencies between available features is disregarded. Despite this, filter methods employ a straightforward search strategy compared to wrapper and embedded methods, giving the benefit of an easier design and faster algorithm.

The linear correlation between features relies on measures of the general characteristics of the training data such as distance, consistency, dependency, information, and correlation [219]. As the degradation data involves continuous features and responses, the correlation between these two variables is significant in determining the characteristics of the features. Examples of correlation-based filter methods are the Pearson correlation coefficient and Spearman correlation coefficient.

Pearson correlation coefficient measures linear correlation between two variables [220]. The resulting coefficient lies in between -1 and 1. A correlation of -1 indicates a perfect negative correlation i.e. when one variable increases, the other decreases, whereas 1 indicates the opposite, and 0 means no linear correlation exists between

the two variables. Obviously, the main disadvantage of Pearson correlation coefficient is that it is only sensitive to a linear relationship. If the relation is non-linear, the Pearson correlation coefficient can be close to zero. The Pearson correlation coefficient, cc_p is defined as:

$$cc_p = \frac{\mathrm{cov}(X,Y)}{\sqrt{\sigma_X{}^2 \sigma_Y{}^2}} \tag{3-9}$$

where cov designates the covariance and σ^2 the variance. Solving for *cov* and σ^2 give cc_p as in equations 3-10 or 3-11.

$$cc_p = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{3-10}$$

$$cc_p = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left(N\sum x_i{}^2 - \left(\sum x_i\right)^2\right)\left(N\sum y_i{}^2 - \left(\sum y_i\right)^2\right)}} \tag{3-11}$$

where x_i is the feature value, y_i is the response value, and N is the number of observations.

The Spearman correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data. The relationships between features and the response could be nonlinear thus it is not always accurate to apply Pearson correlation coefficient. The Spearman correlation coefficient is more appropriate when the data points seem to follow a curve instead of a straight line. Also, it is less sensitive to the effects of outliers [221]. The Spearman correlation coefficient is actually a Pearson correlation coefficient based on the rank of the variables. Thus, the interpretation of the Spearman coefficient is similar to Pearson's but in this case, is based on the rank of the variables. Hence for variables with a monotonic trend, a perfect increasing trend yields a coefficient of 1 while a perfect decreasing trend is -1. Therefore, equation 3-9 is rewritten as follows for determining the Spearman correlation coefficient, r_s:

$$cc_s = \frac{\mathrm{cov}(rg_X, rg_Y)}{\sqrt{\sigma^2_{rg_X} \sigma^2_{rg_Y}}} \tag{3-12}$$

where *rg* is the ranking of the variables. If the data does not have tied ranks, the Spearman coefficient can be simplified to:

$$cc_s = 1 - \frac{6\sum \Delta_i^2}{N\left(N^2 - 1\right)} \tag{3-13}$$

where $\Delta_i = rg(x_i) - rg(y_i)$, the difference between the two ranks of each observation.

3.4.1.2 Wrapper Methods

Unlike filter methods, wrapper techniques require a predetermined learning algorithm and use its performance on the provided features to identify the relevant ones [218]. Hence, wrapper methods rank features according to their relevance to the model. Wrapper methods consider the selection of a set of features as a search problem, where different combinations are prepared, evaluated and compared to other combinations. In this way, feature dependencies are considered for model development. In comparison to the filter method, the wrapper approach is computationally expensive. Although this approach can improve model performance, overfitting may also be expected.

In the wrapper method, heuristic strategies are normally applied. Examples of simple algorithms are forward selection and backward elimination. Forward selection starts with an empty set of features while backward elimination begins with the full set of features. In each iteration, forward selection keeps adding the feature which best improves the model until the addition of a new variable does not improve the performance of the model. Backward elimination on the other hand, removes the least significant feature at each iteration which improves the performance of the model. The elimination process is repeated until no improvement is observed in the model performance. In terms of interdependent features, backward elimination seems to be better than the former [110].

The selection of either a new additional or removal feature for both wrapper methods requires a search algorithm. Best-first search is a simple heuristic search algorithm that sorts the queue for feature evaluation according to their ranking [218]. The ranks

of continuous types of features could be derived from filter methods. This algorithm chooses a subset of features with the top score ignoring the number of features in the subset. Hence, models derived from this approach may suffer from the *curse of dimensionality*.

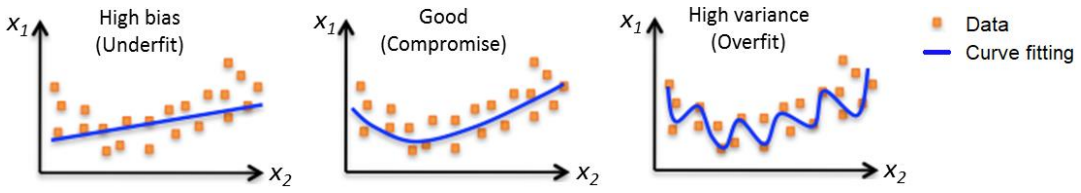
3.4.1.3 Embedded Methods

Embedded methods differ from the wrapper methods in the way feature selection and learning algorithm interact. In embedded methods, the feature selection is incorporated as a part of the learning process thus splitting the data into training and testing sets is unnecessary [110]. For instance, decision trees are iteratively developed by splitting the data based on appropriate features. Each feature is given a value depending on its importance for the classification task, and the final decision is made by calculating the value of decision nodes [222]. This built-in mechanism of feature selection makes embedded methods less prone to overfitting.

In addition, embedded methods bridge the gap between filter and wrapper methods. The filter method is applied in this approach to obtain a number of candidate subsets. The best candidate subset is then verified through the wrapper method. Therefore, the embedded model usually achieves a comparable accuracy to both filter and wrapper methods. However, for a small size of training data, a filter method can be expected to perform better than an embedded method [222].

Regularisation techniques embed feature selection into the learning algorithm indirectly by improving the performance of the model. Applying curve fitting to the set of data in Figure 3-2 may result in either a good model, an underfitted model or an overfitted model. An underfitted model shown in Figure 3-2a does not really capture the underlying trend of the data thus results in a high bias model and destroys the accuracy of the model. This is mostly due to the low complexity of the model. Figure 3-3 shows the prediction error from the training and testing set. The underfitted model that has high bias yields unacceptable errors that are much greater than the desired performance for both sets. Nevertheless, a model with high complexity may fit the training data too well (low error) as shown in Figure 3-2c. However, the training data may contain noise and inaccurate data, hence, the model

might not really fit a new data set. This will negatively impact the model's ability to generalise and thus yields a very high error for the test data in Figure 3-3. The very large difference between the error in training and testing data shows that overfitting results in a high variance model. Therefore, there should be a trade-off between a high bias model and a high variance model so that a compromise model can be developed as shown in Figure 3-2b.



(a) (b) (c)

Figure 3-2. Model fitting based on the complexity of the model

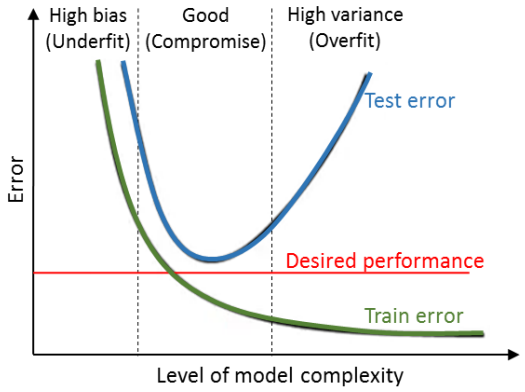


Figure 3-3. Model performance based on prediction error of the training and testing sets

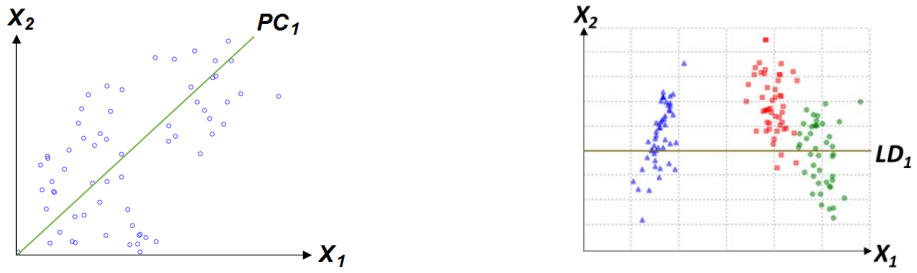
In order to overcome the overfitting problem, the regularisation approach adds a degree of bias to the regression model with considerably less variance. This type of approach penalises the coefficient magnitude of the features as well as minimising the error between the predicted and the true values. There are two common techniques for this approach i.e. Ridge and Lasso (least absolute shrinkage and selection operator) regressions [223]. These approaches differ in the penalisation mechanism, where the former adds a penalty equivalent to the square of the magnitude of coefficients, while the latter to the absolute value of the magnitude of coefficients.

3.4.2 Feature Transformation

The number of extracted features determines the dimensionality of the data set. As the dimensionality increases, the represented space grows exponentially requiring substantially more axes to adequately sample the spaces [224]. With the growing space, the data becomes increasingly sparse. In contrast to feature selection, feature transformation mathematically determines the most important dimensions to be kept while ignoring the rest. The idea is to find a smaller subset of dimensions that captures the most variation in the data. This reduces the dimensionality of the data while eliminating irrelevant features, making successful analysis simpler.

A common technique for this approach is principal component analysis (PCA) [11], [223]. The goal of PCA is to map the data from the original high dimensional space to a lower dimensional space that captures as much of the variations in the data as possible. PCA aims to find the most useful subset of dimensions to summarise the data. Figure 3-4a illustrates how two different features, X_1 (x-axis) and X_2 (y-axis), are reduced to only one feature: the first principal component, PC_1. In order to do this, PCA finds a new coordinate system, in which the first principle component, PC_1, captures most of the variation in the data. The second principle component, which is orthogonal to the first one, captures the second greatest variance, and for a greater number of features, the list continues. In PCA, the original data sample is mapped to this new set of dimensions with minimal loss in information.

Despite being a useful technique for reducing the dimensionality of a dataset, PCA can be difficult to interpret since the new dimension in the transformed data no longer has actual physical meaning [219]. The PCA approach is often compared with linear discriminant analysis (LDA) [223]. PCA treats the entire data as a whole and ignores the class labels (unsupervised learning) while LDA tries to explicitly model the difference between class labels (supervised learning). Instead of transforming the features into a different space, LDA draws a decision region between the given classes. For example, three classes are given in Figure 3-4b (blue, red and green) for two features, X_1 (x-axis) and X_2 (y-axis). As we can see, LD_1 gives a linear transformation to the data which maximises the separation of three classes.



(a) (b)

Figure 3-4. The illustration of (a) PCA and (b) LDA approach

3.5 Model Fitting

Predictive models can be categorised into two major classes: end-of-life predictions and future behaviour predictions [225]. Prognostic modelling falls into the former class while forecasting applications are the latter. For prognostics, the selected features need to be monotonic and they will reach an end point due to the degradation process e.g. insulation thickness. Hence, a prognostic model tries to fit a function with the aim of predicting when the threshold value will be reached. Forecasting parameters on the other hand, are unlikely to be monotonic but continuous over time with no threshold value e.g. wind speed and ambient temperature. Forecasting aims to make predictions at some time interval in the future instead of predicting the RUL or TTF.

In the literature, prognostic approaches are frequently separated into three categories: physics-based, data-driven and hybrid approaches [216], [226]–[229]. Physics-based uses mathematical equations to predict the physics governing failure thus requires knowledge of failure mechanisms, properties of the material and operating conditions. Data-driven on the other hand relies completely on the analysis of the historical and current data. This data provides the critical threshold for RUL estimation. Finally, a hybrid approach is a mixture of both physics-based and data-driven techniques for a more robust and accurate prediction. Several data-driven methods are discussed here to be considered as the learning algorithm for estimating the failure time from PD data.

3.5.1 Regression Analysis

Regression analysis is a modelling technique that investigates the relationship between a dependent (response) and one or more independent variables (features) [230]. There are various kinds of regression techniques available to make predictions. The simplest method is linear regression. In a simple linear regression model, a single response measurement, *Y*, is related to a single feature, *X*, for each observation. This model is represented by a linear function:

$$\hat{y}_i = A + Bx_i + e_i \tag{3-14}$$

where $\hat{y}_i$ is the predicted response, *A* is the intercept, *B* is the slope of the linear model, and *e* is the error term. In most problems where more than one feature is available, this expands to the multiple regression function:

$$\hat{y}_i = A + B_j x_{i,j} + e_i \tag{3-15}$$

for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, F$ where *N* is the number of observations and *F* is the number of features. For a nonlinear response, the best fit could be a curve with the power of the independent variable more than 1. The polynomial regression equation is given by:

$$\hat{y}_i = A + B_1 x_i + B_2 x_i^{\,2} + \ldots + B_g x_i^{\,g} + e_i \tag{3-16}$$

where *g* is called the degree of the polynomial. However, an attempt to fit a higher degree polynomial for a lower error may yield an overfitted model.

Other examples of nonlinear functions include exponential, inverse power, Gaussian, logarithmic, trigonometric, and Lorenz curves. Some functions, such as the exponential and inverse power, can be transformed to a linear equation hence linear regression could be performed. The exponential and inverse power law have been widely applied in insulation life models as discussed in Section 2.8. Both functions in equations 2-21 and 2-22 respectively, are rewritten here for further transformation into the linear forms. Considering a negative relationship, the exponential function is given by:

$$\hat{y}_i = k_1 \exp\left[-k_2 x_i\right]$$

and the inverse power function is given by:

$$\hat{y}_i = k_1 x_i^{-k_2}$$

The transformation of these equations are given by the equations 3-17 and 3-18 respectively.

$$\ln \hat{y}_i = \ln k_1 - k_2 x_i \tag{3-17}$$

$$\log \hat{y}_i = \log k_1 - k_2 \log x_i \tag{3-18}$$

For the linear representation therefore, a semilog scale is used for the exponential and a log-log scale for the inverse power function. The *A* and the *B* of the regression function are usually estimated using Least Squares fitting. This technique finds the best-fitting curve by minimizing the sum of the squares between the measurements and the model. Forward selection and backward elimination (explained in Section 3.4.1.2) can be applied for multivariate analysis, which for the regression approach is known as stepwise regression.

3.5.2 Artificial Neural Network

The idea of artificial neural network (ANN) originates from the biology of the human brain. An ANN consists of input layer, one or several hidden layers and output layer of processing elements [231]. A large set of training data allows the ANN to learn by example. The processing element comprises a node and a set of weights. The ANN learns an unknown function by adjusting its weights with observations of input and output. Through the cooperation of many of such processing elements (neurons), the resulting network structure becomes capable of learning very complex functions.

One prognosis model using an ANN is often called the *time delay neural network* (TDNN). This employs a sliding window approach for the input and output data which are usually given in a time series [232]–[235]. For a time series, *t* of length *m,* and user-defined subsequence length, *l*, all possible subsequences of *t* can be found by sliding a window of size *l* across *t* [236]. In Matlab, the neural network toolbox offers three types of TDNN. First is the *nonlinear input-output* that predicts a series

of outputs, $y(t)$ based on l past values of an external input series, $x(t)$. However, it is more accurate to use both past values of the external input and $y(t)$ (if available) as the input for *nonlinear autoregressive with external input* (NARX). Lastly, if no other information is available, the *nonlinear autoregressive* (NAR) technique can predict based on l past values of $y(t)$.

3.5.3 Autoregressive Moving Average

Autoregressive moving average (ARMA) is a combination of autoregressive (AR) and moving average (MA) techniques. ARMA is often used for short-term prediction in time series analysis where the data is usually periodically repeated and non-monotonic [237]. The key feature of the technique is the serial dependence between the measured values i.e. the current measures are potentially influenced by previous observations. The ARMA technique requires observation data to be weakly stationary with constant mean and variance over time [237]. For the case of a non-stationary process or apparent trend, a differencing approach is commonly employed to the variables and this generalisation technique is known as autoregressive integrated moving average (ARIMA). The applications of ARIMA are mainly forecasting of future behaviour and without considering threshold values e.g. load forecasting [238], electricity prices [239], and wind speed prediction [240]. Nevertheless, there is also evidence of its use for prognostic applications such as in [241].

3.6 Model Validation

Traditionally, the holdout approach is employed to increase the reliability of the performance measure which the available data set is split into training and testing sets [242]. The learning algorithm is then applied only to the training set while the test set is used to evaluate the model. However, the evaluation may depend on the choice of the sample sets. The model may yield very poor performance for an unfortunate split when all the test samples are outliers to the training samples. But, if some of the outliers are treated as training samples, the performance could be improved. This method is simple, but for a small number of samples a high bias

model can be expected because every sample may be needed for model building [242].

The estimation of the holdout method can be improved through the cross validation (CV) approach [223]. In *k*-fold cross validation, the data set is divided into *k* subsets, where one of the *k* subsets is used as the test set and the other $k - 1$ subsets are combined together to form a training set. The evaluation is repeated *k* times for a different subset as the test samples. The performance is determined from the average error of all *k* test samples. In this way, all samples are eventually used for both training and testing. Every sample is tested once and trained on $k - 1$ times. The variance of the error is reduced as *k* is increased.

When *k* is equal to the number of available data points, it is called leave-one-out cross validation (LOOCV). This means that the learning algorithm is applied to the $k - 1$ samples with one sample left as the test sample. With a greater value of *k*, longer computation time is needed since the learning process is repeated *k* times. This highlights the advantage of the holdout method over cross validation. LOOCV may result in a low bias model but with a high variance [242]–[244]. The study in [244] recommends 10-fold CV over LOOCV as moderate *k* values ($k = 10, 11, 12, \ldots, 20$) have smaller variance. The variance is also considerably small when compared to smaller *k* values ($k = 2, 3, 4, 5$).

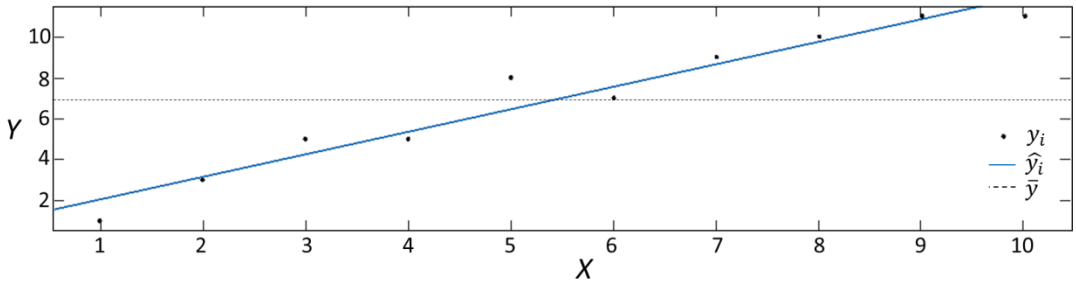
The bootstrap method is a smoothed version of cross validation [243], [245], [246]. In this method, the training set has the same size as the original set. That means, when a number of samples are selected for testing and the remaining samples are the training set, some of the training samples will be duplicated so that the sample size equals the original size. The original bootstrap method is described in [246] to have a low variance but a high bias. Therefore, a modified version called .632 bootstrap was proposed to decrease the bias. Still, the modified bootstrap can be unstable for a small number of samples, hence another version called .632+ bootstrap was proposed [243]. However, for a large number of features, LOOCV and *k*-fold CV might outperform this latest version [242].

3.7 Model Performance

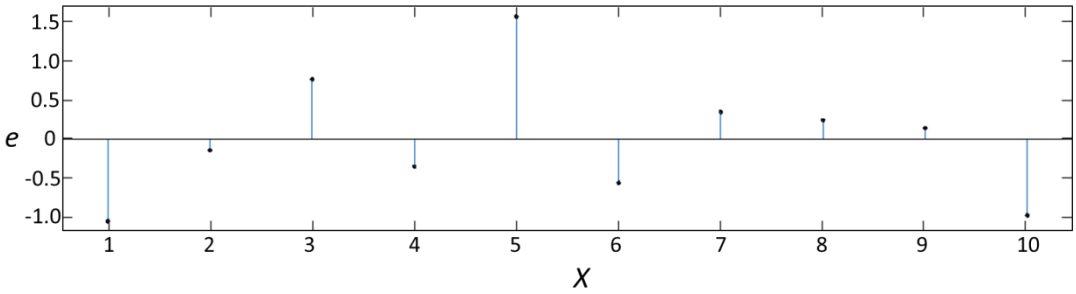
For a regression model, prediction of future values can be performed through extrapolation. A well-fitting regression model results in predicted values close to the observed data values. The characteristics of these two values can be measured using goodness-of-fit (GOF) statistics [246]. These statistical measures are mostly based on the three parameters shown in Figure 3-5a i.e. the actual response, y_i, the average of y_i, $\bar{y}_i$, and the fitted response, $\hat{y}_i$. The difference between the actual and fitted response gives the residual of the model which is the error term in equations 3-14 to 3-16.

$$e_i = y_i - \hat{y}_i \tag{3-19}$$

The residuals are usually visualised graphically to illustrate their behaviour so that the relationship between the model and the actual data can be observed. The model is said to fit the data well if the residuals appear randomly around zero. Figure 3-5b shows the residual plot of Figure 3-5a.



(a) Linear regression



(b) Residual plot

Figure 3-5. Example of a best fitting for a simple linear regression

The three response parameters in Figure 3-5a are used to compute the sum of squares due to error, SSE, the sum of squares of the regression, SSR, and the sum of squares about the mean, SST as follows:

$$SSE = \sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2 \tag{3-20}$$

$$SSR = \sum_{i=1}^{N} \left(\hat{y}_i - \bar{y} \right)^2 \tag{3-21}$$

$$SST = \sum_{i=1}^{N} \left(y_i - \bar{y} \right)^2 \tag{3-22}$$

where N is the total number of observations and $SST = SSE + SSR$. The SSE gives the total deviation of the fitted values from the actual values. From SSE, the mean squared error, MSE and root mean squared error, RMSE are computed using equations 3-23 and 3-24 respectively. These two measures are examples of GOF metrics. The coefficient of determination, R^2, is another GOF measure that can be determined using equation 3-25. *R*-squared ranges between 0 and 1, where 1 means the model explains all the variation of the independent variables.

$$MSE = \frac{SSE}{N} \tag{3-23}$$

$$RMSE = \sqrt{MSE} \tag{3-24}$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3-25}$$

R^2 assumes that every independent variable in the model helps to explain variation in the dependent variable [246]. However, some independent variables may not contribute and should be penalised. To cater for this, R^2 is modified to adjusted R^2:

$$R_{adj}{}^2 = 1 - \left(1 - R^2\right) \frac{N-1}{N-F-1} \tag{3-26}$$

where N is the size of the sample and F is the number of independent variables. Every time a new parameter is added to the model, R^2 will increase irrespective of

how well they are correlated to the dependent variable. Conversely, R_{adj}^{2} increases only when the correlation exists.

A survey of the metrics used to evaluate model performance has been undertaken in [225], which classified the metrics into four subcategories: accuracy, precision, robustness and trajectory. MSE and RMSE contribute to the accuracy of the model performance as they measure the closeness of the predicted value to the actual one. The precision reveals how closely the predictions are clustered together and can be represented by standard deviation, σ. Robustness measures the sensitivity of the predictions with changes of algorithm parameter variations or external disturbances. Meanwhile, the definition trajectory is not clearly defined. As a result, four metrics are proposed in [247] that evaluate the model algorithm systematically, one metric at a time.

The first metric, *prognostic horizon* (PH), determines the difference between the time index when the predictions first meet the specified performance criteria, i and the time index of actual end of life, EoL as given in equation 3-27. The specified performance criteria can be the $\pm\alpha$-bounds as shown in Figure 3-6. The time indexes when RUL predictions of algorithm 1 and 2 enter the $\pm\alpha$-bounds are denoted as i_1 and i_2 respectively. Inserting the time indexes into equation 3-27 results in PH_1 that is larger than PH_2. This means algorithm 1 has better performance since its predictions enters the $\pm\alpha$-bounds earlier than algorithm 2.

$$PH = EoL - i \tag{3-27}$$

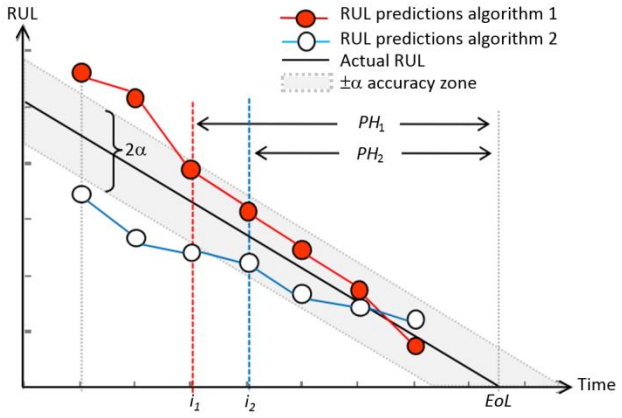


Figure 3-6. Prognostic horizon [248]

If an algorithm yields a reasonable prognostic horizon, the next metric, *α-λ accuracy*, can identify whether the predictions, *r*, fall within desired error margins (specified by parameter α) of the actual RUL, r_*, at specific time instances (specified by the parameter λ) by giving a binary output i.e. true or false to the condition in equation 3-28. Referring to Figure 3-7, the accuracy zone is given as $2\alpha r_*(i_\lambda)$ at any value of i_λ, where the lower and upper boundaries are $r_*(i_\lambda)\cdot(1-\alpha)$ and $r_*(i_\lambda)\cdot(1+\alpha)$ of equation 3-28 respectively. This means only predictions, $r(i_\lambda)$, that fall within the cone of accuracy in Figure 3-7 satisfy the α-λ accuracy metric. A positive result can be further evaluated by comparing α-λ accuracy with the actual RUL using the *relative accuracy* (RA) metric in equation 3-29.

$$r_*(i_\lambda)\cdot(1-\alpha) \le r(i_\lambda) \le r_*(i_\lambda)\cdot(1+\alpha) \tag{3-28}$$

$$RA = 1 - \frac{\left|r_*(i_\lambda) - \langle r(i_\lambda)\rangle\right|}{r_*(i_\lambda)} \tag{3-29}$$

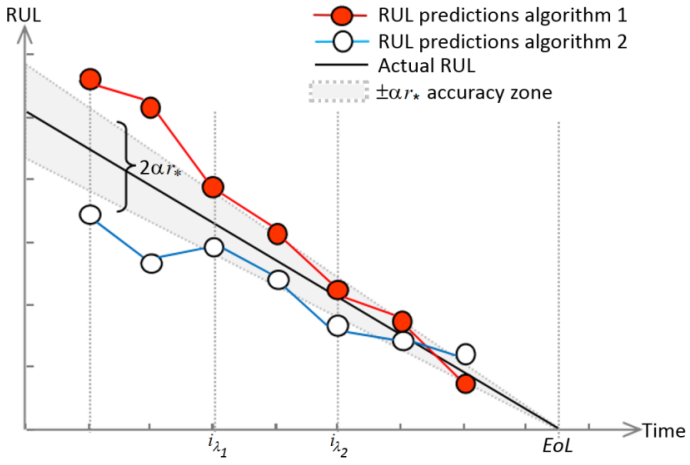


Figure 3-7. Concept of *α-λ* accuracy [248]

The RA is expressed at a specific time thus not represents the general behaviour of the algorithm. Evaluating the RA at multiple time instances requires an aggregate measure i.e. *cumulative relative accuracy* (CRA) which can be expressed as:

$$CRA = \frac{1}{|s_\lambda|}\sum_{i\in s_\lambda} w(r(i))RA_\lambda \tag{3-30}$$

where $w(r(i))$ is a weight factor as a function of RUL at all time indices, s_λ is the set of all t_λ when a prediction is made and $|s_\lambda|$ is the cardinality of the test.

Finally, when all the three metrics are satisfied, the *convergence* metric in equation 3-31 quantifies how fast the algorithm improves with time.

$$C_M = \sqrt{\left(x_c - t_p\right)^2 + {y_c}^2} \tag{3-31}$$

where C_M is the Euclidean distance between the centre of mass (x_c, y_c) and (t_P, 0). The centre of mass of the area under the curve, $M(i)$ is defined as follows:

$$x_c = \frac{1}{2}\frac{\sum_{i=P}^{\text{EoUP}}\left(t_{i+1}^2 - t_i^2\right)M(i)}{\sum_{i=P}^{\text{EoUP}}\left(t_{i+1} - t_i\right)M(i)} \tag{3-32}$$

$$y_c = \frac{1}{2}\frac{\sum_{i=P}^{\text{EoUP}}\left(t_{i+1} - t_i\right)M(i)^2}{\sum_{i=P}^{\text{EoUP}}\left(t_{i+1} - t_i\right)M(i)} \tag{3-33}$$

where EoUP is the time index for last useful prediction made.

3.8 Summary and Conclusion

In failure prognosis, physics-based models might have a direct relation with the degradation mechanism. However, the understanding of the underlying process may not fully relevant in practice, hence limiting its implementation. Instead, data-driven approach could be applied to historical degradation data for the development of prognostic model. Employing PD data as the degradation measure in predicting the lifetime of insulation is a concept of Type III prognostic algorithm. Type III or degradation based algorithm applies an indirect measure under the consideration of operating condition to characterise the lifetime.

The step-by-step procedure in developing a prognostic model is generally comprises six stages. The first stage is pre-processing the degradation data which may include

normalisation, standardisation or outliers detection. Normalisation or standardisation is necessary when features with varying scales are grouped together. Outliers in data set may be due to variability in the measurement or experimental errors. Identifying the outliers and assessing their impacts would help the decision either to include or remove the outliers.

Second, evaluating the suitability of the extracted features as prognostic parameter would indirectly aid the third stage, i.e. feature reduction. This can be done by selecting features based on three qualities metrics: monotonicity, prognosability, and trendability.

In data-driven technique, model selection involves the process of model fitting either through regression analysis or machine learning. Regression is the simplest technique for prognostic study, hence should be prioritised in developing the model depending on the behaviour of the selected features.

Next, validation of a prognostic model can be done by applying the model to a new set data. For a small size of data, cross validation or bootstrap technique is preferred than holdout since all samples can be used for both training and testing.

Finally, the performance of the prognostic model should be evaluated based on predictions from the test set data. Four model performance metrics (prognostic horizon, α-λ performance, relative accuracy, and convergence) proposed in [247] can be considered for evaluating the performance of the model.

This step-by-step procedure is the guideline applied in this thesis, for developing a prognostic model using PD data from electrical treeing experiment, which contribute to failure prognosis in cable monitoring.

Chapter 4
Identifying Prognostic Parameter from PD Data

4.1 Introduction

Automated PD data analysis systems have been shown to correctly diagnose defects causing PD [10], [79], [91], [96], [122], [131], [212], [213]. This thesis explored the extension of the concept towards predicting evolution of the defect; moving beyond diagnostics towards prognostics. This research work employed PD data from electrical treeing experiment described in the next section which incorporated both fundamental (50 Hz) and composite waveforms. In reference to the published works on PD diagnosis, a total of 31 features are extracted from the PD data using PRPDA and PSA. In this thesis, the behaviour of the features over time is observed and evaluated using three performance metrics i.e. *monotonicity*, *prognosability* and *trendability*. The extracted features might be redundant or not relevant. Therefore, filter technique, which is the simplest feature selection approach, is employed to reduce the size of the features. The relationship between the behaviour of the selected PD feature and the harmonic properties is identified and to be considered in the model development.

4.2 Electrical Treeing Experiment

The prognostic model developed in this work is based on electrical treeing experiment done by Sanjay Bahadoorsingh at the University of Manchester in 2008 [18]. An in-house test facility and PD instrumentation was developed under the Electrical Energy and Power Systems group in School of Electronic and Electrical Engineering, University of Manchester to investigate the influence of harmonics on electrical tree growth and breakdown times [173], [249], [250] as well as the influence on partial discharge patterns due to electrical treeing [175].

4.2.1 Test Waveforms

As discussed in Section 2.7, Montanari and Fabiani [165] have formulated three measures to describe the amplitude and waveshape of composite waveforms, i.e. peak parameter, $K_{p,}$ RMS parameter, K_{rms}, and waveshape parameter, K_s as in equations 2-18 to 2-20 respectively. Among the three parameters, K_p was found to be the most prevailing in lifetime reduction [164]–[167] as the harmonic distortion may increase the peak voltage. The change in peak however, does not necessarily changes the RMS voltage, hence, K_{rms} is less dominant than K_p. An increase in peak voltage will increase the voltage slope, *du/dt*, which is related to K_s [251]. In [170], the increase in *du/dt* was claimed to increase PD activity thus K_s was found to be more dominant than K_p. In contrast, research in [251] found that K_s was the least dominant among the three parameters.

Due to the contradict ideas on the impact of K_s on the life reduction, Bahadoorsingh then investigated the influence of only K_s parameter (equation 2-20) to the insulation lifetime, i.e. $K_p = K_{rms} = 1$. The investigation also considers the THD (equation 2-17) which is also a waveshape measure. Both K_s and THD are rewritten here to aid the interpretation of composite waveforms explained in this chapter.

$$THD = \sqrt{\sum_{h=2}^{N}\left(\frac{U_h}{U_1}\right)^2}$$

$$K_s = \sqrt{\sum_{h=1}^{N} h^2 \left(\frac{U_h}{U_1}\right)^2}$$

The research incorporated the use of both fundamental (50 Hz) and composite waveforms. The composite waveforms were characterised by harmonic components superimposed onto the fundamental. Six composite waveforms were selected based on analysis of many composite waveforms with varied K_s and THD values. These composite waveforms are produced using a peripheral component interconnect (PCI) based arbitrary waveform generator (AWG) and an amplifier that capable of reproducing waveforms with an output in the range of ± (20-30) kV. The peak voltages of all waveforms are kept constant at 14.4 kV. At this voltage, the recorded PD is assumed to have originated internally from treeing activity and/or externally due to imperfect boundaries in the physical setup [18]. The mathematical expressions of the fundamental, u_7 and the six composite waveforms, $u_1, u_8, u_9, u_{11}, u_{12}$ and u_{13} are given in equations 4-1 to 4-7 where the subscripts refer to wave group detailed in Table 4-1.

$$u_1(\theta) = 10.25\sqrt{2}\left(\sin\theta + 0.4\sin 3\theta\right) \tag{4-1}$$

$$u_7(\theta) = 10.18\sqrt{2}\left(\sin\theta\right) \tag{4-2}$$

$$u_8(\theta) = 9.7\sqrt{2}\left(\sin\theta + 0.05\sin 5\theta\right) \tag{4-3}$$

$$u_9(\theta) = 10.4\sqrt{2}\left(\sin\theta + 0.05\sin 7\theta\right) \tag{4-4}$$

$$u_{11}(\theta) = 9.35\sqrt{2}\left(\sin\theta + 0.178\sin 7\theta\right) \tag{4-5}$$

$$u_{12}(\theta) = 10.15\sqrt{2}\begin{pmatrix}\sin\theta + 0.032(\sin 5\theta + \sin 7\theta + \sin 11\theta + \\ \sin 13\theta + \sin 23\theta + \sin 25\theta\end{pmatrix} \tag{4-6}$$

$$u_{13}(\theta) = 10.18\sqrt{2}\begin{pmatrix}\sin\theta + 0.02(\sin 5\theta + \sin 7\theta + \sin 11\theta + \\ \sin 13\theta + \sin 23\theta + \sin 25\theta\end{pmatrix} \tag{4-7}$$

The fundamental 50 Hz signal is represented by Wave 7. Among the composite waveforms, Wave 1 represents the most distorted waveform with 40% of THD due to the 3rd harmonic. This very high THD value is in reference to work reported in [30], [174]. The value intentionally exceeds the maximum allowable THD in IEC 61000-2-4 [172] in order to highlight the impact of voltage distortion on the electrical tree growth since no significant impact was found for lower THDs. Waves 8 and 9 employ the 5th and 7th harmonic orders respectively with THD = 5%. The 7th harmonic order is also employed by Wave 11 but with higher THD and K_s compared to Wave 9. Both Wave 12 and 13 utilize six harmonic orders i.e. the 5th, 7th, 11th, 13th, 23rd and 25th with Wave 12 having higher THD and K_s than Wave 13. All the waveforms are plotted in Figure 4-1 and Figure 4-2.

Table 4-1. Properties of the seven test waveforms

Wave	***Properties***				
	Harmonic Order	**% Magnitude per Harmonic**	**Phase**	**K_s**	**THD%**
1	3	40.0	0	1.56	40.00
7	1	0.0	0	1.00	0.00
8	5	5.0	0	1.03	5.00
9	7	5.0	0	1.06	5.00
11	7	17.8	0	1.60	17.80
12	5, 7, 11, 13, 23, 25	3.2	0	1.60	7.85
13	5, 7, 11, 13, 23, 25	2.0	0	1.27	5.00

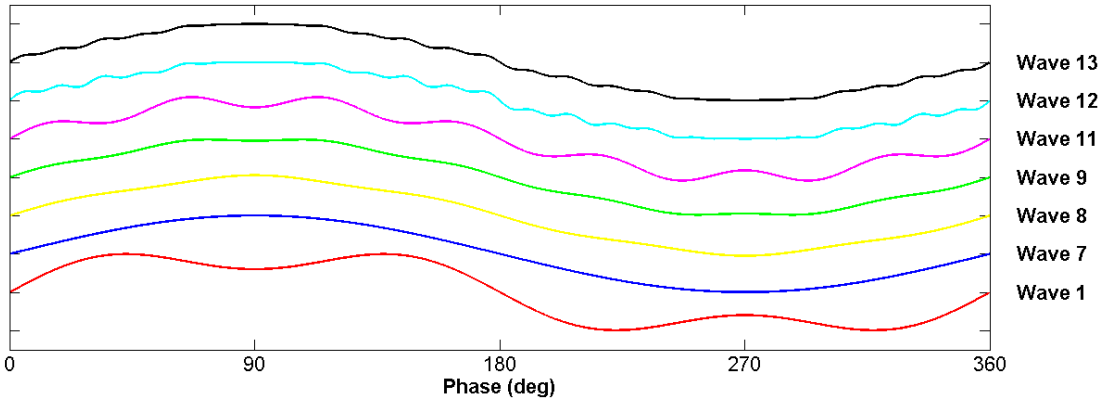


Figure 4-1. Illustration of the test waveforms defined in Table 4-1

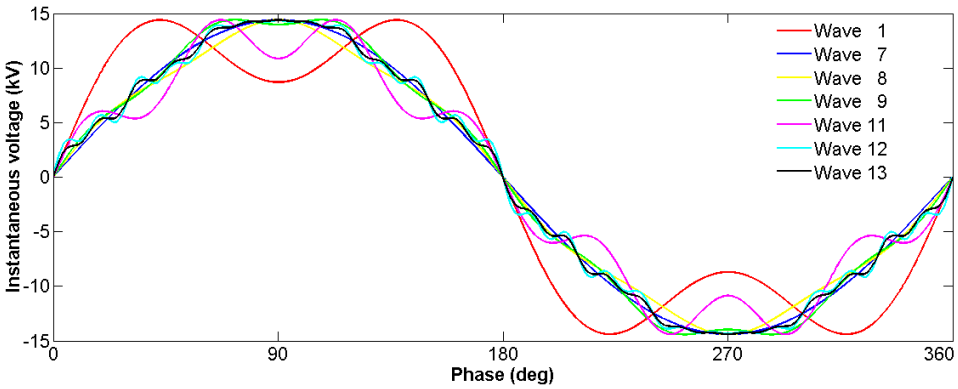


Figure 4-2. Test waveforms with 14.4kV voltage peak

4.2.2 Insulation Sample

Research in [18] employed epoxy resins produced by Huntsman as the insulation sample in electrical treeing experiments. The insulation sample is prepared by mixing the epoxy resin (Araldite LY 5052) and an amine hardener (HY 5052) using a weight ratio of 100:38. This mixture was vacuumed for almost an hour to remove embedded bubbles. The epoxy resin is encased in a hollow acrylic square cube of 25 mm side. The samples are tested using a point-plane test arrangement. The hypodermic needle with a tip radius of 3 μm, length of 51 mm and thickness of 1.1 mm was set with a plane separation gap of 2 ± 0.5 mm as shown in Figure 4-3. Most hypodermic needles are coated with a transparent lubricant to reduce frictional forces permitting easy movement of the needle under human skin [252]. This lubricant coating is removed by soaking the needles in Silstrip for 72 hours. The commercially available chemical Silstrip derives its name from its silicone stripping attributes [252].

