

THE CHARACTERISATION AND AUTOMATIC CLASSIFICATION OF TRANSMISSION LINE FAULTS

Ulrich Minnaar

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Electrical Engineering
UNIVERSITY OF CAPE TOWN

February 2014

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

DECLARATION

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for another degree at any other university or institution.

ULRICH MINNAAR

September 2013

ACKNOWLEDGEMENTS

I would like to acknowledge the following people for their encouragement and support:

- My supervisors, Prof. Trevor Gaunt and Prof. Fred Nicolls, for their guidance and feedback, as well as their expertise
- My colleagues at Eskom for the role they have played in my development as well as the many useful conversations which have helped shape my understanding of power systems
- My family, for teaching me the value of education, supporting me and giving me the freedom to learn in my own way over the years
- And especially to my wife Desiré, for the patience and support during the many sacrificed evenings and weekends.

ABSTRACT

A country's ability to sustain and grow its industrial and commercial activities is highly dependent on a reliable electricity supply. Electrical faults on transmission lines are a cause of both interruptions to supply and voltage dips. These are the most common events impacting electricity users and also have the largest financial impact on them. This research focuses on understanding the causes of transmission line faults and developing methods to automatically identify these causes.

Records of faults occurring on the South African power transmission system over a 16-year period have been collected and analysed to find statistical relationships between local climate, key design parameters of the overhead lines and the main causes of power system faults. The results characterize the performance of the South African transmission system on a probabilistic basis and illustrate differences in fault cause statistics for the summer and winter rainfall areas of South Africa and for different times of the year and day. This analysis lays a foundation for reliability analysis and fault pattern recognition taking environmental features such as local geography, climate and power system parameters into account.

A key aspect of using pattern recognition techniques is selecting appropriate classifying features. Transmission line fault waveforms are characterised by instantaneous symmetrical component analysis to describe the transient and steady state fault conditions. The waveform and environmental features are used to develop single nearest neighbour classifiers to identify the underlying cause of transmission line faults. A classification accuracy of 86% is achieved using a single nearest neighbour classifier. This classification performance is found to be superior to that of decision tree, artificial neural network and naïve Bayes classifiers.

The results achieved demonstrate that transmission line faults can be automatically classified according to cause.

CONTENTS

DECLARATION.....	ii
ACKNOWLEDGEMENTS.....	iii
ABSTRACT	iv
CONTENTS.....	v
LIST OF FIGURES	viii
LIST OF TABLES	x
ACRONYMS.....	xii
1. INTRODUCTION.....	1
1.1 Improving Transmission Reliability	2
1.2 Objective and Research Hypothesis	3
1.3 Thesis Structure	4
2. CAUSES OF TRANSMISSION LINE FAULTS.....	6
2.1 Characteristics of Major Fault Causes.....	8
2.2 Waveform Characterisation for Event Classification.....	17
2.3 Statistical Pattern Recognition and Classification.....	22
2.4 Classification of Power System Events	24

2.5	Conclusions.....	26
3.	ANALYSIS OF TRANSMISSION LINE FAULTS IN RELATION TO WEATHER AND CLIMATE	28
3.1	Database and Management	29
3.2	Transmission System Results and Discussion	32
3.3	Establishing the Statistical Significance of variations in fault frequency	45
3.4	Analysis of a single class of faults	47
3.5	Conclusions.....	49
4.	WAVEFORM CHARACTERISATION OF TRANSMISSION LINE FAULTS	50
4.1	Waveform Characterisation.....	51
4.2	Identifying the start and end of a fault	55
4.3	Waveform Characteristics	58
4.4	Discussion	67
4.5	Conclusions.....	69
5.	CLASSIFYING TRANSMISSION LINE FAULTS.....	71
5.1	Introduction.....	71
5.2	Feature Selection and Classification	73
5.3	Comparing Classifier Performance.....	89

5.4	Discussion and Conclusions.....	91
6.	CONCLUSIONS.....	93
6.1	Assessing the hypothesis.....	94
6.2	Contributions.....	97
6.3	Significance of Results.....	98
6.4	Future Work.....	100
	REFERENCES.....	101
	APPENDIX A: FAULT FREQUENCY DATA FOR SOUTH AFRICAN TRANSMISSION LINES.....	111

LIST OF FIGURES

Figure 2.1: Two different flashover mechanisms and paths are demonstrated (Vosloo et al. 2009)	9
Figure 2.2: Description of a Pattern Recognition System (Duin, et al, 2002).....	22
Figure 3.1: Ground flash density (N_g) map with transmission lines of South Africa for 2006-2010 (Eskom, 2010).....	30
Figure 3.2: Transmission Line Fault Causes - 12229 faults from 132 kV to 765kV.	32
Figure 3.3: Fault frequency per 100 km per year per rainfall activity area	34
Figure 3.4: Average 400kV line fault frequency statistics by fault cause per 100 km per year	35
Figure 3.5: Season and time dependent frequency of bird streamer faults on the South African 400 and 275kV networks.....	38
Figure 3.6: Fault frequency on 400kV, 275kV and 220kV lines fitted with birdguards	40
Figure 3.7: Season and time dependent frequency of fire-caused faults on the South African 400kV network	41
Figure 3.8: Season and time dependent frequency of pollution faults on the South African 400kV network.....	42

Figure 3.9: Season and time dependent frequency of lightning faults on the South African 400kV network.....	43
Figure 3.10: Season and time dependent frequency of lightning faults on the South African 400kV network in the thunderstorm activity region normalised to $N_g=1$	44
Figure 3.11: Lightning faults on 400kV transmission lines in the thunderstorm rainfall activity area (Group 1- Poorly Performing Lines, Group 2 – Rest of Lines) with Y = number of faults per Kilometres N_g and R^2 =coefficient of determination..	48
Figure 4.1: Simulink model-Discrete Symmetrical Components.....	54
Figure 4.2: Stages of a fault measurement.....	56
Figure 4.3: Zero Sequence Rate of Change of Current	61
Figure 4.4: Zero sequence current at $\frac{1}{2}$ cycle and one cycle after fault initiation ...	62
Figure 4.5: Sequence Component Fault Current Time Constant.....	66
Figure 5.1: Classification Accuracy.....	83
Figure 5.2: F-measure for fault causes.....	84
Figure 5.3: F-measure using contextual features.	85
Figure 5.4: F-measure using waveform and contextual features combined by rule 1.	87
Figure 5.5: F- measure using waveform and contextual features combined by rule 2	88

LIST OF TABLES

Table 2.1: Causes of Transmission faults.....	7
Table 2.2: Bird Streamer Event Characteristics associated with interested sector of electricity transmission.	11
Table 2.3: 4 by 4 matrix representing fault statistics (Herman & Gaunt, 2010).....	16
Table 4.1: Faulted phases according to underlying cause.....	59
Table 4.2: Basic Statistics for Maximum Rate of Change of Current.....	60
Table 4.3: Basic Statistics for Maximum Sequence Voltage during Fault.....	62
Table 4.4: Basic Statistics for sequence currents at $\frac{1}{2}$ cycle and one cycle after fault initiation.	63
Table 4.5: Basic Statistics for Maximum Sequence Currents.	63
Table 4.6: Fault Resistance calculations.	64
Table 4.7: Fault Resistance.....	65
Table 4.8: Basic Statistics for Fault Insertion Phase Angle.	65
Table 4.9: Basic Statistics for Sequence Component Fault Current Time Constant.	67
Table 4.10: Statistical Significance of Causes Influencing Waveform Features	69
Table 5.1: Features ranked by F-statistic	76

Table 5.2: Confusion Matrix..... 80

Table 5.3: Classification performance using wrapper selection methods 90

ACRONYMS

1-NN	Single nearest neighbour
AFIS	Advanced fire information system
ANOVA	Analysis of variance
ANN	Artificial neural network
FIPA	Fault insertion phase angle
GIS	Geospatial information system
MODIS	Moderate resolution imaging spectroradiometer
Ng	Ground flash density
RMS	Root mean square
SEVIRI	Spinning enhanced visible and infrared imager

CHAPTER 1

1.INTRODUCTION

This chapter motivates the reasons for undertaking the research and defines the objectives and scope of the thesis.

Modern society is dependent on an electrical supply that is both reliable (Alvehag & Soder, 2011) and compatible with the needs of equipment connected by utility customers (Cigre, 2011).

A country's ability to sustain and grow its industrial and commercial activities is highly dependent on a reliable electricity supply. Supply interruptions and voltage dips are the two most common events impacting customers and also have the largest financial impact on customers. These events disrupt commercial activities and manufacturing processes, resulting in decreased output and profitability (Chowdhury & Koval, 2000).

Long duration interruptions have the greatest impact, as borne out by major events such as the New York and Italian blackouts of 2003 as well as the extensive interruptions that occurred in South Africa in early 2008. However, shorter interruptions and voltage dips also have an economic impact on customers due to processes or services being interrupted (Cigre, 2011).

Transmission lines play an important role in providing an electricity supply to customers that is both reliable and within voltage dip compatibility levels of customer equipment. Faults on transmission lines are a root cause of both

interruptions and voltage dips (Cigre, 2011). The focus of this thesis is on the characterisation and classification of transmission line faults with the aim of improving the reliability of transmission power lines.

1.1 Improving Transmission Reliability

Historically, deterministic criteria have been used for the planning and design of transmission systems. Reliability can be improved by increasing capital and operational and maintenance expenditure to reduce the frequency and duration of faults; this however risks over-investment and consequently higher electricity tariffs to be paid by customers (Chowdhury & Koval, 2000). The need for a reliable electricity supply is balanced by the need to minimise the cost of operating the transmission system.

Deterministic methods do not account for the stochastic nature of failures, customer demand or power system behaviour (Chowdhury & Koval, 2000). Probabilistic techniques for power system simulation and analysis have been developed to account of the stochastic nature of power system behaviour (Edimu, et al, 2011).

Decreasing the time taken to restore a line after a fault has occurred is another way in which transmission system operators can improve network reliability. During the course of a typical fault on a transmission line, the network control operators will estimate the probable location of a fault based on available system information and measurements. A field operator is dispatched to determine the location and cause of a fault prior to restoring the line (Xu, et al, 2005). This process may be

completed within a few minutes or may take several hours to complete. In the event where the root cause is uncertain, an extensive section of line may be patrolled prior to the line being restored.

Understanding the causes of line faults and the impact these have on the reliability of the transmission system can play an important role in decision-making for 1) planning and design, 2) maintenance and 3) operation of the network to improve reliability.

Automatic classification of faults according to cause has primarily been explored as a means to reduce the time it takes to restore a distribution line to service (Xu & Chow, 2006) as well as identifying the cause of power quality disturbances i.e. voltage dips (Bollen, et al, 2007).

1.2 Objective and Research Hypothesis

Based on the identification of the benefits of a fault classification system and the work that has been done towards such an approach, a hypothesis arises to the effect that:

Transmission line faults can be classified according to their underlying event cause using statistical pattern recognition techniques; however this requires knowledge of the external environment influencing the event.

The overall aims of the research is to 1) improve understanding of the impact that the climate and environment has on the causes and frequency of faults on the South African transmission network; 2) identify electrical fault waveform

characteristics relevant to identifying fault causes and 3) ultimately automate the classification of transmission line faults using statistical pattern recognition techniques.

To achieve this aim, the following research questions are addressed:

- What are the primary causes of faults on transmission lines on the South African transmission network and how do they impact the fault frequency performance?
- Can significant variables related to interruption performance be identified?
- Can event characteristics be identified that are relevant features for automatically classifying transmission line faults according to underlying cause? If so, which characteristics are these?
- Can faults be classified using only electrical waveform characteristics?
- What classification performance is achieved?

1.3 Thesis Structure

Chapter 2 describes the primary causes of transmission line faults on the South African transmission system. **Chapter 3** presents an analysis method relating frequency of faults on overhead lines to local climate. Fault analysis by time-of-day and time-of-year (season) is presented. The statistical significance of the differences between mean fault frequencies for fault causes, climate, time of day and season is established. **Chapter 4** investigates the characterisation of

measured fault waveforms. **Chapter 5** discusses the classification of transmission line faults according to underlying causes and **Chapter 6** concludes with a summary of the findings and an assessment of the research hypothesis.

CHAPTER 2

2.CAUSES OF TRANSMISSION LINE FAULTS

This chapter discusses the existing literature pertaining to transmission line faults and their analysis. Statistical pattern recognition is reviewed, including the characterisation and classification of power system events.

The geographic location of a power system plays an important role in the frequency and causes of faults to which it is exposed (Pahwa, et al , 2007).

Vosloo investigated fault causes on the South African transmission system and concluded that the majority of transmission network faults are “...*in one way or another connected to natural phenomena such as weather and climate or occurs as a consequence thereof.*” (Vosloo H , 2005). In 2004 a list of primary fault causes and sub-categories was introduced by Eskom to allow analysis of faults that could be traced to the root cause of faults (Vosloo H , 2005), as listed in table 2.1.

Table 2.1: Causes of Transmission faults.

Primary Category	Sub-Category
Bird	Streamer
	Pollution
	Nest
Fire	Veld
	Cane
	Refuse
	Fynbos
	Reed
Lightning	
Pollution	Bird pollution
	Fire
	Industrial
	Marine
Tree Contact	Alien
	Indigenous
Unclassified	
Other	

Faults due to 'other' causes include events due to occurrences such as failure of hardware, poor workmanship, tree contact, impact of foreign objects, theft and vandalism.

Wind and lightning have been identified as two major weather-related causes of outages (Alvehag & Soder , 2011). The Eskom classification, in contrast, does not

include wind as a major cause of faults in South Africa as the transmission system is only rarely affected by extreme wind conditions with a low frequency of occurrence. The primary categories of fault causes identified in in table 2.1, while not an exhaustive list of all possible fault causes, provide an appropriate list of fault causes commonly occurring on the South African transmission network. The classifications used by a utility will consider faults that occur on their network (Pahwa, et al , 2007). The major fault causes identified are birds, lightning, fire and pollution (Vosloo H , 2005).

2.1 Characteristics of Major Fault Causes

2.1.1 Bird

Birds predominantly cause flashovers on power lines in three ways i.e. bird streamers, pollution and electrocution (Vosloo, et al, 2009) along two different flashover paths illustrated in figure 2.1.