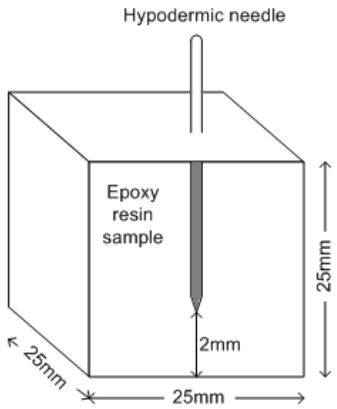


Figure 4-3. Schematic of epoxy resin sample [252]

4.2.3 Partial Discharge Measurement

The developed in-house test facility includes a partial discharge measuring system with the ability to plot PRPD [18]. The PD instrumentation system monitored and recorded electrical PD activity in accordance with the IEC 60270, High Voltage Test Techniques – Partial Discharge Measurements standard [75] over the frequency range 100 to 400 kHz. The system can be divided into two parts: hardware and software as shown in Figure 4-4.

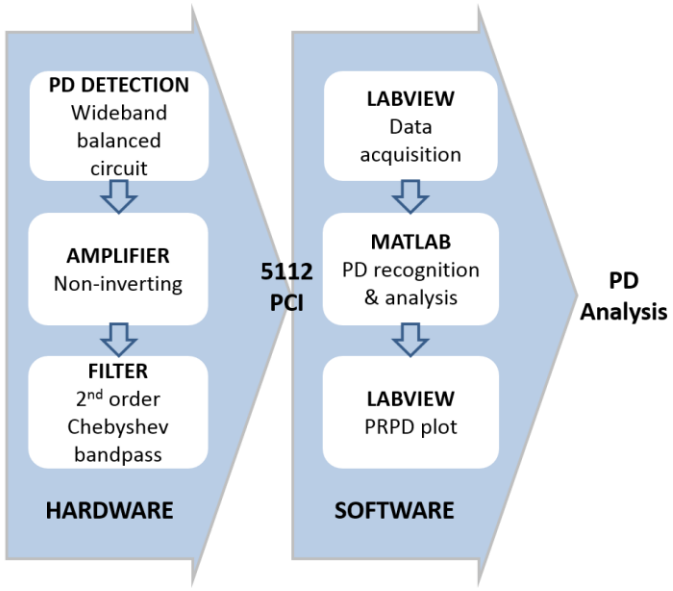


Figure 4-4. Overview of partial discharge measuring system

The hardware part includes the PD detection circuit, amplifier circuit and filtering circuit. Electrical PD detection is utilised in accordance to the IEC 60270 standard [75]. Instead of using a straight detection circuit, a wideband balanced circuit is employed which reduces the disturbances from the range of 25 pC-30 pC to 10 pC-11 pC. The disturbances might be due to the high voltage amplifier when energized at zero voltage, external high voltage tests in the vicinity, internal noise from the measuring system, switching operations in adjacent circuits and radio transmissions etc [75]. This arrangement yielded a minimum detectable level of 5.5 pC at a multiplier of 50 pC/V. The detected PD pulses are amplified using a non-inverting amplifier of variable gain of approximately 100. The filter circuit implements the 2nd order Chebyshev bandpass filter with frequency response of 100 kHz to 400 kHz.

The software part contains data acquisition, PD recognition and analysis and finally PRPD plot. The peripheral component interconnect (5112 PCI) from National Instruments (NI) is used for data acquisition of input signal from PD measuring system. Through the LabVIEW environment, data from the PCI is processed and analysed in Matlab and finally exported back to the LabVIEW platform for PRPD plotting. The partial discharge points were superimposed onto a reference waveform plot and continuously updated to produce an output display as shown in Figure 4-5. The *x*-axis represents the phase locations in degrees whilst the *y*-axis is formatted in pC for partial discharge activity. For each PD, the apparent charge and excitation waveform phase position are logged in LabVIEW Measurement (.lvm) format. Discharges are collected as bursts of activity (hereafter called PD burst or PD pattern). The raw PD data were collected at a sampling rate of 5 MSamples/s over periodic 80 ms windows of continuous capture. This data is sufficient for PRPD and PSA approaches.

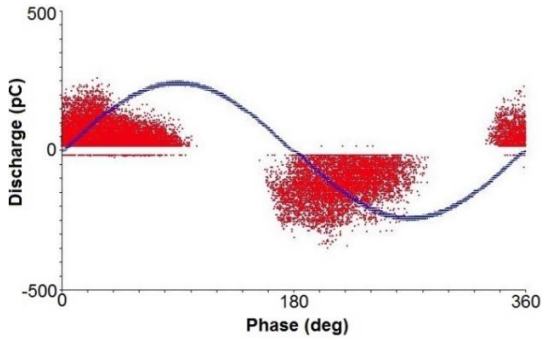


Figure 4-5. PRPD plot from the PD instrumentation [18]. Red points represent the apparent charges while the blue line represents the applied voltage.

4.2.4 Experimental Procedure

The electrical treeing experiment involved a total of 42 samples with six samples tested under the influence of each (of the seven) composite waveforms [18]. The plan of the experiment is summarised in Figure 4-6. The first stage is to initiate electrical treeing under the influence of 50 Hz at 18.0 kV peak within 30 minutes. This initiation is indicated by an electrical tree of length 10 μm. If no tree was initiated, that sample was discarded and a new sample was setup for testing. After 30 minutes, the composite waveform was increased to 14.4 kV peak and maintained for a

maximum of 2.5 hours as the excitation waveform. During this second stage, partial discharge readings were recorded continuously in 5 minute batches and visual images of tree growth were captured for every 2 minutes. If the breakdown occurs during this stage, the time was recorded, otherwise, a ramp process was invoked in the final stage increasing the voltage at rate of 1 kVrms/min from 10.6 kV – 24.8 kV peak. Should the sample breakdown within ten minutes, the breakdown voltage was recorded.

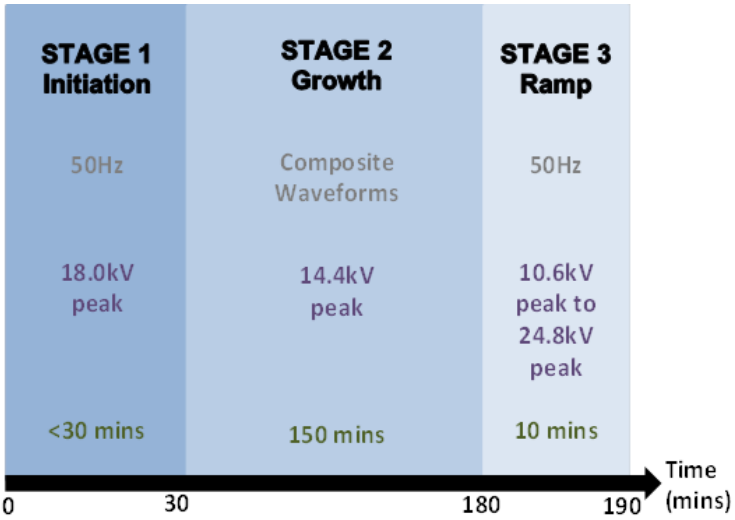


Figure 4-6. General plan for each sample under test [18]

4.2.5 Sample Properties

The exercise explained in Section 4.2.4 was originally to comprehensively analyse the influence of power quality (as a function of the composite waveforms due to THD and K_s) on electrical tree growth and breakdown times for all 42 samples tested [173], [249], [250]. However, in this research, the prognostic modelling is derived from the results of the samples that breakdown in Stage 2. Out of 42, only 31 samples reached breakdown within 2.5 hours. The time that the first tree branch touching the ground electrode was also recorded. Table 4-2 shows the identification of the 31 samples and both the touch ground time, t_{gnd} and the breakdown time, t_{bd}. The samples are labelled according to the group of the tested waveform, the batch of the samples and the sample number as shown in Figure 4-7. From 31 samples, one sample from each wave group was randomly picked as the test samples with the remaining 24 samples as the training.

Table 4-2. Breakdown time of 31 samples [18], and training/testing split

Wave	*Stage 2 RMS (kV)*	*Stage 2 Peak (kV)*	*Sample ID*	*Stage 2 Touch Ground Time (min)*	*Stage 2 Breakdown Time (min)*	*Training*	*Testing*
1	11.04	14.4	01T346	40	115		√
			01T354	32	118	√	
			01T382	42	129	√	
			01T426	33	60	√	
7	10.18	14.4	07T332	50	93	√	
			07T343	45	103	√	
			07T355	45	115		√
			07T383	43	130	√	
8	9.71	14.4	08T365	35	147	√	
			08T374	41	118	√	
			08T392	48	140		√
9	10.40	14.4	09T325	30	122	√	
			09T344	48	123	√	
			09T381	37	107		√
			09T412	47	138	√	
			09T421	-	71	√	
11	10.14	14.4	11T333	42	108	√	
			11T345	45	143	√	
			11T372	40	129		√
			11T391	45	120	√	
			11T425	37	58	√	
12	10.21	14.4	12T342	43	123	√	
			12T373	40	113	√	
			12T394	40	138	√	
			12T424	38	77		√
13	10.19	14.4	13T324	30	93	√	
			13T331	53	117	√	
			13T363	47	125		√
			13T376	48	122	√	
			13T393	42	149	√	
			13T423	48	94	√	

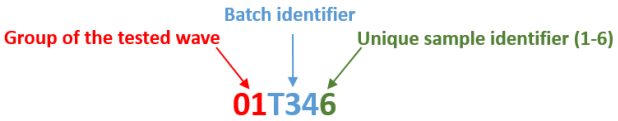


Figure 4-7. Format of sample ID

4.3 Analysis of Partial Discharge Data

This section describes the process of extracting PD features using PRPDA and PSA approaches. A total of 31 features were extracted i.e. 24 from PRPDA and 7 from PSA. These features are mostly used for failure diagnosis. In this thesis, the diagnostics concept is extended for predicting the TTF. Hence, the behaviour of these features over time is analysed as well as the impact of the harmonics on the features' behaviour.

4.3.1 Phase-Resolved Partial Discharge Analysis

This work employed the phase-resolved technique described in Section 2.4.2.1. A total of 24 statistical features were extracted using equations 2-6 to 2-9 considering both positive and negative half cycles. Figure 4-8 summarizes the flow of PRPDA reported in this work with more detailed description in the following subsections.

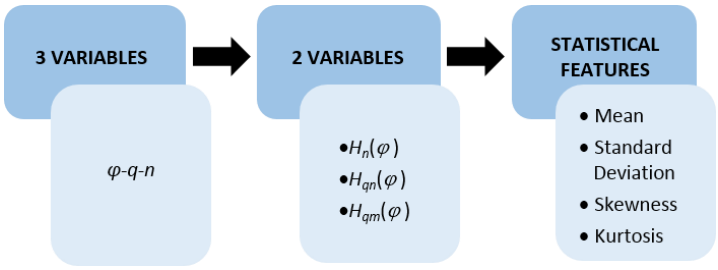


Figure 4-8. The flow of PRPDA

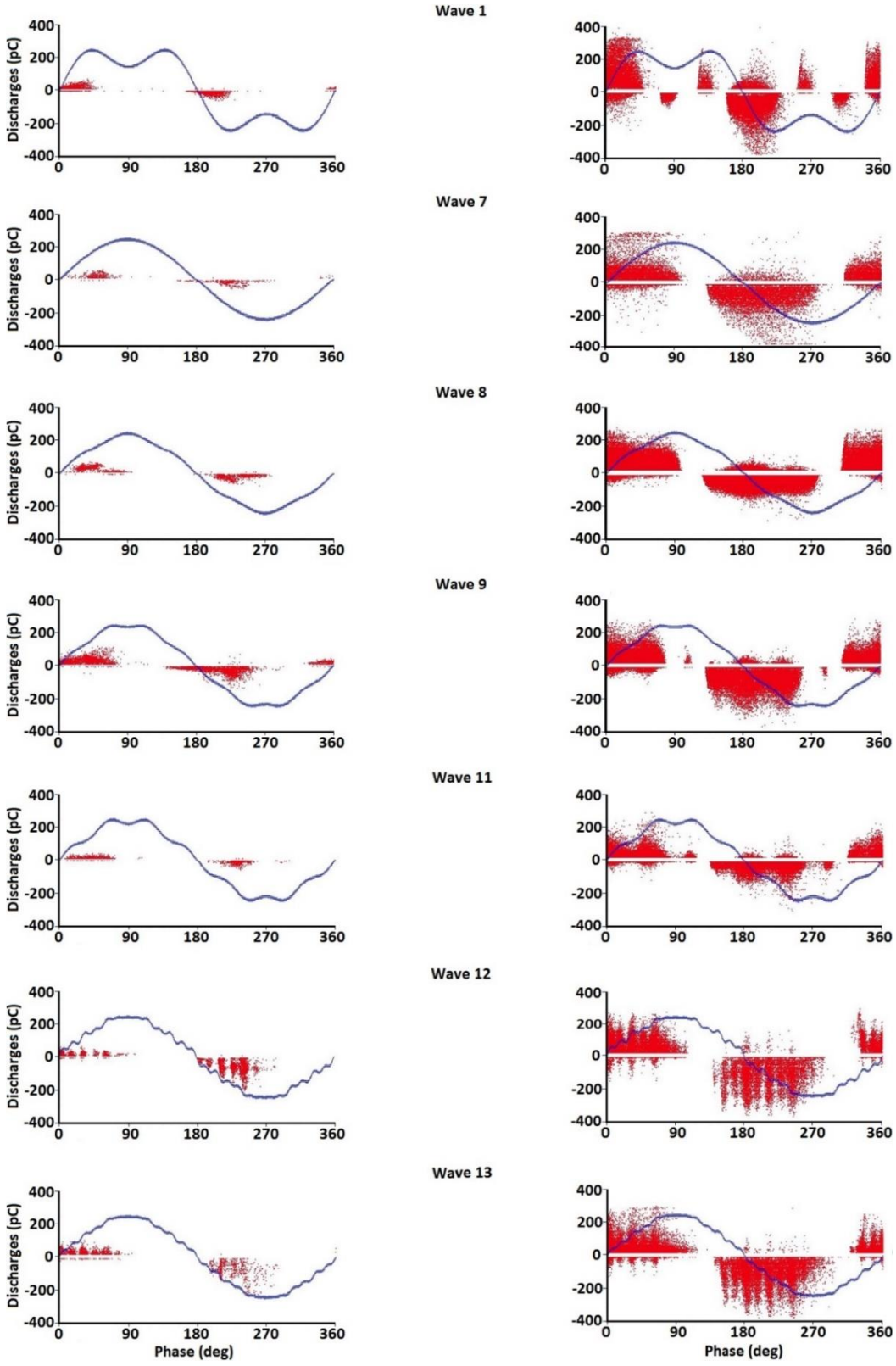
4.3.1.1 *φ-q-n* Plots

Phase-resolved analysis focuses on phase distribution thus the three important parameters, *φ-q-n*, were plotted in two axes, *φ-q*, where the pulse repetition rate is illustrated by the intensity of discharges. The PRPD patterns of all training samples were observed from the start of treeing until breakdown occurs to get the general idea of how the pattern changes.

Figure 4-9 shows the *φ-q-n* plots of treeing samples, taking one sample from each wave group during early and late growth in 5-minute batches. Generally, two PD clusters can be seen, i.e. the positive PDs which dominantly occurred along the rising edges of the voltage waveform and the negative PDs along the falling edges. These two PD clusters are growing in both axes directions throughout the treeing process but remain separated because no PD activity occurs after the peaks i.e. dead zones. This can be seen in all test waveforms except for Wave 1 due to high distortion. In the literature, the PD occurrence is mostly reported in the first and third quadrant. However, during the late growth, we can see the PD slightly enters the second and fourth quadrant. Those PD might be due to the residual charges in the PD site [161].

Among the six composite waveforms, only Wave 8 has two peaks per voltage cycle, resembling the fundamental, Wave 7. With low THD and K_s, the resultant PRPD looks similar to the fundamental. On the other hand, the PRPD pattern of other composite waveforms which have four peaks per voltage cycle differ from Wave 7 and 8 showing clearer harmonic influence. Although the PRPD patterns of Wave 12 and 13 are greatly affected (due to more than one harmonic), the resultant PD can still be separated into two clusters. New PD clusters seem to grow in Wave 9 and 11 (7^{th} harmonic) at around 110° and 290° where the rising edge of the second peak is. The distortion, however, is not high enough to induce negative PD at the falling edge of the first peak as shown by the six PD clusters of Wave 1 (3^{rd} harmonic). Hence, only 4 clusters can be seen in Wave 9 and 11. The extra two clusters however can be merged to the original clusters since their phase occurrence is within the range of the original cluster.

The most affected PRPD pattern, Wave 1, has the highest THD and slightly lower K_s than Wave 11 and 12. This means the impact of THD on the resultant PD clusters is greater than K_s. Also, it can be concluded that the waveshape of the PRPD pattern is greatly influenced by the harmonic order. Waveforms with higher harmonic order will have more rising and falling edges thus resulting in more distorted PRPD patterns.



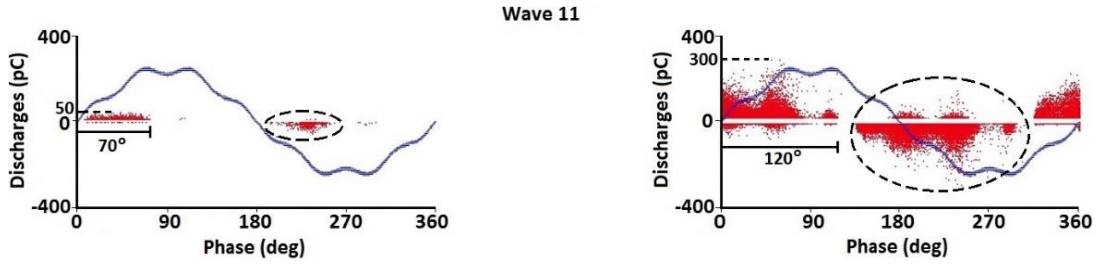
(a) Early growth (b) Late growth

Figure 4-9. The PRPD pattern from each wave group during (a) early growth and (b) late growth. Red points represent the apparent charge while the blue line represents the applied voltage.

Prior to breakdown, three changes can be spotted from Figure 4-9 as listed below:-

- The number of PD pulses per voltage cycle increases
- The maximum PD magnitude increases
- The phase distribution widens

The listed changes are shown graphically in Figure 4-10 taking Wave 11 from Figure 4-9 as an example. The number of PD pulses is represented by the red points which are sparsely dispersed during the early growth and turn denser towards breakdown as indicated with the circles. The maximum discharge can be seen to increase from 50 pC to 300 pC as well as phase distribution from 70° to 120° (positive PDs). The *φ-q-n* plots of all samples were then transformed to the two-variable plots as described in the next section before extracting the statistical features.



(a) Early growth (b) Late growth

Figure 4-10. The changes of PRPD pattern for a sample tested with Wave 11. Red points represent the apparent charge while the blue line represents the applied voltage.

4.3.1.2 Two-Variable Plots

The two-variable plots were derived from the *φ-q-n* plots for the convenience of comparison between the number of PD pulses, *n*, and the discharge magnitude, *q*. In reference to [111], the discharge magnitude, *q*, is represented by the average and maximum values. Thus, three 2D phase distributions were selected for further analysis:

- The distribution of the number of discharges, $H_n(\varphi)$
- The distribution of the maximum values of discharges, $H_{qmax}(\varphi)$
- The distribution of the average values of discharges, $H_{qn}(\varphi)$

To generate those three plots, the one voltage cycle of 360° is divided into 60 buckets giving each bucket 6° to smooth out some of the phase variability of PD. To illustrate this, one-quarter (0° - 90°) of voltage cycle with 15 buckets is considered as shown in Figure 4-11. For each bucket, the number of discharges, H_n, the maximum value of discharge, H_{qmax} and the average value of discharge, H_{qn} are determined where H_{qmax} and H_{qn} are using absolute values. These values are plotted in Figure 4-12 showing the three 2D phase distributions in reference to Figure 4-9.

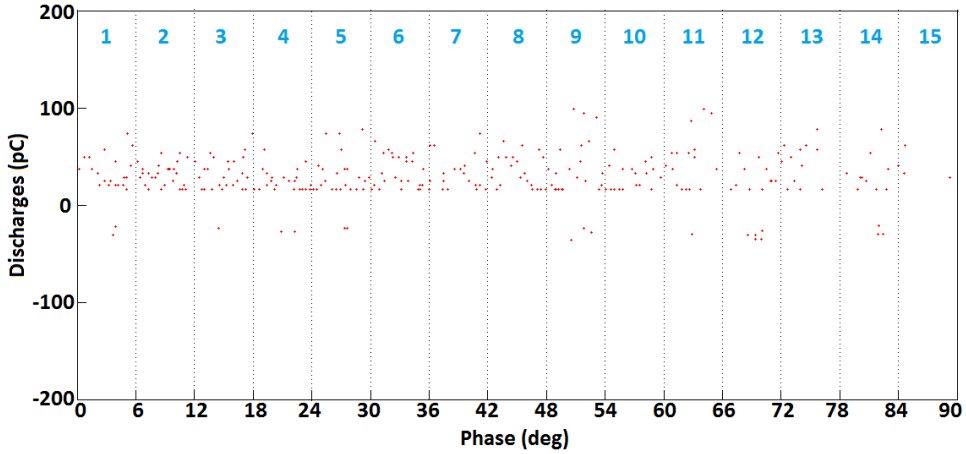
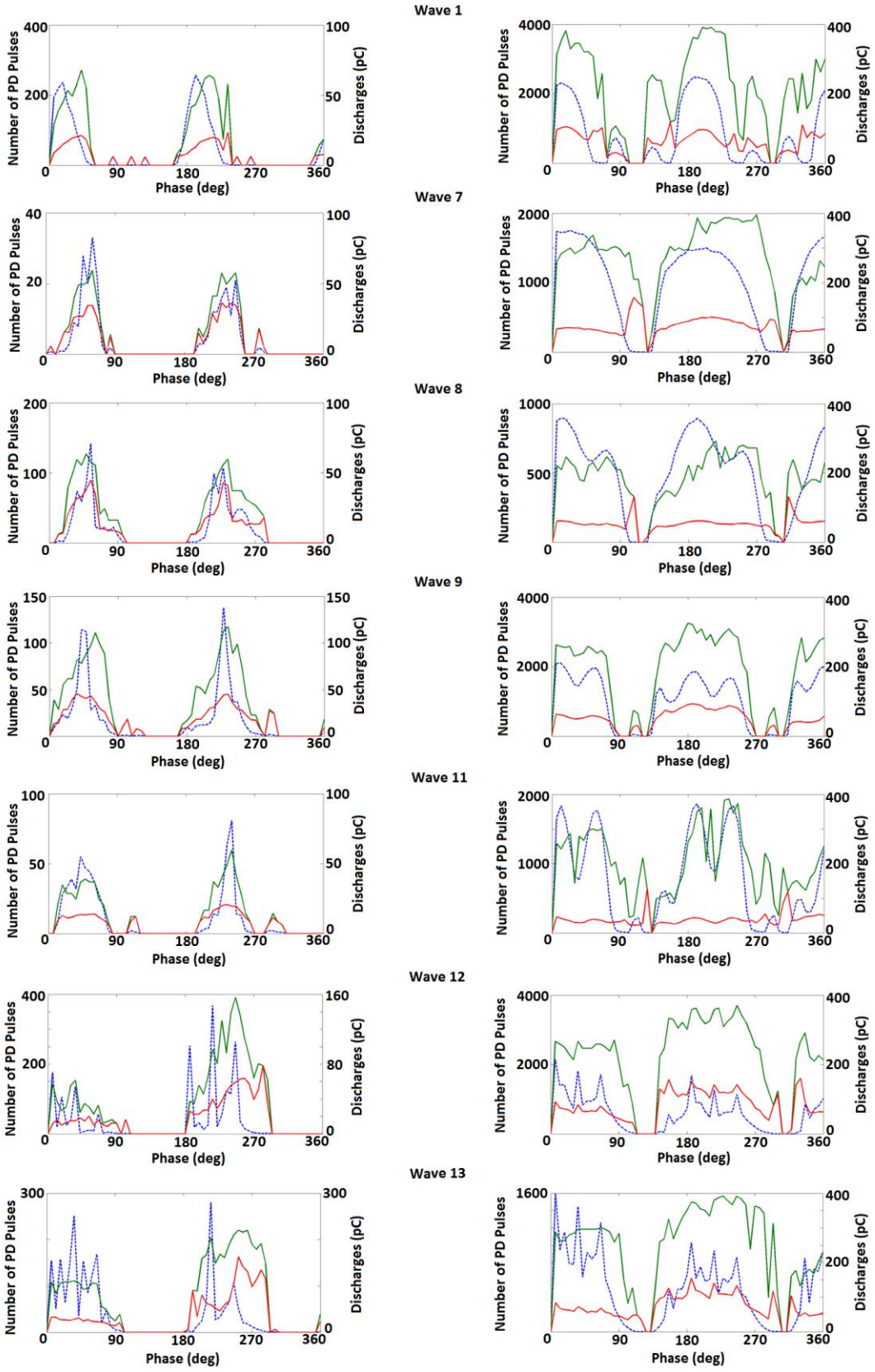


Figure 4-11. Fifteen buckets in one-quarter voltage cycle

As mentioned previously in Section 4.3.1.1, the positive and negative PD clusters in Figure 4-9 are clearly separated because of no PD activity after the peaks which gives gaps between the clusters. Some of these gaps however, are not clearly shown in the 2D distributions especially when the gaps became narrower during the late growth as shown in Figure 4-12b. The number of peaks of the non-sinusoidal waveforms and the dead zone can been seen clearest in the $H_n(\varphi)$ distribution. When comparing the $H_{qmax}(\varphi)$ with $H_{qn}(\varphi)$ distributions, the patterns look similar except that $H_{qn}(\varphi)$ can sometime gives patterns where the positive and negative clusters are difficult to be identified as in the late growth of Wave 1 and Wave 11. The $H_{qn}(\varphi)$ also behaves differently in the late growth of Wave 7 and 8 with a sudden rise before or after a plateau. It can be concluded that $H_{qn}(\varphi)$ is the least similar to the φ-q-n plots in Figure 4-9.



(a) Early growth (b) Late growth

Figure 4-12. H_n (dashed blue line), H_{qmax} (solid green line) and H_{qn} (solid red line) distributions of Figure 4-9.

4.3.1.3 Statistical Features

The three changes to the PRPD pattern observed in Section 4.3.1.1 are examined through four statistical moments [91]–[94] i.e. mean, μ, standard deviation, σ, skewness, *Sk*, and kurtosis, *Ku*.

The changes in the number of PD pulses can be tracked from the μ and σ of the $H_n(\varphi)$ distribution and changes in maximum PD magnitude from the μ and σ of the $H_{qmax}(\varphi)$ and $H_{qn}(\varphi)$. Both μ and σ are expressed in equations 2-8 and 2-9 respectively where the input variables, x_i, and the number of samples, N, would be the properties of the discharges ignoring the phase values.

In order to investigate the correlation between the changes in phase distribution and the PD pulse height (either $H_n(\varphi)$, $H_{qmax}(\varphi)$ or $H_{qn}(\varphi)$), the skewness and kurtosis of the three 2D phase distributions were determined. Moreover, it is mentioned in [91] that discharge parameters of a single defect can be fairly well described by a normal distribution process including skewness and kurtosis. In this case, the phase values are treated as grouped data, x_i, where PD pulse heights are the frequencies, f_i. The mean, standard deviation, skewness and kurtosis of a grouped data can be determined using equations 4-8 to 4-11 respectively. The kurtosis is in the form of excess kurtosis which is relative to normal distribution kurtosis.

$$\mu = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \tag{4-8}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i \left(x_i - \mu\right)^2}{\sum_{i=1}^{n} f_i}} \tag{4-9}$$

$$Sk = \frac{\sum_{i=1}^{n} f_i \left(x_i - \mu\right)^3}{\sigma^3 \sum_{i=1}^{n} f_i} \tag{4-10}$$

$$Ku = \frac{\sum_{i=1}^{n} f_i \left(x_i - \mu\right)^4}{\sigma^4 \sum_{i=1}^{n} f_i} \tag{4-11}$$

As a result, for each 2D plot, four statistical moments are calculated for positive and negative PDs separately, giving a total of 24 features as tabulated in Table 4-3.

Table 4-3. The total of 24 statistical features from PRPDA

2D PRPD Distributions	***Statistical Features***	
	Positive PD	**Negative PD**
$\boldsymbol{H_n(\varphi)}$	H_n *pos* μ	H_n *neg* μ
	H_n *pos* σ	H_n *neg* σ
	H_n *pos sk*	H_n *neg sk*
	H_n *pos ku*	H_n *neg ku*
$\boldsymbol{H_{qmax}(\varphi)}$	H_{qmax} *pos* μ	H_{qmax} *neg* μ
	H_{qmax} *pos* σ	H_{qmax} *neg* σ
	H_{qmax} *pos sk*	H_{qmax} *neg sk*
	H_{qmax} *pos ku*	H_{qmax} *neg ku*
$\boldsymbol{H_{qn}(\varphi)}$	H_{qn} *pos* μ	H_{qn} *neg* μ
	H_{qn} *pos* σ	H_{qn} *neg* σ
	H_{qn} *pos sk*	H_{qn} *neg sk*
	H_{qn} *pos ku*	H_{qn} *neg ku*

As mentioned previously, the phase range is widening throughout the tree growth. Referring to the $H_n(\varphi)$ distribution in Figure 4-12, we can see the positive PD clusters of Wave 7 widening from 0° - 90° in early growth to 300° - 120° in late

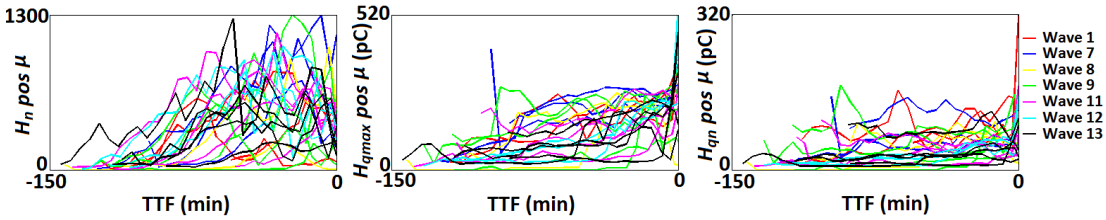
growth. No PD was recorded after 120° (slightly after the positive peak) until the negative PDs start to occur at around 140° until 300° (slightly after the negative peak). Wave 8 has a similar pattern to Wave 7 thus, was expected to have the same phase range. Although Wave 9 and 11 have extra clusters of positive and negative PDs (see *φ-q-n* plots in Figure 4-9 and 2D plots in Figure 4-12), both have the same phase range as Wave 7 and 8 since the extra clusters are within the range. For the cases of Wave 12 and 13, no extra PD cluster appears, therefore they result in the same phase range. The special case only applies on Wave 1 which has three clusters of positive and negative PDs, thus, giving a different phase range. Table 4-4 summarises the phase ranges of all test waveforms.

The PRPD pattern of each sample is recorded for every 5 minutes until breakdown occurs. In order to identify the prognostic indicators, every 5 minute 2D PRPD pattern is represented by the 24 statistical features in Table 4-3. The mean, standard deviation, skewness and kurtosis of the positive and negative 2D phase distributions of all 24 training samples are shown in Figure 4-13 to Figure 4-20 respectively. The features are plotted in reference to the TTF as indicated in equation 4-12 where t_i is the current time and t_{bd} is the breakdown time. This is to aid observation on the trend when reaching breakdown point. The colour of the line graph represents the test waveform group.

$$TTF_i = t_i - t_{bd} \tag{4-12}$$

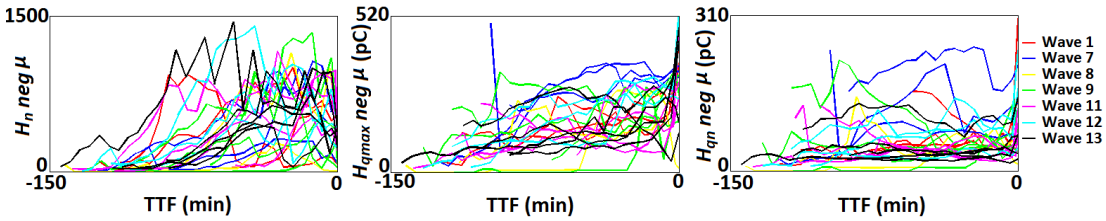
Table 4-4. The phase ranges of positive and negative cycles for each test waveform

Wave	***Phase Range (°)***	
	Positive Cycle	**Negative Cycle**
1	0 – 59 108 – 155 240 – 287 336 – 359	60 – 107 156 - 239 288 – 335
7, 8, 9, 11, 12, 13	0 – 119 300 – 359	120 – 259



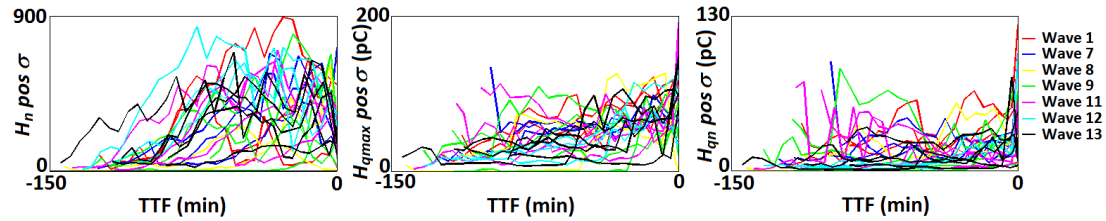
(a) (b) (c)

Figure 4-13. Mean of positive (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



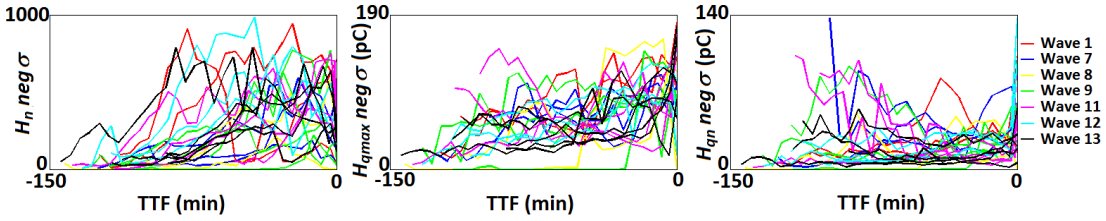
(a) (b) (c)

Figure 4-14. Mean of negative (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



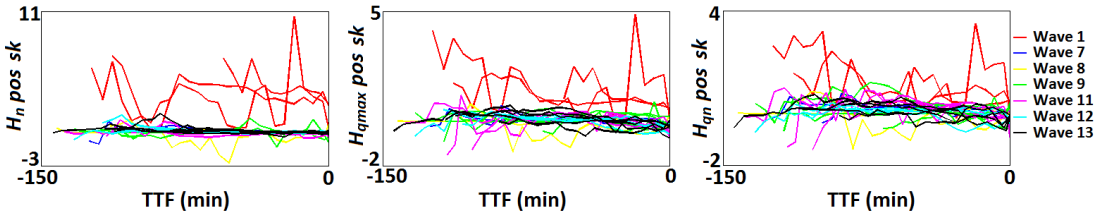
(a) (b) (c)

Figure 4-15. Standard deviation of positive (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



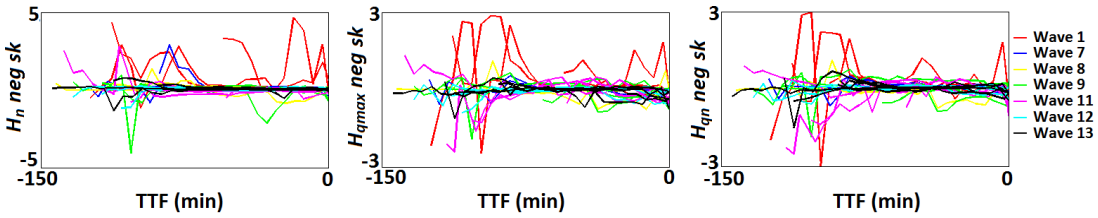
(a) (b) (c)

Figure 4-16. Standard deviation of negative (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



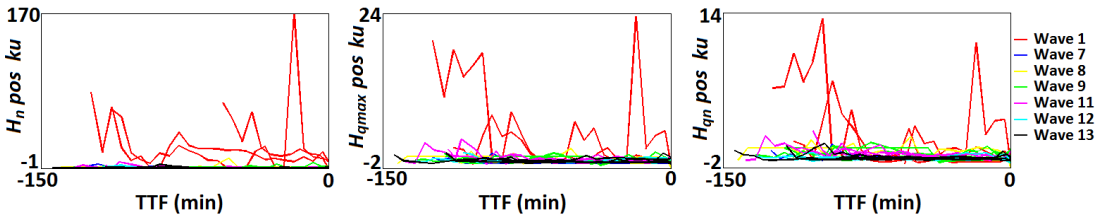
(a) (b) (c)

Figure 4-17. Skewness of positive (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



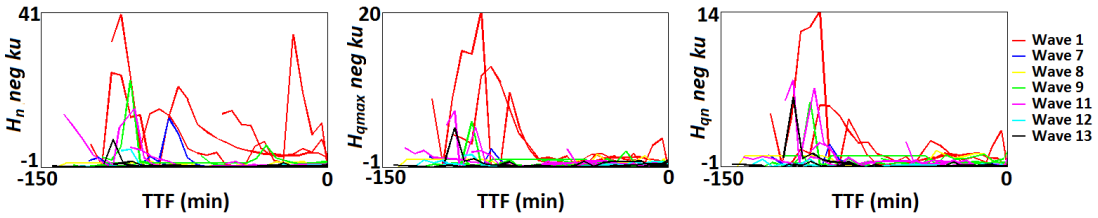
(a) (b) (c)

Figure 4-18. Skewness of negative (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



(a) (b) (c)

Figure 4-19. Kurtosis of positive (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$



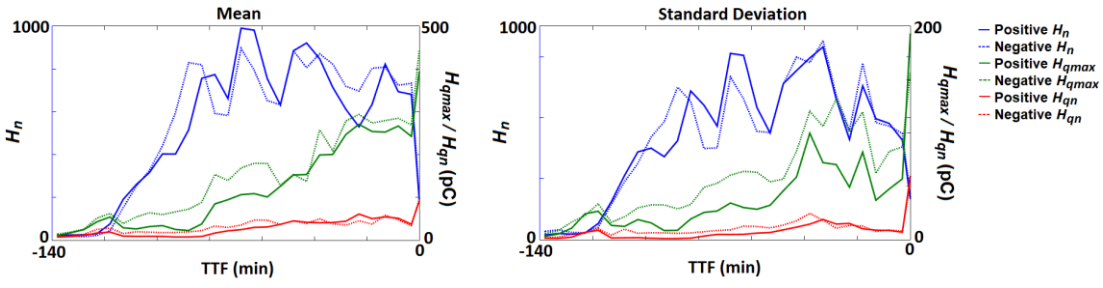
(a) (b) (c)

Figure 4-20. Kurtosis of negative (a) $H_n(\varphi)$, (b) $H_{qmax}(\varphi)$ and (c) $H_{qn}(\varphi)$

The effect of harmonics can be seen in the skewness (Figure 4-17 and Figure 4-18) and kurtosis (Figure 4-19 and Figure 4-20) rather than the mean (Figure 4-13 and Figure 4-14) and standard deviation (Figure 4-15 and Figure 4-16) with Wave 1 showing the most deviating pattern. This is because Wave 1 has the most distorted waveform resulting in a different phase range in Table 4-4 compared to other test waveforms.