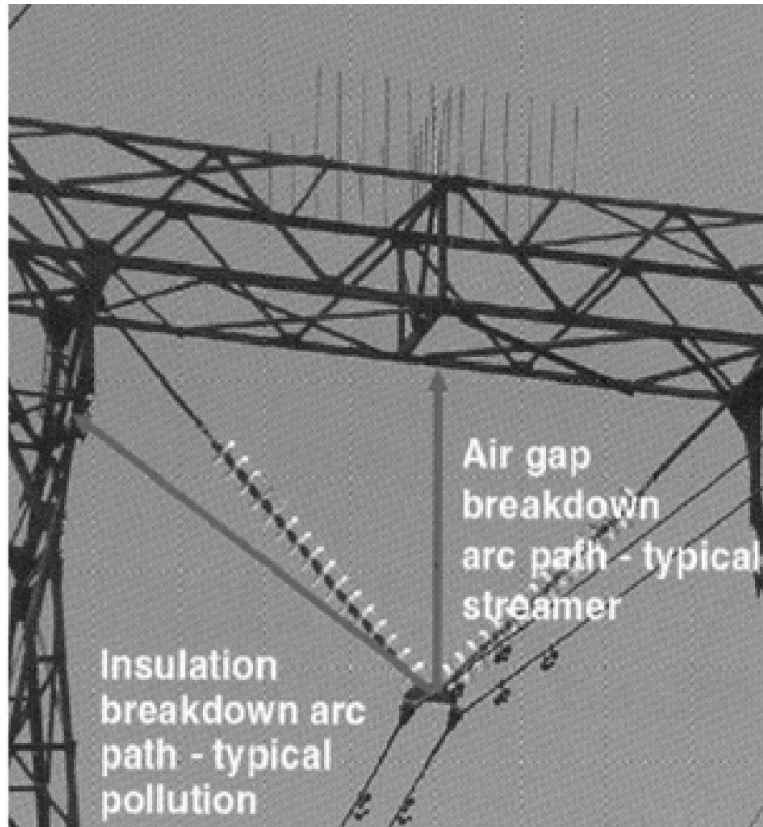


Figure 2.1: Two different flashover mechanisms and paths are demonstrated (Vosloo et al. 2009)

2.1.2 Bird Streamer

Bird streamers were first identified as a cause of unknown transmission line faults in California in the 1920's (Michener, 1928). Flashovers are caused by large birds (vultures, herons, hadeda, ibis and the bigger raptors) excreting long streamers which short circuit the air gap between the structure and the conductor (Van Rooyen, et al, 2003).

Flashovers using simulated streamers have been successfully reproduced under laboratory conditions in the USA (West, Brown and Kinyon 1971) and South Africa (Burger & Sarduski, 1995).

Experiences with bird streamer flashovers have been documented by Burnham in Florida (Burnham, 1995), who provided a list of characteristics associated with bird streamer occurrence. Single-phase-to-ground faults due to bird streamers have also been reported on Turkey's 420kV transmission lines (Iliceto, et al, 1981) and on South Africa's transmission and distribution networks (Van Rooyen, et al, 2003). Birdguards (anti-perching devices) have been employed extensively as a solution on transmission towers in Turkey (Iliceto, et al, 1981) and Eskom implemented a national program to fit birdguards on transmission lines throughout South Africa on lines with a high frequency of bird streamer faults.

In table 2.2 the characteristics identified in (Burnham, 1995) in are associated with three spheres of electricity transmission planning and operation i.e. network planning, network control and field services.

Table 2.2: Bird Streamer Event Characteristics associated with interested sector of electricity transmission.

Interested sphere		Event characteristics
Field services		<ul style="list-style-type: none"> • Presence of large bodied birds • A lack of natural roosting spots such as trees • Presence of dead or injured birds near structures after an outage • Outages which can be explained by bird behaviour and structure design: <ul style="list-style-type: none"> ○ Birds prefer outside end of crossarms ○ Birds avoid high voltage stress ○ Birds avoid side of structure facing parallel lines ○ Birds prefer side of structure facing water, lakes, swamps canals, fields etc. ○ Structure must offer roost above energized parts ○ Short air gaps are more susceptible • Features of flashed insulators/hardware/structure: <ul style="list-style-type: none"> Flashed insulator with dropping residue ○ Absence of flashmark on insulator ○ flashmark on crossarm or conductor hardware or only one end of an insulator
		<ul style="list-style-type: none"> • Instantaneous relay actions with successful reclosure, limited to one or two per night tending to occur in the same area
Network planning	Network control	<ul style="list-style-type: none"> • Bimodal temporal distribution of outages — distinct peaks at 06:00 and 22:00 • Seasonal pattern related to presence of birds or their feeding habits

One of the key attributes of bird streamer faults is a clear diurnal and seasonal pattern of occurrence. Although the variation has operational significance, it also affects planning because the time-based distribution of incidents affects the probability of multiple outages at the same time.

The diurnal and seasonal patterns associated with bird streamers provide an indication of characteristics by which this type of fault may be identified by 1) operators or 2) classification systems.

2.1.3 Bird Pollution

Streamers from smaller birds do not bridge the air gap on towers; instead these cause a pollution coating to build-up along the insulator string. Unlike the streamer mechanism that bridges the air gap and initiates faults immediately, the polluted insulators flash over along the insulator surface when appropriate wetting occurs some time later (Macey, et al, 2006).

2.1.4 Electrocutation

The interactions between birds and power lines differ according to the voltage of the line. Faults due to the electrocution of birds bridging the conductors-to-tower air gap by the wings and body occur primarily at voltages of and below 132kV where clearances are smaller than on higher voltage lines (Van Rooyen, et al, 2003). An implication of this is that fewer faults due to the electrocution of birds would be expected on transmission networks when compared to distribution networks.

2.1.5 Lightning

Rainfall in South Africa is generally divided into two seasons namely 1) winter rainfall in the western and south west part of the country and 2) summer rainfall in the central, northern and eastern regions. Rainfall in the southern part of the country is distributed throughout the year (Jandrell , et al, 2009)

Summer rainfall in South Africa is generally associated with summer thunderstorms and lightning i.e. convectional rainfall (Jandrell, et al, 2009). The lightning incidence is very low for the winter rainfall region, which is characterised by frontal activity, and in the all-year rainfall region in the south. Lightning, therefore has a seasonal pattern of occurrence, primarily occurring during summer.

Eskom previously operated a LPATS (Lightning Position and Tracking System) system to detect lightning. Since 2006 this has been replaced by FALLS (Fault Analysis and Lightning Location System) operated by the South African Weather Services due to the previous system reaching the end of its life and not meeting operational requirements (Evert & Schulze , 2005).

2.1.6 Fire

Air normally acts as an isolation medium between live conductors and the ground due to its di-electric properties. During a fire the properties of air change due to smoke and particles that occur between the lines and the ground, possibly resulting in a flashover. Three theoretical models have been proposed to explain the reasons for this and these relate to 1) reduced air density, 2) presence of conductive particles in the air and 3) conductivity of the flames (Sukhnandan & Hoch, 2002).

Faults are commonly caused by veld and sugar cane fires under lines and line servitudes are cleared of vegetation to reduce the risk of fires under transmission lines. Veld fires affecting transmission lines commonly occur during the winter months in South Africa (Vosloo H, 2005).

To minimise the problem of line faults due to fires Eskom uses the Advanced Fire Information System (AFIS) (Davies, et al, 2008). AFIS utilises the data from the Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA's Terra and Aqua earth orbiting satellites as well as the SEVIRI (Spinning Enhanced Visible and Infrared Imager) sensor to detect fires. The AFIS system alerts users of fires near transmission infrastructure, archives fire events and allows access and retrieval of the archive via a web-based application. This has led to the following benefits (Mcfarren & Frost, 2009):

- Improved management of flashovers
- Better overview of fires and
- Increased planning decision-support for vegetation management close to transmission lines.

2.1.7 Environmental Pollution of Insulators

The pollution flashover mechanism is mainly a function of the properties of the insulator surface. For hydrophilic surfaces, such as glass and ceramics, the surface wets completely so that an electrolytic film covers the insulator; hydrophobic surfaces, such as silicone rubber, cause the water to bead into

separate droplets preventing the formation of a continuous layer (Macey, et al, 2006).

The two main pollution processes are (Macey, et al, 2006):

- Pre-deposited pollution of salt and industrial particulates that accumulates over time and needs to be wetted to form a conducting electrolyte. Although the effect of bird deposits is similar to this effect, it's classified under birds because it helps to identify mitigation approaches.
- Instantaneous pollution that is already a conducting electrolyte.

Methods and techniques to assess the severity of pollution at a particular site include: surface deposit technique to determine the level of pollution on an insulator, directional dust deposit gauges, site severity classification and automated insulator pollution monitoring (Macey, et al, 2006).

2.1.8 Analysing Transmission Line Faults

Transmission line fault and reliability statistics have been reported as faults per year (faults/year) as well as being scaled by line length i.e. faults per 100 kilometres of line per year (faults/100km/year) (Chowdhury & Koval , 2000). Assessing and reporting faults in this manner provides useful information to utilities for: 1) establishing performance benchmarks for future reliability 2) developing reliability criteria and design standards and 3) identify deteriorating line performance over time and identifying lines/networks that require reinforcement (Chowdhury & Koval, 2000).

Bird streamer, fire and lightning faults have distinct seasonal variations of occurrence, which affects the seasonal reliability performance of transmission networks. Seasonal performance has been reported in the USA for the Mid-Continent Power Pool by reporting statistics for winter and summer (Chowdhury & Koval, 2008).

Herman and Gaunt (2010) introduce a time dependent characterisation of interruptions that recognises the impact of time-of-day and season on fault cause. A 4 by 4 matrix characterising interruptions by time-of-day and season is presented that represents statistics for frequency according to six-hourly time blocks and seasons.

Table 2.3: 4 by 4 matrix representing fault statistics (Herman & Gaunt, 2010)

S/I	00 - 06	06 - 12	12 - 18	18 - 24
Season 1	$\mu_{11}; \sigma_{11}$	$\mu_{12}; \sigma_{12}$	$\mu_{13}; \sigma_{13}$	$\mu_{14}; \sigma_{14}$
Season 2	$\mu_{21}; \sigma_{21}$	$\mu_{22}; \sigma_{22}$	$\mu_{23}; \sigma_{23}$	$\mu_{24}; \sigma_{24}$
Season 3	$\mu_{31}; \sigma_{31}$	$\mu_{32}; \sigma_{32}$	$\mu_{33}; \sigma_{33}$	$\mu_{34}; \sigma_{34}$
Season 4	$\mu_{41}; \sigma_{41}$	$\mu_{42}; \sigma_{42}$	$\mu_{43}; \sigma_{43}$	$\mu_{44}; \sigma_{44}$

Table 2.3 illustrates fault statistics (e.g. fault frequency) represented in a 4 by 4 matrix as mean (σ) and standard deviation (μ). This type of characterisation allows a time-dependant probabilistic approach to be used for network reliability analysis (Edimu, et al, 2013).

2.2 Waveform Characterisation for Event Classification

Characterisation of event waveforms (e.g. voltage dips) is applied to reduce data, interpret and characterize events for analysis and management of power quality (IEC, 2008). Methods proposed and used for this purpose include the ABC classification (Bollen, 2003) and the South African NRS048-2 voltage dip characterisation (NERSA, 2007). The aim of characterisation here is to describe events with a limited number of parameters (Koch, et al, 2001).

Characterisation can also be conducted with the aim of doing automatic classification of disturbances (Bollen, et al, 2007). The aim of characterisation is then *“to find common features that are likely related to specific underlying causes in power systems”* (Bollen, et al, 2009). A number of signal processing techniques, including root mean square (rms), Fourier and wavelet transforms (Fernandez & Rojas, 2002), have been used to extract features and characterize events for automatic classification. These techniques are used to address a number of issues that are relevant to characterising disturbances, including: finding the fundamental voltages/currents and their harmonics; detecting transition points in waveforms; waveform segmentation and feature extraction (Gu & Styvaktakis, 2003).

A number of methods have been applied for the characterisation of voltage dip events. Characterisation for voltage dips is usually done to reduce data, interpret and characterise events (Minnaar, et al, 2010). Methods applied for segmentation and characterisation of voltage dip waveforms events include the RMS method, Fourier and S-transforms, wavelet analysis, multi-resolution S-Transform as well as the Park vector (Gargoom, et al, 2005).

2.2.1 RMS Method

The general equation used to calculate RMS is:

$$V_{rms}(t_k) = \sqrt{\frac{1}{N} \sum_{t=t_k-N+1}^{t_k} v^2(t)}$$

Event identification via RMS is done by comparing change in magnitude with a predetermined threshold. Application of RMS requires simple signal processing and is recognised as being very efficient. It is widely used in power quality instruments that monitor RMS

The importance of phasor information is recognised and introduced via a method to deduce phasors from RMS voltages for analysis purposes (Bollen et al, 2003).

For analysis purposes a method of segmentation based on rate of change is introduced (Styvaktakis et al, 2002) and finds application in a classification system based on RMS voltage only.

2.2.2 Short-Time Fourier Transform

The short time Fourier transform of a signal $v[k]$ is:

$$V_{STFT}(m, \omega) = \sum_k v[k] \cdot w[k-m] e^{-j\omega k}$$

Where $\omega=2\pi n/N$, N is the length of $v[k]$, $n=1, \dots, N$, and $w[k-m]$ is a selected window that slides over the analysed signal

The STFT has limitations due to its fixed window length, which has to be chosen prior to the analysis. This drawback is reflected in the achievable frequency resolution when analysing non-stationary signals with both low and high-frequency components (Gargoom et al, 2005).

2.2.3 Multi-resolution S-Transform

The S-Transform is described as being either a phase-corrected version of the wavelet transform or a variable window Short Time Fourier transform that simultaneously localizes both real and imaginary spectra of the signal (Perez & Barros, 2006). It is defined by convolving the analysed signal, $v[k]$, with a window function. The S-transform of a discrete signal $v[k]$ can be calculated as:

$$V_{ST}\left[k, \frac{n}{N}\right] = \sum_{m=0}^{N-1} V\left[\frac{m+n}{N}\right] \cdot e^{\frac{-2\pi^2}{n^2}} \cdot e^{j\omega k}$$

Where k, m and $n = 0, 1, \dots, N-1$ and $V[m+n/N]$ is the Fourier transform of the analysed signal $v[k]$ $\omega = 2\pi n/N$, N is the length of $v[k]$

2.2.4 Park Vector – DQ Transform

Park's vector is based on the instantaneous vector sum of all of the three phase vectors (v_1, v_2, v_3). The Park transform finds general application in the field oriented control of induction motors. The vector components (v_d, v_q) are given by (Gargoom et al, 2005):

$$v_D = \sqrt{\frac{2}{3}}v_1 - \sqrt{\frac{1}{6}}v_2 - \sqrt{\frac{1}{6}}v_3,$$

$$v_Q = \sqrt{\frac{1}{2}}v_2 - \sqrt{\frac{1}{6}}v_3.$$

2.2.5 Wavelet Analysis

The wavelet transform is based on the decomposition of a signal into daughter wavelets derived from the translation and dilation of a fixed mother wavelet. The general formula is given by:

$$V_{WT} = \int_{-\infty}^{+\infty} v(t) \cdot \varphi_{a,b}^*(t) \cdot dt$$

The most popular applications of wavelets are (Fernandez & Rojas, 2002):

- Power system protection
- Power quality
- Power system transients
- Partial discharge
- Load forecasting
- Power system measurement

2.2.6 Application of Characterisation Methods

The majority of studies have focussed on identifying characteristics by which various events may be classified according to the disturbance type e.g. dips, transient and swells (Gu & Styvaktakis, 2003). It has been recognised that while this work is relevant for developing methods, its practical value is limited (Bollen, et al, 2007). A further challenge in waveform characterisation is a shortage of data with many studies utilising synthetic data for characterisation and classification

purposes, leading to results with limited application to real-world scenarios (Bollen, et al, 2007).

Characterising events according to their underlying causes is of more practical value, as it allows utilities to respond appropriately to individual events as well as put solutions in place to mitigate against specific causes, but limited work has been published in this regard. Characterisation and classification of faults due to causes internal to the power system (e.g. transformer energizing, load changes and motor starting) has been explored and characterised via rms and Kalman filtering (Styvaktakis, et al, 2002). Characterisation via rms has the following disadvantages: dependency on window length and time interval for updating values, the magnitude and duration of an event may vary with a change in selection of these two parameters; no phase angle and no point-on-wave for start of event information is available (Gu & Styvaktakis, 2003).

Characterisation and analysis of external faults based on the voltage and current waveforms has been investigated for causes such as lightning, tree and animal contact and cable faults (Barrera, et al, 2012). Features are obtained from voltage and current waveforms recorded at distribution substations (12.47kV) for a relatively small sample of 180 events. Features obtained through waveform characterisation by Barrera et al (2012) include: maximum zero sequence current and voltage, fault insertion phase angle (FIPA), maximum change of current and voltage magnitudes (phase and neutral) as well as maximum arc voltage.

2.3 Statistical Pattern Recognition and Classification

Pattern recognition can be described as the science of information procedures that are able to classify, describe and label measurements (Jonker, et al, 2003)

A traditional description of a pattern recognition system includes stages such as: sensing, data pre-processing (e.g. segmentation), feature extraction and classification (Duin, et al, 2002). This is illustrated in Figure 2.2

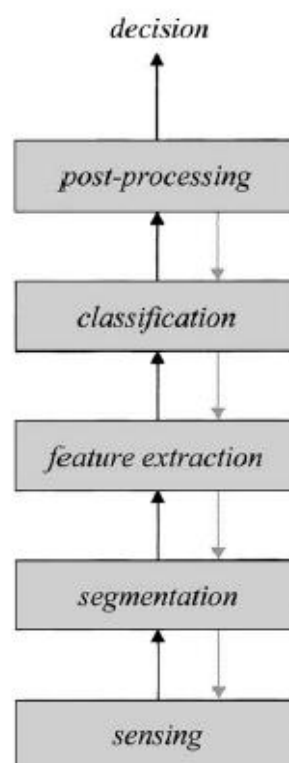


Figure2.2: Description of a Pattern Recognition System (Duin, et al, 2002)

Issues that need to be addressed when developing a statistical pattern recognition system may include: choice of features and feature extraction, training classifiers

with small and unbalanced data sets and making use of prior knowledge (Duin, et al, 2002).

Design of classifiers may be done along a number of strategies depending on the information available. For applications where the class-conditional probabilities are known, the optimal Bayes decision rule can be used for classifier design. Parametric problems are those where the form of the class conditional densities is known (e.g. Gaussian) but some parameters are unknown. Where the form of the class-conditional probabilities are unknown then it is described as a non-parametric problem and the density function can either be estimated or the decision boundary directly constructed from the training data (e.g. k-nearest neighbour). For the majority of real-world classification problems, the underlying cause probabilities and class-conditional probabilities are unknown (Jain, et al, 2000).

Irrespective of the classifier used, the performance depends on the number of training samples available and values of these samples. The goal of a classification is not to maximise the classification on the training set but to generalise its classification performance to data which was not included in the training set.

2.4 Classification of Power System Events

Extensive work has been conducted internationally towards developing pattern recognition techniques for recognition of power system events. This includes identifying the presence and classifying the faulted phases e.g. single-phase-to-ground fault or phase-to-phase (Ziolowski, et al, 2007) and fault location (Mora-Florez, et al, 2009). These studies make use of both simulated and measured data from fault recorders on power systems. Much less work has been conducted with the aim of identifying the underlying causes of events (Gu & Styvaktakis, 2003). One such study uses the CN2 induction algorithm to determine a set of classification rules to identify four causes (lightning, tree, cable and animal caused faults) for distribution networks (Barrera, et al, 2012).

A number of efforts have been made to classify events that occur on distribution power systems according to their root causes. Work focusing on identifying animal-caused faults on distribution systems includes: using discrete wavelet transforms in combination with artificial immune systems (Xu & Chow, 2008), Bayesian networks (Teive, et al, 2011), artificial neural networks (Chow, et al, 1993) and fuzzy systems (Meher & Pradhan, 2010). Artificial neural networks (ANN) and linear regression have been used to classify tree- and animal-caused faults based on distribution utility outage data (Xu & Chow, 2006). Other methods applied to automatically diagnose the root cause of faults include support vector machines (Barrera, et al, 2012), expert systems for classifying events from measurements i.e. voltage step change, transformer energizing (Styvaktakis, et al, 2002), as well as linear discriminant analysis (Cai & Chow, 2009). The majority of these studies have

focused on outages and faults on distribution networks, with very few studies aimed at root cause identification of transmission line faults (Ravikumar, et al, 2008).

2.4.1 Feature Subset Selection

Feature subset selection is the process of finding a subset of the original features that generates a classifier with the highest possible accuracy. This is done by selecting from an existing set of features as opposed to constructing new features (Kohavi & John, 1997). Feature subset selection may be done using domain knowledge to select ad hoc features or by feature selection methods, which can be divided into 3 categories (Guyon & Elisseeff, 2003):

1) filter methods select/rank features based on criteria that are independent of a classifier (this often results in features which are not optimised for a specific classifier); 2) wrapper methods select an optimal subset of features tailored to a given classifier and 3) embedded methods that conduct feature selection as part of the process of training a specific classifier (i.e. the training and feature selection cannot be separated).

Feature selection is usually applied to meet one or more of the following objectives (Guyon, 2008):

- Improve classification performance
- Reduce computational requirements
- Reduce data storage requirements
- Reduce cost of future measurements
- Improve data or model understanding.

2.5 Conclusions

This review has identified birds, lightning, fire and pollution as the primary causes of faults on South African transmission lines. Seasonal patterns of occurrence are recorded for fire, lightning and bird streamer faults, with bird streamers also showing a distinct diurnal pattern of occurrence.

Analysis to recognise the impact of season and time-of-day has been found for faults occurring in the USA and South Africa. A 4 by 4 matrix for time-dependent analysis is highlighted.

The characterisation of fault waveforms to identify fault causes has been shown to have practical value; however efforts in this direction have been limited due to a shortage of measurement data. Characterisation has been conducted for internally caused faults as well as distribution faults with limited datasets of voltage and current measurements. The following features have been identified as applicable for characterising fault waveforms: maximum zero sequence current and voltage, fault insertion phase angle (FIPA), maximum change of current and voltage magnitudes (phase and neutral) as well as maximum arc voltage.

A description of a statistical pattern recognition system has been presented along with a discussion on factors that need to be considered in the design thereof: choice of features, feature extraction as well as the training and design of the classifier.

Work conducted to classify faults according to cause has used a number of classifiers including: artificial neural networks, the CN2 algorithm, fuzzy systems, expert systems, support vector machines and linear discriminant analysis.

CHAPTER 3

3. ANALYSIS OF TRANSMISSION LINE FAULTS IN RELATION TO WEATHER AND CLIMATE

This chapter presents an analysis method relating frequency of faults on overhead lines to local climate. Fault analysis by time-of-day and time-of-year (season) is presented and the statistical significance of the differences between mean fault frequencies for fault causes, climate, time of day and season is established.

This analysis proposes the characterisation of power system performance by weather and climate. Since many of the causes of faults are characterised by seasonal or diurnal variation, the mean and variance of the frequency of faults can be analysed for a 4 by 4 matrix of time-of-year or season, and time-of-day (Herman & Gaunt, 2010). The steps of the analysis are:

- Associate each transmission line to a rainfall area i.e. thunderstorm or frontal
- Characterise transmission system fault records according to the major cause types
- Characterise faults by time-of-day and seasonal quadrants.

For all causes of faults, the fault incidence is generally scaled by the line length, representing the exposure to the stress or cause of fault. For this case the grid management regions names adopted in Eskom serve as geographic indicators that can be used to link individual lines to the regional severity as depicted in maps.

3.1 Database and Management

Fault data records were collated in spreadsheet format with standardised line descriptions, geospatial information system (GIS) number assigned to lines, line length, line design and operating voltages, start date and time of event, original fault description, assigned fault cause and sub-cause, and Eskom Transmission grid region. The issues and processes documented for gathering and cleaning up the fault data on the Eskom transmission network for analysis purposes include (Vosloo 2005):

- A lack of knowledge concerning fault mechanisms in earlier years
- Incorrect application of line naming convention
- Changes in line configuration
- Vague descriptions by operators, e.g. storm, where there is no indication of whether the fault is due to a lightning strike or wetting of a polluted insulator and consequent fault
- Recording of “fire” as a cause where investigation records indicated that the fire had been put out prior to the fault occurring.

The data was complemented with data from the Eskom GIS database on transmission lines which includes geographic co-ordinates for each transmission tower (longitude, latitude, height above sea level).

The event spreadsheet was inspected to ensure that all relevant fields were populated, naming conventions were standardised and the GIS number assigned

matched the line description. The GIS number assigned to lines was utilised as the reference for consistent linkage of events to lines.

The ground flash density (N_g) map displayed in figure 3.1 illustrates N_g in 20 km squares. Ground flash density per line is calculated for each square, and each tower within a square is assigned a value for flash density equal to that for the square within which it is located. The flash density per line is calculated by averaging N_g across the towers that comprise a transmission line.

For the purpose of this analysis, the country was divided into two areas indicating the dominant nature of rainfall activity that is present i.e. frontal rainfall activity (F) occurring during winter and thunderstorm rainfall activity (T) occurring during summer.

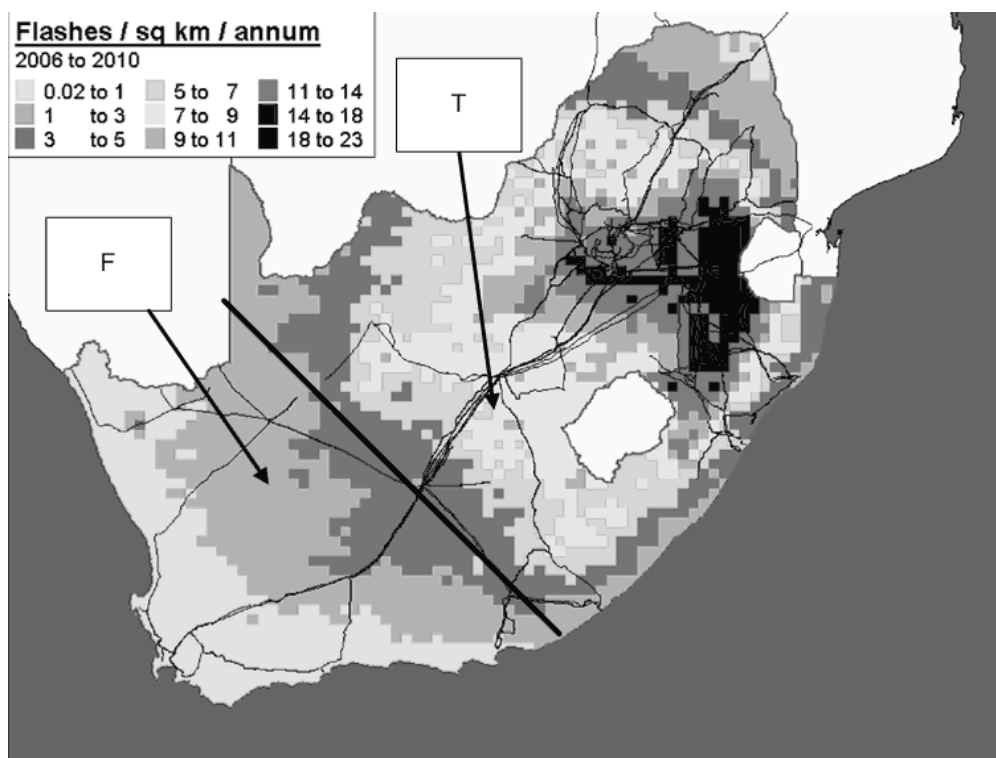


Figure 3.1: Ground flash density (N_g) map with transmission lines of South Africa for 2006-2010 (Eskom, 2010).

A database was constructed that:

- grouped Eskom's transmission grid regions according to the thunderstorm and frontal activity areas illustrated in figure 3.1. The geographic areas assigned to the management regions can be divided neatly with the southern and western grid management regions (which cover the western lower lying parts of South Africa) falling in the frontal activity areas and the other four grid management regions falling into the thunderstorm activity area;
- linked the transmission lines with grid regions; and
- linked fault data with respective transmission lines.

The construction of the database in this manner ensures that 1) data integrity is kept with respect to line and regional data, and 2) event data can be analysed with respect to geographic context.

The original faults dataset consisting of 12229 events was reduced to 11753 for analysis due to:

- Events occurring on lines without GIS numbers assigned, and
- Events occurring on lines not described in the lines dataset extracted from the GIS database.

3.2 Transmission System Results and Discussion

3.2.1 Fault Data

A total of 12229 faults occurred on the Eskom transmission overhead lines during the period 1993 to the end of 2009. Figure 3.2 illustrates the breakdown of fault causes according to the primary categories in table 2.1. This fault data was collected by field operators, reviewed and cleaned up as described in section 3.1. It forms the basis for a) the fault analysis conducted and b) the source of data for automatic classification.

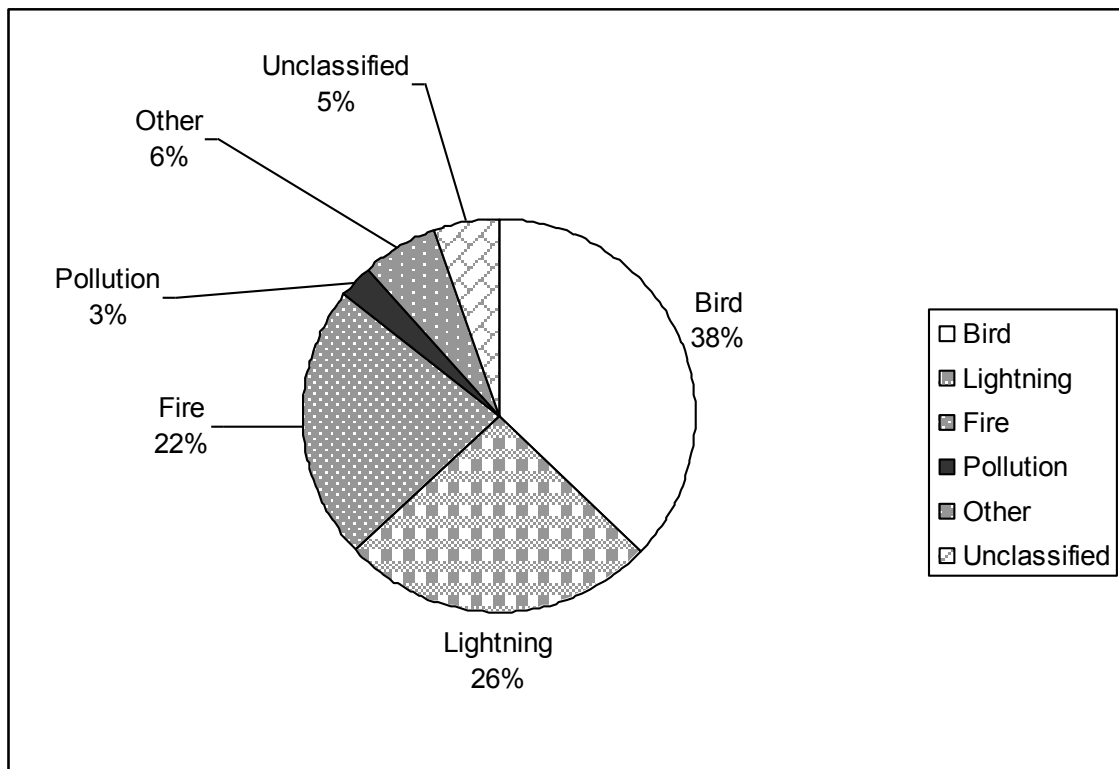


Figure 3.2: Transmission Line Fault Causes - 12229 faults from 132 kV to 765kV.