Overall, positive and negative PDs show a similar trend for all four features. For each 2D phase distribution, the mean and standard deviation seem to have similar behaviour either positive or negative PD pulses. In $H_n(\varphi)$ distribution, both features rise rapidly during initial growth and start to fluctuate during intermediate growth before a sudden decrease just before breakdown occurs. Thus, the increase in PD occurrence (i.e. $H_n(\varphi)$) highlighted in Section 4.3.1.1 can only be claimed for a part of the tree growth due to the sudden fall near to breakdown. The same behaviour has been reported in [161] for surface tracking where the number of PD is increased throughout the deterioration but decreases rapidly right before breakdown occurs.

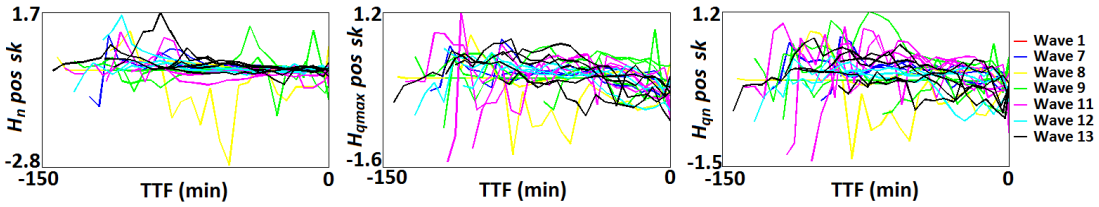
Interestingly, during the sudden decrease of $H_n(\varphi)$, mean and standard deviation of $H_{qmax}(\varphi)$ and $H_{qn}(\varphi)$ experience a sudden increase after a period of gradual increase. To add more, the increase in PD magnitude ($H_{qmax}(\varphi)$ and $H_{qn}(\varphi)$) is proven throughout the growth with a slow rise at the start and change rapidly near the failure. This behaviour can be seen clearer in Figure 4-21 with the plots of mean and standard deviation of the 2D phase distributions of a sample tested with Wave 11.



(a) (b)

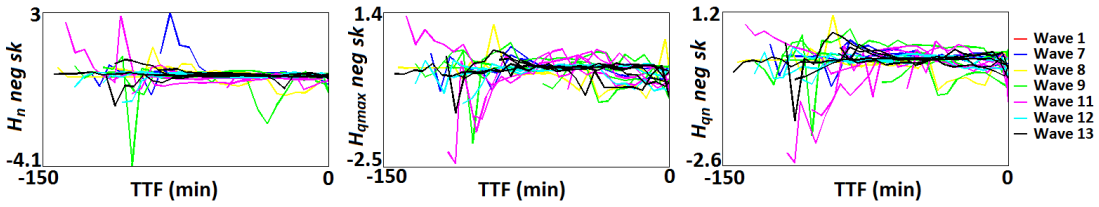
Figure 4-21. Positive and negative $H_n(\varphi)$,$H_{qmax}(\varphi)$ and $H_{qn}(\varphi)$ distributions of a sample tested with Wave 11 (a) mean, and (b) standard deviation

It is difficult to observe the behaviour of skewness and kurtosis in Figure 4-17 and Figure 4-18 because the range of Wave 1 is dominant compared to others. Therefore, the skewness and kurtosis are replotted ignoring the Wave 1 as shown in Figure 4-22 to Figure 4-25. Despite the small ranges of the skewness and kurtosis, we can still observe that the early growth of $H_n(\varphi)$ is more positive compared to late growth. However, some of the samples behave differently in $H_{qmax}(\varphi)$ and $H_{qn}(\varphi)$ distributions. Although phase distribution is found in Section 4.3.1.1 to increase throughout the tree growth, this observation only considered the phase itself without correlating with the PD numbers or magnitude. The measure of skewness and kurtosis perhaps can give the impact of the PDs in the phase distribution. Thus, the skewness and kurtosis of pure and lightly distorted waveforms give only a slight change while the greater harmonic distortion result in a greater changes.



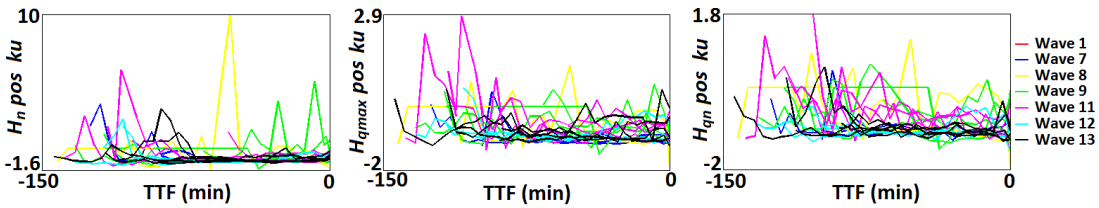
(a) $H_n(\varphi)$ (b) $H_{qmax}(\varphi)$ (c) $H_{qn}(\varphi)$

Figure 4-22. Replot of skewness of positive PDs in Figure 4-17 ignoring Wave 1



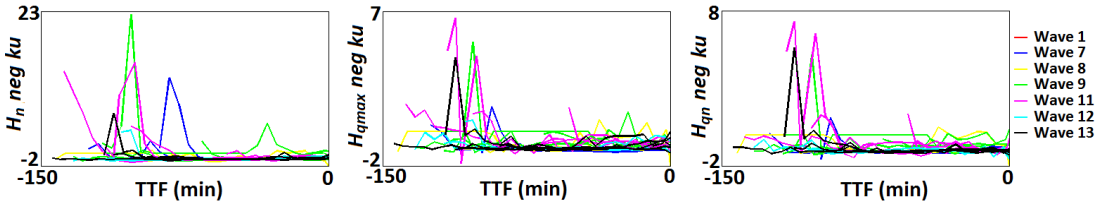
(a) $H_n(\varphi)$ (b) $H_{qmax}(\varphi)$ (c) $H_{qn}(\varphi)$

Figure 4-23. Replot of skewness of negative PDs in Figure 4-18 ignoring Wave 1



(a) $H_n(\varphi)$ (b) $H_{qmax}(\varphi)$ (c) $H_{qn}(\varphi)$

Figure 4-24. Replot of kurtosis of positive PDs in Figure 4-19 ignoring Wave 1



(a) $H_n(\varphi)$ (b) $H_{qmax}(\varphi)$ (c) $H_{qn}(\varphi)$

Figure 4-25. Replot of kurtosis of negative PDs in Figure 4-20 ignoring Wave 1

4.3.2 Pulse Sequence Analysis

It is mentioned in [96], [253], that the voltage differences do not occur at random but in specific sequences due to space charges built up by the discharge process. Thus, the sequence of the voltage difference between two consecutive PD pulses, *du*, is an important parameter in PD mechanism. This parameter is calculated by first examining the instantaneous voltage, $u(t)$, using one of equations 4-1 to 4-7, with consideration of the waveform group i.e. Wave 1, 7, 8, 9, 11, 12 or 13. The *du* is then determined using equation 4-13 where *n* refers to the current PD pulse and *n*-1 to the previous one.

$$du_n = u_n - u_{n-1} \tag{4-13}$$

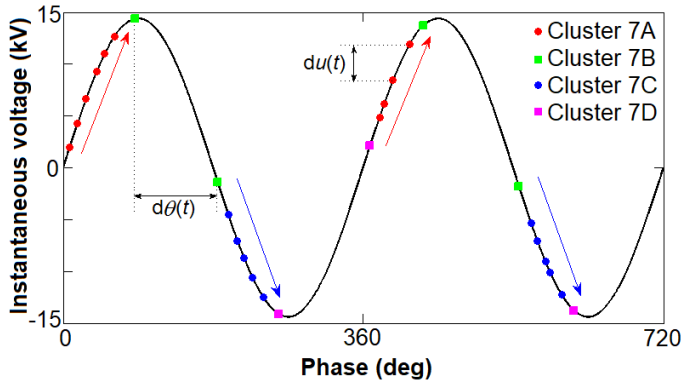
The behaviour of *du* is illustrated in the plots of Figure 4-26 showing the PD pulses overlaid on the three cycles of instantaneous voltage for a sample tested under Wave 7 (fundamental) at early tree growth (Figure 4-26a) and immediately before breakdown (Figure 4-26b). From the figure, we can see the PD occurrence is higher during the late growth compared to the initial stage as discussed previously in Sections 4.3.1.1 and 4.3.1.3 which results in a decrease of the *du*. This also applies to the phase difference between two consecutive PD pulses, $d\theta$, and can be obtained by replacing *u* in equation 4-13 with θ.

$$d\theta_n = \theta_n - \theta_{n-1} \tag{4-14}$$

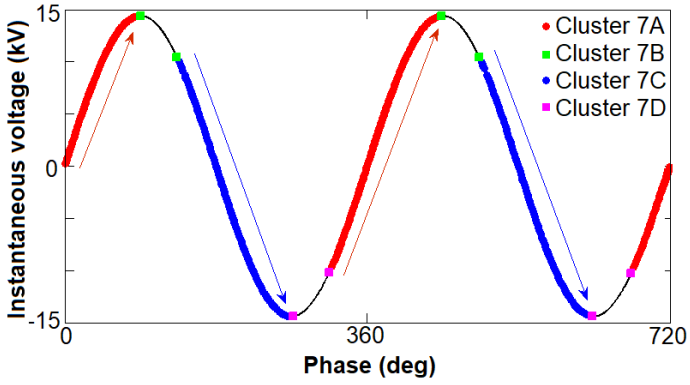
Both *du* and $d\theta$ can be divided into two groups i.e. small and large. The small group comprises PDs from clusters 7A and 7C while the large group comprises PDs from

clusters 7B and 7D. These four clusters are grouped according to the polarity and magnitude of *du* depending on the consecutive voltages:

- All PD pulses in the rising edges form cluster 7A which results in positive and smaller *du*.
- The last pulse around the positive peaks and the first pulse in the falling edges form cluster 7B which results in negative and greater *du*.
- All PD pulses in the falling edges form cluster 7C which results in negative and smaller *du*.
- The last pulse around the negative peaks and the first pulse in the rising edges form cluster 7D which results in positive and greater *du*.



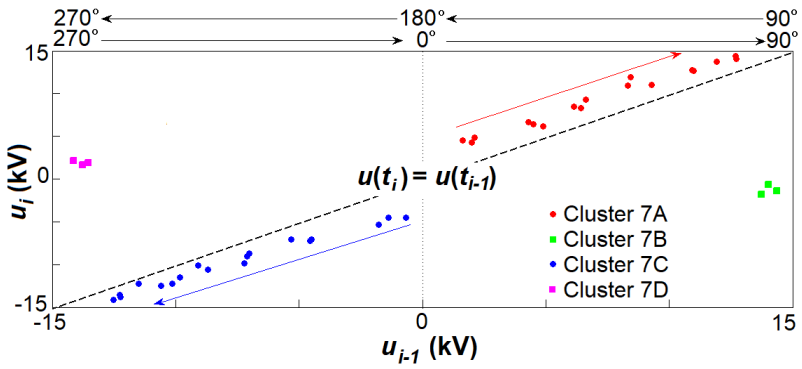
(a)



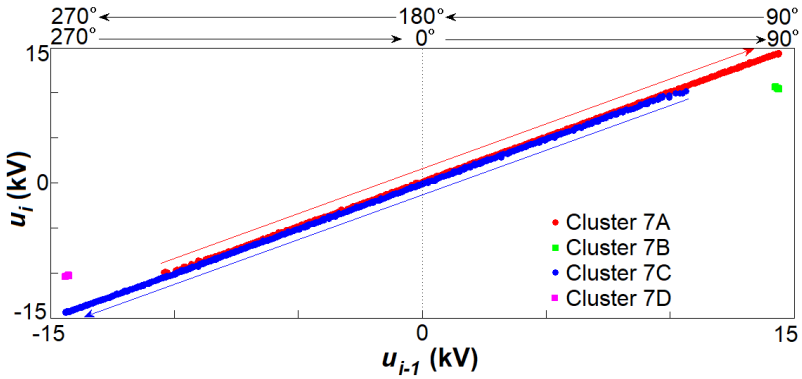
(b)

Figure 4-26. The instantaneous voltage, $u(t)$, of PD occurrence overlaid on Wave 7 during (a) early growth and (b) late growth

The behaviour of *du* can be observed more clearly in the u_n vs u_{n-1} plot, shown in Figure 4-27. Clusters A and C contain the majority of the PD pulses which occur after a relatively small voltage change. The small *du* values results in PDs which appear near the 45° line where $u_n = u_{n-1}$. The decrease of *du* towards breakdown yields a smoother 45° line of PDs in cluster A and C as shown in Figure 4-27b. In contrast, clusters B and D arise from the relatively large *du* around the peak voltage, since PD stops occurring when *du* changes from positive to negative or vice versa. The large values of *du* form clusters farther from the 45° line but moving towards the line as the electrical treeing approaches breakdown. Prior to breakdown, the PD pulses fill almost the entire phase range resulting a very small voltage change i.e. $du_n \approx du_{n-1}$, thus, forming a 45° line. Throughout the tree growth, the four clusters in Figure 4-27a merge to form the dominant diagonal line in Figure 4-27b although sometimes with distinct PDs from cluster B and D.



(a) Early growth



(b) Late growth

Figure 4-27. The consecutive voltage plot, u_n vs u_{n-1}, of a sample tested with Wave 7 during (a) early growth and (b) late growth

The discussed characteristics of *du* also apply to $d\theta$, therefore both features should be considered as prognostic indicators. The ratio of these key changes, $du/d\theta$, represents the slope between the consecutive pulses and can be obtained by dividing equation 4-13 with equation 4-14 i.e.

$$\frac{du_n}{d\theta_n} = \frac{u_n - u_{n-1}}{\theta_n - \theta_{n-1}} \tag{4-15}$$

In this study, three parameters are considered for further investigation i.e. *du*, $d\theta$ and $du/d\theta$ and are described in Section 4.3.2.2.

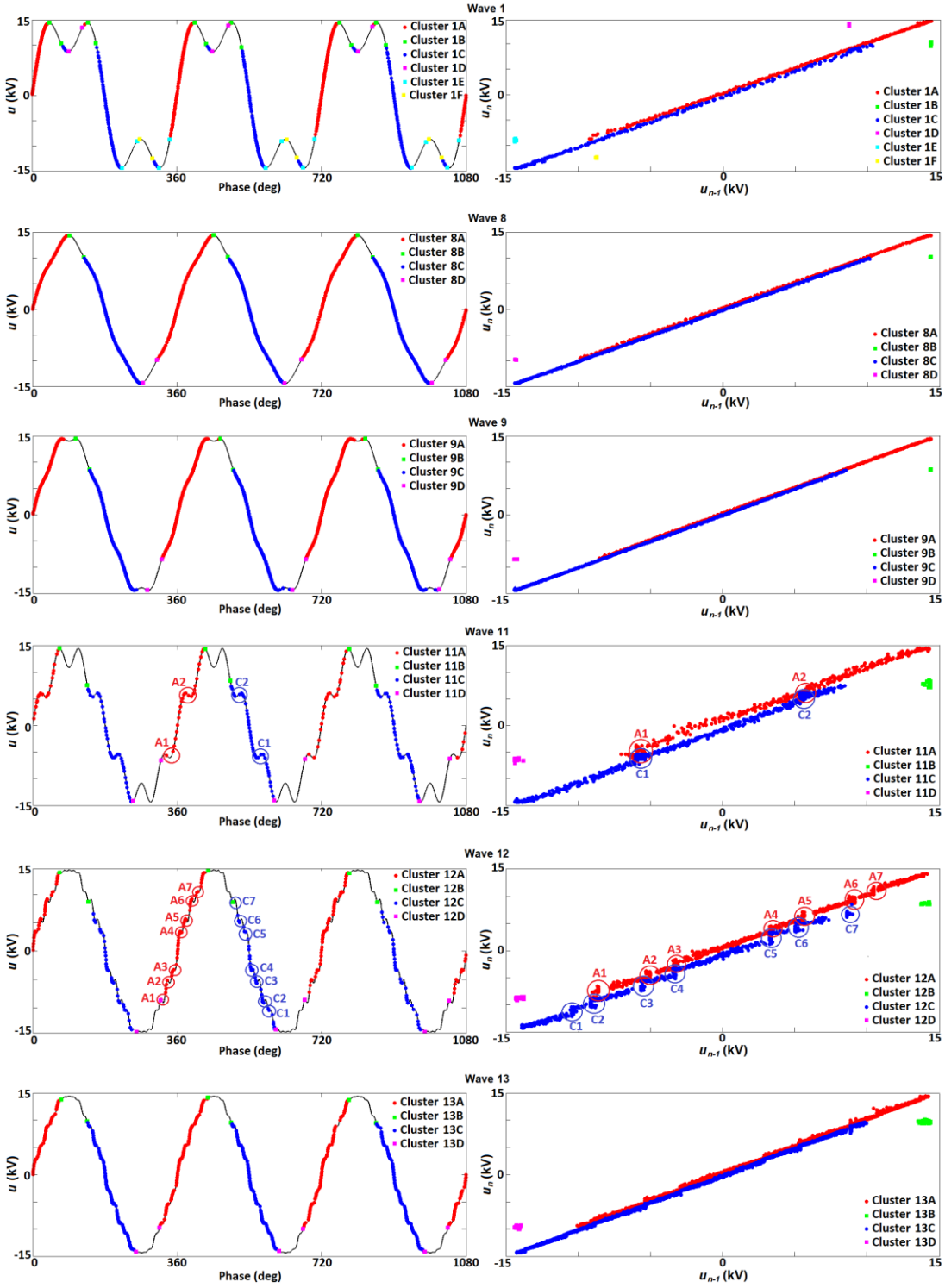
4.3.2.1 Effect of Harmonics

The distorted test waveforms show slightly different behaviour in PD compared to Wave 7. The plots of instantaneous voltage and u_n vs u_{n-1} of test waveforms 1, 8, 9, 11, 12 and 13 are shown in Figure 4-28. Among the six non-sinusoidal test waveforms, samples tested with Wave 1 show the greatest difference, having six clusters compared to the four clusters of all others. This is because Wave 1 has an extra two peaks per voltage cycle where the polarity change in *du* occurs more frequently compared to Wave 7, giving more points of non-activity in the plot (six points compared to two for waveform 7). However, the general trend of *du* decreasing towards breakdown remains. Although Wave 9 and 11 also have 4 peaks, their THD values are smaller than Wave 1, thus giving a smaller range of *du* values between the two peaks in either the positive or negative half cycle. These contribute to clusters 9A and 9C respectively rather than forming new clusters.

Since Wave 8 has the smallest THD and K_s compared to the other harmonic waveforms, the u_n vs u_{n-1} plot looks very similar to the fundamental Wave 7. Wave 9 also shows the same characteristics as Wave 7 and 8, having also only a 5% THD. Wave 11 contains the 7^{th} harmonic like Wave 9, but with greater THD and K_s. Thus, it has extra features in the u_n vs u_{n-1} plot. It can be seen that around ±5 kV of cluster 11A and 11C, the *du* changes polarity from positive to negative (A1 and A2) and vice versa (C1 and C2). Those changes however do not generate a new cluster of points, but instead form an extra feature marked by circles in the u_n vs u_{n-1} plot.

The samples tested with Waves 12 and 13 show similar characteristics to Wave 11 but at different voltages, depending on the voltage at which the polarity of *du* changes. As we can see in plots of Wave 12, the changes occur at seven spots in cluster 12A and seven spots of cluster 12C. The same applies to Wave 13, but is not clearly shown in both plots because the changes are very small due to smaller THD and K_s.

Some of the points discussed here repeat the discussion in 4.3.1.1 showing that phase-resolved and pulse sequence representations are closely related. Therefore, the same conclusion can be made here. First, waveform with high THD value may increases the number of PD cluster. Second, no deterministic relationship can be concluded between K_s and PD cluster. However, compared to PRPDA, the harmonic order does not seem to affect the u_n vs u_{n-1} plot since *du* is not related to the waveshape.



(a) Instantaneous voltage of PD occurrence (b) Voltage of consecutive PD

Figure 4-28. The instantaneous voltage and the u_n vs u_{n-1} plots of non-sinusoidal samples

4.3.2.2 PSA Features

The PSA features are examined for each 5-minute data file as in PRPDA. Three promising features, du, $d\theta$ and $du/d\theta$, are calculated using equations 4-13 to 4-15 for every consecutive PD in the 5-minute data. Finally, the mean of each feature is calculated using equation 2-8 to represent each batch of 5-minute data until the failure.

The du and $d\theta$ features were calculated from the small group only, from the large group only, and from all PD pulses ignoring which group of waveform they belong to. The small group represents PD pulses in clusters A and C of Figure 4-28 while the large group represents clusters B and D (Wave 1 includes clusters E and F). However, the $du/d\theta$ behaves differently and thus was calculated from all data only. In total, 7 potential features have been extracted from PSA as tabulated in Table 4-5. The behaviour of these 7 PSA features can be observed in Figure 4-29 to Figure 4-31. The colours of the plots indicate the wave group as in the PRPDA approach in Section 4.3.1.3.

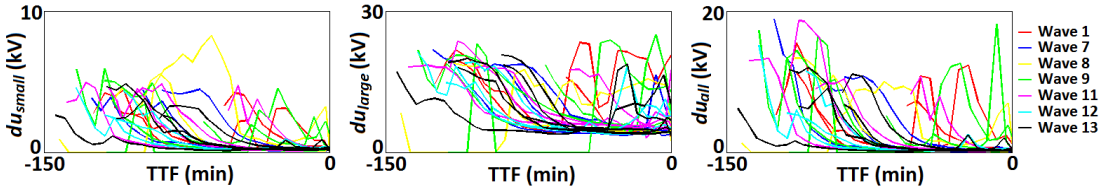
Table 4-5. The features and subfeatures from PSA

Features	***Subfeatures***
du	du_{small}
	du_{large}
	du_{all}
dθ	$d\theta_{small}$
	$d\theta_{large}$
	$d\theta_{all}$
du/dθ	-

Interestingly, du and $d\theta$ show a very similar pattern of decreasing exponentially throughout the tree growth. Voltage derivative however shows a different pattern with an increase during the early growth but then decreases slowly towards the breakdown. Also, it appears visually that PSA features show a smoother behaviour compared to PRPDA in Figure 4-13 to Figure 4-20. No distinct difference is

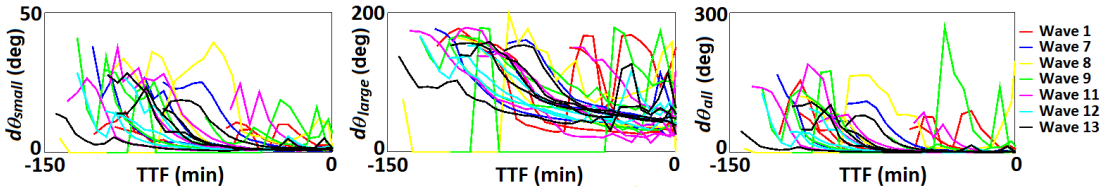
observed between the harmonic groups in du and $d\theta$ features. However, as for $du/d\theta$ in Figure 4-31, it can be seen that samples with the same harmonic group trend together which might be useful for classification purposes.

Two samples from Wave 8 and 9 were found to behave differently where no PD activity was recorded for more than 30 minutes as indicated with the arrows in Figure 4-31. Although this is an expected behaviour of PD, however, this may affect the modelling process as these samples were not really in the same trend as the other samples. These two samples, 08T365 and 09T325, obviously are the outliers in the prognostic model reported in this work. The elimination of the outliers will be discussed in the next chapter.



(a) du_{small} (b) du_{large} (c) du_{all}

Figure 4-29. The subfeatures of du (a) du_{small} (b) du_{large} and (c) du_{all}



(a) $d\theta_{small}$ (b) $d\theta_{large}$ (c) $d\theta_{all}$

Figure 4-30. The subfeatures of $d\theta$ (a) $d\theta_{small}$ (b) $d\theta_{large}$ and (c) $d\theta_{all}$

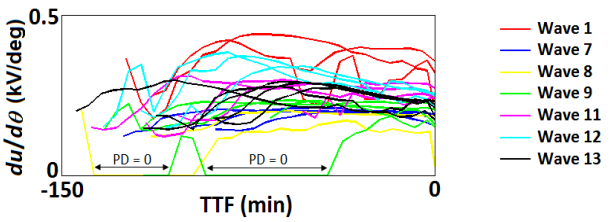


Figure 4-31. The voltage derivative, $du/d\theta$

4.4 Feature Reduction

A total of 24 samples have been analysed using the PRPDA and PSA approaches in order to extract potential features as described in Section 4.3. As a result, 24 features are extracted from PRPDA and 7 from PSA. In order to keep the originality of the features, feature selection approach is employed rather than feature transformation. Filter technique, the simplest approach for feature selection, is applied to reduce the size of the features. The selection is based on the three prognostic performance metrics, described in Section 3.3.

The total score of three prognostic performance metrics introduced in [31], i.e. *monotonicity*, *prognosability*, and *trendability* as indicated in equations 3-6 to 3-8 respectively, is considered for feature reduction. The total metric is given as:

$$Total = Monotonicity + Prognosability + Trendability \quad (4\text{-}16)$$

For *trendability*, the quantile-quantile (Q-Q) plot is performed for every sample combination and Pearson's correlation coefficient is determined using equation 3-9. The final value is given by the smallest absolute correlation as indicated in equation 3-8.

The score of each metric and the total score of all metrics for all 31 features are tabulated in Table 4-6. Each feature is ranked based on the sum of the prognostic metrics. The PRPDA features are numbered from 1 to 24 while PSA features from 25 to 31. Each of these metrics ranges from zero to one, one indicating a very high score on that metric and zero indicating that the parameter is not suitable according to that metric. The prognostic metrics and the total metric in Table 4-6 are plotted in Figure 4-32 to Figure 4-35 in order to compare the metric scores between PRPDA and PSA. PRPDA features from the same statistical quantity are grouped together, i.e. mean, μ, standard deviation, σ, skewness, *Sk*, and kurtosis, *Ku* where the features are labelled according to the feature's number in Table 4-6.

Table 4-6. Prognostic metrics of PRPDA and PSA features

	No	*Features*	*Monotonicity*	*Prognosability*	*Trendability*	*Total*	*Rank*
PRPDA	***1***	H_n *pos* μ	0.27	0.42	0.0100	0.70	22
	2	H_n *neg* μ	0.31	0.50	0.0016	0.81	14
	3	H_{qmax} *pos* μ	0.29	0.59	0.0053	0.89	9
	4	H_{qmax} *neg* μ	0.28	0.58	0.0036	0.86	12
	5	H_{qn} *pos* μ	0.19	0.33	0.0016	0.52	28
	6	H_{qn} *neg* μ	0.21	0.34	0.0021	0.55	27
	7	H_n *pos* σ	0.24	0.36	0.0055	0.61	25
	8	H_n *neg* σ	0.30	0.40	0.0010	0.70	22
	9	H_{qmax} *pos* σ	0.16	0.51	0.0014	0.67	24
	10	H_{qmax} *neg* σ	0.19	0.51	0.0090	0.71	21
	11	H_{qn} *pos* σ	0.16	0.36	0.0008	0.52	28
	12	H_{qn} *neg* σ	0.11	0.40	0.0013	0.51	30
	13	H_n *pos sk*	0.12	0.45	0.0015	0.57	26
	14	H_n *neg sk*	0.09	0.79	0.0064	0.89	9
	15	H_{qmax} *pos sk*	0.13	0.68	0.0026	0.81	14
	16	H_{qmax} *neg sk*	0.09	0.70	0.0002	0.79	16
	17	H_{qn} *pos sk*	0.10	0.67	0.0097	0.78	17
	18	H_{qn} *neg sk*	0.12	0.62	0.0056	0.75	20
	19	H_n *pos ku*	0.13	0.85	0.0021	0.98	8
	20	H_n *neg ku*	0.10	0.89	0.0004	0.99	7
	21	H_{qmax} *pos ku*	0.11	0.77	0.0049	0.88	11
	22	H_{qmax} *neg ku*	0.09	0.74	0.0032	0.83	13
	23	H_{qn} *pos ku*	0.11	0.65	0.0028	0.76	19
	24	H_{qn} *neg ku*	0.08	0.69	0.0021	0.77	18
PSA	**25**	du_{small}	0.39	0.93	0.0200	1.32	4
	26	du_{large}	0.52	0.68	0.0052	1.21	5
	27	$\boldsymbol{du_{all}}$	0.50	0.95	0.0028	**1.45**	**1**
	28	$d\theta_{small}$	0.50	0.93	0.0015	1.43	2
	29	$d\theta_{large}$	0.63	0.74	0.0034	1.37	3
	30	$d\theta_{all}$	0.56	0.63	0.0046	1.19	6
	31	$du/d\theta$	0.03	0.37	0.0085	0.41	31

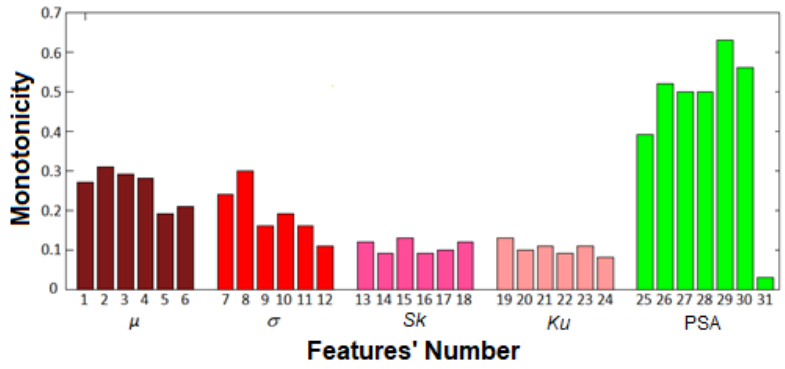


Figure 4-32. Monotonicity scores

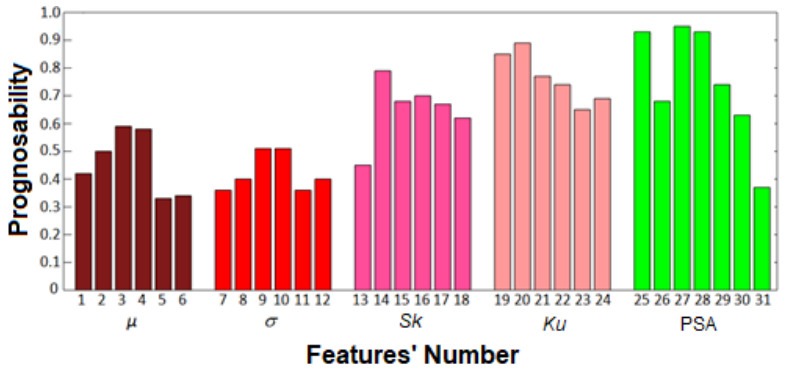


Figure 4-33. Prognosability scores

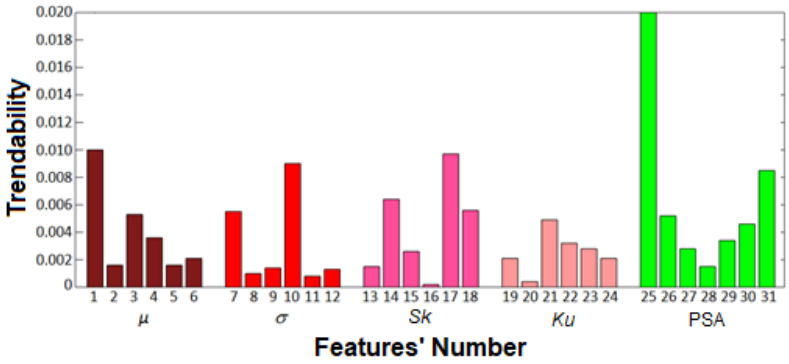


Figure 4-34. Trendability scores

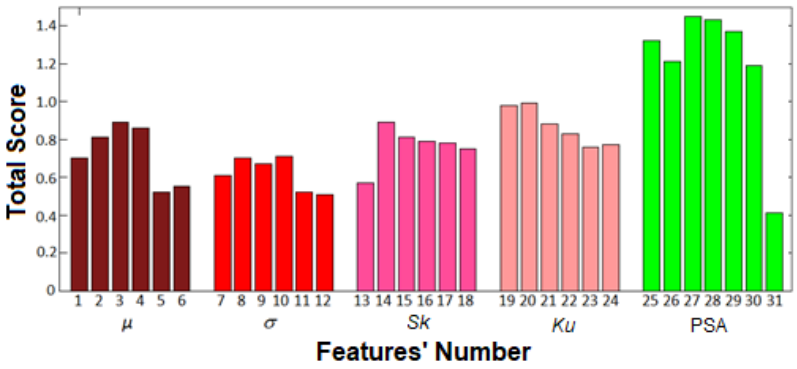


Figure 4-35. Total scores of monotonicity, prognosability and trendability

From the scores in Table 4-6 and bar graphs in Figure 4-32 to Figure 4-35, we can conclude that PSA features have better score than PRPDA for all three prognostic metrics as well as the total metrics. The great difference between these two approaches can be seen in monotonicity scores in Figure 4-32 indicating more consistent behavior among PSA features. According to [31], monotonicity is an important feature of a prognostic parameter because it is generally assumed that systems do not undergo self-healing, which would be indicated by a non-monotonic parameter. However, the monotonic trend of PSA features is mostly observed in the first half of the tree growth and becomes constant towards breakdown, and results in scores that are less than 0.7. This is due to the sequential data during the constant stage yields a $d/dx = 0$, thus not counted for monotonicity score in equation 3-6.

Taking du_{all} feature (first rank) of sample 08T374 as an example, the calculation of individual monotonicity score is shown graphically in Figure 4-36. The (+) indicates positive d/dx where the second is greater than the first while (-) indicates negative d/dx and (=) indicates zero d/dx. By referring to equation 3-6, the total of 24 data points yields 13 of negative d/dx and 2 of positive d/dx giving a monotonicity of 0.48. For the case of PRPDA features, μ and σ results in higher score compared to *Sk* and *Ku* due to their monotonic trend in the early stage.

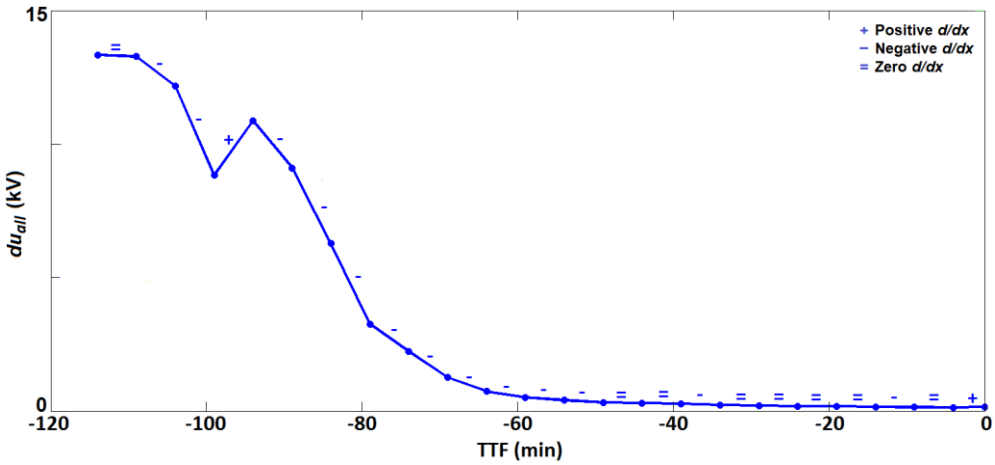


Figure 4-36. The positive and negative d/dx for determining the monotonicity score of du_{all} of sample 08T374.

Most of the 24 samples behave similarly with PSA features being more monotonic compared to PRPDA features. However, two samples, 08T365 and 09T325, show

very different behaviour compared to other samples and result in a much lower score. This is due to the lack of PD activity at certain time intervals which results in a greater number of $d/dx = 0$ and almost equal number of positive and negative d/dx as can be seen in the du_{all} feature of 09T325 in Figure 4-37. Since the number of positive and negative d/dx is the same, hence, the monotonicity score is 0.

Among the seven PSA features, $du/d\theta$ behaves differently where the monotonicity score is even smaller than most of the PRPDA features. As we can see in Figure 4-31, there is no certain direction of $du/d\theta$ with an increase at the early growth but a decrease towards breakdown. After all, the average of the total monotonicity scores of each feature are calculated and tabulated in Table 4-6 and compared in Figure 4-32.

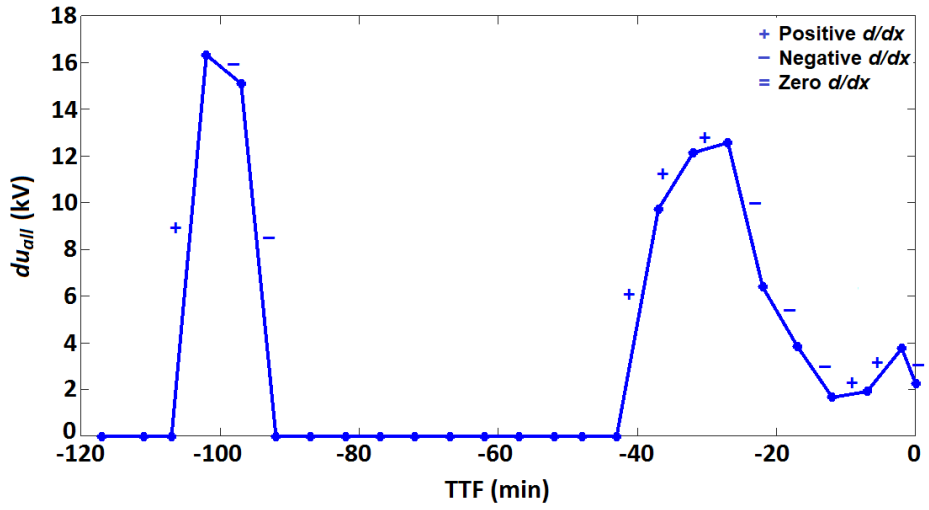
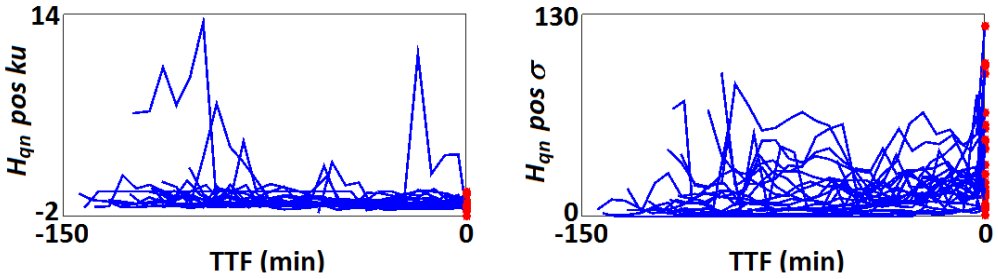


Figure 4-37. The positive and negative *d*/*dx* for determining the monotonicity score of du_{all} of 09T325

Prognosability in equation 3-7 encourages well-clustered failure values, i.e. small standard deviation of failure values, and large parameter ranges. This means, prognosability only relies on the first (start of treeing) and the last (breakdown) feature data ignoring what happens in between. This is the reason that skewness and kurtosis in the PRPDA features have a greater score than mean and standard deviation as shown in Figure 4-33. Even though skewness and kurtosis of samples from Wave 1 behave differently, the failure values are clustered well with other samples as indicated with the red * in Figure 4-38a (taking H_{qn} *pos ku* as an example)

disregarding which harmonic group they belong to. The H_{qn} *pos* σ (comparing standard deviation with kurtosis) in Figure 4-38b on the other hand, has greater parameter range but greater standard deviation of failure values, thus results in lower prognosability score.