The four most significant individual causes of faults are birds, lightning, fire and pollution, together causing 89% of all faults.

The average fault frequency for the two rainfall areas, as well as the entire country, depicted in table 3.2 shows that lines located within the summer rainfall area of South Africa have a higher fault frequency than those within the frontal activity area.

Table 3.2: Fault Frequency Statistics.

Rainfall Activity	Total Line Length (km)	Fault frequency (faults/100km/year)	Std Dev
Thunderstorm	18416	2.856	0.649
Frontal	7906	1.581	0.265
Whole country	26322	2.458	0.451

The fault frequency statistics for the major voltage levels are presented in figure 3.3.

These results provide indicative values of fault performance for transmission lines at the respective voltage levels within the identified rainfall areas for planning future networks. Analysis is conducted along voltage levels due to differences in design features at respective voltage levels e.g. clearances, which may influence performance with respect to different fault types.

3.2.2 Fault Frequency by Rainfall Activity Area – 400kV lines

Lines operated at 400kV comprise the bulk of the South African transmission system with extensive exposure to both frontal and thunderstorm activity. Figure 3.3 illustrates quite clearly that the average fault frequency for 400kV lines in the

thunderstorm activity area is significantly higher than those located in the frontal rainfall areas.

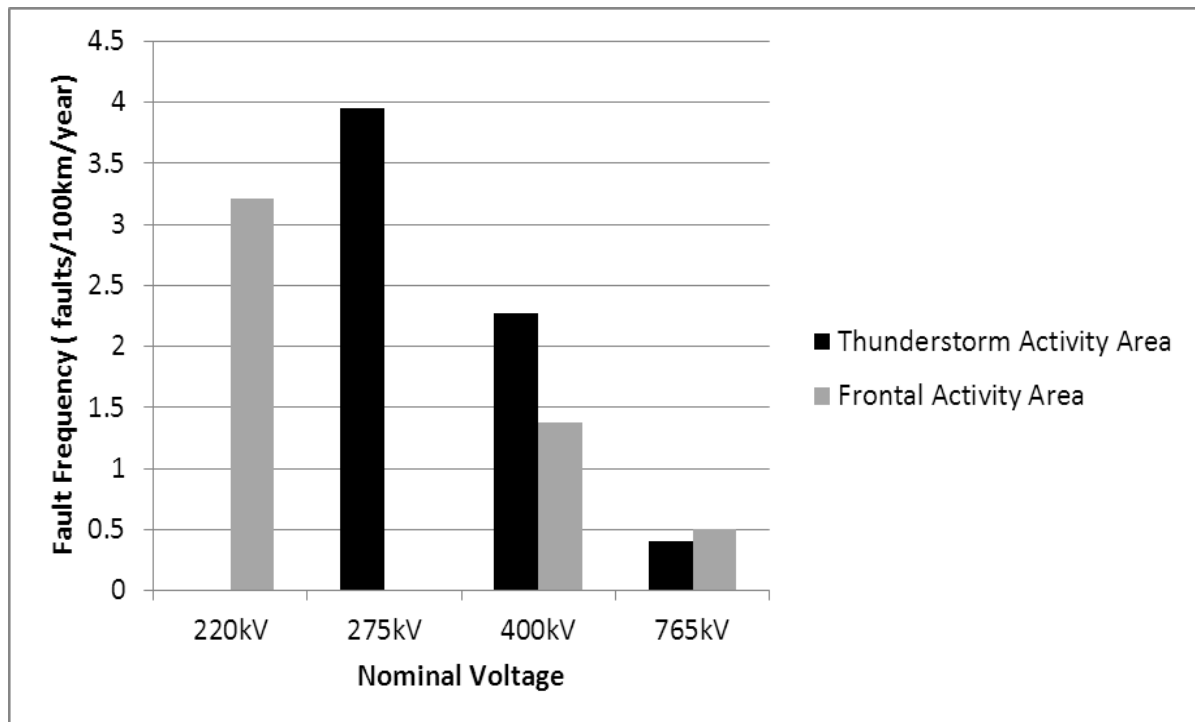


Figure 3.3: Fault frequency per 100 km per year per rainfall activity area

Visually inspecting the fault frequency of 400kV lines based on the primary fault causes shown in figure 3.4, indicates the following:

- The fault frequency for faults caused by fire and lightning is higher in the thunderstorm rainfall area than in the frontal rainfall area
- The fault frequency of pollution faults is higher in the frontal rainfall area compared with the thunderstorm rainfall activity
- There are insignificant differences in fault frequency for bird streamers and “other” fault causes between the two rainfall areas

- using the dominant rainfall activity as a climatic indicator for characterising faults appears to be useful because rainfall and lightning are associated together in the thunderstorm region, and rainfall and fire are indirectly related in that the absence of rain during the winter months leads to dry conditions with high levels of light combustible fuels e.g. dry grass.

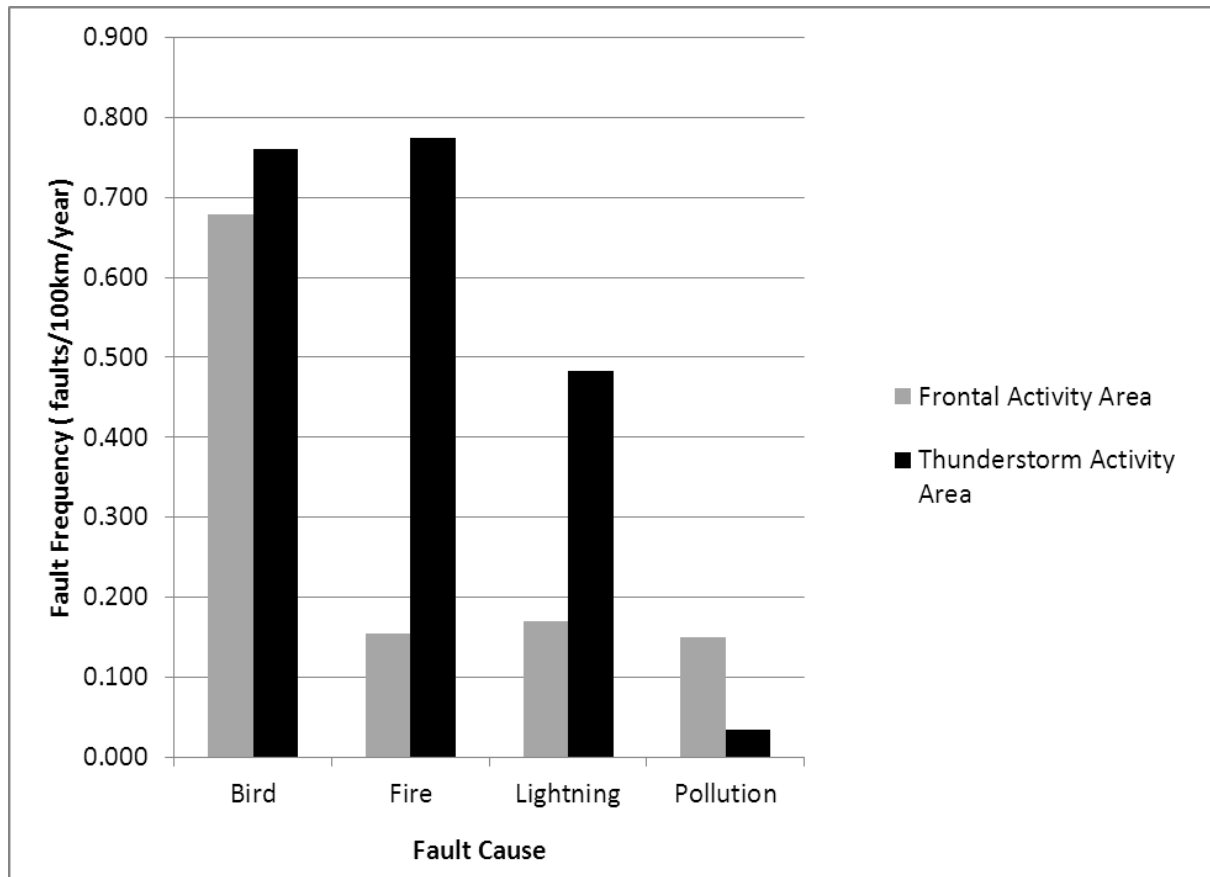


Figure 3.4: Average 400kV line fault frequency statistics by fault cause per 100 km per year

The frequency of faults due to fire, lightning and pollution are directly related to the existing local climate as represented by the nature of the rainfall activity in the area. Further, this analysis provides a clear indicator of the fault causes for which local

climate plays a much smaller role in the frequency with which they occur e.g. bird streamers.

This analysis identifies where relationships exist between local climate and causes of power system faults. The impact of local climate on transmission line performance is illustrated by the differences in overall line performance and the line performance for individual fault causes in the two climatic regions.

3.2.3 Time-of-day and time-of-year analysis

The time-of-day and time-of-year analysis for networks and specific fault causes can be graphically represented using bar charts that associate seasonal and time-of-day intervals with interruption indices in a similar manner to the 4 by 4 matrix intervals proposed for system reliability studies (Herman & Gaunt, 2010). Using a limited number of fault causes and time-season categories ensures sufficient events for statistically significant samples while achieving a useful distinction between them. Table 3.3 presents the diurnal and seasonal fault frequency data (mean and standard deviation) for the 400kV lines in a 4 by 4 matrix.

Table 3.3: Season and time dependant characterisation (mean and standard deviation) of the frequency of bird streamer faults on the South African 400 kV networks

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400 kV				
<i>Season 1: Jan-Mar</i>	0.104; 0.032	0.026; 0.018	0.003; 0.006	0.04; 0.024
<i>Season 2: Apr-Jun</i>	0.11; 0.055	0.056; 0.022	0.004; 0.006	0.061; 0.035
<i>Season 3: Jul-Sep</i>	0.064; 0.025	0.032; 0.014	0.005; 0.008	0.037; 0.02
<i>Season 4: Oct-Dec</i>	0.078; 0.034	0.009; 0.006	0.005; 0.007	0.027; 0.014

For ease of illustration, the 4 by 4 matrix data is represented as bar graphs in figures 3.5-3.10 of this chapter.

Figure 3.3 illustrates the diurnal and seasonal patterns commonly associated with bird streamers on the 400 and 275kV networks.

3.2.4 Bird Streamer

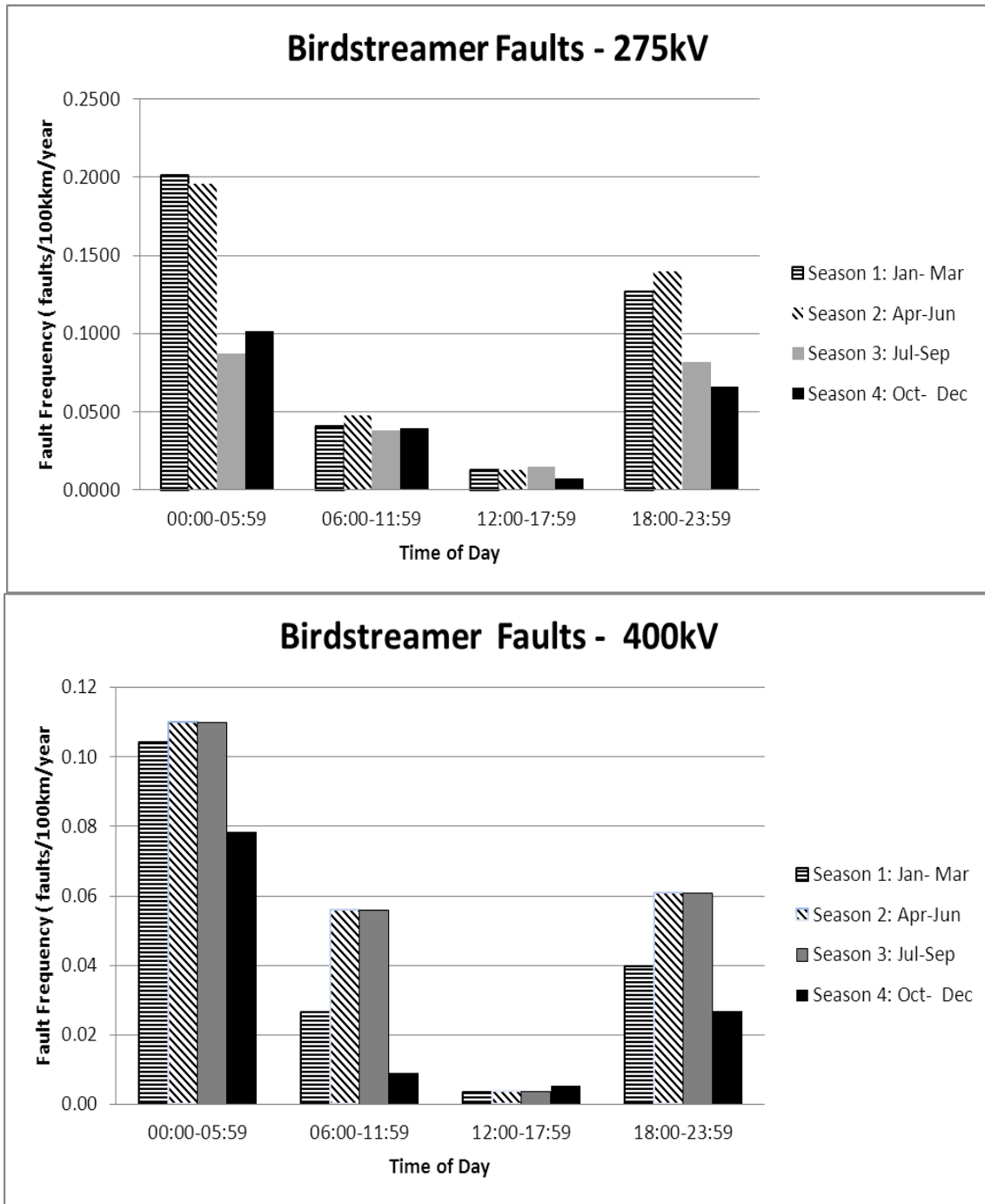


Figure 3.5: Season and time dependent frequency of bird streamer faults on the South African 400 and 275kV networks

Fault frequency on the 400kV network is at its peak of 0.11 faults/100km/year from January to September in the early hours of the morning until 06:00.

Table 3.3 and figure 3.5 present the diurnal and seasonal mean and standard deviation of fault frequency/100km/year for bird streamers on 400kV and 275kV transmission lines.