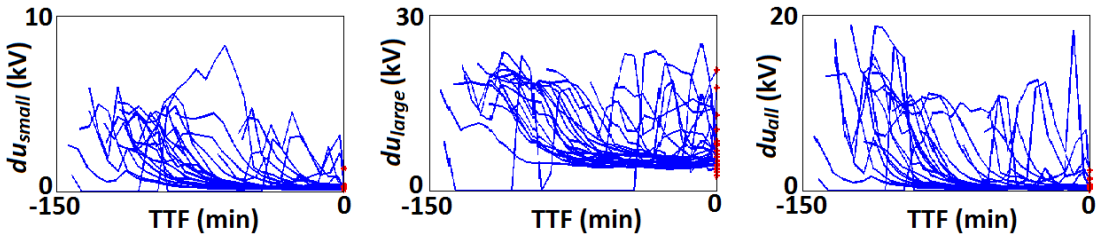


(a) σ = 0.4, range = 1.0
prognosability = 0.65

(b) σ = 33.7, range = 32.9
prognosability = 0.36

Figure 4-38. The distribution of failure values (indicated by *) of (a) H_{qn} *pos ku*, and (b) H_{qn} *pos* σ

Most of PSA features exhibit well-clustered failure values based on plots in Figure 4-29 to Figure 4-31 with du_{small}, du_{all} and $d\theta_{small}$ in Figure 4-33 having relatively higher scores than the rest. The distribution of the failure values among *du* subfeatures can be compared in Figure 4-39. Although the parameter ranges of *du* subfeatures are considered small, with a very small standard deviation, du_{small} and du_{all} yield high prognosability scores. The du_{large} however, scored less than du_{all} and du_{small} due to higher standard deviation, σ.



(a) σ = 0.2, range = 3.2
prognosability = 0.93

(b) σ = 4.3, range = 11.4
prognosability = 0.68

(c) σ = 0.5, range = 9.1
prognosability = 0.95

Figure 4-39. The distribution of failure values (indicated by *) of *du* subfeatures in Figure 4-29

Trendability is defined in [31] as a measurement of how well each parameter in a population is described by the same underlying function. This measurement is performed in this thesis using Q-Q plot. The Q-Q plot is a plot of the sorted quantiles of one data set against the sorted quantiles of another data set [254]. It is used to visually inspect the similarity between the underlying distributions of two data sets. If the two distributions are similar, then the points would lie close to the linear identity line, $y = mx + c$. Since the tree samples give data of different sizes, thus, the quantiles are selected to correspond to the sorted values from the smaller data set. The quantiles for the larger data set are then interpolated as well as the features' values. The quantiles, q of each feature are determined using equation 4-17 where i is the number of the quantile, N is the size of the feature, t is the recorded time in minutes, t_0 is the start time of treeing thus equals to 0, and t_{bd} is the failure time [254].

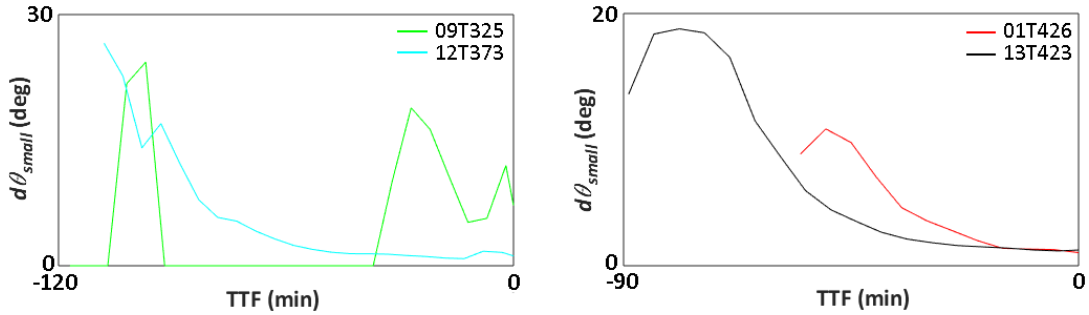
$$q_i = \frac{t_i + t_{i-1}}{2t_{bd}}; i = 1, \dots, N \tag{4-17}$$

For example, using equation 4-17, the quantiles of a feature recorded every 5 minutes and for which failure occurs at 25 minutes (t = 5, 10, 15, 20 and 25 minutes) are obtained as q = 0.1, 0.3, 0.5, 0.7 and 0.9.

From visual observation, the behaviour of PSA features looks more consistent than PRPDA. Therefore, it is expected that trendability among the samples in PSA features is better than PRPDA. However, outliers may affect the trendability scores since this measure only considers the minimum correlation coefficient as indicated in equation 3-8. The comparison between a good and a poor trendability can be observed in Figure 4-40 to Figure 4-42.

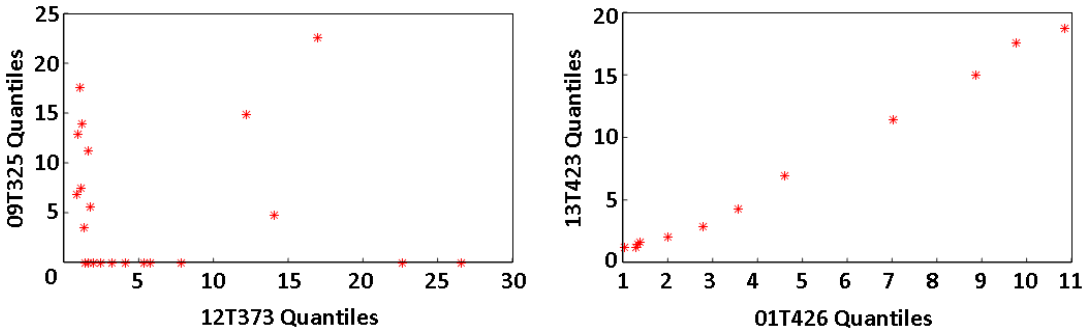
In Figure 4-40a, we can see that the $d\theta_{small}$ decreases exponentially for sample 12T373 but behaves randomly for sample 09T325, thus results in very low Pearson's correlation coefficient, cc_p (refer equation 3-9). In contrast, both samples in Figure 4-40b show declining behaviour hence yield a very high cc_p. It can be seen in Figure 4-41 that only the Q-Q plot of high cc_p shows a linear identity.

The trendability is given by the smallest absolute correlation coefficient. The Pearson's correlation coefficients between 09T325 and other samples are shown in Figure 4-42a and can be compared with the scores between 01T426 and other samples in Figure 4-42b. As can be seen, the cc_p of 09T325 with other samples is lower than 0.5. Sample 01T426 on the other hand, has only three correlations that are lower than 0.5 i.e. with 01T354, 08T365 and 09T325 which are the potential outliers. The cc_p with other samples are mostly greater than 0.8 showing a great strength of relationship between 01T426 and other samples. As a final point, the harmonics order does not seem to influence the correlation between samples.



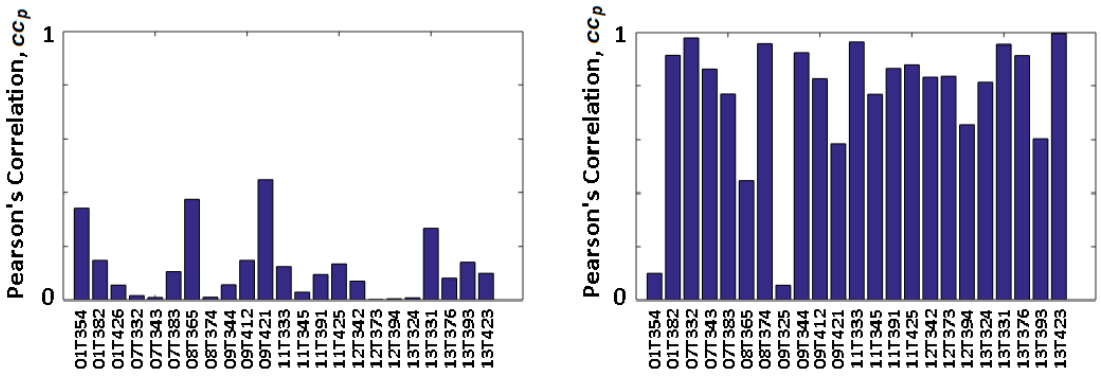
(a) cc_p = 0.0015 (b) cc_p = 0.9964

Figure 4-40. The relationship between two samples for $d\theta_{small}$ feature, (a) the lowest correlation (b) the highest correlation



(a) cc_p = 0.0015 (b) cc_p = 0.9964

Figure 4-41. The QQ-plots of samples in Figure 4-40 (a) 09T325 and 12T373 (b) 13T423 and 01T426

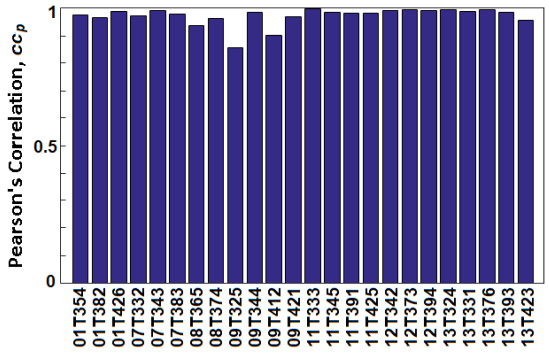


(a) 09T325 (b) 01T426

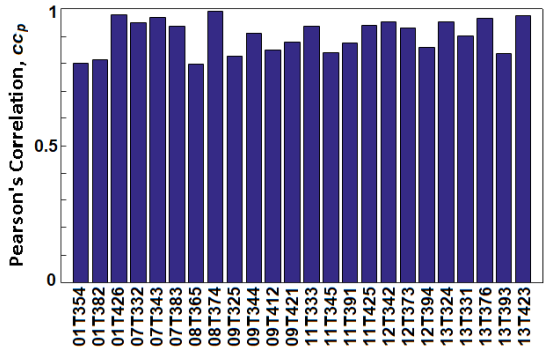
Figure 4-42. The Pearson's correlation between (a) 09T325 and other 23 samples, and (b) 01T426 and other 23 samples

4.5 Feature Selection

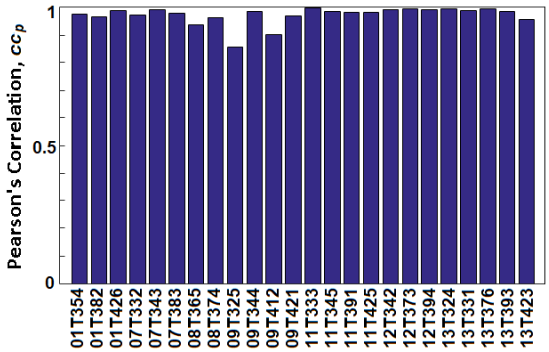
Referring to Table 4-6, the first two ranks features have very close scores i.e. 1.45 and 1.43. Both features are from PSA i.e. du_{all} and $d\theta_{small}$. The individual metric scores of both features are also very close. As can be seen in Figure 4-29c and Figure 4-30a, both features have very similar in characteristics. Taking both redundant features as the key parameter in the model development may cause overfitting as mentioned earlier in Section 3.4 . Features that behave similarly will have high correlation measure. Therefore, in this section, the Pearson's correlation between the top three features (du_{all}, $d\theta_{small}$ and $d\theta_{large}$) are calculated using equation 3-10 or 3-11 and are plotted in Figure 4-43. As can be seen in Figure 4-43a, the correlation between du_{all} and $d\theta_{small}$ is very high, nearly 1, except for 09T325. For the third rank feature, $d\theta_{large}$, some of the samples yield lower correlation, around 0.8 (see Figure 4-43b and 4-43c), but still a good number for the correlation measure. Considering the following rank-3 to rank-6 features may also introduce redundancy as the features are from the same subset, i.e. du and $d\theta$. Rank-7 feature on the other hand, is from PRPD, but the total of three metrics score is much lower compared to PSA features. Therefore, only the first rank feature, du_{all}, is selected for the development of the prognostic model in Chapter 5. Modelling with a single parameter leads to a simpler model and is also easier to visualise. This is very useful especially when the approach proposed in this thesis is new to cable monitoring.



(a) Pearson's correlation between du_{all} and $d\theta_{small}$



(b) Pearson's correlation between du_{all} and $d\theta_{large}$



(c) Pearson's correlation between $d\theta_{small}$ and $d\theta_{large}$

Figure 4-43. The Pearson's correlation between du_{all}, $d\theta_{small}$ and $d\theta_{large}$ for all 24 samples

4.6 Effect of Harmonics on du_{all}

As mentioned previously in Section 4.3.2.2, there is no deterministic relationship between the harmonic order and the behaviour of du_{all}. Here, the plots in Figure 4-29c are replotted which correspond to the growth time instead of the TTF. Also, the samples are grouped according to the harmonic order, THD and waveshape factor, K_s as shown in Figure 4-44. Again, no deterministic relationship can be determined between either harmonic order, THD or K_s and du_{all} since the growth rates (du_{all}/du_{all}) of the samples in the same group are varies. This finding supports the judgement of the previous research [18]. Therefore, all the samples are treated equally in the prognostic modelling without discriminating the harmonics content.

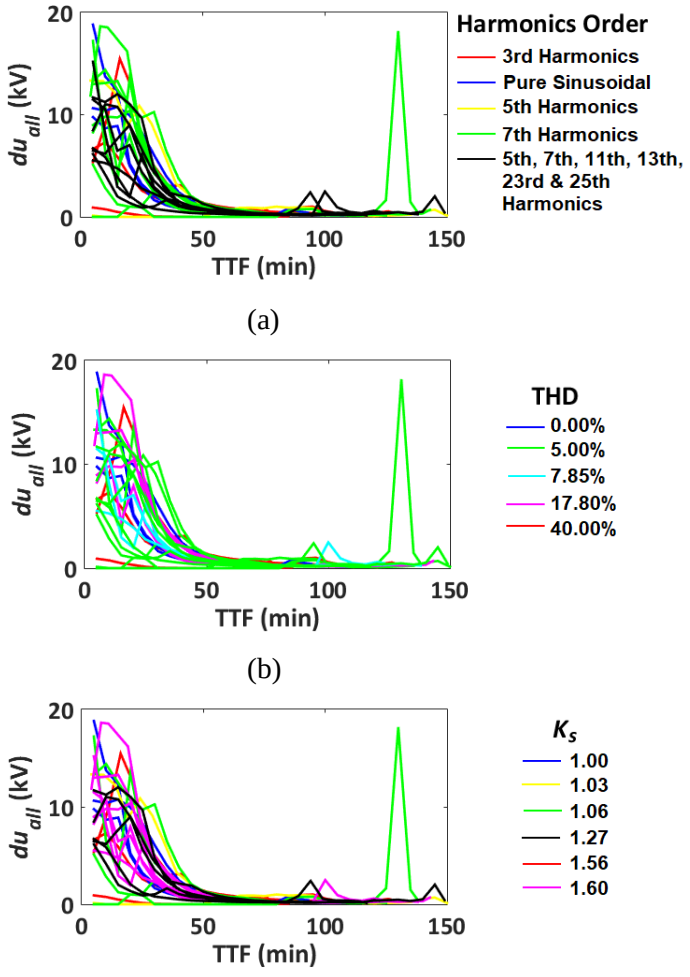


Figure 4-44. The du_{all} feature with samples grouped in (a) harmonics order, (b) THD, and (c) K_s

4.7 Summary and Conclusion

Identification of potential prognostic parameters is one of the crucial components in prognostic modelling. The parameters are expected to show a monotonic behaviour that represents the degradation process of the system. In this chapter, a total of 31 features have been extracted with 24 of them from PRPDA and 7 from PSA. These features are mostly used for diagnostics which the concept was extended here for prognostic modelling.

Observation on PRPD patterns from the start tree growth until breakdown occurred, revealed three important changes i.e. the increase in number of PD, PD magnitude and phase occurrence. These changes were represented by four statistical measures (mean, standard deviation, skewness and kurtosis) on three 2D phase distributions ($H_n(\varphi)$, $H_{qmax}(\varphi)$ and $H_{qn}(\varphi)$). The statistical measures were determined separately for positive and negative half cycles, giving a total of 24 PRPD features. It was shown visually that the positive and negative PDs of all features behaved similarly towards the breakdown.

The PRPD pattern was greatly influenced by the harmonic order. Waveforms with higher harmonic order will have more rising and falling edges thus result in more distorted PRPD patterns. High THD was found to increase the number of PD clusters. However, no deterministic relationship can be concluded for waveshape factor, K_s. It was mentioned in [18] that the peak voltage employed in this thesis might be too high for K_s to have a dominating influenced. Among the PRPD features, only skewness and kurtosis were influenced by harmonics. The increased in the number of PD clusters due to high THD resulted in highly distorted PRPD pattern thus increased the skewness and kurtosis.

From PSA, it was found that voltage change, *du* decreased exponentially throughout the tree growth. The *du* can be divided into two groups i.e. du_{small} and du_{large}, while the total *du* is denoted as du_{all}. A change in *du* resulted in a similar change in $d\theta$. Hence, the same features can be extracted from the phase occurrence, i.e. $d\theta_{small}$,

$d\theta_{large}$ and $d\theta_{all}$. Lastly, the seventh PSA feature was derived from the ratio of the key changes, $du/d\theta$.

Similar to PRPDA, there was no deterministic relationship between K_s and PSA pattern. High THD (40%) also influenced the PSA pattern (u_n vs u_{n-1} plot) by adding two more clusters to the existing PD clusters. Nonetheless, the harmonic order did not change the number of clusters except for extra features in the existing PD clusters.

Throughout the tree growth, PSA features showed a smoother behaviour than PRPDA with clearer monotonic trend. No distinct difference was observed between the harmonic groups in all PD features except for $du/d\theta$.

The performance of 31 PD features were evaluated using three performance metrics i.e. monotonicity, prognosability and trendability. The total score of these metrics was used to rank each feature i.e. the top rank has the highest score. For PRPDA, mean and standard deviation features yielded better score in monotonicity compared to skewness and kurtosis, and vice versa for prognosability metric. PSA features scored remarkably for prognosability and were comparable to skewness and kurtosis from PRPDA. For monotonicity, PSA features only showed a monotonic trend during the first half of total growth thus yielded a moderate score, but still higher than PRPDA.

Although PSA features dominate the top ranks in Table 4-6, only one feature is selected for the model development. The reason for this, is the high correlation between the PSA features. Features with similar behaviour may become redundant and caused overfitting to occur. Therefore, only the first rank feature, du_{all}, is selected for the development of prognostic model in Chapter 5.

The three prognostic performance metrics not only assist in features selection but also in identifying the outliers. So far, two samples, 08T365 and 09T325, were spotted to show different behaviour with very low monotonicity score. These two samples also yield low scores in trendability analysis as shown in Figure 4-42.

Another sample, 01T354, gains much lower score than 09T325 in Figure 4-42 hence is also suspected to be an outlier. The identification process is based on the individual scores of monotonicity and trendability which is discussed in detail in the next chapter.

Finally, no unique characteristic is observed to differentiate the effect of harmonics on the electrical tree growth. Grouping the samples according to the THD and K_s also did not reveal any relationship. Thus, all harmonic groups are treated equally in the next chapter with no discrimination.

Chapter 5
Building a Prognostic Model Using the Selected PD Feature

5.1 Framework of the Developed Model

Little work has been reported regarding insulation lifetime prediction using PD data, even though PD monitoring is essential for condition monitoring in power equipment. Published life models are mostly based on the Weibull analysis of failure data which has no information on the degradation process. Alternatively, this research work proposes a framework for developing a prognostic model using PD data from electrical treeing experiment. The PD data is analysed and evaluated, in order to find the most suitable prognostic parameter. As explained in Section 4.7 , a prognostic model is developed based on a degradation parameter, du_{all} applying the simplest approach, curve fitting. The framework of the model development and evaluation is given in Figure 5-1.

First of all, the behaviour of the first ranked feature, du_{all} is visualised to identify the outliers and to aid the decision of selecting the model's algorithm. Next, the outliers are identified from the individual sample score of the *monotonicity* and *trendability* metrics. It is important to know the reason for different behaviour before deciding whether to keep or remove the outliers. The simplest option is considered for fitting the data i.e. curve fitting technique. At this stage, additional algorithm can be

introduced in order to improve the model's predictions. Then, the properties of the model are determined. This includes the threshold value where the TTF is predicted. Validation of the model is done by applying the model on the testing samples. Both holdout and cross-validation techniques are employed and compared. The performance of the model is finally revealed using the two metrics proposed in [247] i.e. *prognostic horizon*, PH and *cumulative relative accuracy*, CRA, and one metric proposed in this thesis, *convergence horizon*, CH.

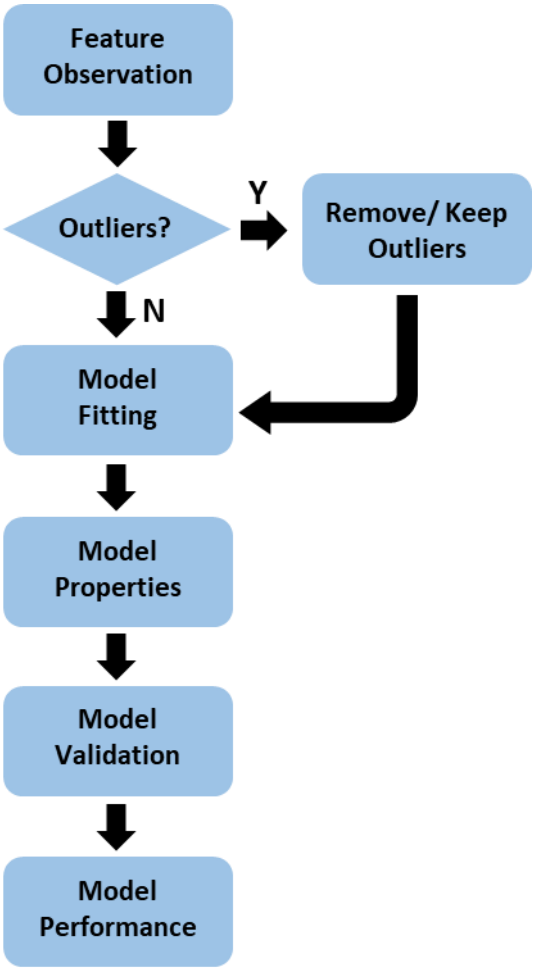


Figure 5-1. The framework of the developed prognostic model

5.2 Outliers

This research work employs a single feature for building a prognostic model for electrical treeing breakdown. The feature with the highest total score of the three prognostic metrics (sum of *monotonicity*, *prognosability* and *trendability*), du_{all}, is chosen. A total of 24 samples are used for training the model. However, some of the samples behave very differently from the others, thus requiring some filtering. From visual observation, most of the samples decrease exponentially and tend towards a constant value towards breakdown. Nevertheless, those samples with unusual behaviour have much less monotonicity, thus do not really trend together. Therefore, the monotonicity and trendability metrics are recalculated individually for each sample in order to verify the outlier samples. The prognosability metric is excluded since it is measured based on population.

The average of the monotonicity score for 24 samples is 0.5 (see Table 4.6) which can be considered as a moderate score due to the consistent du_{all} values during the second half of samples' lifetime. In this section, all 24 samples are tabulated in Table 5-1 with the individual monotonicity and trendability scores. For the case of trendability, the correlation between one sample to each of the other 23 samples is taken and averaged. From the monotonicity scores, two samples have been identified to have the lowest scores i.e. 08T365 and 09T325. Both samples also yield low scores for trendability as well as sample 01T354. Although the monotonicity of 01T354 is closer to the average, the trendability score is much lower compared to other samples. These low scores are highlighted in red and the negative sign for monotonicity indicates a more dominant decreasing pattern.

Table 5-1. Monotonicity and correlation of training samples for du_{all} feature

Wave	*Sample*	*Monotonicity*		*Pearson's Correlation of Trendability*			
		24 samples	21 samples	24 samples		21 samples	
				Average	Lowest	Average	Lowest
1	01T354	-0.39	-	0.22	0.016	-	-
	01T382	-0.36	-0.36	0.74	0.128	0.82	0.51
	01T426	-0.73	-0.73	0.79	0.049	0.87	0.64
7	07T332	-0.78	-0.78	0.78	0.042	0.86	0.59
	07T343	-0.55	-0.55	0.83	0.053	0.91	0.79
	07T383	-0.56	-0.56	0.76	0.123	0.83	0.66
8	08T365	-0.07	-	0.45	0.111	-	-
	08T374	-0.48	-0.48	0.81	0.032	0.89	0.66
9	09T325	-0.00	-	0.12	0.003	-	-
	09T344	-0.42	-0.42	0.81	0.035	0.89	0.64
	09T412	-0.41	-0.41	0.60	0.097	0.66	0.49
	09T421	-0.50	-0.50	0.67	0.168	0.73	0.49
11	11T333	-0.48	-0.48	0.81	0.144	0.89	0.67
	11T345	-0.21	-0.21	0.79	0.005	0.87	0.60
	11T391	0.46	0.46	0.80	0.122	0.88	0.62
	11T425	-0.55	-0.55	0.72	0.016	0.80	0.55
12	12T342	-0.54	-0.54	0.81	0.058	0.88	0.67
	12T373	-0.55	-0.55	0.82	0.034	0.91	0.77
	12T394	-0.37	-0.37	0.69	0.013	0.76	0.51
13	13T324	-0.50	-0.50	0.81	0.003	0.89	0.67
	13T331	-0.48	-0.48	0.79	0.196	0.85	0.59
	13T376	-0.50	-0.50	0.82	0.157	0.90	0.69
	13T393	-0.25	-0.25	0.68	0.160	0.74	0.52
	13T423	-0.61	-0.61	0.80	0.120	0.88	0.59
	Average	-0.41	-0.50	0.71	0.003	0.84	0.62

To justify those three samples as being the outliers, the average and the lowest correlations are recalculated using only 21 samples excluding those three. From the results, we can see a great difference in the lowest correlation between these two populations (Figure 5-2). Figure 5-3 shows the behaviour of all 24 samples with red solid lines representing the three outliers while the blue solid lines represent the other 21 samples. The difference in the behaviour of the three outliers is due to the lack of PD activities in some of the time interval and thus causes the du_{all} values to become greater or zero (when no PD is recorded). Different behaviour will result in a different model when the curve fitting technique is employed. Since the key parameters of the developed model are determined as an average from all samples (explained later in Section 5.5), considering the outliers will definitely alter the average parameters and is not representative of most of the samples. Therefore, a simpler approach is considered by removing the outliers in the training algorithm. This limitation can be improved in the future when more data is available to allow treatment on the outliers before the training process.

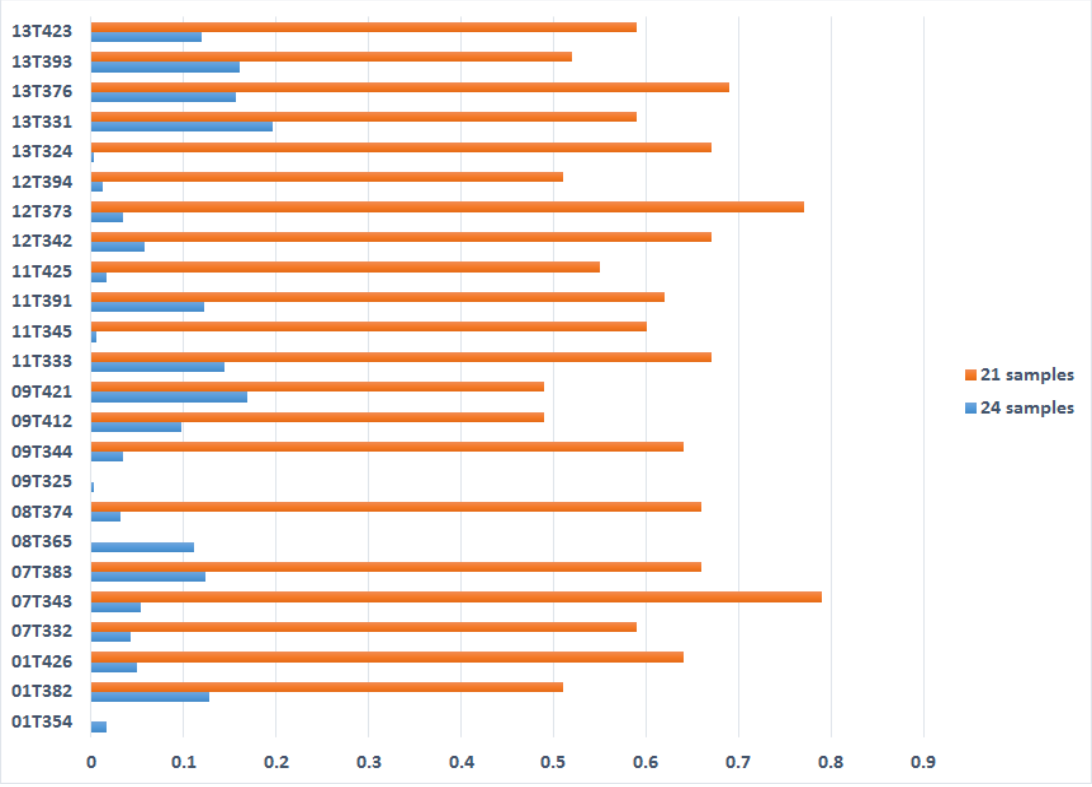


Figure 5-2. Comparison on the lowest Pearson's correlation trendability score between the sets of 24 and 21 samples

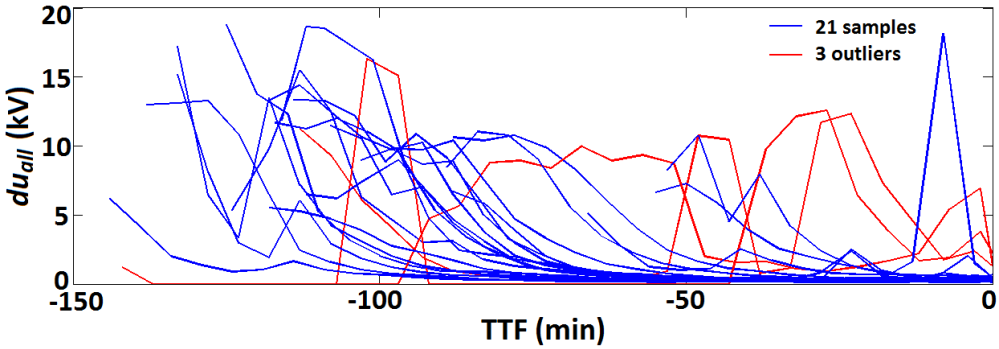


Figure 5-3. du_{all} plots of 21 samples and 3 outliers

5.3 Characteristics of du_{all}

As outlined in the experimental plan in Figure 4-6, the inception of electrical treeing is performed under the influence of 50Hz at 18.0kV peak within 30 minutes. Only initiated electrical trees of 10µm are counted for Stage 2 where the composite waveform was maintained at 14.4 kV peak for a maximum of 2.5 hours. The analysis of PD data in Chapter 4 only considers the tree growth in Stage 2, and this section discusses the correlation between PD data and the electrical treeing mechanism.

From visual observation of Figure 5-3, no distinct pattern is detected at the start of Stage 2 before du_{all} decreases exponentially. When the first tree branch touches the ground electrode, du_{all} reaches a consistent value before breakdown. Hence, the behaviour of du_{all} can be divided into three regions as shown in Figure 5-4, taking sample 09T412 as an example.

The tree image in the first region shows an electrical tree about to grow, resulting in a decreasing trend of du_{all} in Region 1 (other samples may experience a different trend). This behaviour may be due to the change of frequency and magnitude of applied voltage from Stage 1 to Stage 2 of the experimental plan. Bahadoorsingh mentioned in [18] that at the initial stage, very faint filaments of light were observed at the needle tip penetrating and eroding into the epoxy resin. The light is described as intermittent prior to electrical tree initiation. The formed tubule gets thicker as the electrical tree begins to grow towards the ground electrode.

In Region 2, all the samples show a declining pattern with a steep drop at the start and becoming more gradual upon entering Region 3. The dashed line in Figure 5-3 indicates the time of the first tree branch touching the ground electrode, t_{gnd}.

In Region 3, du_{all} remains constant with all samples experiencing a return tree growth, i.e. the tree branches grew from the ground electrode towards the needle tip (see tree image in Region 3 of Figure 5-4). Notably, the return growth occurs after the original tree traversed the insulation gap. A similar observation was documented in [13], [27]–[29]. This return tree grows through the existing tubules from the ground electrode, and results in thicker channels. At the same time, the downward growth also slowly widens with fine tubules. Thus, the tree is more likely to develop thicker channels rather than spreading to the entire sample. It was reported in [13] that the tree growth was also influenced by the hot ionised gas plasma in the channels. These ions look for the shortest path to create and sustain the breakdown arc. Since breakdown did not occur when the original tree growth traversed the insulation gap, this suggests the original growth is non-conducting while return growth is conducting. However, there is no clear indicator to justify the starting time of the return tree.

The small du_{all} values in Region 3 depict a higher repetition rate of discharges. This can be due to the increase of ionisation in the tubules when the discharges erode the ground electrode [13]. The behaviour of du_{all} of all samples can be observed in Figure 5-5 and Figure 5-6 in the next section. Some of the samples show fluctuations just before breakdown, as shown in Figure 5-4. The increase of du_{all} just before breakdown might be due to the decrease of discharge activity when a tree channel suffers partial breakdown. This implies full breakdown is about to happen. It is mentioned in [18] that the discharge activity is restored gradually with further development of the conductive channels to create favourable conditions for the breakdown arc. The activity of the discharges then increases, immediately prior to the luminous breakdown event.

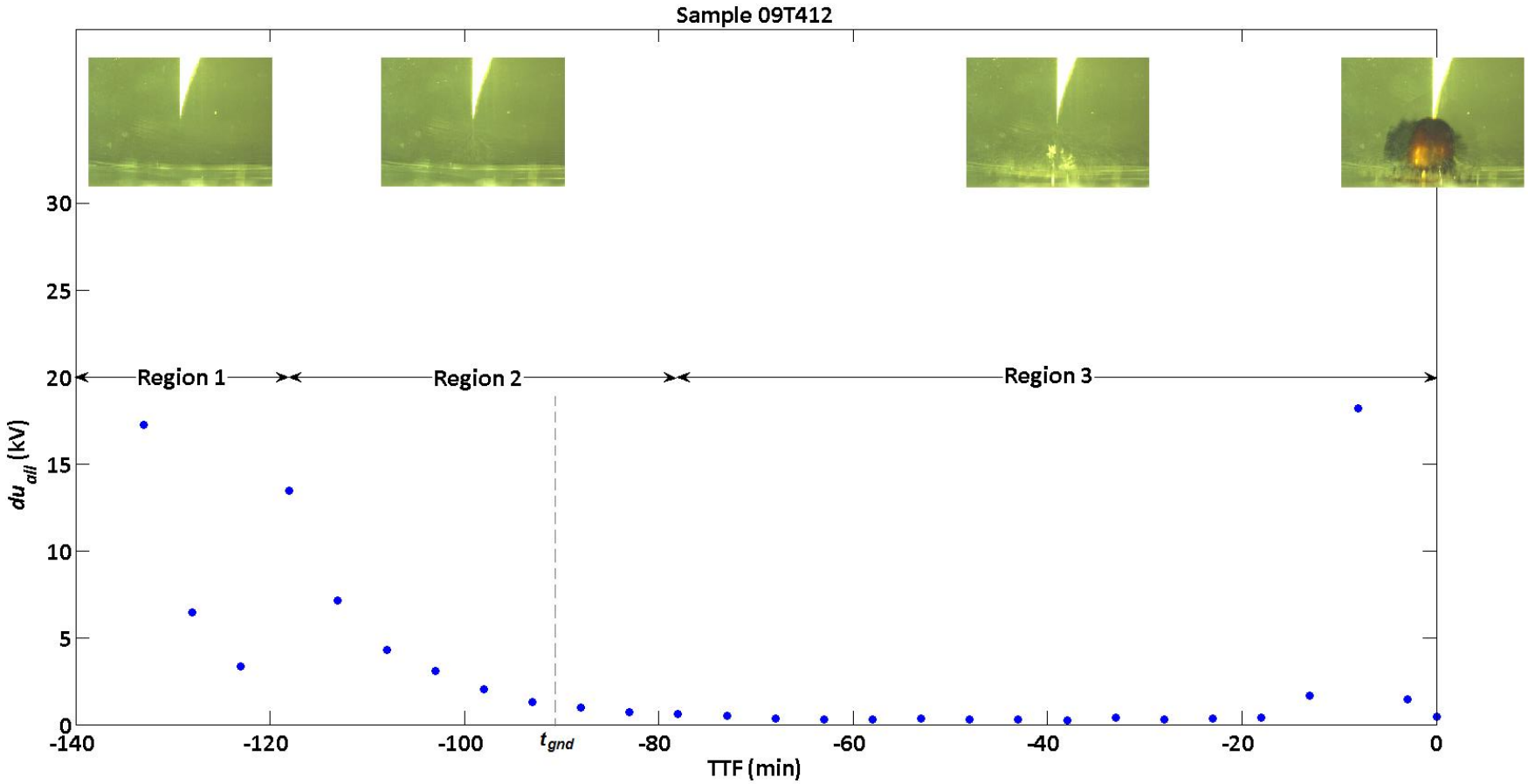


Figure 5-4. du_{all} feature and tree images of sample 09T412

5.4 Applying Curve Fitting to du_{all}

Curve fitting is one of the simplest approaches that can be applied in model fitting. Since the stress endurance from accelerated life testing mostly obeys the exponential or inverse-power law [178], both these models are considered for fitting the du_{all}. Figure 5-5 plots the du_{all} exponential fit and inverse power fit for all 21 training samples. A clearer view of the points in Region 3 can be seen in Figure 5-6 by applying a logarithmic scale on the *y*-axis of Figure 5-5. As can be seen, the exponential fits better than the inverse power and can be verified by the GOF values in Table 5-2. Two of the statistical measures discussed in Section 3.6 are computed to evaluate the performance of the fits i.e. adjusted coefficient of determination, $R_{adj}^{\ 2}$ and root mean squared error, *RMSE*. Based on the GOF values, the exponential curve is considered for the prognostic model rather than the inverse power curve.

Table 5-2. The GOF of exponential and inverse-power fits

Curve Fitting	*Adjusted R-Square*	*RMSE*
Exponential	0.84	1.33
Inverse power	0.69	2.02

Referring to Figure 5-5, the inconsistency of du_{all} values in Region 1 occurs for a very short period i.e. 10 to 20 minutes after the tree initiated. In this region, the samples do not really have a common behaviour. It should also be noted that not all samples experienced the random behaviour thus not all samples have a clear Region 1 in Figure 5-5. In contrast, all samples trend together in Region 2 and 3 where the du_{all} points decrease exponentially before remaining constant. One approach to modelling three distinct regions of behaviour is to fit three piecewise functions. However, if three piecewise functions were considered here, the model would have no predictive power because du_{all} does not trend towards a specific value at the end of Region 1 and the start of Region 2. Since du_{all} is mostly constant in Region 3, a piecewise model would be equal to a constant value with no predictive power. Hence, similarly to Region 1, no individual fitting is required for this region.