Bird streamer faults on the 275kV network have a greater frequency of occurrence than on the 400kV network. The peak frequency for the 275kV network is nearly double the peak frequency occurring on the 400kV network. The underlying causes for this could be related to the tower design and clearance distances used on the towers. This is supported by findings indicating that increased vertical clearance between the conductor and the tower results in fewer bird streamer faults (Vosloo et al. 2009).

Eskom embarked on a project to install birdguards on transmission lines with a high incidence of faults due to bird streamers from 2000-2002. The impact of these birdguards is shown in figure 3.6, which illustrates the sharp drop in bird streamer faults on the transmission lines which have been fitted with these devices. Birdguards are typically fitted to feeders with a higher exposure to birds, resulting in more bird-related faults when compared to the general population of feeders on the transmission network.

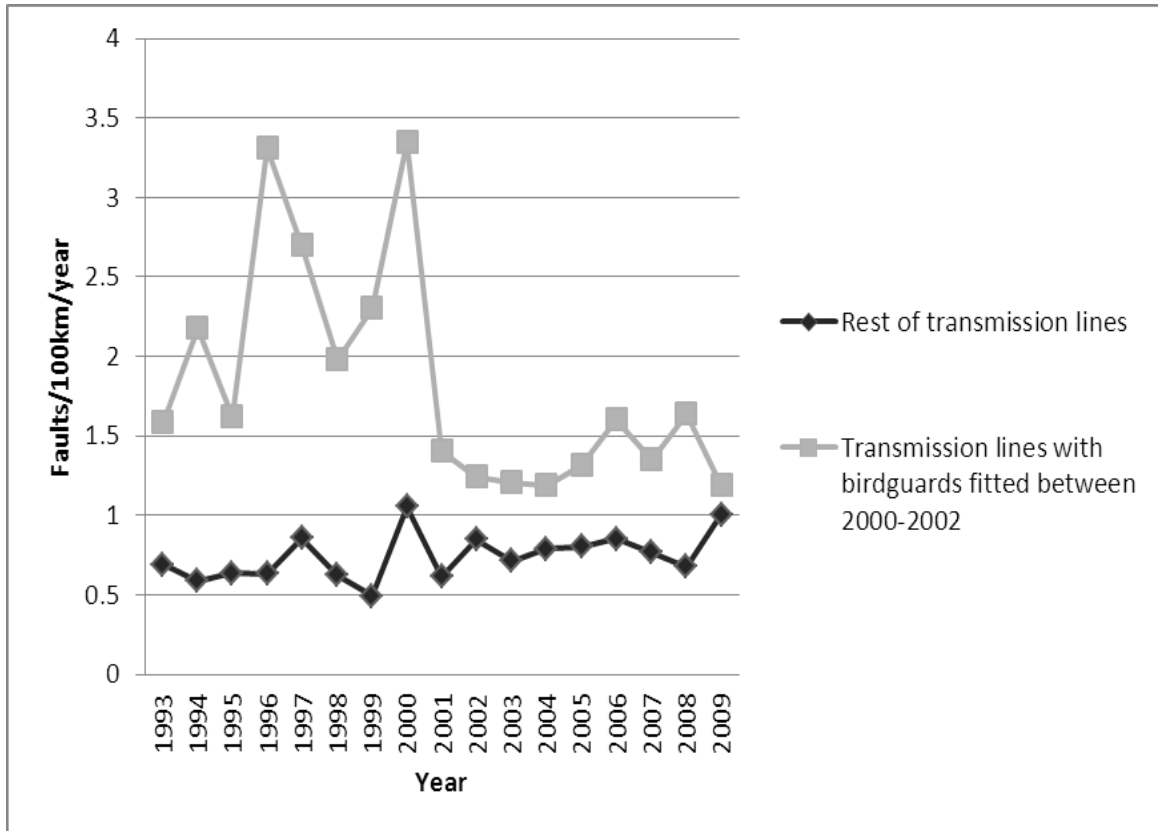


Figure 3.6: Fault frequency on 400kV, 275kV and 220kV lines fitted with birdguards

Figure 3.6 illustrates that the average fault frequency for these lines declined from 2.38 faults/100km/year before birdguards were fitted to 1.35 faults/100km/year after birdguards were installed from 2000 onwards. The performance of lines where it was deemed necessary to install birdguards was significantly higher than the rest of the lines on the transmission network.

3.2.5 Fire

Results are illustrated in figure 3.7 for the 400kV network that falls within the thunderstorm rainfall activity region.

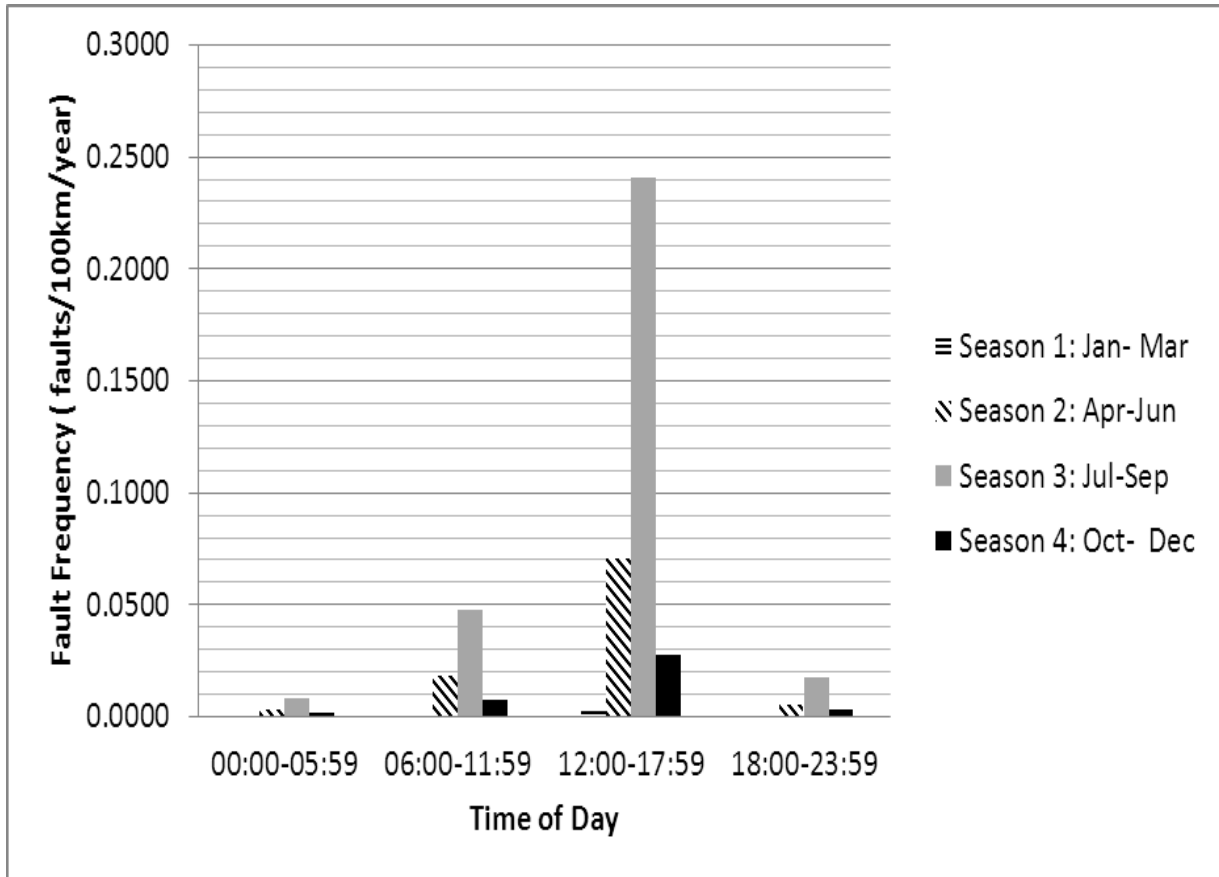


Figure 3.7: Season and time dependent frequency of fire-caused faults on the South African 400kV network

The fault frequency statistics indicate, as expected, higher levels during the drier winter months from April to September with the peak of these faults falling in the three months from July to September. The seasonal and diurnal patterns of fire-caused faults for the 275kV network (not shown) are similar to those for the 400kV network.

3.2.6 Pollution

The results for insulator pollution-caused faults on the 400kV network are illustrated in figure 3.8, showing that the highest incidence occurs in the early morning hours during the period January to March.

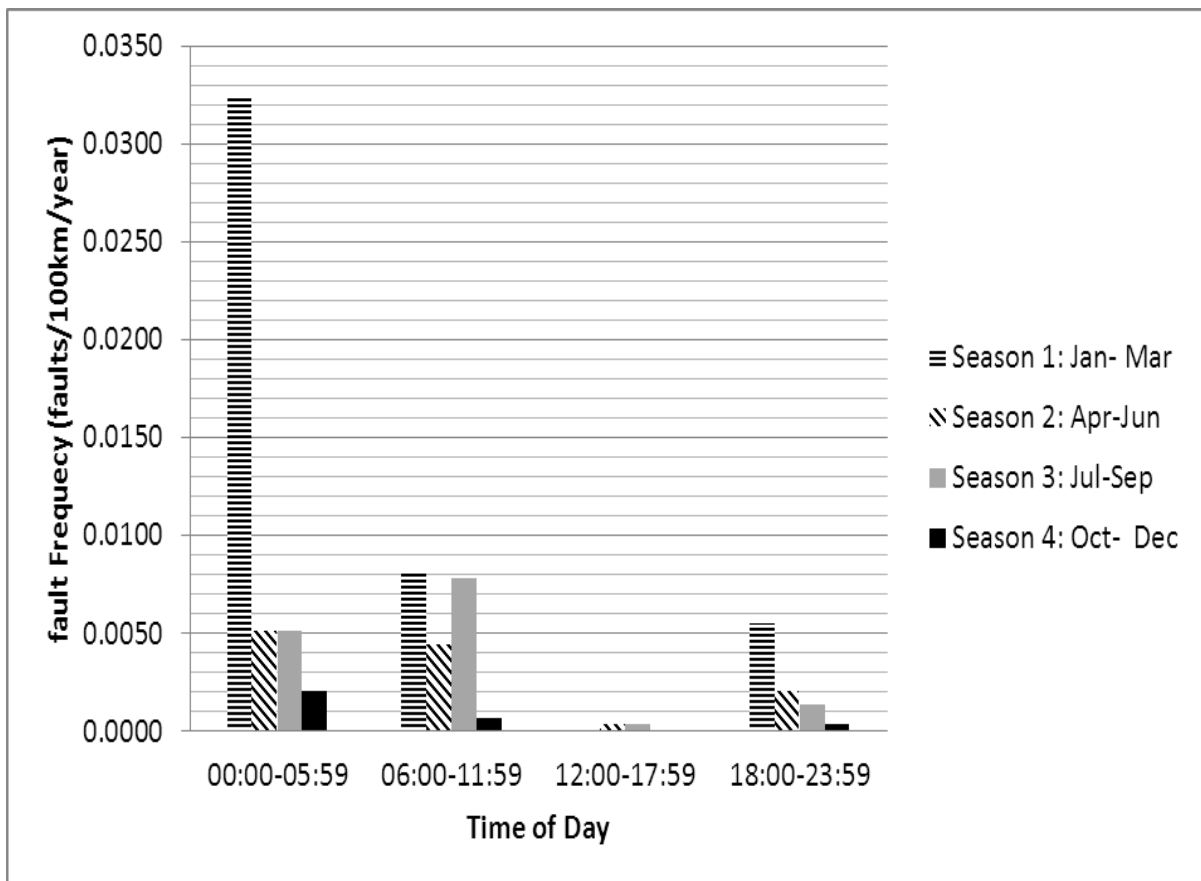


Figure 3.8: Season and time dependent frequency of pollution faults on the South African 400kV network

3.2.7 Lightning

The results of the time dependent characterisation of faults due to lightning for 400kV networks across South Africa are presented in figure 3.9. The results indicate higher levels of lightning initiated faults in afternoons and evenings during the periods Season 1 (January to March) and Season 4 (October to December).

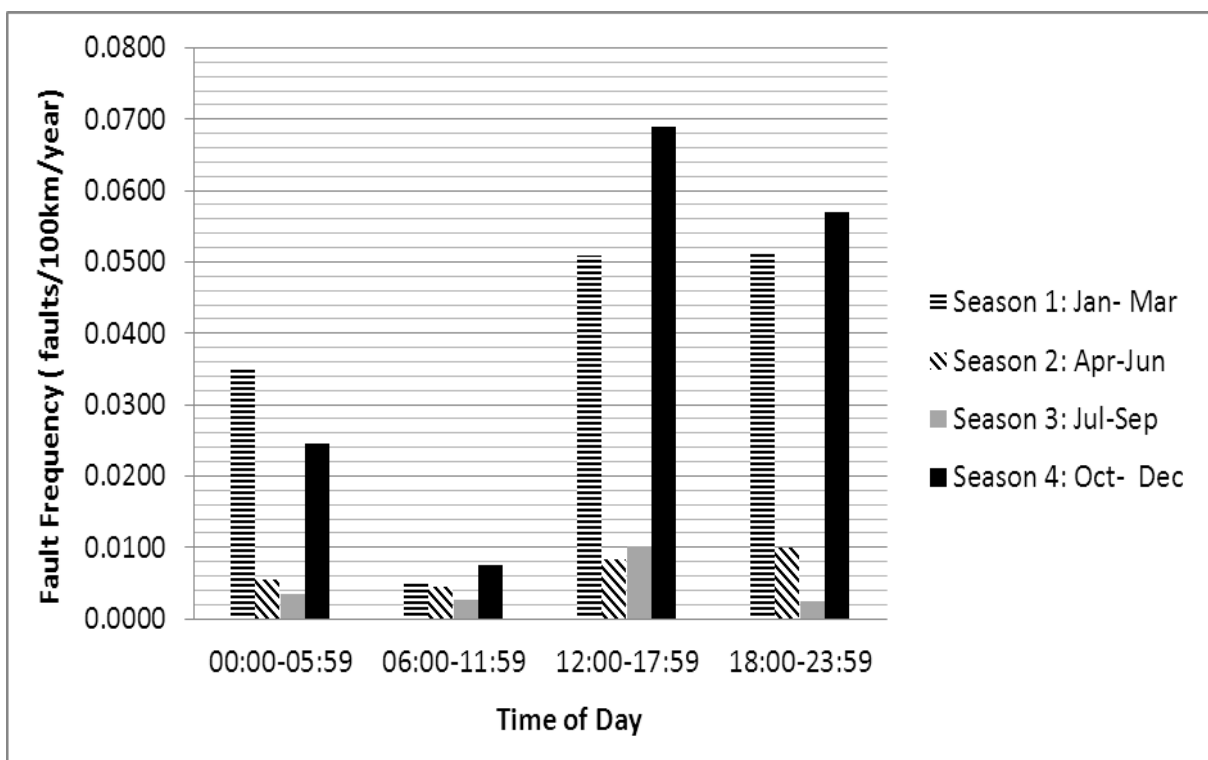


Figure 3.9: Season and time dependent frequency of lightning faults on the South African 400kV network

These months coincide with the summer thunderstorms in South Africa and the results indicate a significant increase in the frequency of faults on the 400kV network, which increase from approximately 0.01 faults/100km/year during the winter months up to 0.069 faults/100km/year during the summer rainfall periods.

3.2.8 Normalisation of the lightning initiated fault data

Allocating scaling factors such as the line length and the lightning incidence (ground flash density, N_g) to all lines in the transmission network, the fault incidence can be identified, normalised to 'per 100 km of line and $N_g=1$ ' for lightning incidence.

The lightning fault data, normalised to $N_g=1$ in this manner, relates frequency of fault directly to the overall lightning exposure of the transmission lines (Minnaar, Gaunt, & Nicolls, 2012). Figure 3.10 illustrates the normalised lightning fault data for the 400kV lines in the thunderstorm activity region of South Africa. The 275kV lines, all located in the thunderstorm activity area, can be analysed in the same manner.

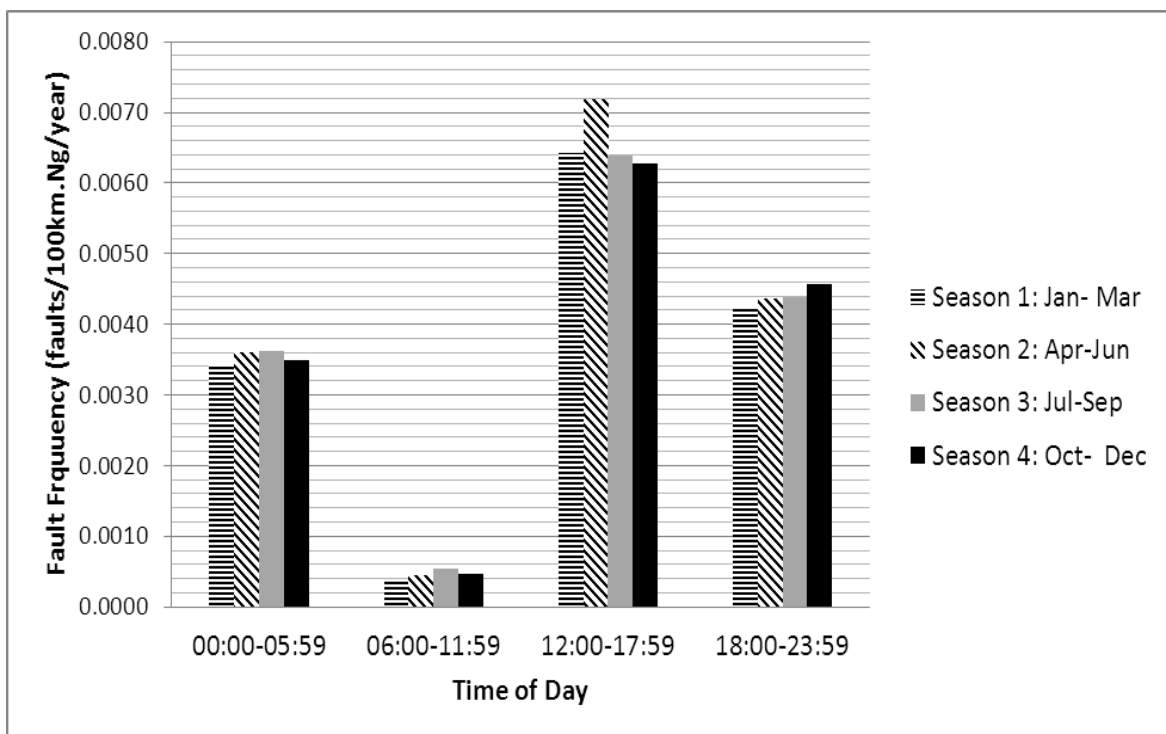


Figure 3.10: Season and time dependent frequency of lightning faults on the South African 400kV network in the thunderstorm activity region normalised to $N_g=1$

A comparison of the two sets of results indicates that for all times of year and time of day, the 275kV network experiences more lightning initiated faults than the 400kV network relative to the exposure of the lines to lightning flashes.

3.3 Establishing the Statistical Significance of variations in fault frequency

Analysing the fault frequency data by rainfall area produces results that indicate variations due to the influence of the climate. Similarly variations are produced due to the time-of-day and season. The statistical significance (to the 0.05 level) of the impact of rainfall area, voltage, time-of-day and season on fault frequency on the 220kV, 275kV and 400kV networks was tested by means of one- and two-way analysis of variance (ANOVA), for which results are given in table 3.4 ('yes' indicates statistically significant).

Table 3.4: Statistical Significance of factors influencing fault frequency

	Voltage level (kV)	Thunderstorm Region		Frontal Region	
Variable					
All faults Combined	220, 275, 400	Yes		Yes	
		Season	Time of Day	Season	Time of Day
Fire	220	na	na	No	No
Lightning	220	na	na	No	Yes
Fire	400	No	No	No	Yes
Fire	275	No	Yes	na	na
Lightning	400, 275	Yes	Yes	Yes	No
Bird Streamers	400	Yes	Yes	Yes	Yes
Bird Streamers	275, 220	No	Yes	No	Yes
Pollution	400, 275, 220	No	Yes	No	No
Other	400, 275, 220	No	No	No	No

The results show that while figure 3.5 indicates a visibly different response by season for bird streamer faults on the 275kV network, this difference is not statistically significant in the present dataset.

The 4 by 4 matrix (as illustrated via bar graphs in figures 3.5-3.10) allows statistically significant classification of fault frequency according to climate, season and time-of-day for the major causes of faults. This is a quite different relationship from considering all faults combined which do not show statistically significant variations. The full dataset of fault frequency on the South African transmission network for 220kV, 275kV and 400kV lines is presented in Appendix A.

The analysis of this dataset provides a statistical analysis of the fault performance of a large transmission system over an extended period. High data volumes are available for fault types which occur frequently across the network or in particular areas. This significant volume of data implies that statistical analysis can be conducted with a high degree of confidence for the entire network.

The classification of fault frequency by climate, season and time-of-day and voltage level for the major causes of faults provides insight into specific areas where fault types that occur less frequently (e.g. pollution) may be analysed in a specific area. The dataset may be relatively sparse on certain fault causes across the entire network, however in the areas where these fault types are prevalent (e.g. pollution in certain coastal areas) the analysis identifies this and allows for remedial action to be taken.

3.4 Analysis of a single class of faults

In addition to the analysis of the characteristic season and time of faults, the dataset allows analysis within a single class of faults. For example, having defined the incidence of lightning faults on 400kV lines in the thunderstorm region in terms of the faults/kmNg to remove the variation caused by line length and lightning intensity, it was found there was significant scatter in the parameters. Figure 3.11 shows that there appear to be at least two (maybe more) families of lines with different failure performance, separated by the dashed line.

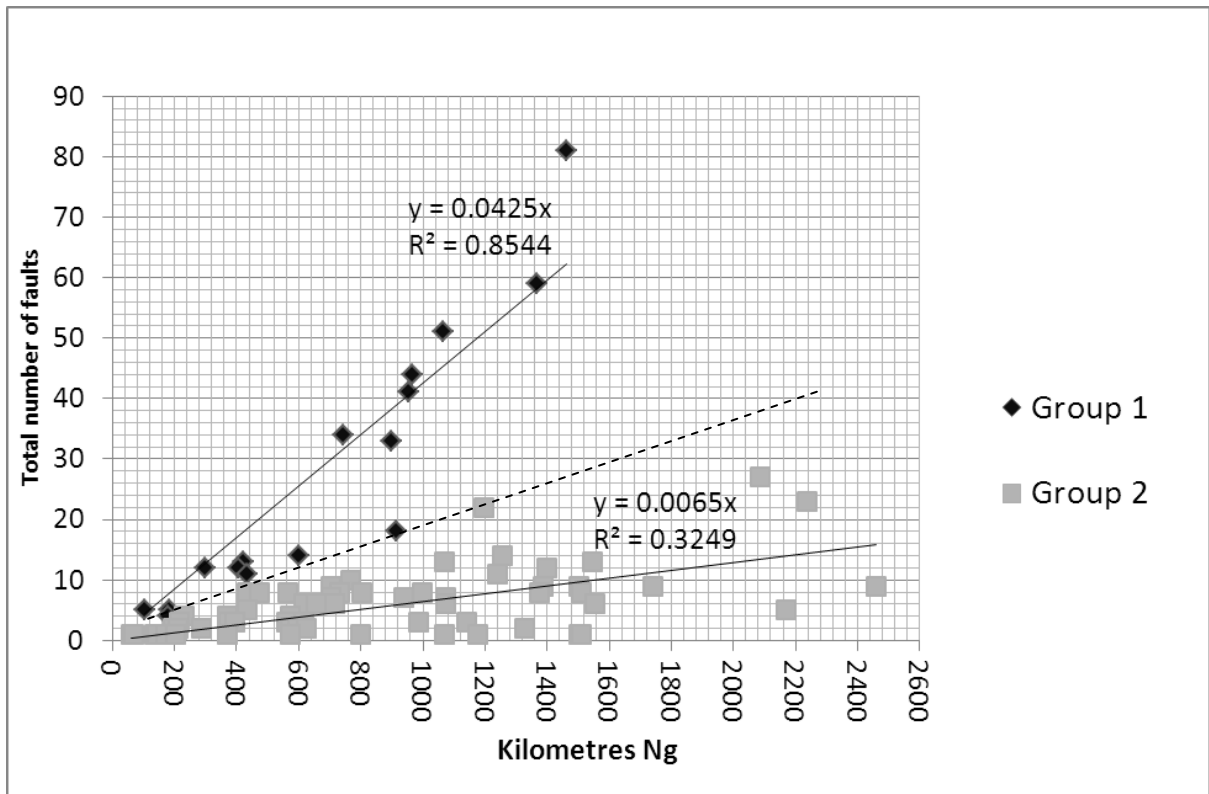


Figure 3.11: Lightning faults on 400kV transmission lines in the thunderstorm rainfall activity area (Group 1- Poorly Performing Lines, Group 2 – Rest of Lines) with Y = number of faults per Kilometres Ng and R^2 =coefficient of determination

The population of lines was split according to the performance to investigate whether there are common factors distinguishing the groups from each other. The analysis shows there are such factors, including altitude and tower footing resistance, indicating the possibility of improving line performance by intervention on existing lines and the specification of alternative tower designs and installation parameters for future lines.

The information of the poorly performing lines (Group 1) was passed on to the Eskom transmission division to investigate remedial actions to improve the fault performance of these lines.

3.5 Conclusions

The primary causes of faults on the South African transmission system are identified as bird streamers (38%), lightning (26%) and fires (22%).

An analysis method is presented that relates the frequency of faults on overhead lines to the climate of the area in which the line is located and the causes of power system faults.

The effectiveness of birdguards in reducing fault frequency ascribed to bird streamers has been clearly demonstrated.

Faults analysed by time-of-day and time-of-year (season) provide fault frequency statistics that represent more information than average annual frequency, taking the network performance into account during different time periods. The statistical significance of the differences between mean fault frequencies for fault causes, climate, time of day and season is established.

The information derived from the collection and analysis of fault frequency data leads to three very different applications. One will be better modelling of the whole system's reliability, using interruption duration and network loading data similarly classified by a 4 by 4 matrix, with substantial implications for both planning and system operations. A second application group will be on design and selection of parameters for specific lines and identifying lines with relatively poor performance needing to be improved. Thirdly, the approach lays a foundation for future work to be conducted into reliability analysis and electrical fault pattern recognition taking local geography, climate and power system parameters into account.

CHAPTER 4

4. WAVEFORM CHARACTERISATION OF TRANSMISSION LINE FAULTS

This chapter describes the characterisation of transmission line fault waveforms by instantaneous symmetrical component analysis for transient and steady state fault conditions. A set of characteristics are developed as a feature set for fault cause classification.

Characterisation and analysis of external faults based on the voltage and current waveforms has been investigated for causes such as lightning, tree and animal contact and cable faults (Barrera, et al., 2012). Several features considered by Barrera et al (2012) have been retained in this study, albeit in a modified form. Maximum zero sequence current and voltage is defined in this study, relative to pre-fault levels. This enables measurements taken at different voltage levels to be considered together, as well as relating these maximum values to pre-fault conditions. The fault insertion phase angle (FIPA) is considered here as the large dataset will give a clear indication of the relationship between fault types and fault peak.

Features utilised by Barrera et al (2012) but not considered here include:

- Maximum change of current and voltage magnitudes (phase and neutral), calculated over a time period one quarter of a cycle before and after the fault initiation. These features were found to be relevant to distinguishing cable faults, which are not considered in this work.

- Maximum arc voltage. This feature is only applicable to single-phase to ground faults.

This chapter describes the characterisation of transmission line fault waveforms using instantaneous symmetrical component analysis. A total of twenty one individual waveform characteristics are extracted for use as input features to classify faults.

4.1 Waveform Characterisation

Fault measurements were available for 2672 out of the 11753 faults in the dataset. Fault measurements were checked to ensure that they provided adequate pre-fault data and measurements were available for all voltage and current channels. The measurement data utilised in this study is taken from 78 digital fault recorders (primarily SIMEAS-R and Siemens P513 devices), on the Eskom transmission network at 220kV, 275kV and 400kV over a period of 13 years from 1995 until 2008. Current and voltage waveforms are sampled at 2500 Hz.

A similar database of fault measurements, sourced from Scottish Power, was utilised by Styvaktakis (2002) for analysis and expert system classification of faults. Fault measurements of this nature are primarily used to determine the correct operation of protection equipment and the underlying causes are not routinely associated with individual measurements.

4.1.1 Management of Waveform Data

The original waveform data recorded on digital fault recorders is stored in the IEEE C37.111 Common Format for Transient Data Exchange (COMTRADE). These files each represent a unique fault event measurement. The following data is extracted for characterisation: sampling rate, start date, start time, faulted phase, distance-to-fault, red phase current and voltage, white phase current and voltage, blue phase current and voltage, neutral phase current and voltage, number of samples and time.

The data is imported into a Matlab structure with each of the abovementioned items stored within the structure. An array of structures is compiled with each individual measurement being a unique structure in the array (file name is associated for identification). This is implemented to enable bulk signal processing of waveform data by repeating the same calculations inside a loop to obtain the desired waveform characteristics.

4.1.2 Symmetrical Component Analysis

The Fortescue symmetrical component transformation is applicable to the steady state conditions that follow the fault transient condition (Fortescue 1918).

The zero (0), positive (1), and negative (2) sequence components of the voltage are given in terms of the phase voltages (a,b,c) by:

$$V_0 = \frac{1}{3}(V_a + V_b + V_c)$$

$$V_1 = \frac{1}{3}(V_a + kV_b + k^2V_c)$$

$$V_2 = \frac{1}{3}(V_a + k^2V_b + kV_c)$$

$$\text{where } k = \frac{1}{2} + \frac{1}{2}j\sqrt{3}$$

and V_a, V_b, V_c are phase voltages.

Lyon (1954) found that the same approach of symmetrical component transformation can also be used to analyse faults in the transient condition. This study utilises instantaneous symmetrical component analysis for feature extraction as it allows a fault to be analysed during both the steady state and transient conditions.