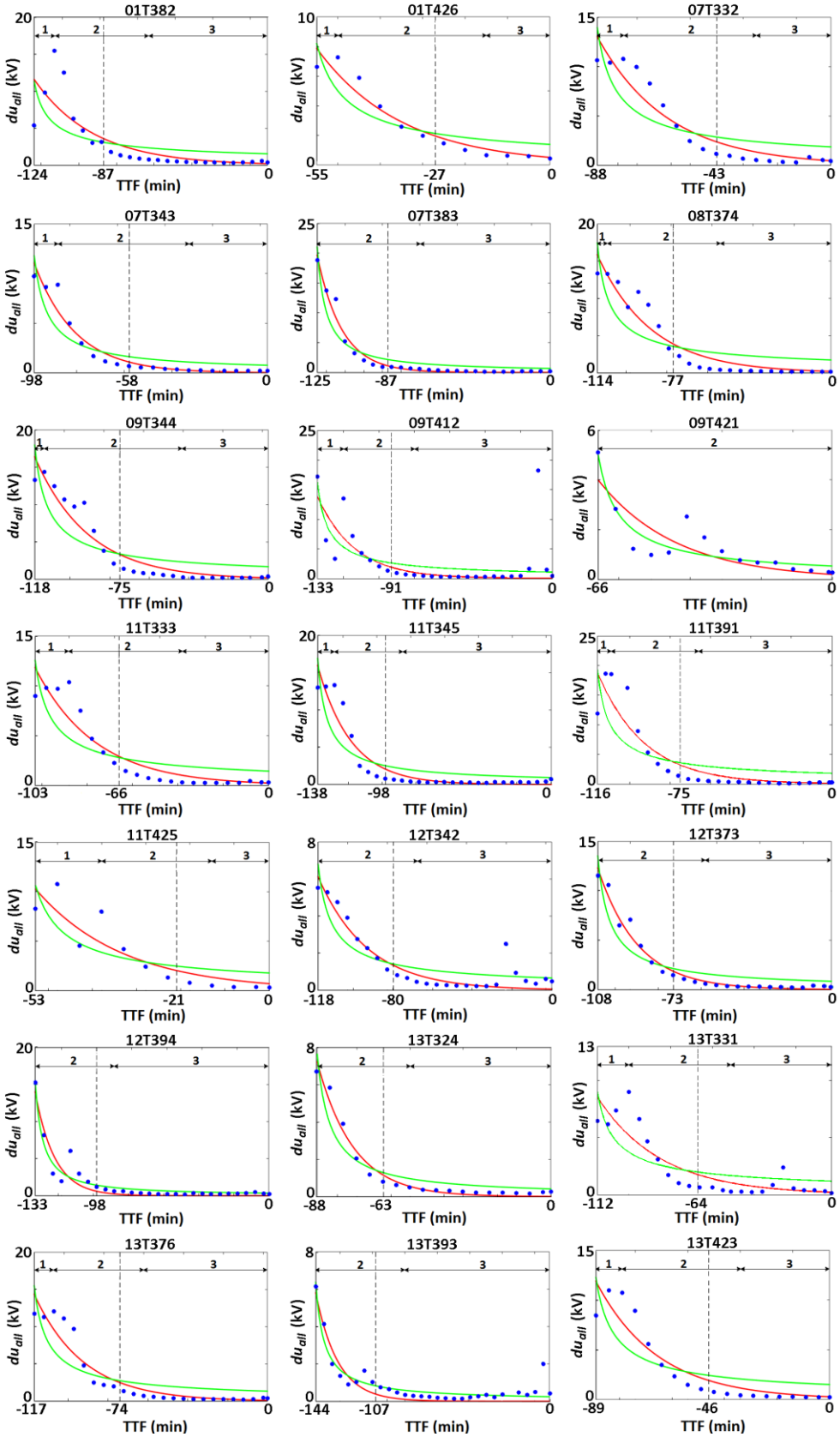


Figure 5-5. The exponential and inverse power fits for du_{all} of 21 samples. Dashed line indicates the t_{gnd}, red lines represent the exponential fits and green lines represent the inverse power fit

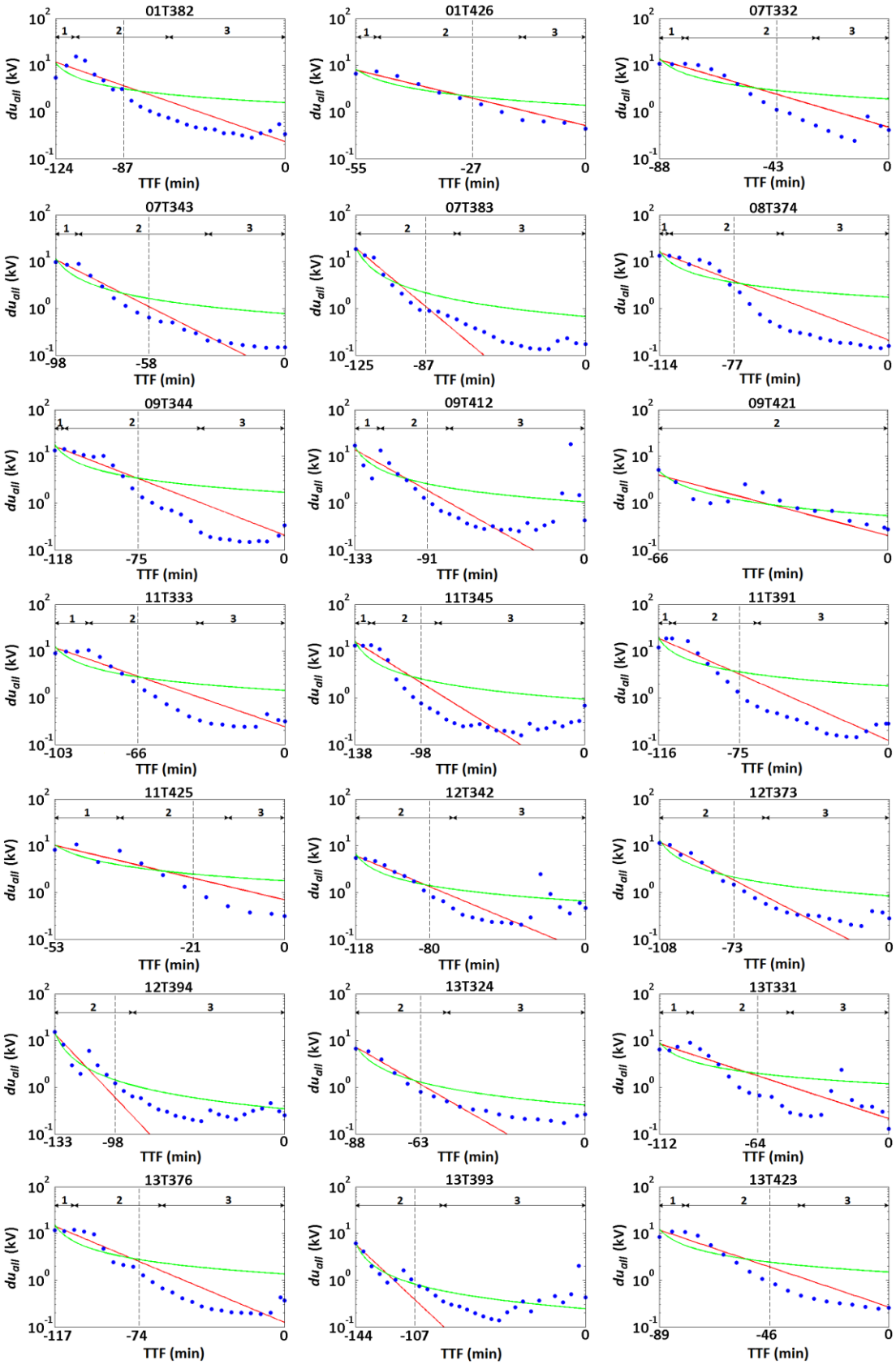
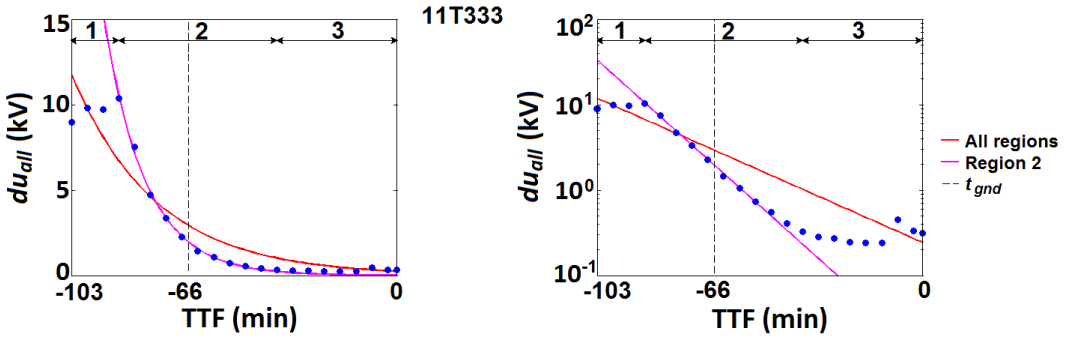


Figure 5-6. The exponential and inverse power fits in semilog scale. Dashed line indicates the t_{gnd}, red lines represent the exponential fits and green lines represent the inverse power fit

Among the three regions, only the du_{all} points in Region 2 show a clear monotonic trend. If only these points were considered for exponential fitting, it is expected to closely meet the du_{all} at the touch ground time, t_{gnd} (see Figure 5-7), which occurs at the 'knee' of the fit. This means, the fit is close to the points in Region 2 but deviates from the points in Region 3 (see Figure 5-7b), thus the fit diverges from the actual breakdown point, $du_{all(bd)}$ (when TTF = 0). This is experienced by samples with no Region 1 e.g. 07T383, 12T342, 12T373, 12T394, 13T324 and 13T393. Referring to Figure 5-6, the inverse power curve seems to fit these samples more closely compared to the exponential.

On the other hand, the fit of sample 09T421 (no Region 1) yields a breakdown point, $du_{fit(bd)}$ close to the $du_{all(bd)}$ due to the nonexistence of Region 3. Among all samples, only this sample has no record of the t_{gnd} because of its very fine tree branches, resulting in a poorer quality image hence the first ground touch is ambiguous. The fit closely meets all the du_{all} points and may include the du_{all} at t_{gnd} if it were known. Therefore, the exponential fits of the samples with no Region 1 may aid the prediction of t_{gnd} better than the breakdown time, t_{bd} when compared to samples with Region 1.

From Figure 5-7, we can see that the fit of all du_{all} points deviates from points in Region 2. This is due to the du_{all} points in Region 1 which actually aid the fit to reach the true breakdown point. This fit yields a higher knee point and lower slope in Figure 5-7a and Figure 5-7b respectively when compared to the fit of only du_{all} points in Region 2. Conversely, du_{all} points in Region 3 do not really affect the fit. This means, fitting du_{all} points in Region 1 and 2 of sample 11T333 will result in an almost identical fit to all du_{all} points. Also, fitting du_{all} points in Region 2 and 3 will result in a similar fit to the du_{all} points in Region 2 only. Therefore, samples with a long Region 3 will push the fitted curve farther from the breakdown point. This can be seen in samples with a long breakdown time in Figure 5-6 e.g. 09T412, 11T345, 12T394 and 13T393. On the contrary, samples with a short breakdown time will have a brief Region 3 e.g. 01T426, 09T421 and 11T425, hence yielding curves that fit closer to the breakdown point.



(a) (b)

Figure 5-7. Comparison between exponential fit of all du_{all} points and exponential fit of only du_{all} points in Region 2 in (a) linear scale and (b) semilog scale

For simplicity, the semilogarithmic (or semilog) scale is applied instead of the linear scale so that the fit can be expressed as a linear equation. Hence, the linear regression can also be considered to fit $\log_{10} du_{all}$ values. The GOF in Table 5-2 is rewritten in Table 5-3 to compare the closeness of the three fits; exponential, inverse power and linear. Among the three fits, exponential shows the highest accuracy. Both linear and inverse power fits have very close accuracy. Since fitting $\log_{10} du_{all}$ using inverse power does not give a linear fit, only exponential and linear fits are considered and compared for model development.

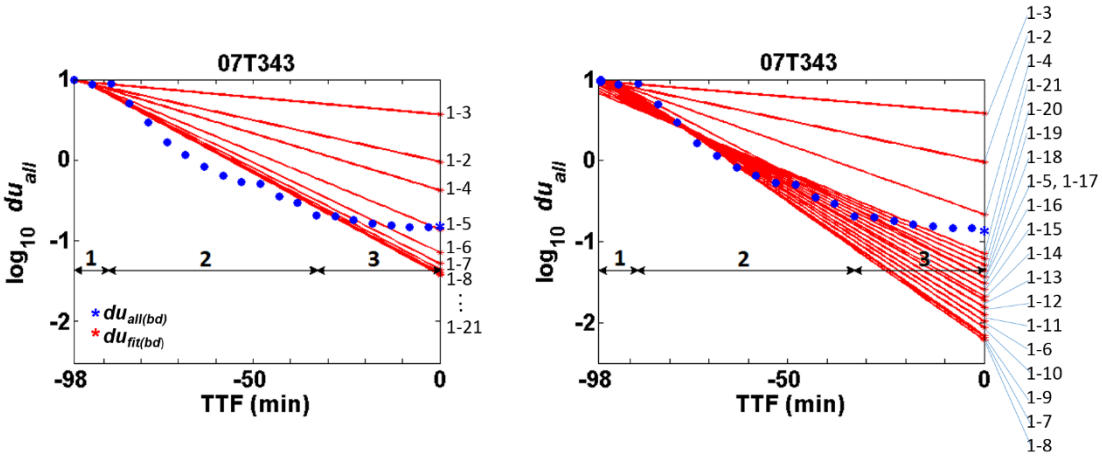
Table 5-3. The GOF of exponential, inverse power and linear fits

Curve Fitting	*Adjusted R-Square*	*RMSE*
Exponential	0.84	1.33
Inverse power	0.69	2.02
Linear	0.71	2.01

The accuracy of the lifetime predictions depends on the availability of the currently available data. Thus, both linear and exponential fits are employed with consideration of the availability of the du_{all} values. Figure 5-8 shows the plots of both fits where the labels indicate which of the du_{all} values are being fitted. For instance, line 1-3 fits the first three du_{all} values. The very small change in du_{all} in Region 1 of sample 07T343 causes the 1-3 line to be close to horizontal, hence it is impossible

for this curve to meet the actual breakdown point in two and a half hours. As the available du_{all} values enter Region 2, the gradient of the fit, m_{fit}, increases. The exponential fit meets the $du_{all(bd)}$ when the first five du_{all} points are available whereas the linear fit requires four du_{all} values. When more data are available and Region 2 becomes more dominant than Region 1, both linear and exponential curves move farther from the $du_{all(bd)}$ point as the gradient of the line keeps increasing. Thus, m_{fit} is one of the main criteria in fitting the du_{all}.

The changes in m_{fit} cause the du_{all} of the fit at TTF = 0, $du_{fit(bd)}$, to vary. Referring to Figure 5-8a, the $du_{fit(bd)}$ of the exponential curve decreases significantly at first and becomes more consistent when more du_{all} values are available. The $du_{fit(bd)}$ of the linear curve also decreases at the initial stage but then keep increasing when the m_{fit} started to decrease (see line 1-9 and above). This indicates that du_{all} values in Region 3 have an influence on linear fits but not on exponential fits.



(a) (b)

Figure 5-8. Fitting the du_{all} of sample 07T343 with consideration of the availability of the data (a) exponential fit (b) linear fit.

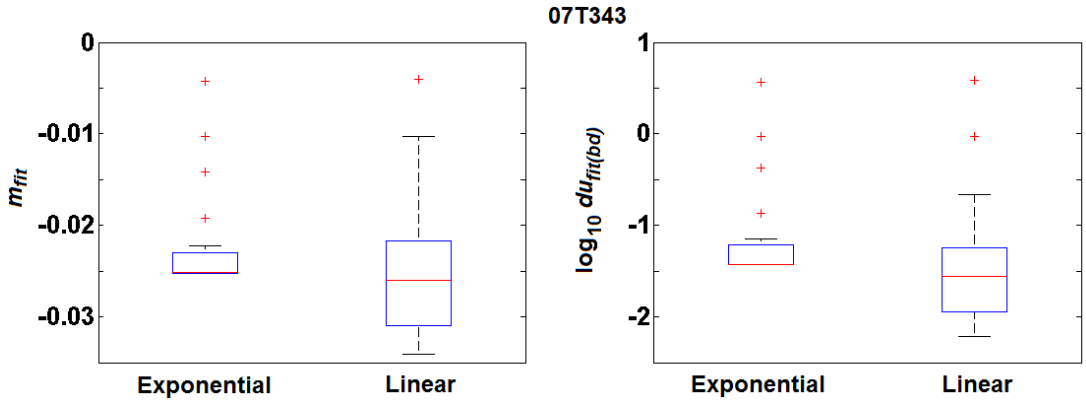
The m_{fit} and $\log_{10}$ $du_{fit(bd)}$ of both exponential and linear fits considering the availability of the du_{all} values of sample 07T343 are given in Table 5-4 and plotted in the boxplots in Figure 5-9. From the boxplots, we can see both m_{fit} and $\log_{10}$ $du_{fit(bd)}$ have the same distribution pattern. For exponential fits, the median of m_{fit} and $\log_{10}$ $du_{fit(bd)}$ are also the minimum values because of the strong skew on the data. In addition, the first four $du_{fit(bd)}$ values that lie beyond the upper whisker are outliers

(where the maximum whisker length is 1.5 of the interquartile range (IQR)). Referring to Table 5-4 and Figure 5-8, both m_{fit} and $du_{fit(bd)}$ of the exponential fits begin to keep their values consistent when the first eight du_{all} values (1-8) are available. This means, the fit of eight du_{all} values is very similar to the fit of all du_{all} values. Therefore, the properties of the prognostic model can be based on the exponential fits in Figure 5-6 which consider all the du_{all} values.

In contrast, the $du_{fit(bd)}$ of linear fits keep changing and do not reach a constant value thus resulting in a more scattered distribution than the exponential fits (higher variance). The closest $du_{fit(bd)}$ value to the $du_{all(bd)}$ is when all du_{all} values are available which violates the purpose of prediction. As an alternative, an average value can be considered for determining the properties of the prognostic model. Based on Figure 5-8b, the average of $du_{fit(bd)}$ might exclude the values from the outliers i.e. curves 1-2, 1-3 and 1-4.

Table 5-4. The m_{fit} and $\log_{10} du_{fit(bd)}$ of the exponential and linear fits of sample 07T343 with consideration on the availability of the du_{all} values

Available du_{all} *points*	*Exponential Fit*		*Linear Fit*	
	m_{fit}	$\text{Log}_{10}\, du_{fit(bd)}$	m_{fit}	$\text{Log}_{10}\, du_{fit(bd)}$
1-2	-0.010	-0.023	-0.010	-0.023
1-3	-0.004	0.564	-0.004	0.582
1-4	-0.014	-0.378	-0.017	-0.664
1-5	-0.019	-0.864	-0.026	-1.438
1-6	-0.022	-1.148	-0.032	-1.986
1-7	-0.024	-1.288	-0.034	-2.177
1-8	-0.025	-1.361	-0.034	-2.218
1-9	-0.025	-1.399	-0.034	-2.164
1-10	-0.025	-1.417	-0.032	-2.062
1-11	-0.025	-1.422	-0.030	-1.904
1-12	-0.025	-1.426	-0.029	-1.813
1-13	-0.025	-1.428	-0.028	-1.724
1-14	-0.025	-1.429	-0.028	-1.675
1-15	-0.025	-1.429	-0.027	-1.597
1-16	-0.025	-1.428	-0.025	-1.517
1-17	-0.025	-1.427	-0.024	-1.439
1-18	-0.025	-1.427	-0.023	-1.360
1-19	-0.025	-1.426	-0.022	-1.286
1-20	-0.025	-1.425	-0.021	-1.211
1-21	-0.025	-1.424	-0.020	-1.146

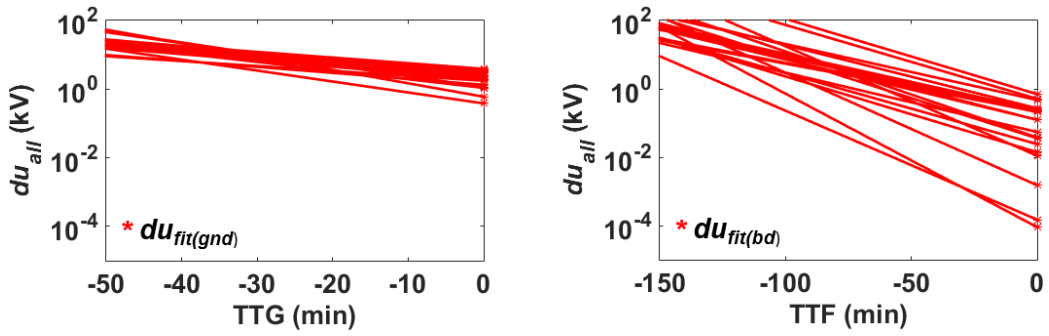


(a) (b)

Figure 5-9. The boxplot of the (a) m_{fit} and (b) $\log_{10} du_{fit(bd)}$ for both exponential and linear fits of sample 07T343

5.5 Determining the Properties of the Model

As mentioned in Section 3.5, a prognostic model requires a threshold value as a prediction point. Based on the analysis of the treeing data, the prediction point can be formulated from the du_{all} of the fit, du_{fit}, either at t_{gnd} ($du_{fit(gnd)}$) or at t_{bd} ($du_{fit(bd)}$). Taking exponential fit as an example, these values can be extracted by gathering all the fittings in Figure 5-6 where the x-axis can be either time-to-ground, TTG, or time-to-failure, TTF, as shown in Figure 5-10a and Figure 5-10b respectively. The $du_{fit(gnd)}$ and $du_{fit(bd)}$ are the y-intercepts of the plots respectively. The smaller range of $du_{fit(gnd)}$ values shows a better quality of prognostic model compared to $du_{fit(bd)}$. However, for a lifetime that is less than 2.5 hours, and some of the TTG values are only a quarter of the TTF, predicting the TTG may be less beneficial. Therefore, the $du_{fit(bd)}$ values were selected as the prediction point of the prognostic models derived from exponential and linear curve fittings.



(a) (b)

Figure 5-10. Exponential fits of all 21 training samples with (a) du_{all} vs TTG, and (b) du_{all} vs TTF

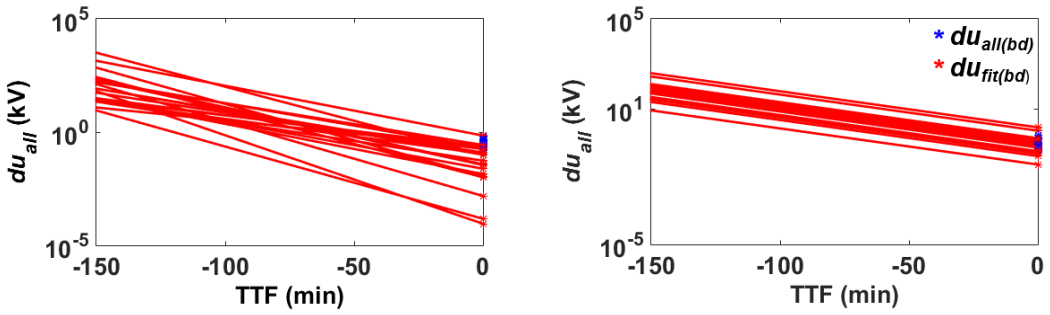
5.5.1 Exponential Fitting

Exponential fits in Figure 5-10b are replotted in Figure 5-11a together with the actual breakdown points, $du_{all(bd)}$. As can be seen, the $\log_{10} du_{all(bd)}$ varies between -1and 0 but the fits have a wider range of $\log_{10} du_{fit(bd)}$ that is between -4 and 0 (also listed in Table 5-5) which depicts a poor quality of prediction point. The highlighted $\log_{10} du_{fit(bd)}$ points in Table 5-5 indicate the ones that are out of the actual breakdown range. Those highlighted samples mostly have high m_{fit} values that are ranged

between -0.017 to -0.039. In order to obtain a closer $du_{fit(bd)}$ value, the m_{fit} therefore should be in a smaller range. Henceforth, an algorithm was designed that would maximise the accuracy of breakdown point predictions by selecting a realistic gradient value. The construction of the algorithm proceeded as follows.

The developed prognostic model aims to predict the TTF based on the current du_{all} value. The available du_{all} values will be fitted with an exponential curve and extrapolated to the threshold point, du_{th}, which is estimated from the $\log_{10} du_{fit(bd)}$ of the training samples. To maximise the accuracy, some of the fits in Figure 5-11a are modified to a new gradient, m_{model}. From the properties of Figure 5-11a in Table 5-4, the m_{model} is set to -0.016 which is the most frequent gradient that results in realistic $du_{fit(bd)}$ values. Consequently, the fits that have gradient higher than -0.16 are the ones that will be modified. The new $du_{fit(bd)}$ values are determined by anchoring the fit at du_{all} when t = 0 or TTF = $-t_{bd}$. Taking sample 12T394 as an example in Figure 5-12, the actual fit that yields the lowest $\log_{10} du_{fit(bd)}$ (i.e. -4), has then increased using this process to -0.9.

All the fits in Figure 5-11a are redrawn in Figure 5-11b and the new $du_{fit(bd)}$ values are tabulated in Table 5-5. As a result, the range of $\log_{10} du_{fit(bd)}$ decreases from 3.89 to 1.65. The fits are modified purposely for determining the du_{th} which is the average of the $\log_{10} du_{fit(bd)}$ of the modified fit, -0.61. Fixing the gradient may cause overfitting thus allowance for m_{fit} values is set from -0.013 to -0.018 based on data in Table 5-5. Finally, the properties of the model are given in Table 5-6 to be applied to the testing samples for validation in the next section.



(a) (b)

Figure 5-11. Exponential fits of all 21 training samples with (a) the actual fits, and (b) the modified fits

Table 5-5. Properties of the fits in Figure 5-11

Sample ID	t_{bd} (min)	*Actual Fits (Figure 5-11a)*		*Modified Fits (Figure 5-11b)*	
		Gradient, m_{fit}	**Log_{10} $du_{fit(bd)}$**	**Gradient, m_{fit}**	**Log_{10} $du_{fit(bd)}$**
01T382	129	-0.014	-0.64	-0.013	-0.60
01T426	60	-0.022	-0.29	-0.016	0.04
07T332	93	-0.016	-0.33	-0.016	-0.29
07T343	103	-0.025	-1.42	-0.016	-0.48
07T383	130	-0.033	-2.82	-0.016	-0.62
08T374	118	-0.017	-0.68	-0.016	-0.62
09T344	123	-0.016	-0.68	-0.016	-0.67
09T412	138	-0.021	-1.62	-0.016	-0.95
09T421	71	-0.020	-0.69	-0.016	-0.42
11T333	108	-0.016	-0.62	-0.016	-0.57
11T345	143	-0.022	-1.85	-0.016	-0.97
11T391	120	-0.019	-0.91	-0.016	-0.57
11T425	58	-0.022	-0.16	-0.016	0.19
12T342	123	-0.017	-1.26	-0.016	-1.09
12T373	113	-0.024	-1.45	-0.016	-0.60
12T394	138	-0.039	-4.05	-0.016	-0.87
13T324	93	-0.032	-1.96	-0.016	-0.47
13T331	117	-0.014	-0.67	-0.014	-0.67
13T376	122	-0.018	-0.90	-0.016	-0.71
13T393	149	-0.032	-3.83	-0.016	-1.46
13T423	94	-0.018	-0.57	-0.016	-0.34
Average	112	-0.022	-1.305	-0.016	-0.61

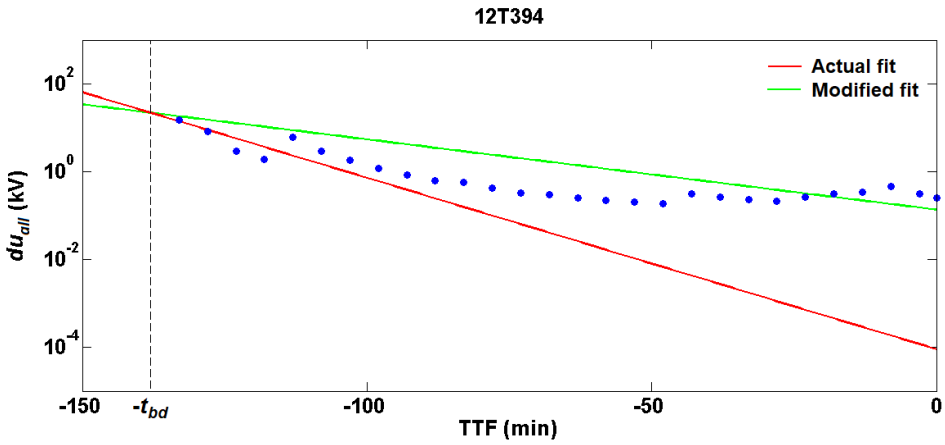


Figure 5-12. The actual and modified fits are anchored at $-t_{bd}$

Table 5-6. Properties of the exponential prognostic model

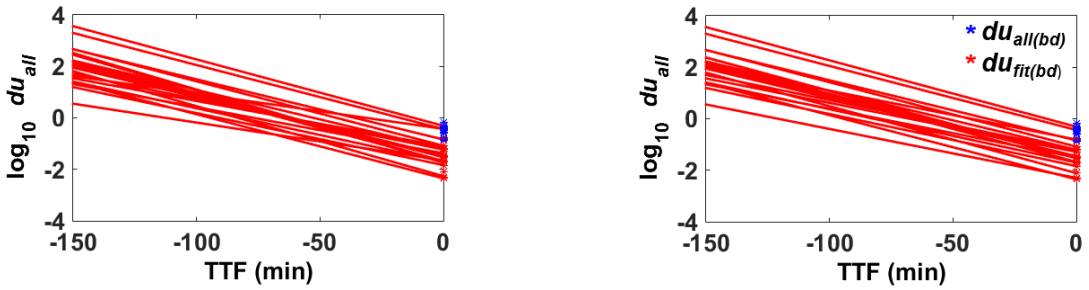
Properties	***Minimum***	***Maximum***	***Range***	***Mean***	***Standard Deviation***
m_{fit}	-0.013	-0.018	0.005	-0.016	0.0006
$\text{Log}_{10}\, du_{fit(bd)}$	-1.46	0.19	1.65	-0.61	0.35
t_{bd}	58 min	149 min	91 min	112 min	25 min

5.5.2 Linear Fitting

As mentioned previously, the linear fit that represents each of the training samples may be taken as an average of the fits considering each availability of du_{all} values. The average linear curves disregard the outliers and are plotted in Figure 5-13a, for which the properties are given in Table 5-7. The same algorithm is applied here for determining the properties of prognostic model.

From Figure 5-11a and Table 5-5, we can see that fits with high m_{fit} causes the $du_{fit(bd)}$ to become lower thus increase the range. As a result, those fits are modified to a new gradient, m_{model}, in order to decrease the range. However, this is not the case for the linear fits in Figure 5-13a. The fits of the maximum and minimum $du_{fit(bd)}$ are in parallel (with almost same value of m_{fit}). Hence, modifying these fits will not decrease the range of $du_{fit(bd)}$.

The m_{fit} varies from -0.014 to -0.028. The range is considered high and should be decreased to avoid unreasonable error in predictions. Nonetheless, modifying the fits may increase the range of $du_{fit(bd)}$, hence the optimum range for m_{fit} would be from -0.019 to -0.026. The fits that are not in the range are highlighted in Table 5-7. Those m_{fit} are modified to the nearest value, either -0.019 or -0.026. All the final fits are replotted in Figure 5-13b. Modifying the fits does not really change the average of $du_{fit(bd)}$ giving du_{th} = -1.41. The properties of the prognostic model are given in Table 5-8.



(a) (b)

Figure 5-13. Linear fits of all 21 training samples with (a) the actual fits, and (b) the modified fits

Table 5-7. Properties of the fits in Figure 5-13

Sample ID	t_{bd} (min)	*Actual Fits (Figure 5-13a)*		*Modified Fits (Figure 5-13b)*	
		Gradient, m_{fit}	**Log_{10} $du_{fit(bd)}$**	**Gradient, m_{fit}**	**Log_{10} $du_{fit(bd)}$**
01T382	129	-0.018	-1.11	-0.019	-1.28
01T426	60	-0.025	-0.44	-0.025	-0.44
07T332	93	-0.023	-0.84	-0.023	-0.84
07T343	103	-0.028	-1.69	-0.026	-1.52
07T383	130	-0.027	-2.26	-0.026	-2.11
08T374	118	-0.025	-1.53	-0.025	-1.53
09T344	123	-0.024	-1.46	-0.024	-1.46
09T412	138	-0.021	-1.70	-0.021	-1.70
09T421	71	-0.014	-0.43	-0.019	-0.81
11T333	108	-0.023	-1.20	-0.023	-1.20
11T345	143	-0.025	-2.35	-0.025	-2.35
11T391	120	-0.026	-1.68	-0.026	-1.68
11T425	58	-0.026	-0.31	-0.026	-0.31
12T342	123	-0.018	-1.38	-0.019	-1.51
12T373	113	-0.023	-1.46	-0.023	-1.46
12T394	138	-0.020	-1.84	-0.020	-1.84
13T324	93	-0.028	-1.74	-0.026	-1.52
13T331	117	-0.020	-1.23	-0.020	-1.23
13T376	122	-0.023	-1.47	-0.023	-1.47
13T393	149	-0.015	-1.65	-0.019	-2.30
13T423	94	-0.025	-1.08	-0.025	-1.08
Average	112	-0.023	-1.37	-0.023	-1.41

Table 5-8. Properties of the linear prognostic model

Properties	***Minimum***	***Maximum***	***Range***	***Mean***	***Standard Deviation***
m_{fit}	-0.019	-0.026	0.007	-0.023	0.003
$\text{Log}_{10}\, du_{fit(bd)}$	-0.31	-2.35	2.04	-1.41	0.52
t_{bd}	58 min	149 min	91 min	112 min	25 min

5.6 Validation of the Model

Seven samples have been randomly selected from each harmonic group to be treated as the testing samples i.e. 01T346, 07T355, 08T392, 09T381, 11T372, 12T424 and 13T363. The samples are used to validate the exponential and linear prognostic models proposed in the previous section. Both holdout and cross-validation methods are employed to compare the effectiveness of the methods on a small sample size.

For testing, the du_{all} values are plotted versus the testing time, t, instead of TTF. Each of the du_{all} values will be considered for fitting as in the description of Figure 5-8 and Table 5-4. Figure 5-14 outlines the flow of the testing procedure. Firstly, the first two du_{all} points are fitted and extrapolated until the du_{th} is reached. The m_{fit} and the time when the fit meets du_{th}, t_{th}, are determined. The procedure is repeated for the next du_{all} value and continued until the last du_{all}. The allowable range of m_{fit} is given in Table 5-6 and Table 5-8 and can be expressed as $m_{fit(min)} \leq m_{fit} \leq m_{fit(max)}$. Thus, three conditions will be treated differently:

- If m_{fit} is less than $m_{fit(min)}$, the fit will be modified to $m_{fit(min)}$ and the new t_{th} will be determined.
- If m_{fit} is between $m_{fit(min)}$ and $m_{fit(max)}$, the t_{th} stays the same.
- If m_{fit} is greater than $m_{fit(max)}$, the fit will be modified to $m_{fit(max)}$ and the new t_{th} will be determined.

The new t_{th} is obtained using the same approach to determine the new $du_{fit(bd)}$ as in the previous section. The fit is modified by anchoring the fit at $t = 0$. Again, referring

to Table 5-6 and Table 5-8, the $t_{bd(est)}$ should be between the $t_{bd(min)}$, 58 and $t_{bd(max)}$, 149 minutes. That means, if t_{th} is lower than 58 minutes, it will be disregarded and the $t_{bd(est)}$ is set to 58 minutes. The same applies if t_{th} is greater than 149 minutes. Next, the TTF is calculated using equation 4-12 for both actual, TTF_{act} and estimated, TTF_{est}. The error, e, is computed as follows:

$$e = TTF_{act} - TTF_{est} \tag{5-1}$$

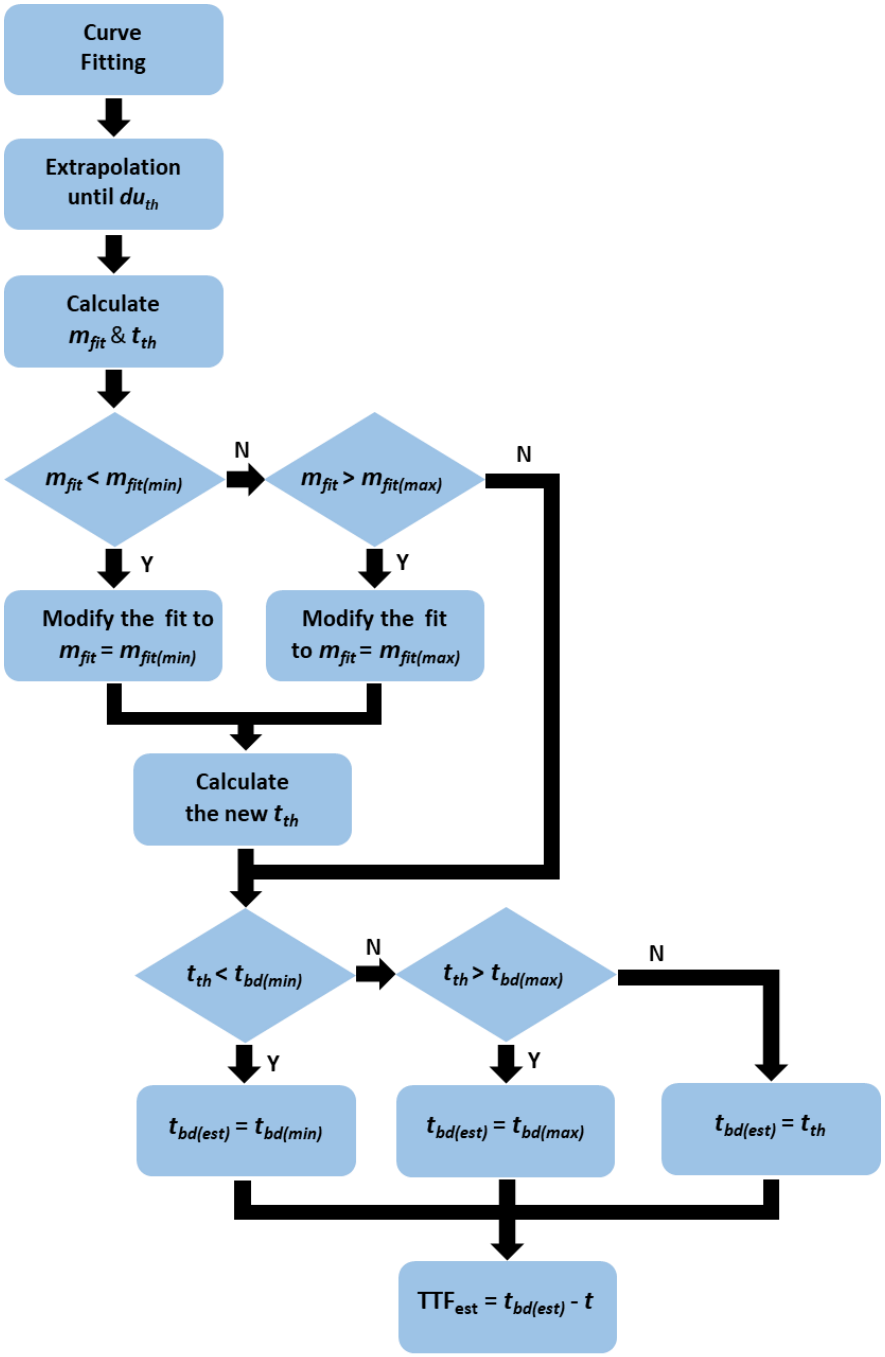


Figure 5-14. The flow of the testing procedure

With only seven test samples, the analysis of the model performance might be limited, especially due to the varying behaviour of the du_{all} trend. Some samples may have the same breakdown time but with different m_{fit}. Also, samples with similar m_{fit} can have a wide range of $du_{fit(bd)}$. Those seven samples may not represent each of the unique characteristics. Therefore, 4-fold cross validation is employed so that each sample can be tested. All the 31 samples (including the 3 outliers, 01T354, 08T365 and 09T325) are divided into four groups. Group 1 has 7 test samples while the other 3 groups have 8 test samples. To start cross validation, Group 1 is used as the test samples for the holdout approach (with Groups 2, 3, and 4 for training). The same training approach is then employed for the other three groups and the properties of the exponential and linear models are given in Table 5-9 and Table 5-10 respectively.