A symmetrical sequence component transformation is implemented in the Matlab Simulink environment, built around the discrete 3-phase sequence analysis block.

The code to calculate waveform characteristics utilises structure arrays in Matlab, so as to make possible the bulk signal processing necessary for 2672 waveforms. The Simulink model calculating sequence components from phase voltage and current data is then called from inside a 'for' loop to calculate the necessary parameters for each individual measurement, which in turn is also stored inside a

second structure array. The same process is then followed to calculate all the characteristics described in this chapter.

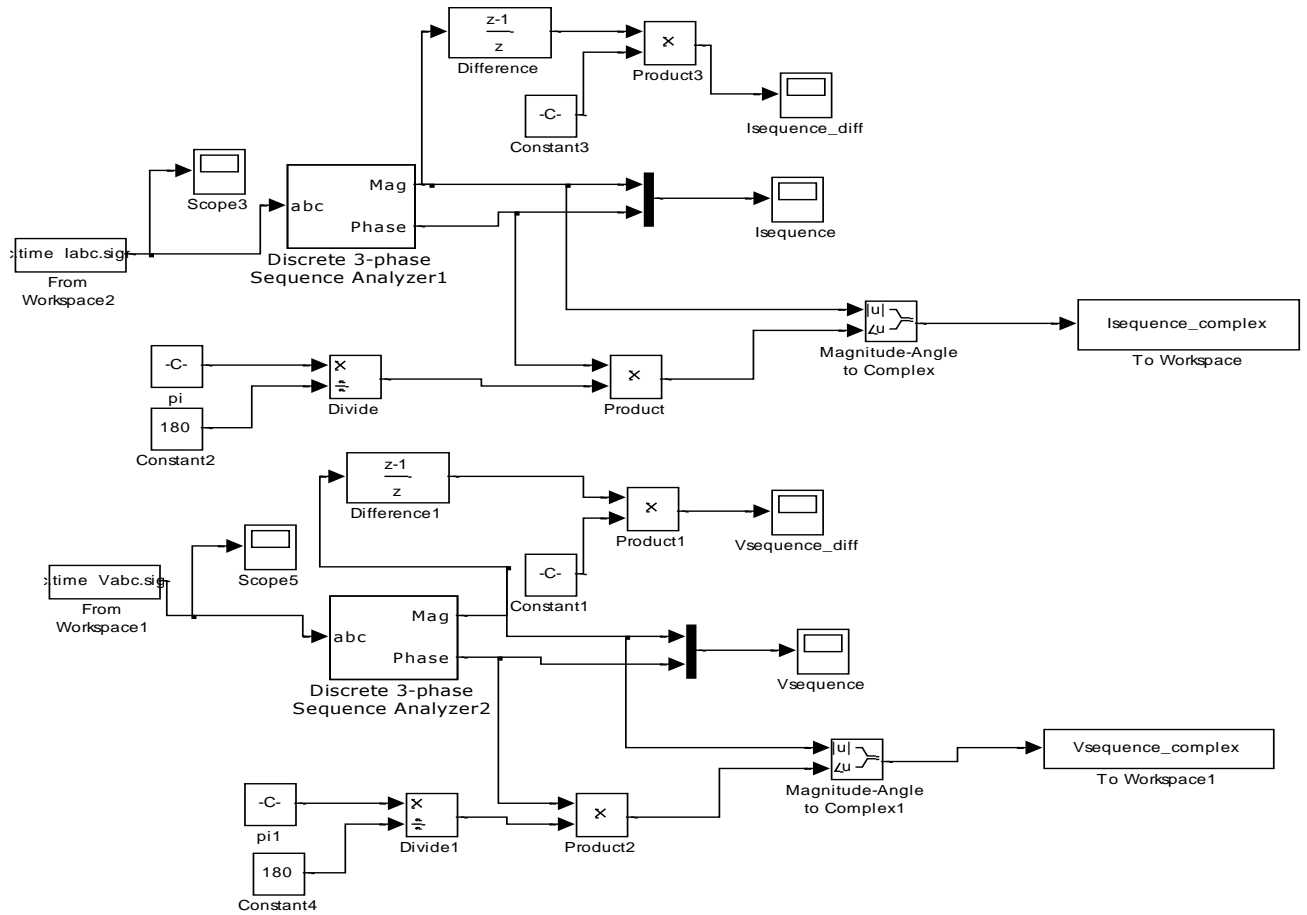


Figure 4.1: Simulink model-Discrete Symmetrical Components.

Figure 4.1 illustrates the model implementation calculating the discrete symmetrical values for voltages and currents, calculated according to Fortescue (1918). The outputs of the Simulink model are discrete waveforms of magnitudes and phase angles for the positive, negative and zero sequence current and voltages. Sequence component currents and voltages are output in complex format and the rates of change for voltage and current sequence components are also exported.

These are used to calculate a range of characteristics for transmission line faults to be used as features for fault cause classification.

4.2 Identifying the start and end of a fault

Identifying the both the start and end of a fault is critical to the success of extracting features from the waveform. The start of the fault separates the pre-fault conditions from the faulted conditions and allows calculation of features only during the faulted segment of the measurement. The three stages of the measurement are illustrated in figure 4.2.

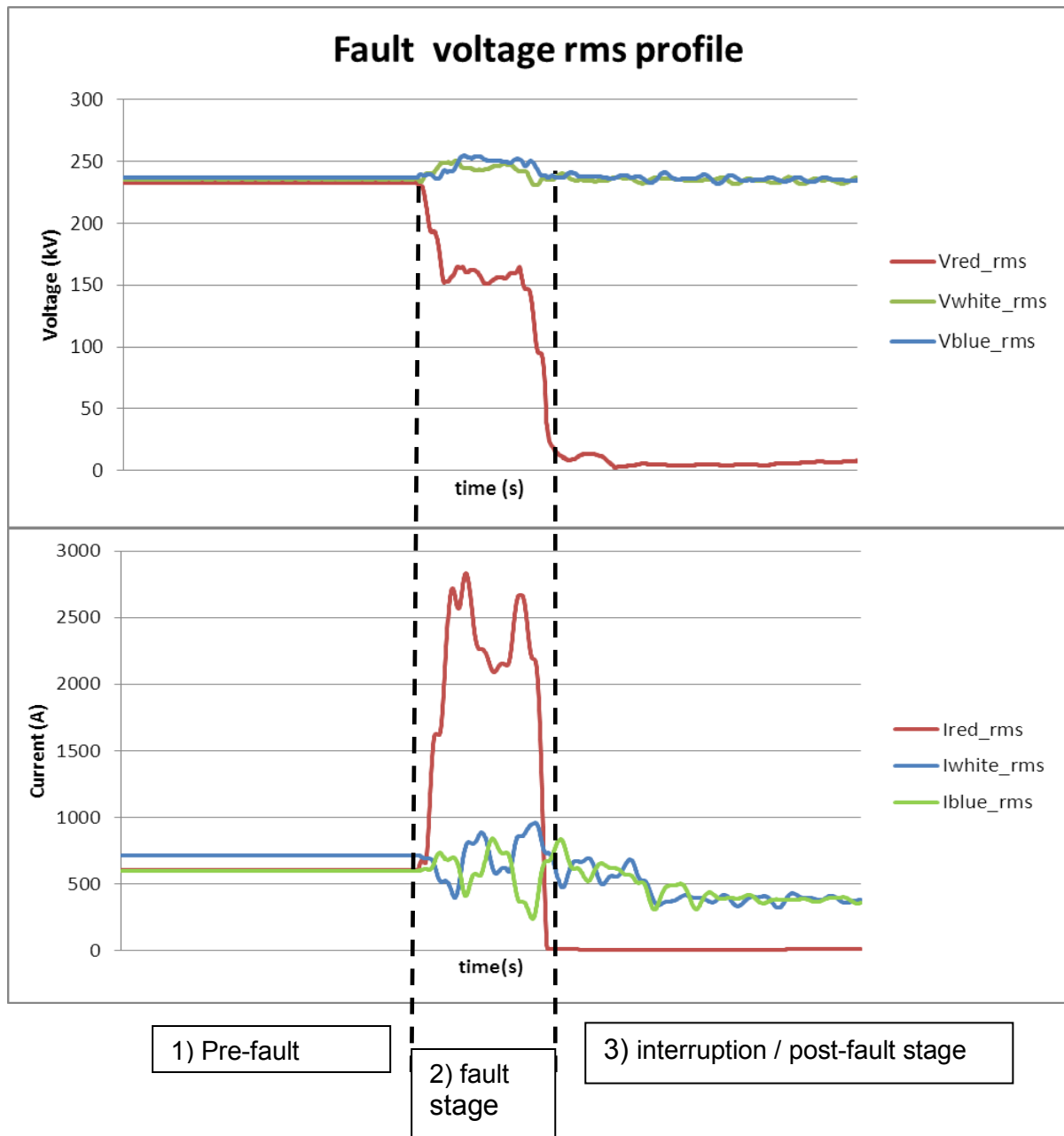


Figure 4.2: Stages of a fault measurement.

Figure 4-2 illustrates the stages of a fault using rms profiles. These are pre-fault (normal steady state operation), the fault and the interruption stage. The scenario illustrated is for a single-phase-to-ground fault.

Fault detection is based on a detection index similar to a rms-base index used for segmentation of rms voltage measurements (Styvaktakis, et al, 2002). The

beginning of the fault is calculated by means of a detection index based on the difference between consecutive values of the positive sequence current:

$$dI = I_{1(n+1)} - I_{1(n)} \dots\dots\dots (1)$$

A threshold is set at 12000 kA per second to trigger the start of a fault event. Features are calculated from the extracted sequence component waveforms.

In the majority of the available measurements, the actual measurement continues beyond the end of the fault (i.e. after the protection has operated), resulting in values that needed to be removed from the measurement. The end of a fault was set using a 40ms moving average of zero sequence current. A fault 'ends' when the zero sequence moving average drops below a threshold of 15% of the peak zero sequence current measured during the fault. Both the fault start and end thresholds were determined by trial-and-error until the start point and endpoint of all the fault measurements in the dataset were successfully captured.

Factors such as the fault level, pre-fault loading level, type and location of load, network configuration or capacitors being switched may have an influence on the resultant fault waveforms measured on the same line. Several of the characteristics calculated from the measurement waveforms are all calculated to compensate for this influence by the use of the following strategies: 1) peak/maximum values are calculated relative to pre-fault values, which provide results with reference to network conditions prior to the fault and enables measurements from different voltage levels to be considered together, 2) maximum rates of change are

calculated to identify the influence of the fault flashover mechanism and 3) characteristics are calculated to compare the magnitude of characteristics at specific time intervals from the start of the fault i.e. the development of the fault.

4.3 Waveform Characteristics

The waveform features extracted are:

- Fault type i.e. single-phase-to-ground, phase-to-phase faults
- Current rate of change (positive, negative and zero sequence sequences)
- Maximum negative sequence voltage
- Maximum Zero sequence voltage
- Maximum sequence current, relative to pre-fault currents (positive, negative, zero sequences)
- Current Magnitude, half cycle after fault initiation (positive, negative, zero)
- Current Magnitude, one cycle after fault initiation (positive, negative, zero)
- Fault resistance (1 and 2 cycles after fault initiation)
- Fault initiation phase angle
- Sequence Component Fault Current Time Constant (positive, negative, zero).

4.3.1 Faulted Phases

This feature is based on the hypothesis that the phases faulted may be a consequence of the physical flashover mechanism of the underlying fault cause. An example of this can be seen in table 4.1, where pollution-caused faults are almost exclusively single-phase-ground-faults.

The faulted phases according to underlying cause for the 220kV, 275kV and 400kV networks are shown in table 4.1. The number of phases faulted is indicated by 'L' and fault to ground is indicated by 'N' so for example a single-phase-to ground fault is indicated as 'L-N' and a three phase fault as 'L-L-L'.

Table 4.1: Faulted phases according to underlying cause.

Faulted Phases	Bird streamer	Fire	Lightning	Other	Pollution	Grand Total
L-N	1070	453	534	261	122	2440
L-L-N	15	33	66	8	1	123
L-L		25	1	2		28
L-L-L	30	4	32	4		70
L-L-L-N	4	3	3	1		11
Grand Total	1119	518	636	276	123	2672

Table 4-1 illustrates that more than 90% of all faults measured on the South African transmission network between 1995 and 2008 were single-phase-to-ground faults.

4.3.2 Maximum Rate of Change of Current (ΔI_{max})

These characteristics are extracted as they provide a picture of the dynamic state of the fault during the initial transient stage.

This component is calculated as the maximum difference between consecutive samples, once the fault has been triggered. This is done for the positive, negative and zero sequence components. The maximum rate of change for each sequence component is treated as an individual feature. The measure is shown in equation 2.

$$\Delta I_{max} = \max(I_{n+1} - I_n) \dots \dots \dots (2)$$

During the initial stages of a fault, current increases rapidly from pre-fault levels until the fault is established. This feature is chosen to establish whether there is any difference due to underlying cause in the rate at which fault current rises during the initial stages of a fault. The mean (μ) and standard deviation (σ) of maximum rate of change of positive, negative and zero sequence currents are shown in table 4.2, denoted as Amperes per millisecond (A/ms).

Table 4.2: Basic Statistics for Maximum Rate of Change of Current

Characteristic	220kV		275kV		400kV	
	μ	σ	μ	σ	μ	σ
ΔI_{max1}	372.86	249.71	653.27	557.28	637.36	658.87
ΔI_{max2}	381.24	251.81	658.02	536.79	673.24	656.88
ΔI_{max0}	380.87	281.40	577.51	516.21	635.07	693.86

It is apparent from table 4.2 that the voltage level at which a fault occurs affects the maximum rate of change of sequence currents. The mean values increase with voltage magnitude at which a fault occurs. While this may be the case, it is observed that there is a high level of variance in the maximum rate of change across all three voltage levels, indicated by the high standard deviation values.

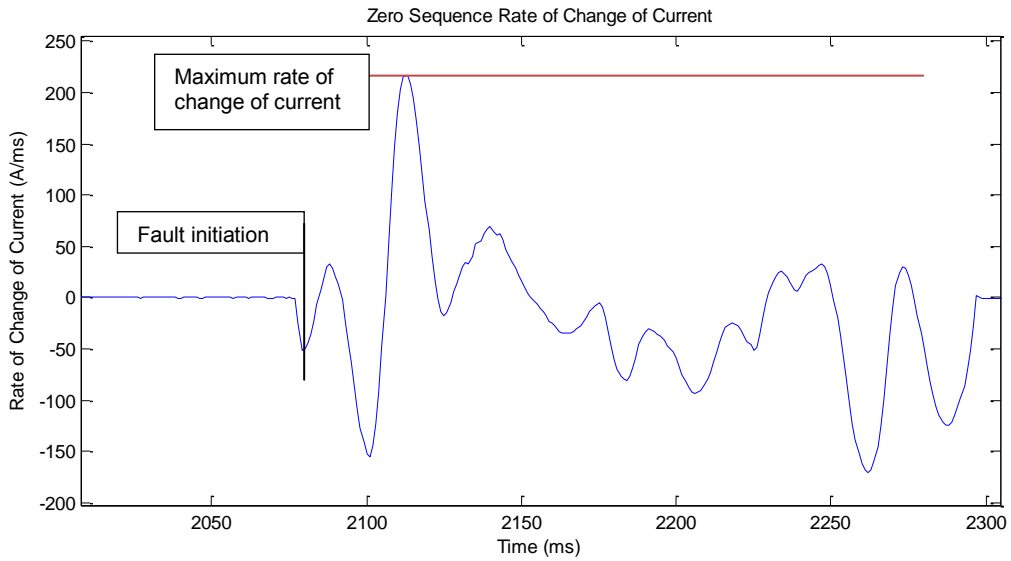


Figure 4.3: Zero Sequence Rate of Change of Current

Figure 4.3 illustrates an example of Rate of Change of Zero Sequence current with the maximum indicated.