Table 5-9. Properties of the exponential model employing 4-fold validation approach

Group	***Properties***	***Minimum***	***Maximum***	***Range***	***Mean***	***Standard Deviation***
1	m_{fit}	-0.013	-0.018	0.008	-	-
	$\mathrm{Log}_{10}\, du_{fit(bd)}$	-1.46	0.19	1.65	-0.61	0.35
	t_{bd}	58 min	149 min	91 min	112 min	25 min
2	m_{fit}	-0.013	-0.018	0.008	-	-
	$\mathrm{Log}_{10}\, du_{fit(bd)}$	-1.46	0.08	1.54	-0.61	0.35
	t_{bd}	58 min	149 min	91 min	114 min	24 min
3	m_{fit}	-0.013	-0.018	0.008	-	-
	$\mathrm{Log}_{10}\, du_{fit(bd)}$	-1.46	0.19	1.65	-0.56	0.36
	t_{bd}	58 min	149 min	91 min	111 min	24 min
4	m_{fit}	-0.013	-0.018	0.008	-	-
	$\mathrm{Log}_{10}\, du_{fit(bd)}$	-1.09	0.08	1.17	-0.58	0.29
	t_{bd}	60 min	143 min	83 min	113min	23 min

Table 5-10. Properties of the linear model employing 4-fold validation approach

Group	*Properties*	*Minimum*	*Maximum*	*Range*	*Mean*	*Standard Deviation*
1	m_{fit}	-0.019	-0.026	0.007	-0.023	0.003
	$\text{Log}_{10}\, du_{fit(bd)}$	-2.35	-0.31	2.04	-1.41	0.52
	t_{bd}	58 min	149 min	91 min	112 min	25 min
2	m_{fit}	-0.019	-0.026	0.007	-0.023	0.003
	$\text{Log}_{10}\, du_{fit(bd)}$	-2.35	-0.31	2.04	-1.46	0.48
	t_{bd}	58 min	149 min	91 min	114 min	24 min
3	m_{fit}	-0.020	-0.026	0.006	-0.023	0.003
	$\text{Log}_{10}\, du_{fit(bd)}$	-2.44	-0.31	2.13	-1.39	0.48
	t_{bd}	58 min	149 min	91 min	111 min	24 min
4	m_{fit}	-0.020	-0.026	0.006	-0.023	0.003
	$\text{Log}_{10}\, du_{fit(bd)}$	-2.35	-0.44	1.91	-1.43	0.44
	t_{bd}	60 min	143 min	83 min	113min	23 min

Table 5-11 lists the predicted values from the exponential model of sample 09T381, a test sample from Group 1. All the fits have m_{fit} greater than -0.018. Modifying the m_{fit} to -0.018 gives a closer prediction, $t_{bd(est)}$ to the actual breakdown, t_{bd} (107 minutes) compared to the t_{th} values from the actual fits. The error, e starts to converge at t = 26 min, where the error becomes consistent. In order to visualise the error, the TTF_{act} and TTF_{est} are plotted together versus the testing time, t (which represents the availability of du_{all} values) in Figure 5-15. The estimated TTF values based on the actual fits (before modification) are also plotted to emphasise the effectiveness of the modified fits.

As mentioned previously, when only du_{all} values in Region 1 are available, the exponential and linear fits may give unreasonable predictions, as shown by the TTF_{est} from the actual fit (dashed line) during the early tree growth in Figure 5-15. The proposed model (solid line) has improved those early predictions of all test samples. For the exponential model, most of the samples start to converge in less than 50 minutes, where the error between $TTF_{act(exp)}$ and $TTF_{est(exp)}$, e, remains unchanged

after this point. In contrast, the linear model does not really have a convergence point as the *e* can change after being constant for a while. This is due to the behaviour illustrated in Figure 5-8b.

Table 5-11. The lifetime prediction of sample 09T381

Available du_{all} ***points***	***t*** **(min)**	m_{fit}	t_{th} **(min)**	$t_{bd(est)}$ **(min)**	TTF_{act} **(min)**	TTF_{est} **(min)**	***e***
1	6	-	-	58	101	52	49
1:2	11	-0.133	19	58	96	127	-31
1:3	16	-0.055	36	138	91	94	-3
1:4	21	-0.041	46	110	86	83	3
1:5	26	-0.034	54	104	81	75	6
1:6	31	-0.032	57	101	76	69	7
1:7	36	-0.031	57	100	71	63	8
1:8	41	-0.031	58	99	66	58	8
1:9	46	-0.031	58	99	61	53	8
1:10	51	-0.030	59	99	56	48	8
1:11	56	-0.030	59	99	51	43	8
1:12	61	-0.030	59	99	46	38	8
1:13	66	-0.030	60	99	41	33	8
1:14	71	-0.030	60	99	36	28	8
1:15	76	-0.030	60	98	31	22	9
1:16	81	-0.030	60	98	26	17	9
1:17	86	-0.030	60	98	21	12	9
1:18	91	-0.030	60	98	16	7	9
1:19	96	-0.030	60	98	11	2	9
1:20	101	-0.030	60	98	6	-3	9
1:21	106	-0.030	60	98	1	-8	9
1:22	107	-0.030	60	98	0	-9	9

Some of the linear models result in similar or better prediction than the exponential, but the random behaviour towards the actual breakdown is a disadvantage. This is even worse when the linear model changes the prediction to a value that is beyond the actual value although more du_{all} values are available e.g. samples 11T391,

12T373 and 13T376 in Figure 5-15. Thus, the prediction of the exponential model is visually more consistent and reliable than the linear model.

The estimation of t_{bd} in the developed model is based on the average of $\log_{10} du_{fit(bd)}$ values (Table 5-8) i.e. du_{th}. From the training data analysis, the $\log_{10} du_{fit(bd)}$ values are depending on the t_{bd} of the samples and can be separated into three groups as follows:

- $\log_{10} du_{fit(bd)}$ values that are near to du_{th}
- $\log_{10} du_{fit(bd)}$ that are much lower than du_{th}
- $\log_{10} du_{fit(bd)}$ values that are much higher than du_{th}

This concept is also experienced by the test samples and hence influenced the performance of the error, e. Based on the predictions from the exponential model in Figure 5-15, we can see that samples with short t_{bd} yield $e < 0$ ($TTF_{est(exp)} > TTF_{act}$) while samples with average and long t_{bd} yield $e < 0$ ($TTF_{est(exp)} < TTF_{act}$). However, samples with long t_{bd} yield very much greater e than the average t_{bd}. Hence, a threshold of 20 is chosen empirically to separate the performance between the average and long t_{bd} based on the relative number of e above and below this value. The three groups to describe the performance of the test samples are as follows:

- E1: $0 \leq e \leq 20$
- E2: $e > 20$
- E3: $e < 0$

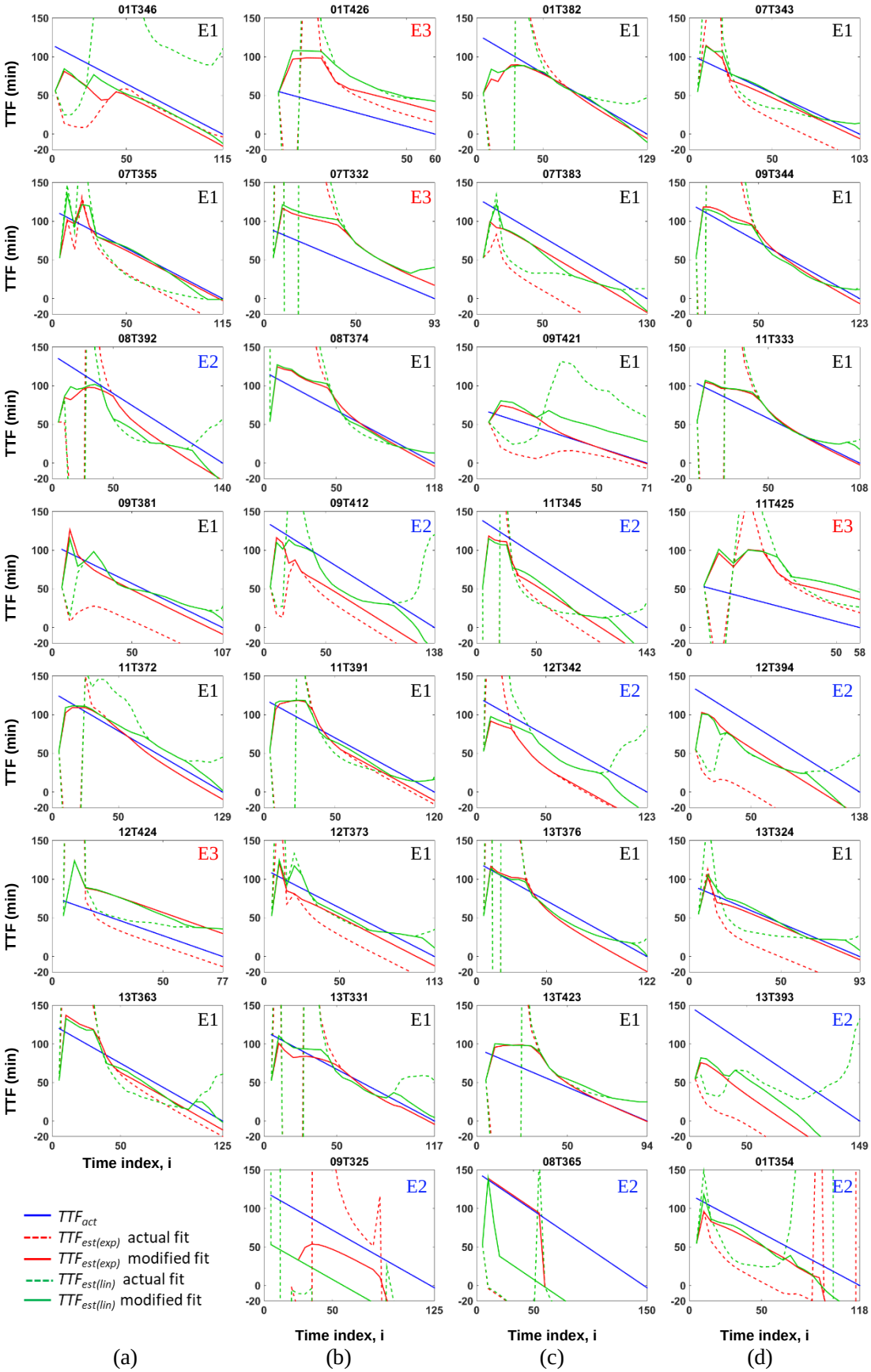
The samples' performances in Figure 5-15 are labelled as E1, E2 and E3 accordingly. Most of the samples fall under E1 category, in which the error is moderate. These samples mostly have the actual t_{bd} from 90 to 130 minutes except for sample 09T421 from Group 3.

Nine samples are categorised as E2, of which three of them are the outliers mentioned in Section 5.2. Five of the samples have a long t_{bd} that ranges from 130 to 150 minutes and one sample (12T342) has E1 range. The modified exponential fit on 12T342 results in a low $t_{bd(est)}$ due to the low value of the first du_{all} which causes the

fit to have a lower *y*-intercept (see Figure 5-6) compared to other samples. The predictions of the E2 group are considered "too early" and not good for prognostics. Too early prediction may lead to early replacement of the asset hence could be a waste of the asset's useful life. Nevertheless, the modified fits of E2 perform better than the actual fits.

Lastly, four samples have *e* under the E3 category due to the low t_{bd} that ranges from 58 to 90 minutes, with sample 07T332 having a slightly longer t_{bd} i.e. 93 minutes. The prediction is beyond the actual value, which gives a penalty to the proposed model. For a short t_{bd} with a high m_{fit} (for example sample 11T425 in Table 5-5), modifying the gradient thus brings the prediction farther causing the modified fits to perform worse than the actual fits.

For the linear model, most of the *e* of the samples can be categorised in the same manner except for samples 09T421, 11T333, 09T381, 11T372 and 13T423. All these samples are under the E1 category where the exponential model results in very small *e*, that is, the $TTF_{est(exp)}$ values are similar to the TTF_{act}. It can be generally concluded that most of the predictions from the linear model, $TTF_{est(lin)}$ values are greater than $TTF_{est(exp)}$ which also applied to those five samples. Hence, the linear model of those samples results in a greater TTF causing the *e* to fall under the E3 category instead of E1.



(a) (b) (c) (d)

Figure 5-15. The plots of TTF of the test samples for the 4-fold cross validation (a) Group 1, (b) Group 2, (c) Group 3, and (d) Group 4

All in all, the proposed algorithm tries to balance the performance between the more extreme cases of samples in E2 and E3. As a result, the algorithm performs the best for samples in E1. Applying cross validation allows better observation of performance on smaller sets of test samples. An unfortunate splitting for the holdout method may not represent the actual performance. For instance, from visual observation of the exponential model, Groups 1 and 3 that have more samples under the E1 category might have better performance than Groups 2 and 4. To add more, Group 4 has the greatest *e* (sample 13T393 with 65 minutes) which lowers the group performance. All the four groups have samples in each category except for Group 3. Hence, applying the holdout method with Group 3 as the test samples may give a biased result. Therefore, the performance of the algorithm applied in this thesis should not be evaluated as an average because it does not reflect the overall performance. Nevertheless, a more reliable analysis can be done statistically by applying a set of performance metrics as described in the next section.

5.7 Performance of the Model

The performance of the four test groups is analysed and compared to evaluate the effectiveness of the cross validation approach over holdout. Both exponential and linear models are compared with the aim to reveal the better model. The analysis also considers the consequences of the three error categories, E1, E2 and E3 on the model's performance. The evaluation is based on two performance metrics introduced in [247] and one metric proposed in this thesis. These three metrics are *prognostic horizon*, PH [247], *convergence horizon*, CH and *cumulative relative accuracy*, CRA [247]. For this latter, the *RA* metric [247] is modified in order to penalise predictions that are beyond the actual lifetime. The sum of the CH and CRA metrics is then used to rank the performance of both exponential and linear algorithms.

5.7.1 Prognostic Horizon

The first metric, prognostic horizon, PH [247], determines how early an algorithm can predict within specified error limits hence can be considered as a robustness measurement. This means, a longer PH is better because earlier reliable prediction provides more time for preventive or corrective action. Therefore, the choice of error bound, α, should depend on the estimated time required to take the action. With an average breakdown time, t_{bd}, around 100 minutes, the error bound in this work is set to 20% of total lifetime.

The equation of PH is given in Chapter 3 (equation 3-27) and is rewritten here using different notation,

$$PH = i_{bd} - i_{\alpha} \tag{5-2}$$

where i_{bd} is the time index at t_{bd} and i_{α} is the first time index when prediction meets the specified criterion for a given α. The time index, i is derived as a fraction of t_{bd} hence $i_{bd} = 1$. All plots in Figure 5-15 are replotted in Figure 5-16 with $\pm\alpha$ bounds as well as the i_{α} for both exponential and linear models. The PH values for all samples are then plotted in Figure 5-17 for comparison.

Generally, both exponential and linear algorithms of samples under the E1 and E2 categories yield $i_{\alpha} \leq 0.2$ except for 01T346, 09T325, 09T421 and 13T393. This results in $PH \geq 0.8$ which means, the algorithms always predict within the desired accuracy zone. On the other hand, the algorithms yield the worst score for samples under E3 since both linear and exponential models never predict within the accuracy zone i.e. $PH = 0$. Only the exponential model on sample 07T332 results in $PH = 0.2$ but this is still not as good as samples under E1. Sample 13T393 is the only sample under the E2 category that has $PH = 0$ and none from E1.

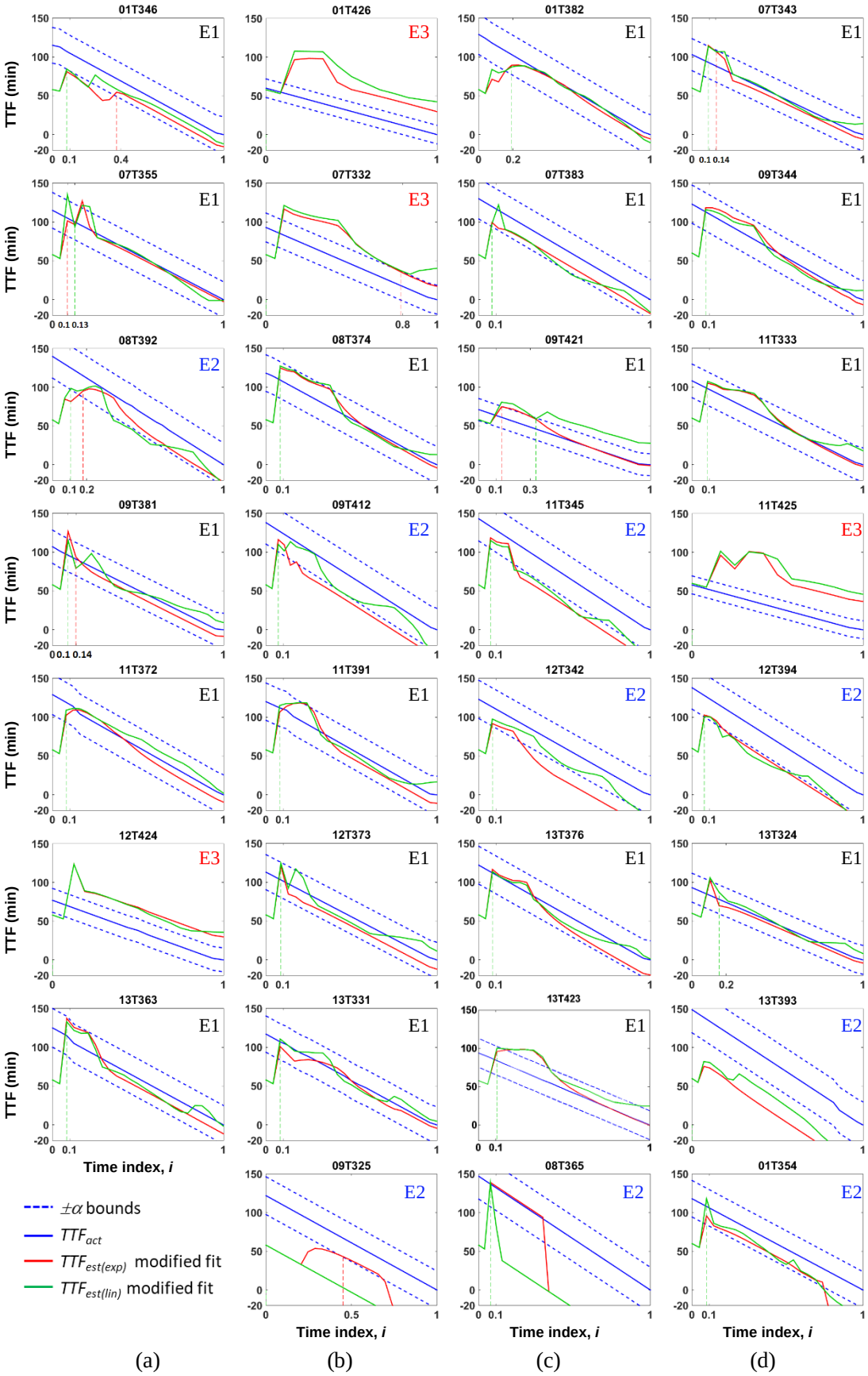


Figure 5-16. The plots of i_α for determining the PH metric of test samples in (a) Group 1, (b) Group 2, (c) Group 3, and (d) Group 4

Comparison between the samples suggests that predictions from both exponential and linear algorithms are not reliable for samples in the E3 category. Although the PH scores are considered high for samples in E1 and E2, it does not reflect the overall performance since it accounts for only the early prediction (based on du_{all} values in Region 1 or early Region 2). Therefore, the score of this metric will not be assessed for overall performance. Comparing the average PH among the four groups, Group 3 yields the highest with PH_{ave} = 0.9. This result is biased since Group 3 has no samples under the E3 category. Hence, the average of group performance does not reflects the actual algorithm's performance. As a final point, both exponential and linear models mostly yield the same PH showing that the proposed algorithm (modified fits) works for both of them. In addition, both models have the same PH average, PH_{ave} = 0.7.

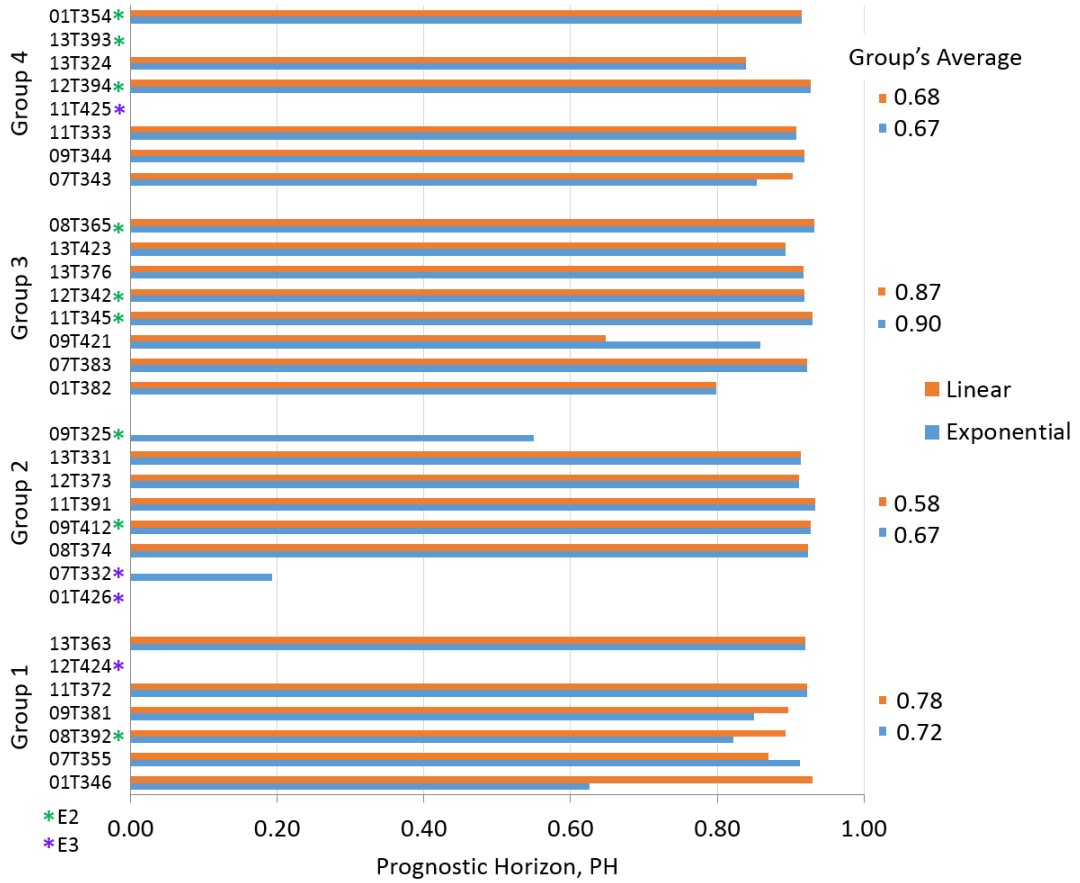


Figure 5-17. The scores of PH metric

5.7.2 Convergence Horizon

The convergence metric introduced in [247] measures the rate of improvement over time of any performance metrics or TTF_{est} itself, in order to reach its perfect score. This means, the best score of convergence will be achieved if the algorithm predicts the actual t_{bd} as early as the prediction starts. Alternatively, convergence may be defined as the time instance that the prediction begins to remain unchanged, i_{CH} until the failure time. After i_{CH}, even if more data are added to the model, the predictions stay the same and result in a TTF_{est} curve that is parallel to the TTF_{act} until the end of life. Therefore, this research work proposed a new metric, *convergence horizon*, CH, which is expressed in a similar way as the *PH* metric:

$$CH = i_{bd} - i_{CH} \tag{5-3}$$

where i_{bd} is the time index at t_{bd} and i_{CH} is the time index when the prediction begins to remain unchanged. In order to maintain the accuracy, the prediction's error should be tied to a $\pm\alpha$-bound. Similar to the PH metric, the $\pm\alpha$-bound is set to 20% of total lifetime that is parallel to the TTF_{act} line.

All the plots in Figure 5-16 are replotted in Figure 5-18 but this time with the indication of i_{CH}. The exponential model works well on all E1 samples with average of CH, CH_{ave} = 0.6, but fail to comply with the error limit set by α in most of the samples under the E2 and E3 categories (CH = 0). Contrariwise, the linear model is only able to converge on three E1 samples with the CH score less than 0.2. This can be seen clearly in Figure 5-19 where the CH scores of all samples are plotted together to aid the comparison.

The inability of the linear algorithm to converge is due to the inconsistent behavior of the linear fit that is illustrated in Figure 5-8b. Unlike the exponential model, the m_{fit} and $du_{fit(bd)}$ of the linear curve keeps changing as time passes and thus causes the predictions to vary. This variation results in a TTF_{est} curve that is not parallel to the TTF_{act} curve around the t_{bd}. Among the four test groups, Group 1 yields the highest average score while the other three groups score similarly.

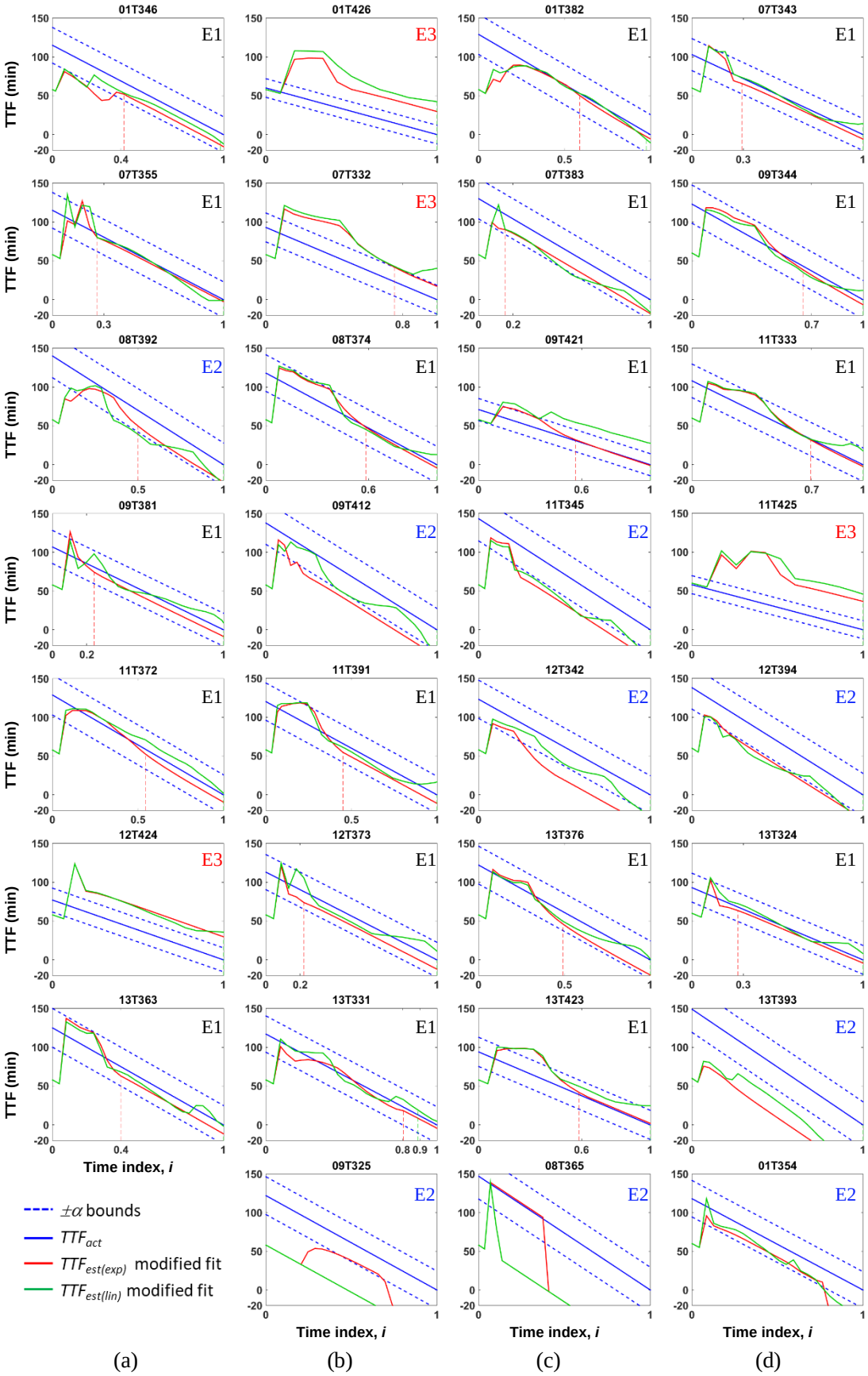


Figure 5-18. The plots of i_{CH} for determining the CH metric of the test samples in (a) Group 1, (b) Group 2, (c) Group 3, and (d) Group 4

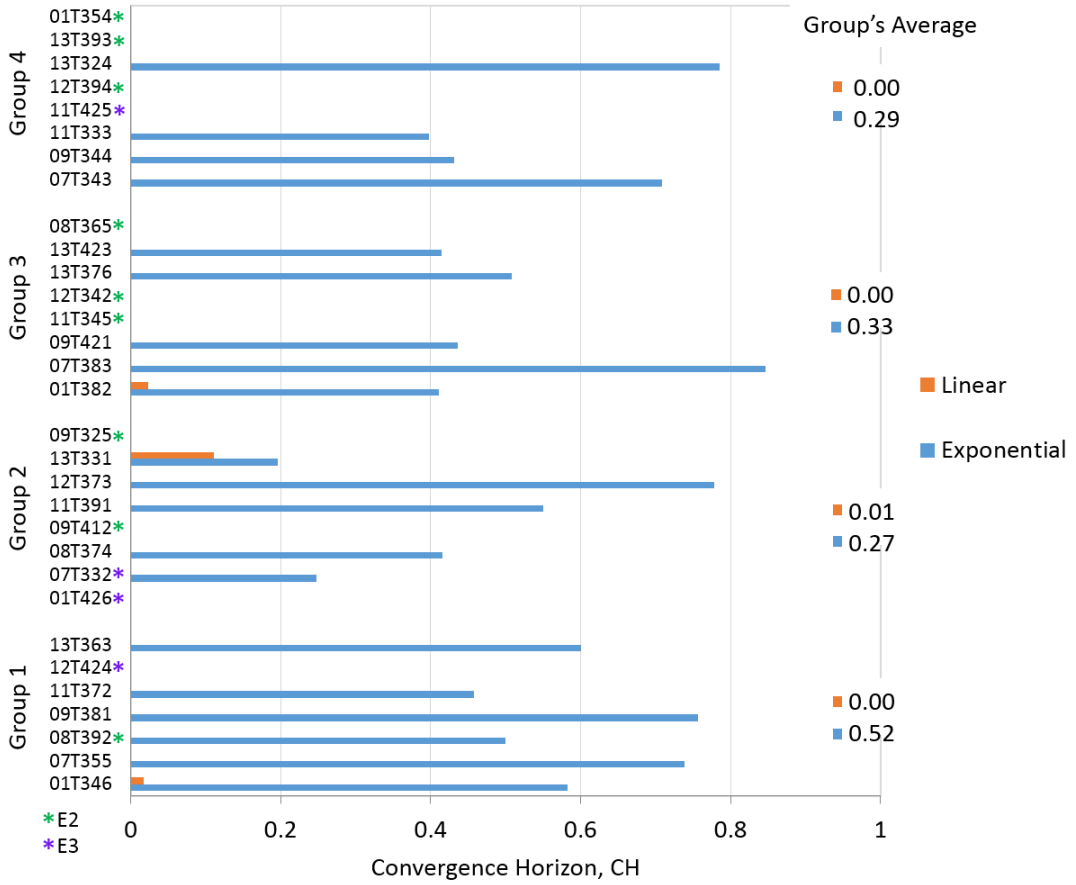


Figure 5-19. The score of CH metric

5.7.3 Relative Accuracy

Relative accuracy, RA is a quantitative measure of a prediction's error relative to the actual value and has been expressed in equation 3-29. In order to give a penalty to the predictions that are beyond the actual values, the equation is modified to equations 5-4 and 5-5, depending on the error of the predictions, e_λ at time index, $i = \lambda$. For the case of $TTF_{est}(i_\lambda) < TTF_{act}(i_\lambda)$, where $e_\lambda > 0$ (refer to equation 5-1), equation 3-29 is modified to the following using the appropriate notations used in this chapter.

$$e_\lambda > 0: \quad RA_\lambda = 1 - \frac{e_\lambda}{TTF_{act}(i_\lambda)} \tag{5-4}$$

There are two scenarios to be considered when $e_\lambda > 0$, which is $e_\lambda \leq TTF_{act}(i_\lambda)$ and $e_\lambda > TTF_{act}(i_\lambda)$. The first scenario could occur if $TTF_{est}(i_\lambda)$ lies between 0 and $TTF_{act}(i_\lambda)$ hence resulting in a positive RA_λ which ranges from 0 to 1. Meanwhile, the second scenario will take place if $TTF_{est}(i_\lambda) < 0$ (the prediction is made after the predicted failure) thus results in a negative RA_λ which is worse than the first case. For the case of $TTF_{est}(i_\lambda) > TTF_{act}(i_\lambda)$ in which $e_\lambda < 0$, equation 3-29 is modified to equation 5-5 so that the RA_λ is always negative.

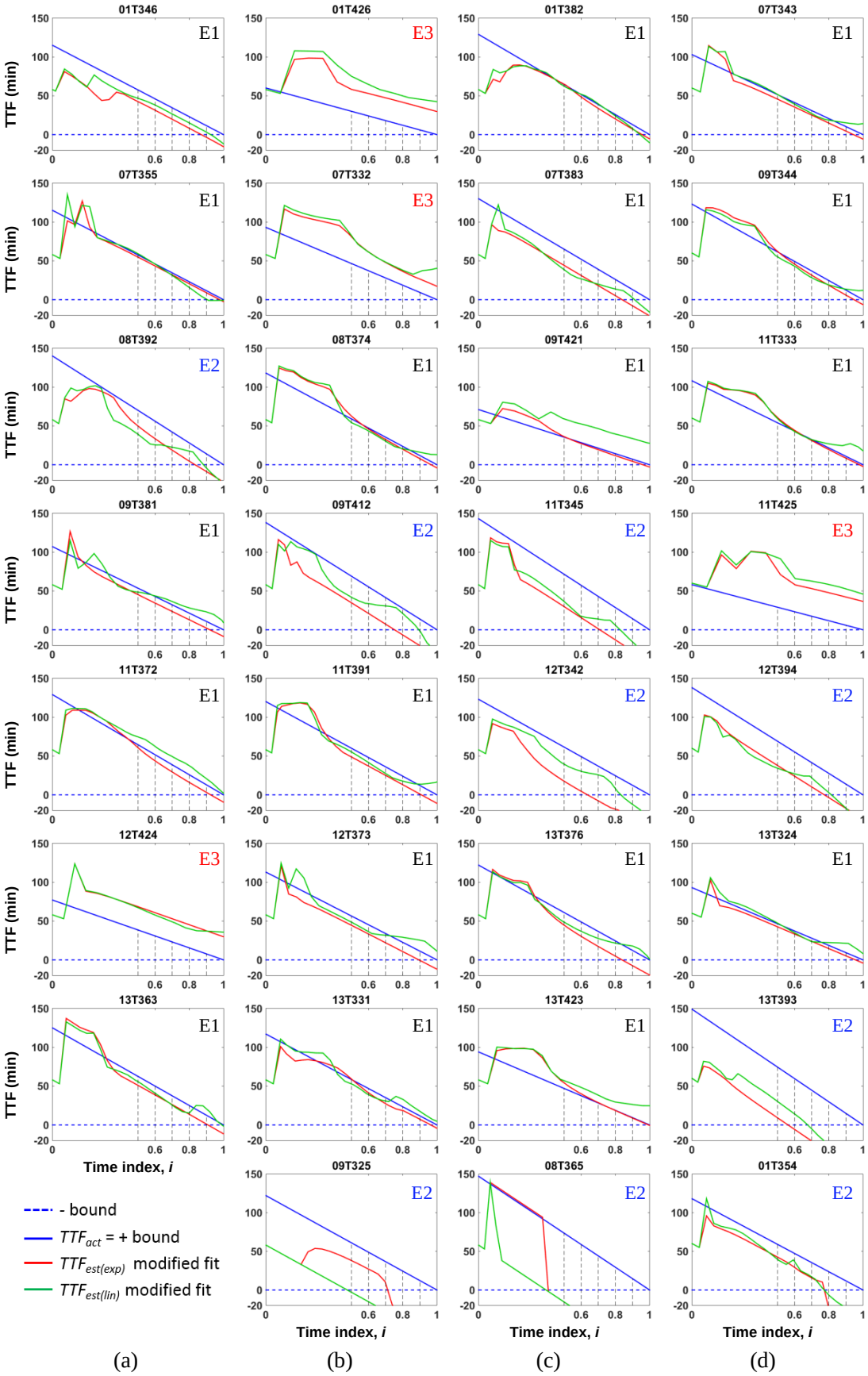
$$e_\lambda < 0: \quad RA_\lambda = \frac{e_\lambda}{TTF_{act}(i_\lambda)} \tag{5-5}$$

Therefore, there are three conditions to be considered i.e. $TTF_{est}(i_\lambda) < 0$, $0 \leq TTF_{est}(i_\lambda) \leq TTF_{act}(i_\lambda)$ and $TTF_{est}(i_\lambda) > TTF_{act}(i_\lambda)$. Only the second condition will give a positive score. The other two conditions result in a negative *RA* which has no boundary.

In this thesis, the RA metrics are evaluated for λ = 0.5, 0.6, 0.7, 0.8 and 0.9, where the predictions should be reliable. Equations 5-4 and 5-5 depict a measure at a specific time. Therefore, for multiple time instances, a normalized weighted sum of RA, that is, the cumulative relative accuracy, CRA is applied and can be expressed as

$$CRA_\lambda = \frac{\sum_{j=1}^{N} w_{\lambda_j} RA_{\lambda_j}}{\sum_{j=1}^{N} w_{\lambda_j}} \tag{5-6}$$

where w_{λ_j} is the weigh factor of the RA at λ_j. It is advised in [247] to give more weight to *RA* that is evaluated at times closer to t_{bd}. However, the right triangle shape error bound and negative RA score has actually given enough weight to the RA. Thus, the weigh factors, w_{λ_j}, for all λ values in this work are assigned to 1. The TTF_{est} of all λ are depicted in Figure 5-20 and the score at each λ can be seen in Figure 5-21 and Figure 5-22. If the RA_λ were evaluated for the exponential model at λ = 0.8, those three conditions mentioned above are equivalent to the E2, E1 and E3 categories respectively. However, this is not the case for other values of λ.



(a) (b) (c) (d)

Figure 5-20. The plots of i_λ for determining RA metric of test samples in (a) Group 1, (b) Group 2, (c) Group 3, and (d) Group 4

Both models perform the worst at λ = 0.9 where the number of samples with positive RA_λ is less than other λ values. The linear fit is much worse since only one sample results in a positive RA_λ. From 30 samples that have negative RA_λ, 18 of them has $TTF_{est}(i_\lambda) > TTF_{act}(i_\lambda)$ of which 14 are from E1 and the remaining 4 from E3. One of the reasons for the 14 samples of E1 to have negative RA_λ is the inconsistent behavior of the linear fit as discussed for the CH metric. Unlike the exponential model, the du_{all} values in Region 3 decrease the m_{fit} of the linear model thus causing the $TTF_{est}(i_\lambda)$ around the t_{bd} (in this case λ = 0.9) to become greater than the $TTF_{act}(i_\lambda)$.

In Figure 5-23, only the positive CRA_λ values are plotted for discussion. It can be seen clearly that the exponential model performs much better than the linear model. The exponential model results in positive CRA_λ for 19 samples which includes all the eighteen E1 samples and one from E2, while the linear model yields positive CRA_λ for 15 samples.

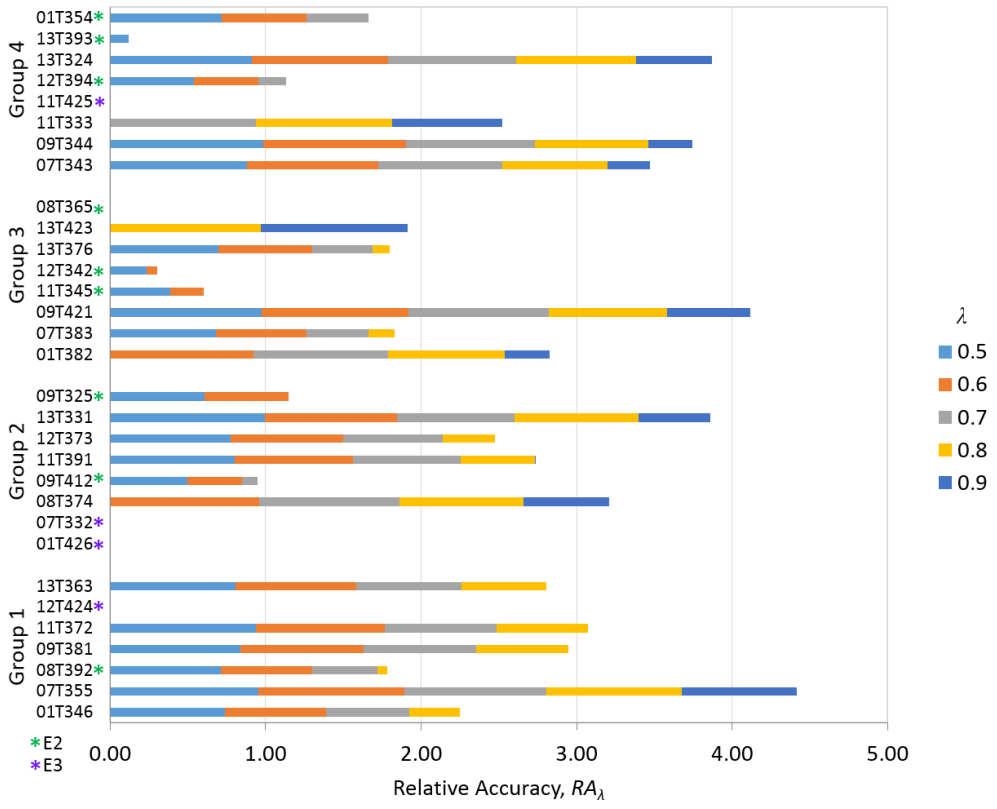


Figure 5-21. The score of RA of exponential model at λ = 0.5, 0.6, 0.7, 0.8, and 0.9

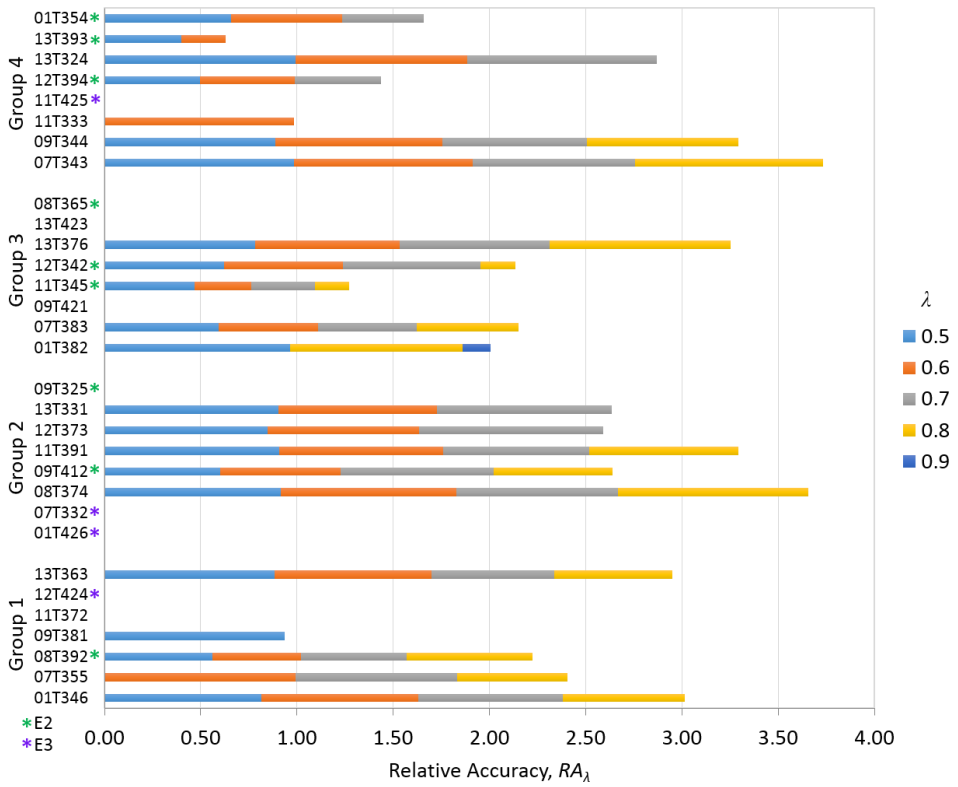


Figure 5-22. The score of RA of linear model at λ = 0.5, 0.6, 0.7, 0.8, and 0.9

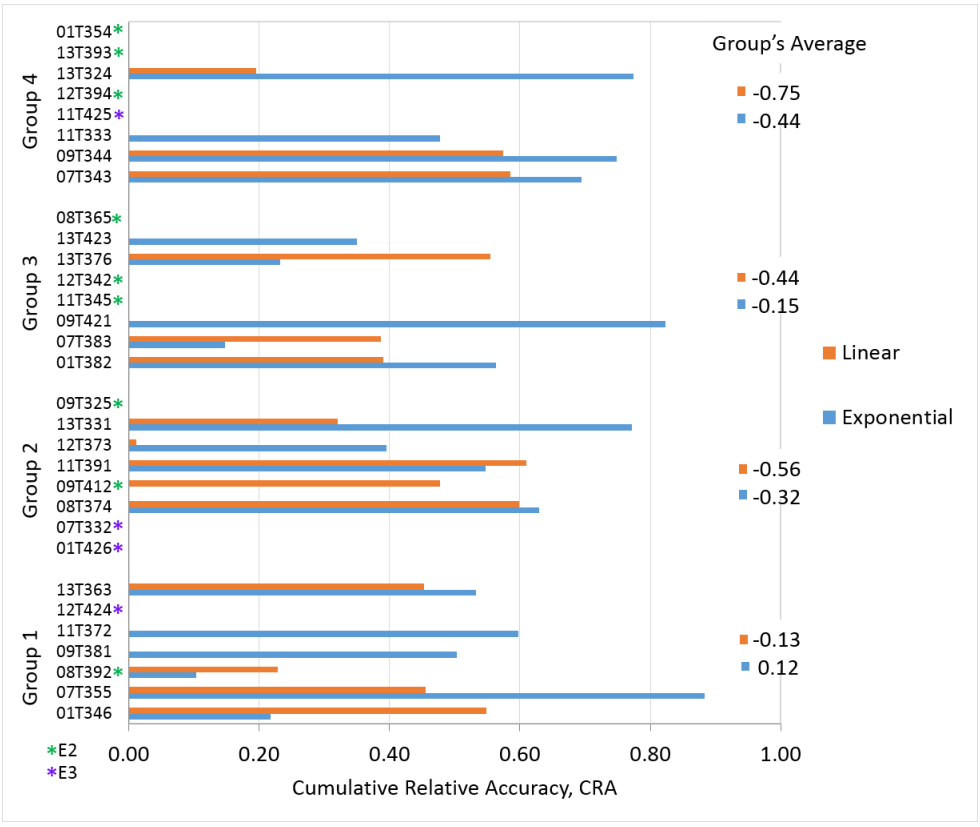


Figure 5-23. The plots of CRA metric

5.7.4 Overall Performance

The performance of the algorithms has been evaluated using three metrics, PH, CH and CRA. The values of all metrics are plotted in Figure 5-24 and Figure 5-25 for the exponential and linear models respectively. However, the overall performance score will exclude the PH metric since it only considers early predictions and does not reflect the entire performance. For instance, sample 12T394 in Figure 5-24 yields *PH* = 0.93 but 0 score for CH and CRA. Nonetheless, an algorithm with a very low score of the PH metric should be disregarded because the algorithm surely will not perform well in the other two metrics. This can be observed on the E3 samples 12T424, 01T426, 07T332 and 11T425 and the only E2 sample, 13T393.

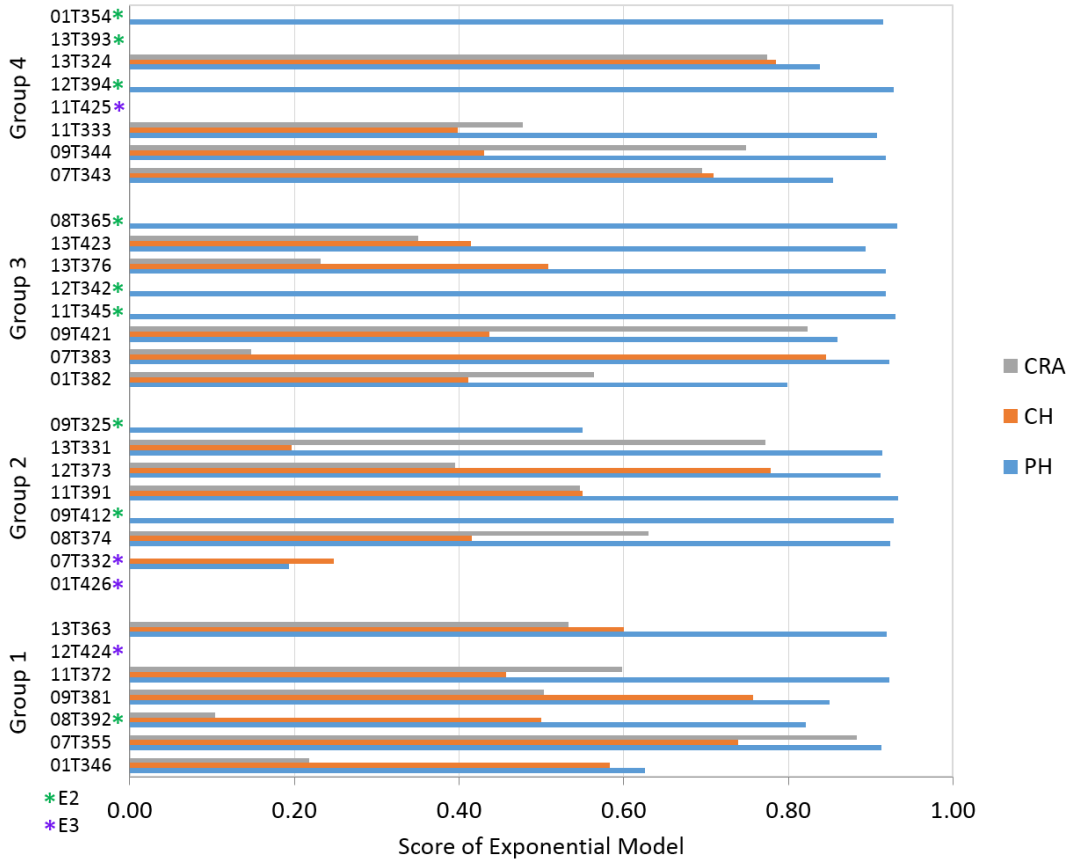


Figure 5-24. The scores of PH, CH and CRA metrics of exponential model

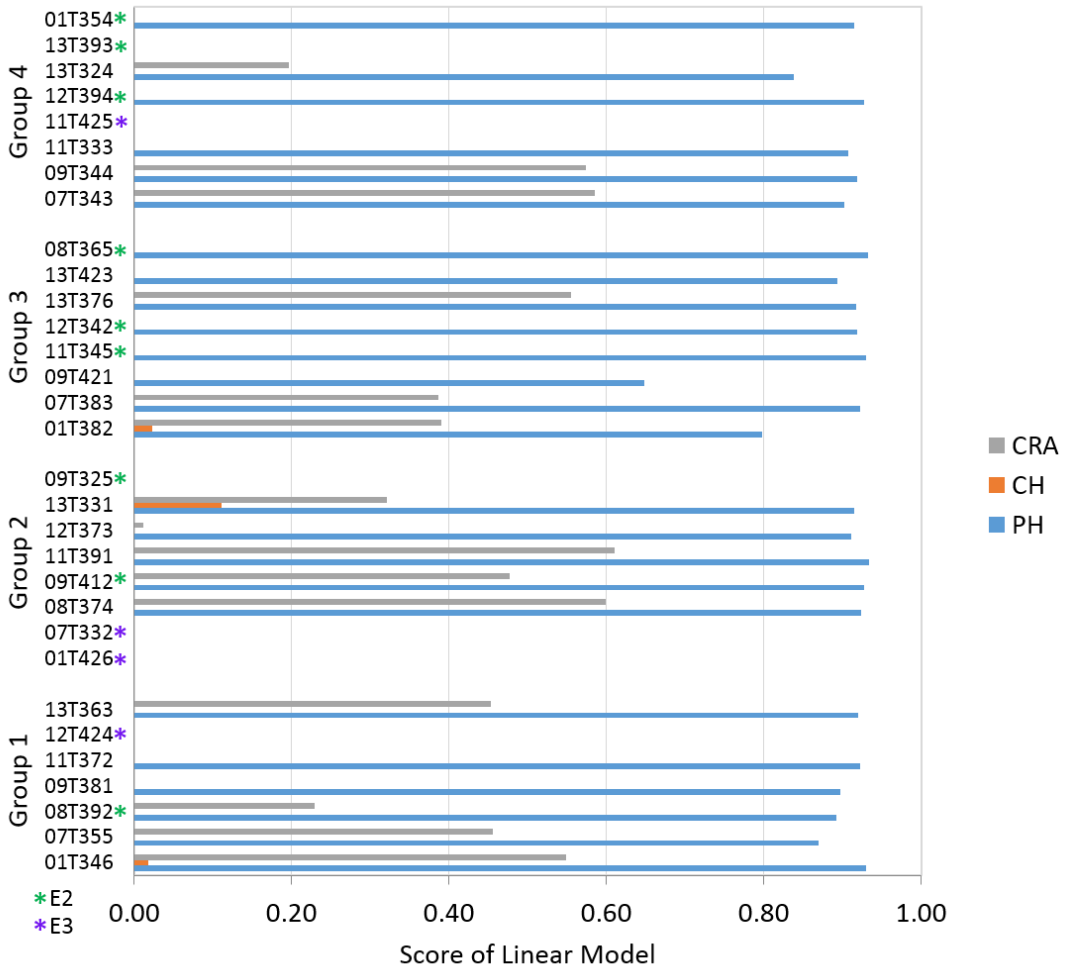


Figure 5-25. The scores of PH, CH and CRA metrics of linear model

The overall performance score is measured as the average of the CH and CRA metrics and is plotted in Figure 5-26. It is clearly shown that the exponential algorithm performs better than the linear. The low score of the linear model in the CH metric really affects its performance. Only samples that have positive CRA yield a positive average score.

The average score of each test group does not represent the real performance of the model. For the holdout method to be acceptable as validation, the test group should have samples from the three *e* categories, E1, E2 and E3. However this only applies to the exponential model since the nonconvergence behaviour of the linear model results in untrustworthy results. Still, the cross validation method assists with a more detailed analysis thus increasing the reliability of the results.

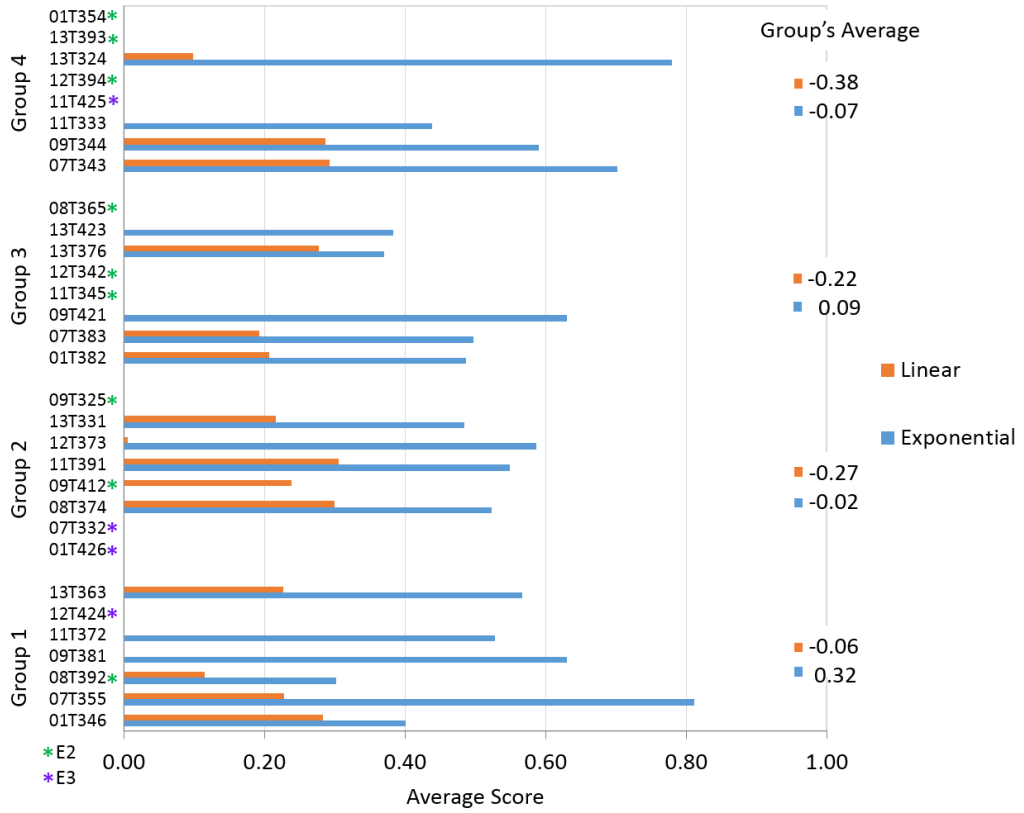


Figure 5-26. The overall score

5.8 Summary and Conclusion

In this chapter, a framework was proposed for the development of a prognostic model using a degradation parameter. From the PD analysis in the previous chapter, du_{all} has been chosen as the degradation parameter in this chapter. Before a model can be developed, any outliers were first identified and removed. Three outliers were recognised through the individual calculation of monotonicity and trendability metrics i.e. 01T354, 08T365 and 09T325. These samples were removed from the 24 training samples since they could poorly influence the properties of the model.

Based on the literature review, both exponential and inverse power models were considered for model fitting. Nonetheless, the exponential showed a better GOF thus was selected for model fitting. The semilog scale was applied rather than linear scale for simplicity. This offers another option of curve fitting; that is, linear regression. Both exponential and linear curves were applied to fit the du_{all} values of the training samples giving two important parameters i.e. m_{fit} and $\log_{10} du_{fit(bd)}$. An algorithm was

designed to reduce the variance of these parameters which also improves the properties of the model. The allowance for m_{fit} was determined and the average of $\log_{10} du_{fit(bd)}$ was regarded as the threshold point, du_{th} giving the properties of the exponential and linear models.

The model was validated by applying it to the test samples. The holdout and cross validation methods were applied and compared. Four test groups were formed to perform 4-fold cross validation. The resulting error, e, of the exponential algorithm can be categorised according to its range i.e. E1: $0 \leq e \leq 20$, E2: $e > 20$ and E3: $e < 0$. However, the error group does not applied to the linear algorithm due to its inconsistency in predictions. Three performance metrics were used to evaluate the performance of both algorithms i.e. PH, CH and CRA. The metrics showed that the exponential algorithm performs better than the linear algorithm.

The resulting error categories from the exponential model were related to the failure time, t_{bd}. The designed exponential algorithm performed the best for samples under the E1 category which has a moderate t_{bd} that ranges from 90 to 130 minutes. This means, the E1 samples mostly have $\log_{10} du_{fit(bd)}$ that is similar to the du_{th}. Consequently, the samples with t_{bd} greater than the E1 range will probably produce an E2 error. This is due to the $\log_{10} du_{fit(bd)}$ values that are smaller than du_{th}. Therefore, when the exponential fit is extrapolated to the du_{th}, the resulting TTF_{est} is much sooner than the TTF_{act} giving e that is greater than 20 minutes. For samples with t_{bd} lower than 90 minutes, the $\log_{10} du_{fit(bd)}$ is much greater than the du_{th} resulting in TTF_{est} that is beyond the TTF_{act}. Accordingly, all E1 samples yield positive scores for the PH, CH and CRA metrics.

On the other hand, the linear algorithm only performs well for the PH metric. The inconsistency in the $\log_{10} du_{fit(bd)}$ values was the reason for the algorithm to not perform well on the CH and CRA metrics.

Lastly, the holdout method can be applied to validate the results, with the condition that the test samples must represent the three error categories. Nevertheless, the CV method allowed more dependable analysis when all samples were treated as test samples.

Chapter 6

Conclusion and Future Work

6.1 Summary and Conclusions

Electrical treeing is a pre-breakdown phenomenon. It is the most likely mechanism of electrical failure in solid insulation. Monitoring of the progress of electrical tree growth is therefore essential. The works related to electrical tree growth modelling were reviewed in Chapter 2. Most of the developed models are physics-based, and relate the tree growth with the activity of discharges within the tree. Although there were evidences for return tree growth, no explicit consideration of the return tree has been made in those models. The discussion of the return tree is very limited, thus the mechanism remains unclear.

The electrical treeing data employed in this thesis showed that breakdown did not occur after the original tree traversed the insulation gap, but continued to grow in the opposite direction. Due to the unclear mechanism of return tree growth, this thesis considered a data-driven approach for building a prognostic model of the electrical treeing phenomenon using the associated PD data.

The common methods for PD analysis as well as extracted PD features were given in Chapter 2. Based on the nature of the employed data, this thesis considered the PRPDA and PSA approaches for extracting PD features in Chapter 4. Twenty-four

features were extracted from the former and seven features from the latter, giving a total of thirty-one features.

Chapter 3 detailed the methods for feature reduction, sample validation and algorithm performance. This chapter introduced three metrics that were proposed in [31] to characterise the suitability of a feature as a prognostic parameter: *monotonicity*, *prognosability*, and *trendability*. Four metrics presented in [247] were also introduced to evaluate the model algorithm: *Prognostic horizon* (PH), *α-λ performance*, *relative accuracy* (RA) and *convergence*.

The filter method is the simplest approach for feature reduction. However, the reviewed measures for the filter method in Chapter 3 were not suitable for prognostic purposes. Instead, this research work employed the three prognostic metrics as measurands for the filter method in Chapter 4. The total score of the three metrics was used to rank the thirty-one features and only the first ranked feature was considered for the prognostic model. The first ranked feature is du_{all}, that resulted in remarkably good scores for prognosability and trendability but scored only moderately for monotonicity. This highlights the contribution of this research work in selecting the best feature for prognostic modelling.

The investigation of the effect of THD and the harmonic order to the du_{all} behaviour revealed no deterministic relation. Therefore, all the test waveforms were regarded equally in the prognosis analysis.

Another contribution of the thesis was the identification of the characteristics of the du_{all} value that correspond to the electrical tree growth. It was shown in Chapter 5 that the du_{all} values can be divided into three regions (Figure 5-4). The du_{all} values were found to behave randomly in Region 1 due to the change of frequency and magnitude of the applied voltage. In Region 2, the du_{all} values show an exponentially declining pattern that represents the fast growth in the treeing propagation stage. The du_{all} values remain constant in Region 3 showing an intense PD activity during the slow growth in the treeing propagation stage as well as the

runaway stage. The consistent du_{all} values were the reason for the moderate score of the monotonicity metric.

The prognostic model reported in this thesis was developed according to a framework proposed in Figure 5-1 in Chapter 5. This contributes to a general framework that can be implemented for developing a prognostic model using a data-driven approach.

The curve fitting approach was used to fit the du_{all} values and both exponential and linear fitting were considered. The model fitting resulted in a wide range of failure values due to the different behaviour of du_{all} in Region 1 and 2. Therefore, an algorithm was designed to decrease the range of the failure values. As a final approach in the modelling, the average of the failure values was regarded as the threshold value of the model. Both exponential and linear models were validated using holdout and 4-cross validation techniques. The proposed algorithm has improved the resultant prediction errors in most of the samples shown in Figure 5-15.

The final contribution was the performance metrics to evaluate prognostic models. The errors between the predicted and actual failure time were evaluated using three performance metrics: *prognostic horizon* (PH), *convergence horizon* (CH) and *cumulative relative accuracy* (CRA). However, only the scores of CH and CRA were counted to rank the models.

The exponential model was found to perform better than the linear model. This is due to the du_{all} characteristic that is closely similar to the exponential curve rather than the linear curve. Nevertheless, the resultant errors showed that the proposed algorithm is only suitable for treeing samples that have a moderate actual failure time. Samples with longer failure time might experience a prediction that is "too early" while the predictions for samples with shorter failure time were most likely to be beyond the actual failure. This was due to the threshold value that was taken as the average of the failure values. Further investigation should be made to improve the error as discussed in the next section.

As a final point, it was appropriate to apply the holdout method to validate the model, since the results from holdout and cross validation were similar.

This thesis therefore, has proposed a new approach in estimating insulation failure time by using PD data as the degradation parameter. Considering the PD behaviour throughout the degradation process aids the understanding of the underlying mechanism. This is an advantage of this approach compared to failure probability approach using Weibull analysis that only employs the TTF data. With the progress of online PD monitoring, the practical deployment of prognostics for cable monitoring could be feasible when the issues related to pulse attenuation and noise data can be minimised.

6.2 Future Work

6.2.1 Treeing Samples with Longer Failure Time

All the samples used in this thesis experienced breakdown within 150 minutes after the inception of an electrical tree. There is a need to validate whether a longer failure time (more than 24 hours) will result in the same characteristics of du_{all} values. Furthermore, the treeing phenomenon that leads to insulation breakdown in electrical plant is obviously much longer than that.

Future work can repeat the framework proposed in this thesis to a new set of samples that has a longer failure time. This can be done by increasing the insulation gap between the needle tip and the ground electrode or reducing the magnitude of the supplied voltage. The work proposed in this thesis can be validated if the du_{all} values from the new sample set can be described similarly.

6.2.2 Prediction on the Touch Ground Time

This research work did not consider the prediction of the time-to-touch-ground, (TTG). For a lifetime that is less than 2.5 hours, and some of the TTG values being only a quarter of the TTF, predicting the TGT may be less beneficial.

However, the prediction of the touch ground time can be considered for samples with a longer failure time. Apart from that, the image recording system should be good enough to aid the identification. With a two-dimensional image, the identification of the ground line can be very difficult.

6.2.3 Hybrid Modelling Approach

In order to increase the validity of the predictions, both data-driven and physics-based approaches can be considered for implementing a hybrid approach. Previous studies relate the discharge activity in the tree channels with the growth of electrical tree. Therefore, the data-driven approach can also consider the growth rate of the tree. Most of the published electrical tree models were given in Section 2.8.4. However, none of the models mentioned the growth of return tree. Further clarification is needed on how these models treat the return tree. Embedding both approaches will give a new insight in understanding the treeing phenomenon.

6.2.4 Sample Size

With a larger sample size, there is a possibility to separate the samples into several classes according to the behavior of du_{all} e.g. the number of du_{all} values in each region or the m_{fit} values. Therefore, the threshold value can be determined separately for each class. In this way, the predictions will become more robust.

6.2.5 Practical Deployment

Most of the published work on PD diagnosis uses offline data. The work presented here might not be appropriate to online PD monitoring due to the differences in the applied stresses between on-line and off-line PD data. Therefore, towards a practical deployment, studies on prognostics of cables should use online PD data. However, online PD monitoring has been associated with issues of signal attenuation and noise. Minimisation of these issues should therefore increase encourage the research in insulation prognosis.

References

[1] K. Abdolall, G. L. Halldorson, and D. Green, "Condition Assessment and Failure Modes of Solid Dielectric Cables in Perspective," *IEEE Transactions on Power Delivery*, vol. 17, no. 1, pp. 18–24, 2002.

[2] S. Bahadoorsingh and S. M. Rowland, "A Framework Linking Knowledge of Insulation Aging to Asset Management," *IEEE Electrical Insulation Magazine*, vol. 24, no. 3, pp. 38–46, 2008.

[3] M. A. Martins *et al.*, "Power Transformer End-of-Life Assessment - Pracana Case Study *," *IEEE Electrical Insulation Magazine*, vol. 27, no. 6, pp. 15–26, 2011.

[4] A. K. S. Jardine, D. Lin, and D. Banjevic, "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[5] S. Alaswad and Y. Xiang, "A Review on Condition-Based Maintenance Optimization Models for Stochastically Deteriorating System," *Reliability Engineering and System Safety*, vol. 157, pp. 54–63, 2017.

[6] ISO55002, "Asset Management — Management Systems —Guidelines for the Application of ISO 55001." p. 32, 2014.

[7] J.-H. Shin and H.-B. Jun, "On Condition Based Maintenance Policy," *Journal of Computational Design and Engineering*, vol. 2, no. 2, pp. 119–127, 2015.

[8] A. Kelen and M. G. Danikas, "Evidence and Presumption in PD Diagnostics," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 780–795, 1995.

[9] R.-N. Wu and C.-K. Chang, "The Use of Partial Discharges as an Online Monitoring System for Underground Cable Joints," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1585–1591, 2011.

[10] P. H. F. Morshuis, G. C. Montanari, and L. Fornasari, "Partial Discharge Diagnostics — Critical Steps towards On-Line Monitoring," in *2014 IEEE PES T&D Conference and Exposition*, 2014, pp. 1–5.

[11] M. Wu, H. Cao, J. Cao, H.-L. Nguyen, J. B. Gomes, and S. P. Krishnaswamy, "An Overview of Sate-of-the-Art Partial Discharge Analysis Tchniques for Condition Monitoring," *IEEE Electrical Insulation Magazine*, vol. 31, no. 6, pp. 22–35, 2015.

[12] R. J. Densley, "Ageing Mechanisms and Diagnostics for Power Cables - An

Overview," *IEEE Electrical Insulation Magazine*, vol. 17, no. 1, pp. 14–22, 2001.

[13] L. A. Dissado and J. C. Fothergill, *Electrical Degradation and Breakdown in Polymers*. Peter Peregrinus Ltd, 1992.

[14] A. Cavallini, M. Conti, G. C. Montanari, C. Arlotti, and A. Contin, "PD Inference for the Early Detection of Electrical Treeing in Insulation Systems," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 11, no. 4, pp. 724–735, 2004.

[15] J. V Champion and S. J. Dodd, "Systematic and Reproducible Partial Discharge Patterns during Electrical Tree Growth in an Epoxy Resin," *Journal of Physics D: Applied Physics*, vol. 29, no. 3, pp. 862–868, 1996.

[16] R. Vogelsang, B. Fruth, T. Farr, and K. Fröhlich, "Detection of Electrical Tree Propagation by Partial Discharge Measurements," *European Transactions on Electrical Power*, vol. 15, no. 3, pp. 271–284, 2005.

[17] N. M. Chalashkanov, S. J. Dodd, L. A. Dissado, and J. C. Fothergill, "Pulse Sequence Analysis on PD Data from Electrical Trees in Flexible Epoxy Resins," in *2011 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, CEIDP*, 2011, pp. 776–779.

[18] S. Bahadoorsingh, "Asset Management and the Role of Power Quality on Electrical Treeing in Epoxy Resin," Ph.D. dissertation, The University of Manchester, United Kingdom, 2009.

[19] G. Chen and C. H. Tham, "Electrical Treeing Characteristics in XLPE Power Cable Insulation in Frequency Range Between 20 and 500 Hz," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 16, no. 1, pp. 179–188, 2009.

[20] L. A. Dissado, S. J. Dodd, J. V Champion, P. I. Williams, and J. M. Alison, "Propagation of Electrical Tree Structures in Solid Polymeric Insulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 4, no. 3, pp. 259–279, 1997.

[21] J. V Champion and S. J. Dodd, "An Approach to the Modelling of Partial Discharges in Electrical Trees," *Journal of Physics D: Applied Physics*, vol. 31, no. 18, pp. 2305–2314, 1998.

[22] S. J. Dodd, "A Deterministic Model for the Growth of Non-Conducting Electrical Tree Structures," *Journal of Physics D: Applied Physics*, vol. 36, no. 2, pp. 129–141, 2003.

[23] G. Bahder, T. Garrity, M. Sosnowski, R. Eaton, and C. Katz, "Physical Model of Electric Aging and Breakdown of Extruded Polymeric Insulated Power Cables," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 6, pp. 1379–1390, 1982.

[24] A. El-Zein, M. Talaat, and M. M. El Bahy, "A Numerical Model of Electrical Tree Growth in Solid Insulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 16, no. 6, pp. 1724–1734, 2009.

[25] R. Bozzo, F. Guastavino, M. Cacciari, A. Contin, and G. C. Montanari, "Stochastic Procedures for the Investigation of Tree Growth in Insulating

Materials for HV Applications," in *Conference Record of the 1994 IEEE International Symposium on Electrical Insulation*, 1994, pp. 269–272.

[26] G. C. Montanari, "Aging and Life Models for Insulation Systems Based on PD Detection," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 4, pp. 667–675, 1995.

[27] D. W. Auckland, J. M. Cooper, and B. R. Varlow, "Factors Affecting Electrical Tree Testing," *Science, Measurement and Technology, IEE Proceedings A*, vol. 139, no. 1, pp. 9–13, 1992.

[28] C. Mayoux, "Degradation of Insulating Materials under Electrical Stress," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 5, pp. 590–601, 2000.

[29] Q. Nie, Y. X. Zhou, X. L. Xing, and Y. S. Wang, "Breakdown Process of Solid Composite Insulation System with Defects under High Frequency," in *2008 Annual Report Conference on Electrical Insulation Dielectric Phenomena*, 2008, pp. 552–554.

[30] R. Sarathi, A. Nandini, and T. Tanaka, "Understanding Electrical Treeing Phenomena in XLPE Cable Insulation under Harmonic AC Voltages Adopting UHF Technique," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 3, pp. 903–909, 2012.

[31] J. B. Coble, "Merging Data Sources to Predict Remaining Useful Life-An Automated Method to Identify Prognostic Parameters," Ph.D. dissertation, University of Tennesse, TN, USA, 2010.

[32] C. A. Harper, *Handbook of Plastics, Elastomers, and Composites*, Fourth. Mc Graw-Hill, 2002.

[33] N. R. Council, *Polymer Science and Engineering: The Shifting Research Frontiers*. Washington, DC: The National Academies Press, 1994.

[34] J. S. T. Looms, *Insulators for High Voltages*. United Kingdom: Peter Peregrinus Ltd, 1988.

[35] M. Amin, M. Akbar, and M. N. Khan, "Aging Investigations of Polymeric Insulators: Overview and Bibliography," *IEEE Electrical Insulation Magazine*, vol. 23, no. 4, pp. 44–50, 2007.

[36] K. N. Mathes, "A Brief History of Development in Electrical Insulation," in *Electrical Electronics Insulation Conference, 1991. Boston'91 EEIC/ICWA Exposition, Proceedings of the 20th*, 1991, pp. 147–150.

[37] G. Teyssedre and C. Laurent, "Advances in High-Field Insulating Polymeric Materials over the Past 50 Years," *IEEE Electrical Insulation Magazine*, vol. 29, no. 5, pp. 26–36, 2013.

[38] A. S. Vaughan, C. D. Green, I. L. Hosier, G. C. Stevens, A. Pye, and J. L. Thomas, "Thermoplastic High Performance Cable Insulation Systems for Flexible System Operation," in *Electrical Insulation Conference (EIC)*, 2015, no. June, pp. 543–546.

[39] T. Andritsch, "Epoxy Based Nanodielectrics for High Voltage DC Applications: Synthesis, Dielectric Properties and Space Charge Dynamics,"

Delft University of Technology, 2010.

[40] T. Tanaka and T. Imai, “Advances in Nanodielectric Materials over the Past 50 Years,” *IEEE Electrical Insulation Magazine*, vol. 1, no. 29, pp. 10–23, 2013.

[41] S. Singha and M. Thomas, “Dielectric Properties of Epoxy Nanocomposites,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 1, pp. 12–23, 2008.

[42] T. W. Dakin, “Application of Epoxy Resins in Electrical Apparatus,” *IEEE Transactions on Electrical Insulation*, vol. EI-9, no. 4, pp. 121–128, 1974.

[43] E. Tuncer, “Dielectric Mixtures - Importance and Theoretical Approaches,” *IEEE Electrical Insulation Magazine*, vol. 29, no. 6, pp. 49–58, 2013.

[44] A. C. Gjaerde, “Multifactor Ageing Models-Origin and Similarities,” *IEEE Electrical Insulation Magazine*, vol. 13, no. 1, pp. 6–13, 1997.

[45] V. K. Agarwal *et al.*, “The Mysteries of Multifactor Ageing,” *IEEE Electrical Insulation Magazine*, vol. 11, no. 3, pp. 37–43, 1995.

[46] A. Kelen, “Ageing of Insulating Materials and Equipment Insulation in Service and in Tests,” *IEEE Transactions on Electrical Insulation*, vol. EI-12, no. 1, pp. 55–60, 1977.

[47] R. A. Jongen, P. H. F. Morshuis, J. J. Smit, A. L. J. Janssen, and E. Gulski, “Failure Analysis of in Service Failed Resin Cable Joints by Means of a Statistical Approach,” *2006 Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, CEIDP*, pp. 517–520, 2006.

[48] G. C. Montanari and L. Simoni, “Aging Phenomenology and Modeling,” *IEEE Transactions on Electrical Insulation*, vol. 28, no. 5, pp. 755–776, 1993.

[49] IEEE1064, “IEEE Guide for Multifactor Stress Functional Testing of Electrical Insulation Systems - Withdrawn.” 1991.

[50] IEC60505, “Evaluation and Qualification of Electrical Insulation Systems - Withdrawn,” Geneva, 2004.

[51] E. O. Forster *et al.*, “Research Needs to Assess the Long-Term Performance of Electrical Insulating Materials and Systems,” in *Task Force on Aging Electrical Insulation*, 1977.

[52] R. J. Densley, “Ageing and Diagnostics in Extruded Insulations for Power Cables,” in *IEEE 5th International Conference on Conduction and Breakdown in Solid Dielectrics*, 1995, pp. 1–15.

[53] R. J. Densley, R. Bartnikas, and B. S. Bernstein, “Multiple Stress Aging of Solid-Dielectric Extruded Dry-Cured Insulation Systems for Power Transmission Cables,” *IEEE Transactions on Power Delivery*, vol. 9, no. 1, pp. 559–571, 1994.

[54] L. Simoni, G. Mazzanti, G. C. Montanari, and L. Lefebre, “A General Multi-Stress Life Model for Insulating Materials with or without Evidence for Thresholds,” *IEEE Transactions on Electrical Insulation*, vol. 28, no. 3, pp. 349–364, 1993.

[55] P. Cygan and J. R. Laghari, “Models for Insulation Aging under Electrical and Thermal Multistress,” *IEEE Transactions on Electrical Insulation*, vol. 25, no. 5, pp. 923–934, 1990.

[56] T. S. Ramu, “On the Estimation of Life of Power Apparatus Insulation under Combined Electrical and Thermal Stress,” *IEEE Transactions on Electrical Insulation*, vol. EI-20, no. 1, pp. 70–78, 1985.

[57] P. Paloniemi, “Multi-Stress Endurance Tests on High-Voltage Motor Insulations, with Equal Acceleration of Each Stress,” *IEEE Transactions on Electrical Insulation*, vol. EI-17, no. 3, pp. 253–261, 1982.

[58] Y. Kako, K. Kadotani, and T. Tsukui, “An Analysis of Combined Stress Degradation of Rotating Machine Insulation,” *IEEE Transactions on Electrical Insulation*, vol. EI-18, no. 6, pp. 642–650, 1983.

[59] T. Tanaka, “Aging of Polymeric and Composite Insulating Materials - Aspects of Interfacial Performance in Aging,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 5, pp. 704–716, 2002.

[60] A. T. Bulinski, S. S. Bamji, and R. J. Densley, “Factors Affecting the Transition from Water Tree to an Electrical Tree,” in *Conference Record of the 1988 IEEE International Symposium on Electrical Insulation*, 1988, pp. 327–330.

[61] G. C. Montanari and G. Mazzanti, “Ageing of Polymeric Insulating Materials and Insulation System Design,” *Polymer International*, vol. 51, no. 11, pp. 1151–1158, 2002.

[62] T. Takada, “Space Charge Formation in Dielectrics,” *IEEE Transactions on Electrical Insulation*, vol. EI-21, no. 6, pp. 873–879, 1986.

[63] G. Chen, Y. Tanaka, T. Takada, and L. Zhong, “Effect of Polyethylene Interface on Space Charge Formation,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 11, no. 1, pp. 113–121, 2004.

[64] K. Y. Lau, A. S. Vaughan, G. Chen, I. L. Hosier, A. F. Holt, and K. Y. Ching, “On the Space Charge and DC Breakdown Behavior of Polyethylene/Silica Nanocomposites,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 1, pp. 340–351, 2014.

[65] G. Mazzanti, G. C. Montanari, and L. A. Dissado, “A Space-charge Life Model for Aging of Polymers,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 6, no. 6, pp. 864–875, 1999.

[66] L. Lan, J. Wu, Y. Yin, X. Li, and Z. Li, “Effect of Temperature on Space Charge Trapping and Conduction in Cross-Linked Polyethylene,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 4, pp. 1784–1791, 2014.

[67] S. S. Bamji, A. T. Bulinski, and M. Abou-Dakka, “Luminescence and Space Charge in Polymeric Dielectrics,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 16, no. 5, pp. 1376–1392, 2009.

[68] M. E. Sepúlveda-garcía, J. M. Martínez-Tarifa, and J. Sanz-Feito, “Electrical Ageing Markers for Polyethylene Insulation Based on Space Charge Accumulation and Apparent Mobility,” *IEEE Transactions on Dielectrics and*

Electrical Insulation, vol. 20, no. 6, pp. 2222–2229, 2013.

[69] X. Wang, N. Yoshimura, Y. Tanaka, K. Murata, and T. Takada, "Space Charge Characteristics in Cross-Linking Polyethylene under Electrical Stress from DC to Power Frequency," *Journal of Physics D: Applied Physics*, vol. 31, no. 16, p. 2057, 1998.

[70] T. Takada, "Acoustic and Optical Methods for Measuring Electric Charge Distributions in Dielectrics," in *Conference on Electrical Insulation and Dielectric Phenomena*, 1999.

[71] A. See, J. C. Fothergill, L. A. Dissado, and J. M. Alison, "Measurement of Space-Charge Distributions in Solid Insulators Under Rapidly Varying Voltage Using the High-Voltage, High-Speed Pulsed Electro-Acoustic (PEA) Apparatus," *Measurement Science and Technology*, vol. 12, no. 8, pp. 1227–1234, 2001.

[72] G. Chen, M. Fu, X. Z. Liu, and L. S. Zhong, "AC Aging and Space-Charge Characteristics in Low-Density Polyethylene Polymeric Insulation," *Journal of Applied Physics*, vol. 97, no. 8, 2005.

[73] Y. L. Chong, G. Chen, H. Miyake, K. Matsui, Y. Tanaka, and T. Takada, "Space Charge and Charge Trapping Characteristics of Cross-Linked Polyethylene Subjected to AC Electric Stresses," *Journal of Physics D: Applied Physics*, vol. 39, no. 8, pp. 1658–1666, 2006.

[74] L. A. Dissado, "Understanding Electrical Trees in Solids: From Experiment to Theory," in *IEEE Transactions on Dielectrics and Electrical Insulation*, 2002, vol. 9, no. 4, pp. 483–497.

[75] IEC60270, "High-Voltage Test Techniques - Partial Discharge Measurements." 2000.

[76] J. H. Mason, "Enhancing the Significance of PD Measurements," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 876–888, 1995.

[77] S. A. Boggs, "Partial Discharge - Part III: Cavity-Induced PD in Solid Dielectrics," *IEEE Electrical Insulation Magazine*, vol. 6, no. 6, 1990.

[78] P. H. F. Morshuis, "Degradation of Solid Dielectrics due to Internal Partial Discharge: Some Thoughts on Progress Made and Where to go Now," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 5, pp. 905–913, 2005.

[79] A. Krivda, "Automated Recognition of Partial Discharges," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 796–821, 1995.

[80] E. Lemke *et al.*, "Guide for Electrical Partial Discharge Measurements on Compliance to IEC 60270," *Electra*, no. 241, pp. 61–67, 2008.

[81] S. Karmakar, N. K. Roy, and P. Kumbhakar, "Detection of Partial Discharges in a High Voltage Equipment," *Journal of Electrical Engineering*, vol. 9, no. 2, pp. 26–31, 2006.

[82] E. Lemke *et al.*, "Practical Aspects of the Detection and Location of Partial

Discharges in Power Cables," *CIGRE WG D1*, vol. 33, pp. 63–69, 2006.

[83] CIGRE, "Guide for Partial Discharge Measurements in Complicance to IEC 60270," in *Working Group D1.33*, 2008, no. December.

[84] M. D. Judd, S. D. J. McArthur, J. R. McDonald, and O. Farish, "Intelligent Condition Monitoring and Asset Management. Partial Discharge Monitoring for Power Transformers," *Power Engineering Journal*, vol. 16, no. 6, pp. 297–304, 2002.

[85] E. Lemke, T. Strehl, S. Markalous, and W. Weissenberg, "Ultra-Wide-Band PD Diagnostics of Power Cable Terminations in Service," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 6, pp. 1570–1575, 2008.

[86] A. G. Sellars, O. Farish, B. F. Hampton, and L. S. Pritchard, "Using the UHF Technique to Investigate PD Produced by Defects in Solid Insulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 3, pp. 448–459, 1995.

[87] W. M. F. Al-Masri, M. F. Abdel-Hafez, and A. H. El-Hag, "Toward High-Accuracy Estimation of Partial Discharge Location," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 9, pp. 2145–2153, 2016.

[88] I. J. Kemp, "Partial Discharge Plant-Monitoring Technology: Present and Future Developments," *IEE Proceedings - Science, Measurement and Technology*, vol. 142, no. 1, p. 4, 1995.

[89] IEC62478, "High-Voltage Test Techniques: Measurement of Partial Discharge by Electromagnetic and Acoustic Methods," 2016.

[90] N. Bakar, A. Abu-Siada, and S. Islam, "A Review of Dissolved Gas Analysis Measurement and Interpretation Techniques," *IEEE Electrical Insulation Magazine*, vol. 30, no. 3, pp. 39–49, 2014.

[91] E. Gulski and F. H. Kreuger, "Computer-Aided Analysis of Discharge Patterns," *Journal of Physics D: Applied Physics*, vol. 23, no. 12, pp. 1569–1575, 1990.

[92] S. M. Strachan, S. E. Rudd, S. D. J. McArthur, M. D. Judd, S. Meijer, and E. Gulski, "Knowledge-Based Diagnosis of Partial Discharges in Power Transformers," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 1, pp. 259–268, 2008.

[93] S. E. Rudd, S. D. J. McArthur, and M. D. Judd, "A Generic Knowledge-Based Approach to the Analysis of Partial Discharge Data," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 17, no. 1, pp. 149–156, 2010.

[94] T. Okamoto and T. Tanaka, "Novel Partial Discharge Measurement Computer-Aided Measurement Systems," *IEEE Transactions on Electrical Insulation*, vol. EI-21, no. 6, pp. 1015–1019, 1986.

[95] M. Hoof and R. Patsch, "Analyzing Partial Discharge Pulse Sequences-A New Approach to Investigate Degradation Phenomena," in *Conference Record of the 1994 IEEE International Symposium on Electrical Insulation*, 1994, pp. 327–331.

[96] R. Patsch and F. Berton, "Pulse Sequence Analysis - A Diagnostic Tool Based

on the Physics Behind Partial Discharges," *Journal of Physics D: Applied Physics*, vol. 35, no. 1, p. 25, 2002.

[97] Suwarno, H. Ichikawa, Y. Suzuoki, T. Mizutani, and K. Uchida, "Partial Discharge Patterns of Electrical Treeing in Polyethylene," in *4th International Conference on Properties and Applications of Dielectric Materials*, 1994, pp. 379–382.

[98] Suwarno, Y. Suzuoki, F. Komori, and T. Mizutani, "Partial Discharges due to Electrical Treeing in Polymers: Phase-Resolved and Time-Sequence Observation and Analysis," *Journal of Physics D: Applied Physics*, vol. 29, no. 11, p. 2922, 1996.

[99] K. Wu, Y. Suzuoki, T. Mizutani, and H. Xie, "Model for Partial Discharges Associated with Treeing Breakdown: I. PDs in Tree Channels," *Journal of Physics D: Applied Physics*, vol. 33, no. 10, pp. 1197–1201, 2000.

[100] E. Gulski, "Discharge Pattern Recognition in High Voltage Equipment," *IEE Proceedings - Science, Measurement and Technology*, vol. 142, no. 1, pp. 51–61, 1995.

[101] L. Hao and P. L. Lewin, "Phase Resolved Partial Discharge Identification using a Support Vector Machine," in *23rd IAR Workshop on Advanced Control and Diagnosis*, 2008, pp. 386–391.

[102] B. Sheng *et al.*, "A Novel On-Line Cable PD Localisation Method based on Cable Transfer Function and Detected PD Pulse Rise-Time," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 4, pp. 2087–2096, 2015.

[103] Y. Gao, B. X. Du, and T. Han, "Effect of Defect Size on PD Pulse Characteristics Measured with High Frequency Current Transformer in XLPE Insulation," in *Proceedings of 2011 International Symposium on Electrical Insulating Materials*, 2011, pp. 173–176.

[104] N. C. Sahoo, M. M. A. Salama, and R. Bartnikas, "Trends in Partial Discharge Pattern Classification: A Survey," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 2, pp. 248–264, 2005.

[105] P. H. F. Morshuis and F. H. Kreuger, "A Relation between Time-Resolved Discharge Parameters and Ageing," in *Sixth International Conference on Dielectric Materials, Measurements and Applications*, 1992, pp. 37–40.

[106] H. R. Mirzaei, A. Akbari, M. Kharezi, and A. M. Jafari, "Experimental Investigation of PD Diagnosis based on Consecutive Pulses Data," in *International Conference on Condition Monitoring and Diagnosis, CMD 2008*, 2008, pp. 1262–1265.

[107] I. D. Bruce, B. G. Stewart, and C. Zhou, "An Examination of the Effectiveness of Pulse Sequence Analysis on PD Data from Artificial Void Defects in PET," in *47th International Universities Power Engineering Conference (UPEC)*, 2012, pp. 1–6.

[108] R. Patsch, J. Menzel, and D. Benzerouk, "The Analysis of PD-Sequences from Different Defects," in *Proceedings of the IEEE International Conference on Properties and Applications of Dielectric Materials*, 2007, pp. 31–35.

[109] E. Gulski, “Computer-Aided Measurement of Partial Discharges in HV Equipment,” *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 969–983, 1993.

[110] I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection,” *Journal of Machine Learning Research (JMLR)*, vol. 3, no. 3, pp. 1157–1182, 2003.

[111] E. Gulski, “Digital Analysis of Partial Discharges,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 822–837, 1995.

[112] S. E. Rudd, “Knowledge-Based Analysis of Partial Discharge Data,” Ph.D. dissertation, University of Strathclyde, 2010.

[113] Warwick Manufacturing Group, *The Use Of Weibull In Defect Data Analysis*. 2007.

[114] G. C. Montanari and M. Cacciari, “A Probabilistic Insulation Life Model for Combined Thermal-Electrical Stresses,” *IEEE Transactions on Electrical Insulation*, vol. EI-20, no. 3, pp. 519–522, 1985.

[115] M. Cacciari, A. Contin, and G. C. Montanari, “Use of a Mixed-Weibull Distribution for the Identification of PD Phenomena,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 6, pp. 1166–1179, 1995.

[116] R. Bozzo, C. Gemme, F. Guastavino, M. Cacciari, A. Contin, and G. C. Montanari, “Aging Diagnosis of Insulation Systems by PD Measurements: Extraction of Partial Discharge Features in Electrical Treeing,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 5, no. 1, pp. 118–124, 1998.

[117] A. Contin, G. C. Montanari, and C. Ferraro, “PD Source Recognition by Weibull Processing of Pulse Height Distributions,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 1, pp. 48–58, 2000.

[118] Y. Han and Y. H. Song, “Using Improved Self-Organizing Map for Partial Discharge Diagnosis of Large Turbogenerators,” *IEEE Transactions on Energy Conversion*, vol. 18, no. 3, pp. 392–399, 2003.

[119] M. K. Abdul Rahman, R. Arora, and S. C. Srivastava, “Partial Discharge Classification using Principal Component Transformation,” *IEE Proceedings - Science, Measurement and Technology*, vol. 147, no. 1, p. 7, 2000.

[120] L. Satish and W. S. Zaengl, “Can Fractal Features be Used for Recognizing 3D Partial Discharge Patterns?,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 3, pp. 352–359, 1995.

[121] F. C. Gu, H. C. Chang, and C. C. Kuo, “Gas-Insulated Switchgear PD Signal Analysis Based on Hilbert-Huang Transform with Fractal Parameters Enhancement,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 4, pp. 1049–1055, 2013.

[122] H.-C. Chen, “Fractal Features-Based Partial Discharge Pattern Recognition using Extension Method,” *Proceedings - 2012 International Conference on Intelligent Systems Design and Engineering Applications, ISDEA 2012*, vol. 6, no. February, pp. 274–277, 2012.

[123] E. Gulski, H. P. Burger, A. Zielonka, and R. Brooks, “Classification of Defects in HV Components by Fractal Analysis of PD Measurements,” *Proceedings of Conference on Electrical Insulation and Dielectric Phenomena - CEIDP ’96*, vol. 2, pp. 484–487, 1996.

[124] L. Satish, E. Gulski, and W. S. Zaengl, “The Use of Fractal Features for Recognition of 3-D Discharge Patterns,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 889–892, 1995.

[125] E. M. Lalitha and L. Satish, “Wavelet Analysis for Classification of Multi-Source PD Patterns,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 1, pp. 40–47, 2000.

[126] H.-G. Kranz, “Diagnosis of Partial Discharge Signals using Neural Networks and Minimum Distance Classification,” *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 1016–1024, 1993.

[127] W. Ziomek, M. Reformat, and E. Kuffel, “Application of Genetic Algorithms to Pattern Recognition of Defects in GIS,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 2, pp. 161–168, 2000.

[128] S. Zhang *et al.*, “Improving Recognition Accuracy of Partial Discharge Patterns by Image-Oriented Feature Extraction and Selection Technique,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, pp. 1076–1087, 2016.

[129] H. Janani, N. D. Jacob, and B. Kordi, “Automated Recognition of Partial Discharge in Oil-Immersed Insulation,” in *2015 Electrical Insulation Conference (EIC)*, 2015, no. June, pp. 467–470.

[130] Z. Ma, C. Zhou, D. M. Hepburn, and K. Cowan, “Fractal Theory Based Pattern Recognition of Motor Partial Discharge,” in *2016 International Conference on Condition Monitoring and Diagnosis*, 2016, pp. 898–901.

[131] E. M. Lalitha and L. Satish, “Fractal Image Compression for Classification of PD Sources,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 5, no. 4, pp. 550–557, 1998.

[132] C. Cachin and H. J. Wiesmann, “PD Recognition with Knowledge-Based Preprocessing and Neural Networks,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 4, pp. 578–589, 1995.

[133] L. Li, J. Tang, and Y. Liu, “Partial Discharge Recognition in Gas Insulated Switchgear Based on Multi-Information Fusion,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 2, pp. 1080–1087, 2015.

[134] L. Satish and B. I. Gururaj, “Partial Discharge Pattern Classification using Multilayer Neural Networks,” *IEE Proceedings A Science, Measurement and Technology*, vol. 140, no. 4, p. 323, 1993.

[135] H. Ma, J. C. Chan, T. K. Saha, and C. Ekanayake, “Pattern Recognition Techniques and Their Applications for Automatic Classification of Artificial Partial Discharge Sources,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 2, pp. 468–478, 2013.

[136] K. Wang *et al.*, “A New Image-Oriented Feature Extraction Method for Partial Discharges,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol.

22, no. 2, pp. 1015–1024, 2015.

[137] K. X. Lai, B. T. Phung, and T. R. Blackburn, “Descriptive Data Mining of Partial Discharge using Decision Tree with Genetic Algorithm,” in *2008 Australasian Universities Power Engineering Conference*, 2008, pp. 1–6.

[138] H. Hirose, M. Hikita, S. Ohtsuka, S.-I. Tsuru, and J. Ichimaru, “Diagnosis of Electric Power Apparatus using the Decision Tree Method,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 5, pp. 1252–1260, 2008.

[139] S. A. Boggs and J. Xu, “Water Treeing-Filled Versus Unfilled Cable Insulation,” *IEEE Electrical Insulation Magazine*, vol. 17, no. 1, pp. 23–29, 2001.

[140] A. T. Bulinski, J.-P. Crine, B. Noirhomme, R. J. Densley, and S. S. Bamji, “Polymer Oxidation and Water Treeing,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 5, no. 4, pp. 558–570, 1998.

[141] R. M. Eichhorn, “Treeing in Solid Extruded Electrical Insulation,” *IEEE Transactions on Electrical Insulation*, vol. EI-12, no. 1, pp. 2–18, 1976.

[142] M. H. Abderrazzaq, “Development of Water Tree Structure in Polyester Resin,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 1, pp. 158–165, 2005.

[143] C. Mayoux, “Aging of Polymeric Insulating Materials in Power Cables,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 4, no. 6, pp. 665–673, 1997.

[144] E. Gulski, H. Putter, and J. J. Smit, “Investigation of Water Treeing - Electrical Treeing Transition in Power cables,” in *2008 International Conference on Condition Monitoring and Diagnosis (CMD)*, 2008, pp. 234–237.

[145] J. V Champion and S. J. Dodd, “The Effect of Voltage and Material Age on the Electrical Tree Growth and Breakdown Characteristics of Epoxy Resins,” *Journal of Physics D: Applied Physics*, vol. 28, no. 2, pp. 398–407, 1995.

[146] J. H. Mason, “The Deterioration and Breakdown of Dielectrics Resulting from Internal Discharges,” *Proceedings of the IEE - Part I: General*, vol. 98, no. 109, pp. 44–59, 1951.

[147] T. Baumann, B. Fruth, F. Stucki, and H. R. Zeller, “Field-Enhancing Defects in Polymeric Insulators Causing Dielectric Aging,” *IEEE Transactions on Electrical Insulation*, vol. 24, no. 6, pp. 1071–1076, 1989.

[148] T. Tanaka and A. Greenwood, “Effects of Charge Injection and Extraction on Tree Initiation in Polyethylene,” *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 5, pp. 1749–1759, 1978.

[149] M. Ieda, “Dielectric Breakdown Process of Polymers,” *IEEE Transactions on Electrical Insulation*, vol. EI-15, no. 3, pp. 206–224, 1980.

[150] N. Yoshimura, F. Noto, and K. Kikuchi, “Growth of Water Trees in Polyethylene and Silicone Rubber by Water Electrodes,” *IEEE Transactions on Electrical Insulation*, vol. EI-12, no. 6, pp. 411–416, 1977.

[151] T. Tanaka, “Charge Transfer and Tree Initiation in Polyethylene Subjected to AC Voltage Stress,” *IEEE Transactions on Electrical Insulation*, vol. 27, no. 3, pp. 424–431, 1992.

[152] G. C. Stone, M. Kurtz, and R. G. van Heeswijk, “The Statistical Analysis of A High Voltage Endurance Test on An Epoxy,” *IEEE Transactions on Electrical Insulation*, vol. E1-14, no. 6, pp. 315–326, 1979.

[153] F. Guastavino and B. Cerutti, “Tree Growth Monitoring by Means of Digital Partial Discharge Measurements,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 10, no. 1, pp. 65–72, 2003.

[154] J. V Champion, S. J. Dodd, and G. C. Stevens, “Analysis and Modelling of Electrical Tree Growth in Synthetic Resins Over a Wide Range of Stressing Voltage,” *Journal of Physics D: Applied Physics*, vol. 27, no. 5, pp. 1020–1030, 1994.

[155] X. Chen, Y. Xu, X. Cao, S. J. Dodd, and L. A. Dissado, “Effect of Tree Channel Conductivity on Electrical Tree Shape and Breakdown in XLPE Cable Insulation Samples,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 3, pp. 847–860, 2011.

[156] X. Chen, Y. Xu, X. Cao, and S. M. Gubanski, “Electrical Treeing Behavior at High Temperature in XLPE Cable Insulation Samples,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 5, pp. 2841–2851, 2015.

[157] K. Wu, Y. Suzuoki, T. Mizutani, and H. Xie, “Model for Partial Discharges Associated with Treeing Breakdown: II. Tree Growth Affected by PDs,” *Journal of Physics D: Applied Physics*, vol. 33, no. 10, pp. 1202–1208, 2000.

[158] J. C. Fothergill, L. A. Dissado, and P. J. J. Sweeney, “A Discharge-Avalanche Theory for the Propagation of Electrical Trees,” *IEEE Transactions on Electrical Insulation*, vol. 1, no. 3, pp. 474–486, 1994.

[159] X. Zheng and G. Chen, “Propagation Mechanism of Electrical Tree in XLPE Cable Insulation by Investigating A Double Electrical Tree Structure,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 3, pp. 800–807, 2008.

[160] X. Chen, Y. Xu, X. Cao, and S. M. Gubanski, “On the Conducting and Non-Conducting Electrical Trees in XLPE Cable Insulation Specimens,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 1, pp. 95–103, 2016.

[161] G. J. Paoletti, “Partial Discharge Theory and Technologies Related to Medium-Voltage Electrical Equipment,” *IEEE Transactions on Industry Applications*, vol. 37, no. 1, pp. 90–103, 2001.

[162] R. Bozzo, G. Coletti, and F. Guastavino, “Studies About the Electrical Treeing Growth Based on the Evolution of the PD Patterns,” in *Conference Record of the 1998 IEEE International Symposium on Electrical Insulation*, 1998, pp. 415–419.

[163] IEEE-Task-Force, “Effects of Harmonics on Equipment,” *IEEE Transactions on Power Delivery*, vol. 8, no. 2, pp. 672–680, 1993.

[164] D. Fabiani and G. C. Montanari, “The Effect of Voltage Distortion on Ageing

Acceleration of Insulation Systems under Partial Discharge Activity," *IEEE Electrical Insulation Magazine*, vol. 17, no. 3, pp. 24–33, 2001.

[165] G. C. Montanari and D. Fabiani, "Searching for the Factors which Affect Self-Healing Capacitor Degradation under Non-Sinusoidal Voltage," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 6, no. 3, pp. 319–325, 1999.

[166] G. C. Montanari and D. Fabiani, "The Effect of Non-Sinusoidal Voltage on Intrinsic Aging of Cable and Capacitor Insulating Materials," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 6, no. 6, pp. 798–802, 1999.

[167] G. C. Montanari, I. Ghinello, and D. Fabiani, "Accelerated Degradation of Capacitor PP Films under Voltage Distortion," in *1998 Annual Report Conference on Electrical Insulation and Dielectric Phenomena (CEIDP)*, 1998, pp. 686–689.

[168] P. Caramia, G. Carpinelli, P. Verde, G. Mazzanti, A. Cavallini, and G. C. Montanari, "An Approach to Life Estimation of Electrical Plant Components in the Presence of Harmonic Distortion," in *Proceedings of International Conference on Harmonics and Quality of Power, ICHQP*, 2000, vol. 3, no. 5, pp. 887–892.

[169] A. Cavallini, D. Fabiani, and G. C. Montanari, "Power Electronics and Electrical Insulation Ssystems - Part 2: Life Modeling for Insulation Design," *IEEE Electrical Insulation Magazine*, vol. 26, no. 4, pp. 33–39, 2010.

[170] G. Mazzanti, G. Passarelli, A. Russo, and P. Verde, "The Effects of Voltage Waveform Factors on Cable Life Estimation Using Measured Distorted Voltages," in *2006 IEEE Power Engineering Society General Meeting*, 2006, pp. 1–8.

[171] M. Florkowski and B. Florkowska, "Distortion of Partial Discharge Images caused by High Voltage Harmonics," *IEE Proceedings - Generation, Transmission and Distribution*, vol. 153, no. 2, pp. 171–180, 2006.

[172] IEC61000-2-4, "Electromagnetic Compatibility." International Electrotechnical Commission, pp. 1–41, 2002.

[173] S. Bahadoorsingh and S. M. Rowland, "An Investigation of the Harmonic Impact on Electrical Tree Growth," in *Conference Record of IEEE International Symposium on Electrical Insulation*, 2010.

[174] R. Sarathi, K. H. Oza, C. L. G. Pavan Kumar, and T. Tanaka, "Electrical Treeing in XLPE Cable Insulation Under Harmonic AC Voltages," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 6, pp. 3177–3185, 2015.

[175] S. Bahadoorsingh, S. Sambeharry, C. Sharma, and S. M. Rowland, "The Statistical Analysis of Harmonic Influenced Electrical Treeing Partial Discharge Data Using the Weibull Distribution," in *Annual Report - Conference on Electrical Insulation and Dielectric Phenomena, CEIDP*, 2011, pp. 772–775.

[176] G. C. Montanari, "A Comparative Investigation of Electrothermal Endurance

Models for Insulating Materials and Systems Ccharacterization," *IEEE Electrical Insulation Magazine*, vol. 13, no. 3, pp. 13–25, 1997.

[177] G. C. Montanari, G. Mazzanti, and L. Simoni, "Progress in Electrothermal Life Modeling of Electrical Insulation During the Last Decades," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 5, pp. 730–745, 2002.

[178] G. C. Montanari, "Notes on Theoretical and Practical," *IEEE Electrical Insulation Magazine*, vol. 29, no. 4, pp. 34–44, 2013.

[179] G. C. Montanari, E. Industriale, F. Ingegneria, and U. Bologna, "Electrical Life Threshold Models for Solid Insulating Materials Subjected to Electrical and Multiple Stresses. Investigation and Comparison of Life Models," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, no. 5, pp. 974–986, 1992.

[180] G. Mazzanti, "Life and Reliability Models for High Voltage DC Extruded Cables," *IEEE Electrical Insulation Magazine*, vol. 33, no. 4, pp. 42–52, 2017.

[181] V. M. Montsinger, "Loading Transformers by Temperature," *Transactions of the American Institute of Electrical Engineers*, vol. 49, no. 2, pp. 776–790, 1930.

[182] T. W. Dakin, "Electrical Insulation Deterioration Treated as a Chemical Rate Phenomenon," *Transactions of the American Institute of Electrical Engineers*, vol. 67, no. 1, pp. 113–122, 1948.

[183] G. C. Montanari and G. Pattini, "Thermal Endurance Evaluation of Insulating Materials: A Theoretical and Experimental Analysis," *IEEE Transactions on Electrical Insulation*, vol. EI-21, no. 1, pp. 69–77, 1986.

[184] J.-P. Crine, "The Compensation Law Revisited. Application to Dielectric Aging," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 26, no. 4, pp. 811–818, 1991.

[185] W. T. Starr and H. S. Endicott, "Progressive Stress-A New Accelerated Approach to Voltage Endurance," *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 80, no. 3, pp. 515–522, 1961.

[186] R. G. Rhudy and H. E. Mazanek, "Accelerated Voltage-Endurance Tests," *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 79, no. 3, pp. 576–579, 1960.

[187] G. L. Moses, "Alternating and Direct Voltage Endurance Studies on Mica Insulation for Electric Machinery," *Transactions of the American Institute of Electrical Engineers*, vol. 70, no. 1, pp. 763–769, 1951.

[188] T. W. Dakin and S. A. Studniarz, "The Voltage Endurance of Cast and Molded Resins," in *13th Electrical/Electronics Insulation Conference (EIC)*, 1977, pp. 318–321.

[189] F. H. Kreuger, "Determination of the Internal Discharge Resistance of Dielectric Materials," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. EI-3, no. 4, pp. 106–114, 1968.

[190] L. Simoni and G. Pattini, “A New Research into the Voltage Endurance of Solid Dielectrics,” *IEEE Transactions on Electrical Insulation*, vol. EI-10, no. 1, pp. 17–27, 1975.

[191] G. C. Montanari and M. Cacciari, “On the Electrical Endurance Characterization of Insulating Materials by Constant and Progressive Stress Tests,” *IEEE Transactions on Electrical Insulation*, vol. 27, no. 5, pp. 1000–1008, 1992.

[192] L. Simoni, “A New Approach to the Voltage-Endurance Test on Electrical Insulation,” *IEEE Transactions on Electrical Insulation*, vol. EI-8, no. 3, pp. 76–86, 1973.

[193] H. Hirose, “A Method To Estimate The Lifetime Of Solid Electrical Insulation,” *IEEE Transactions on Electrical Insulation*, vol. EI-22, no. 6, pp. 745–753, 1987.

[194] L. Simoni, “Geometrical Approach to Multistress Endurance of Electrical Insulation,” in *IEEE International Symposium on Electrical Insulation*, 1990, pp. 8–14.

[195] T. Suzuki, K. Kishi, T. Uozumi, K. Yatsuka, N. Yasuda, and T. Fukui, “Study on V-t Characteristics for XLPE Cable,” in *Proceedings of IEEE/PES Transmission and Distribution Conference*, 1994, pp. 192–199.

[196] Q. Zhuang, P. H. F. Morshuis, X. Chen, J. J. Smit, and Z. Xu, “Life Prediction of A Full-Scale Transformer Winding Insulation Through Statistical Analysis of AC Voltage Endurance Test Data,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 2, pp. 460–471, 2012.

[197] K. Wu and L. A. Dissado, “Model for Electrical Tree Initiation in Epoxy Resin,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 4, pp. 655–668, 2005.

[198] L. Niemeyer, L. Pietronero, and H. J. Wiesmann, “Fractal Dimension of Dielectric Breakdown,” *Physical Review Letters*, vol. 52, no. 12. pp. 1033–1036, 1984.

[199] L. Pietronero and H. J. Wiesmann, “Stochastic Model for Dielectric Breakdown,” *Journal of Statistical Physics*, vol. 36, no. 5–6, pp. 909–916, 1984.

[200] H. J. Wiesmann and H. R. Zeller, “A Fractal Model of Dielectric Breakdown and Prebreakdown in Solid Dielectrics,” *Journal of Applied Physics*, vol. 60, no. 5, pp. 385–390, 1986.

[201] L. A. Dissado, P. J. J. Sweeney, and J. C. Fothergill, “The Field Dependence of Electrical Tree Growth,” in *Sixth International Conference on Dielectric Materials, Measurements and Applications*, 1992, pp. 13–16.

[202] L. A. Dissado and P. J. J. Sweeney, “An Analytical Model for Discharge Generated Breakdown Structures,” in *Proceedings of the 4th International Conference on Conduction and Breakdown in Solid Dielectrics*, 1992, pp. 328–332.

[203] A. Contin, E. Gulski, M. Cacciari, and G. C. Montanari, “Inference of PD in Electrical Insulation by Charge-Height Probability Distribution: Diagnosis of

Insulation System Degradation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 5, no. 1, pp. 110–117, 1998.

[204] M. D. Noskov, A. S. Malinovski, M. Sack, and A. J. Schwab, "Self-Consistent Modeling of Electrical Tree Propagation and PD Activity," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 6, pp. 725–733, 2000.

[205] M. D. Noskov, M. Sack, A. S. Malinovski, and A. J. Schwab, "Measurement and Simulation of Electrical Tree Growth and Partial Discharge Activity in Epoxy Resin," *Journal of Physics D: Applied Physics*, vol. 34, no. 9, pp. 1389–1398, 2001.

[206] H. Z. Ding and B. R. Varlow, "Thermodynamic Model for Electrical Tree Propagation Kinetics in Combined Electrical and Mechanical Stresses," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 1, pp. 81–84, 2005.

[207] Y. Nyanteh, L. Graber, S. Srivastava, C. Edrington, D. Cartes, and H. Rodrigo, "A Novel Approach Towards the Determination of the Time to Breakdown of Electrical Machine Insulating Materials," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 1, pp. 232–240, 2015.

[208] ISO13381-1, "Condition Monitoring and Diagnostics of Machines — Prognostics —Part 1: General Guidelines," *International Standard, ISO*. 2015.

[209] J. W. Hines and A. Usynin, "Current Computational Trends in Equipment Prognostics," *International Journal of Computational Intelligence Systems*, vol. 1, no. 1, pp. 94–102, 2008.

[210] J. B. Coble and J. W. Hines, "Prognostic Algorithm Categorization with PHM Challenge Application," in *2008 International Conference on Prognostics and Health Management*, 2008.

[211] S. Liu, Y. Wang, and F. Tian, "Prognosis of Underground Cable via Online Data-Driven Method With Field Data," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7786–7794, 2015.

[212] F. H. Kreuger, E. Gulski, and A. Krivda, "Classification of Partial Discharges," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 917–931, 1993.

[213] Suwarno, "Partial Discharges in High Voltage Insulations. Mechanism, Patterns and Diagnosis," in *2014 IEEE International Conference on Electrical Engineering and Computer Science (ICEECS)*, 2014, pp. 369–375.

[214] E. Ferreira, "Model Selection in Time Series Machine Learning Applications," Ph.D. dissertation, University of Oulu, Finland, 2015.

[215] B. Zhang, L. Zhang, and J. Xu, "Degradation Feature Selection for Remaining Useful Life Prediction of Rolling Element Bearings," *Quality and Reliability Engineering International*, vol. 32, no. 2, pp. 547–554, 2016.

[216] K. Javed, R. Gouriveau, and N. Zerhouni, "State of the Art and Taxonomy of Prognostics Approaches , Trends of Prognostics Applications and Open Issues Towards Maturity at Different Technology Readiness Levels," *Mechanical Systems and Signal Processing*, vol. 94, pp. 214–236, 2017.

[217] A. Jovic, K. Brkic, and N. Bogunovic, “A Review of Feature Selection Methods with Applications,” in *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200–1205.

[218] R. Kohavi and G. H. John, “Wrappers for Feature Subset Selection,” *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, 1997.

[219] J. Tang, S. Alelyani, and H. Liu, “Feature Selection for Classification: A Review,” in *Data Classification: Algorithms and Applications*, 2014, pp. 37–64.

[220] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, vol. 2. 2009.

[221] C. Croux and C. Dehon, “Influence Functions of the Spearman and Kendall Correlation Measures,” *Statistical Methods and Applications*, vol. 19, no. 4, pp. 497–515, 2010.

[222] T. N. Lal, O. Chapelle, J. Weston, and A. Elisseeff, “Embedded Methods,” in *Feature Extraction. Foundations and Applications*, 2006.

[223] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.

[224] E. R. Pacheco, *Unsupervised learning with R*. Packt Publishing Ltd, 2015.

[225] A. Saxena *et al.*, “Metrics for Evaluating Performance of Prognostic Techniques,” in *2008 International Conference on Prognostics and Health Management*, 2008.

[226] H. M. Elattar, H. K. Elminir, and A. M. Riad, “Prognostics: A Literature Review,” *Complex & Intelligent Systems*, vol. 2, no. 2, pp. 125–154, 2016.

[227] M. Krupa, “Methods of Technical Prognosis - Review,” *International Scientific Journal*, pp. 20–23, 2013.

[228] F. Peysson, M. Ouladsine, H. Noura, J. B. Leger, and C. Allemand, “New Approach to Prognostic System Failures,” in *Proceedings of the 17th World Congress. The International Federation of Automatic Control*, 2008, pp. 12861–12866.

[229] O. E. Dragomir, R. Gouriveau, F. Dragomir, E. Minca, and N. Zerhouni, “Review of Prognostic Problem in Condition-Based Maintenance,” in *IEEE Control Systems Society. European Control Conference*, 2009, pp. 1587–1592.

[230] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*. John Wiley & Sons, 2013.

[231] K. L. Priddy and P. E. Keller, *Artificial Neural Networks. An Introduction*. Washington, USA: SPIE Press, 2005.

[232] A. K. Mahamad, S. Saon, and T. Hiyama, “Predicting Remaining Useful Life of Rotating Machinery Based Artificial Neural Network,” *Computers and Mathematics with Applications*, vol. 60, no. 4, pp. 1078–1087, 2010.

[233] X. Laisheng, “A Sliding-Window Modeling Approach for Neural Network,” *International Journal of Control and Automation*, vol. 7, no. 8, pp. 117–130, 2014.

[234] W. Teng, X. Zhang, Y. Liu, A. Kusiak, and Z. Ma, “Prognosis of the Remaining Useful Life of Bearings in a Wind Turbine Gearbox,” *Energies*, vol. 10, no. 1, 2017.

[235] S. A. Asmai, A. S. H. Basari, A. S. Shibghatullah, N. K. Ibrahim, and B. Hussin, “Neural Network Prognostics Model for Industrial Equipment Maintenance,” in *Proceedings of the 2011 11th International Conference on Hybrid Intelligent Systems*, 2011, pp. 635–640.

[236] E. Keogh, J. Lin, S.-H. Lee, and H. Van Herle, “Finding the Most Unusual Time Series Subsequence: Algorithms and Applications,” *Knowledge and Information Systems*, vol. 11, no. 1, pp. 1–27, 2007.

[237] J. I. Aizpurua and V. M. Catterson, “Towards a Methodology for Design of Prognostic Systems,” in *Annual Conference of the Prognostics and Health Management Society*, 2015, pp. 1–13.

[238] N. Amjady, “Short-Term Hourly Load Forecasting Using Time-Series Modeling with Peak Load Estimation Capability,” *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 498–505, 2001.

[239] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, “ARIMA Models to Predict Next-Day Electricity Prices,” *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, 2003.

[240] H. Liu, H.-Q. Tian, and Y.-F. Li, “Comparison of Two New ARIMA-ANN and ARIMA-Kalman Hybrid Methods for Wind Speed Prediction,” *Applied Energy*, vol. 98, pp. 415–424, 2012.

[241] Y. Zhou and M. Huang, “Lithium-Ion Batteries Remaining Useful Life Prediction Based on A Mixture of Empirical Mode Decomposition and ARIMA Model,” *Microelectronics Reliability*, vol. 65, pp. 265–273, 2016.

[242] A. Molinaro, R. Simon, and R. M. Pfeiffer, “Prediction Error Estimation: A Comparison of Resampling Methods,” *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005.

[243] B. Efron and R. Tibshirani, “Improvements on Cross-Validation: The .632+ Bootstrap Method,” *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 548–560, 1997.

[244] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995, vol. 14, no. 12, pp. 1137–1143.

[245] B. Efron and R. Tibshirani, “Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy,” *Statistical Science*, vol. 1, no. 1, pp. 54–77, 1986.

[246] B. Efron, “Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation,” *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 316–331, 1983.

[247] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, “Metrics for Offline Evaluation of Prognostic Performance,” *International Journal of Prognostics and Health Management*, vol. 1, no. 1, pp. 4–23, 2010.

[248] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, “Evaluating Prognostics Performance for Algorithms Incorporating Uncertainty Estimates,” in *2010 IEEE Aerospace Conference*, 2010.

[249] S. Bahadoorsingh, R. Balliram, C. Sharma, and S. M. Rowland, “Analytical Results of Harmonic Influenced Electrical Tree Growth Images,” in *Conference Record of IEEE International Symposium on Electrical Insulation*, 2012, pp. 1–4.

[250] S. Bahadoorsingh and S. M. Rowland, “Investigating the Impact of Harmonics on the Breakdown of Epoxy Resin through Electrical Tree Growth,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 16, no. 4, pp. 511–514, 2009.

[251] A. Cavallini, D. Fabiani, G. Mazzanti, G. C. Montanari, and A. Contin, “Voltage Endurance of Electrical Components Supplied by Distorted Voltage Waveforms,” in *Conference Record of IEEE International Symposium on Electrical Insulation*, 2000, pp. 73–76.

[252] S. Bahadoorsingh and S. M. Rowland, “Investigating the Influence of the Lubricant Coating on Hypodermic Needles on Electrical Tree Characteristics in Epoxy Resin,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 17, no. 3, pp. 701–708, 2010.

[253] R. Patsch and M. Hoof, “Electrical Treeing - Physical Details Obtained by Pulse Sequence Analysis,” in *1995 EEE 5th International Conference on Conduction and Breakdown in Solid Dielectrics*, 1995, pp. 344–349.

[254] S. Rob, *Introduction to Data Mining for the Life Sciences*. Springer Science & Business Media, 2012.