4.3.3 Maximum Negative (V_{max2} / Zero (V_{max0}) Sequence Voltage during Fault

These features provide an indication of the degree of unbalance (Barrera et al. 2012) during a fault. The maximum negative and zero sequence voltages during a fault are calculated relative to their respective pre-fault values:

$$V_{max} = \max\left(\frac{V_{fault_max}}{V_{pre-fault}}\right) \dots\dots\dots (3)$$

The mean (μ) and standard deviation (σ) values for maximum zero and negative sequence voltage is shown in table 4.3

Table 4.3: Basic Statistics for Maximum Sequence Voltage during Fault

Characteristic	220kV		275kV		400kV	
	μ	σ	μ	σ	μ	Σ
$V_{\max 2}$	60.38	14.37	70.05	20.01	102.88	26.63
$V_{\max 0}$	61.15	21.40	70.03	29.01	116.06	44.40

4.3.4 Sequence Component Currents - $\frac{1}{2}$ Cycle ($I_{1(0.5)}$) and One Cycle Values ($I_{1(1)}$)

The value of each of the positive, negative and zero sequence component currents are measured at time points of $\frac{1}{2}$ cycle and one whole cycle after the fault is initiated. Figure 4.4 illustrates the zero sequence current components at half-cycle and one cycle after fault initiation.

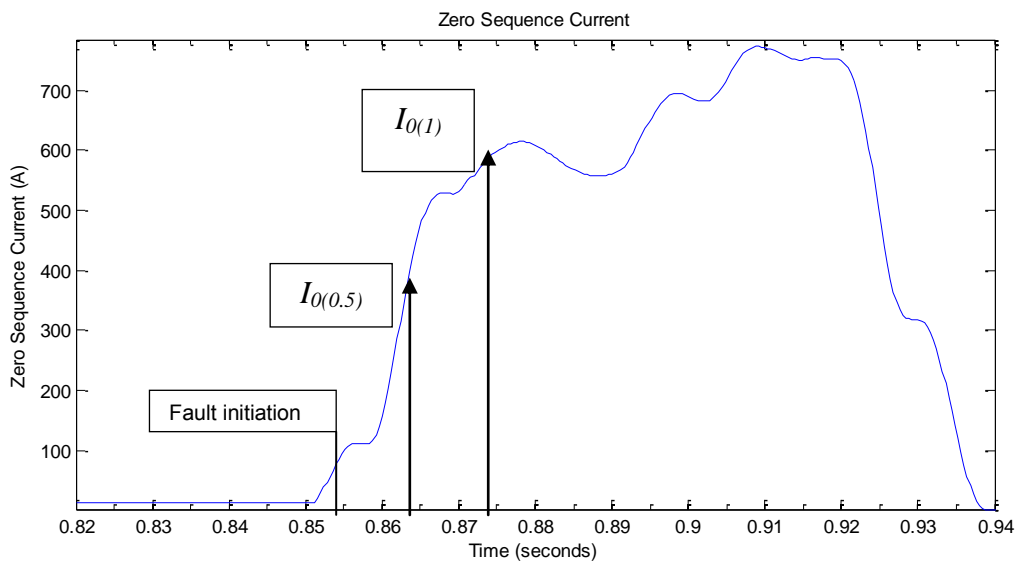


Figure 4.4: Zero sequence current at $\frac{1}{2}$ cycle and one cycle after fault initiation

Table 4.4 shows the mean (μ) and standard deviation (σ) values at $\frac{1}{2}$ cycle and one whole cycle for positive, negative and zero sequence currents.

Table 4.4: Basic Statistics for sequence currents at ½ cycle and one cycle after fault initiation.

Characteristic	220kV		275kV		400kV	
	μ	σ	μ	σ	μ	σ
$I_{1(0.5)}$	3.98	2.66	4.54	9.32	2.63	7.66
$I_{1(1)}$	7.59	5.39	8.42	17.73	4.62	14.94
$I_{2(0.5)}$	96.27	123.43	103.23	208.64	40.98	54.14
$I_{2(1)}$	193.88	215.39	189.14	401.56	77.32	116.98
$I_{0(0.5)}$	252.51	251.90	134.07	221.23	116.16	232.06
$I_{0(1)}$	529.84	537.35	267.69	510.81	217.59	450.86

4.3.5 Maximum Sequence Current, Relative to Pre-Fault Currents

Peak values of sequence currents are calculated relative to the pre-fault current levels on the feeder.

$$I_{max} = \max\left(\frac{I_{fault_max}}{I_{pre-fault}}\right) \dots\dots\dots(4)$$

This gives an indication of the total state of change in current relative to the state of the network/load prior to fault occurrence. This may provide more information than peak values alone as these values may be impacted by the state of the network. Table 4.5 shows the mean (μ) and standard deviation (σ) values for maximum positive, negative and zero sequence currents.

Table 4.5: Basic Statistics for Maximum Sequence Currents.

Characteristic	220kV		275kV		400kV	
	μ	σ	μ	σ	μ	σ
I_{1max}	8.31	5.86	9.62	20.17	8.77	156.99
I_{2max}	211.88	225.03	217.06	436.95	91.97	152.96
I_{0max}	585.02	591.39	307.67	571.11	248.97	502.72

4.3.6 Fault Resistance (R_{fault})

The underlying mechanism by which a fault is formed and the medium along which fault current moves differs for each of the major fault causes. Bird streamer fault current flows via the liquid streamer while fault currents due to fires are conducted via air and smoke particles. The resistivities of these mediums differ significantly. Fault resistance is calculated one and two cycles after the initiation of the fault. The line resistance from the point of measurement to the fault is based on the fault impedance values used for the protection settings for each line. The equations used to calculate fault resistance are based on the fault calculations (Duncan Glover and Sharma 2002) for each fault type, shown in Table 4.6:

Table 4.6: Fault Resistance calculations.

Fault Type	Equation
Single Phase to Ground Faults (L-N)	$R_{fault} = \left Re(Z_{fault}) - Re(Z_{linetofault}) \right $ $= \left Re\left(\frac{V_0 + V_1 + V_2}{3I_1}\right) - Re(Z_{line} \times Distancetofault) \right \dots (5)$
Phase-to-Phase-to-Ground Faults (L-L-N)	$R_{fault} = \left Re(Z_{fault}) - Re(Z_{linetofault}) \right $ $= \left Re\left(\frac{V_0 - V_1}{3I_0}\right) - Re(Z_{line} \times Distancetofault) \right \dots\dots\dots(6)$
Phase –to-Phase Faults (L-L)	$R_{fault} = \left Re(Z_{fault}) - Re(Z_{linetofault}) \right $ $= \left Re\left(\frac{V_1 - V_2}{3I_1}\right) - Re(Z_{line} \times Distancetofault) \right \dots\dots\dots(7)$
Three Phase Faults (L-L-L/L-L-L-N)	$R_{fault} = \left Re(Z_{fault}) - Re(Z_{linetofault}) \right $ $= \left Re\left(\frac{V_1}{I_1}\right) - Re(Z_{line} \times Distancetofault) \right \dots\dots\dots(8)$

The fault resistance statistics are shown in table 4.7. Resistance is calculated in ohms.

Table 4.7: Fault Resistance.

	220kV		275kV		400kV	
Characteristic	μ	σ	μ	σ	μ	σ
Rfault_onecycle	20.30	14.73	36.23	148.70	41.99	118.06
Rfault_twocycle	17.94	13.55	31.69	130.15	37.88	124.74

4.3.7 Fault Insertion Phase Angle (FIPA)

This is a feature calculated based on the premise that external faults are inserted at the peak of the waveform (Barrera et al. 2012). In this instance, FIPA is calculated with reference to the last zero-crossing prior to the start of the fault. For multi-phase faults, FIPA is assumed to be the phase angle closest to the peak.

Table 4.8: Basic Statistics for Fault Insertion Phase Angle.

	220kV		275kV		400kV	
Characteristic	μ	σ	μ	σ	μ	σ
FIPA	115.55	41.43	125.61	85.43	119.20	75.02

Table 4.8 illustrates that the average fault insertion angle (in degrees) takes place after the voltage peak. This is reasonably consistent across voltage levels, however the level of variance differs significantly with voltage level. The mean values indicate that FIPA does not differ due to the fault cause and usually occurs approximately 30° after the waveform peak.

4.3.8 Sequence Component Fault Current Time Constant (τ)

The Sequence Component Fault Current Time Constant treats the fault waveform response in a similar manner to a first order linear time-invariant system. The time constant is calculated as the time taken from fault initiation to 0.63 of the difference between maximum fault current and pre-fault current for each sequence component. The time constant for each sequence component current is treated as an individual feature. This feature is selected as it gives an indication of the dynamic response of the transmission line/network to the fault in question.

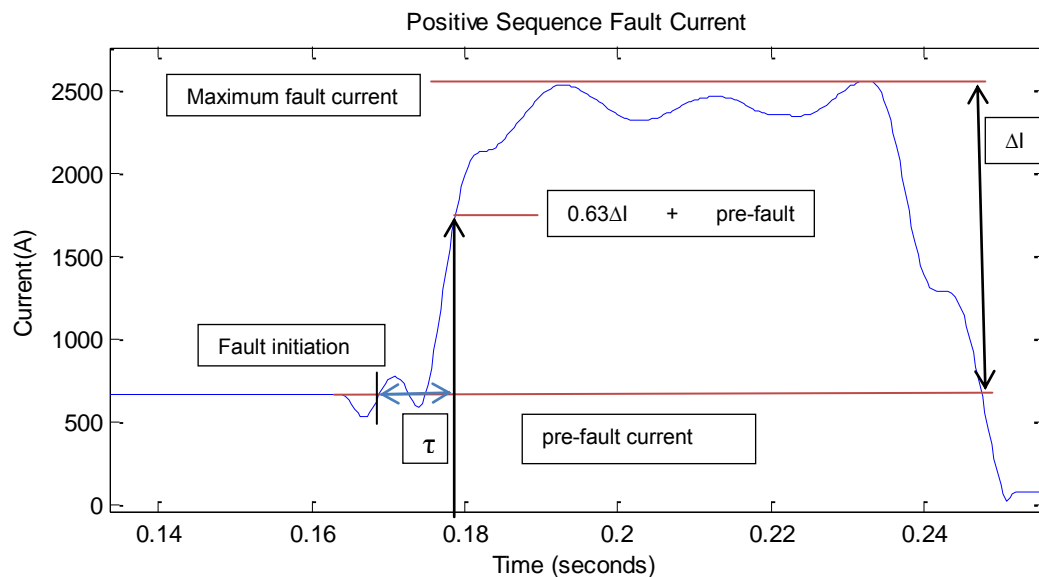


Figure 4.5: Sequence Component Fault Current Time Constant.

Figure 4.5 illustrates the calculation of the sequence component fault current time constant.

Table 4.9: Basic Statistics for Sequence Component Fault Current Time Constant.

Characteristic	220kV		275kV		400kV	
	μ	σ	μ	σ	μ	σ
Positive current Time Constant (τ_1)	12.81	1.48	26.85	70.01	22.72	71.41
Negative current Time Constant (τ_2)	12.85	1.61	25.92	70.21	21.80	76.20
Zero current Time Constant (τ_0)	13.46	7.14	22.31	46.26	20.27	63.65

Table 4.9 provides shows the mean and standard deviation values of the Sequence Component Fault Current Time Constants in milliseconds. These values are very similar across sequences phases; however they differ across voltage levels, indicating that it is influenced by network design and conditions.

4.4 Discussion

The mean and standard deviation values for the extracted features provide indicative values for the extracted features on 220kV, 275kV and 400kV transmission networks. It is observed that the calculated mean values for all the features extracted differ significantly (excluding FIPA) between voltage levels. This indicates the impact of the power system on the magnitudes and changes in voltages and currents in response to faults. The voltage at which a fault occurs should therefore be included as a feature to analyse measurement data.

A second observation is that a high level of variation is present across the majority of the features. In many of these cases this may well be a function of a number of

causes as the measurements are drawn from a large network (>28000km of lines) where the state of the network may differ significantly even for lines at the same voltage level.

4.4.1 Statistical Significance of Waveform Features

Analysing the waveform feature data by fault cause indicates that the fault cause does influence the majority of these features. The statistical significance (to the 0.05 level) of the causes (bird streamer, fire, lightning, pollution and other) on single-phase-to-ground faults occurring on 220kV, 275kV and 400kV networks was tested by means of analysis of variance (ANOVA), for which results are given in table 4.10 ('yes' indicates statistically significant). Single-phase-to-ground faults were chosen as they represent more than 90% of all faults on the transmission network.

Table 4.10: Statistical Significance of Causes Influencing Waveform Features

FEATURE	400kV	275 kV	220kV
ΔI_{max1}	yes	yes	yes
ΔI_{max2}	yes	yes	yes
ΔI_{max0}	yes	no	yes
V_{max2}	yes	no	no
V_{max0}	yes	no	no
$I_{1(0.5)}$	yes	no	yes
$I_{1(1)}$	yes	no	yes
$I_{2(0.5)}$	yes	yes	no
$I_{2(1)}$	no	no	yes
$I_{0(0.5)}$	no	no	yes
$I_{0(1)}$	yes	no	yes
I_{1max}	yes	no	yes
I_{2max}	no	no	yes
I_{0max}	yes	no	yes
Rfault_onecycle	yes	yes	yes
Rfault_twocycle	yes	yes	yes
FIPA	no	no	no
Positive current Time Constant (τ_1)	yes	yes	no
Negative current Time Constant (τ_2)	no	yes	no
Zero current Time Constant (τ_0)	no	yes	no

Table 4.10 indicates statistically significant differences across voltage levels for the majority of the waveform features extracted, with the exception of FIPA. This indicates that fault causes may be differentiated by a combination of these features.

4.5 Conclusions

A total of 21 waveform characteristics are extracted from 2672 measured fault waveforms. These characteristics are calculated along three main strategies: 1) peak/maximum values are calculated relative to pre-fault values, which provide results with reference to network conditions prior to the fault and enables

measurements from different voltage levels to be considered together, 2) maximum rates of change are calculated to identify the influence of the fault flashover mechanism and 3) characteristics are calculated to compare the magnitude of characteristics at specific time intervals from the start of the fault. It is shown that the waveform features extracted have statistically significant differences in mean values based on the underlying fault cause.

The voltage level at which a fault occurs influences the majority of the features extracted. The voltage level therefore is an important feature to include in utilising the available measurement data.

The development of this characterised waveform dataset addresses some of the key concerns raised by Gu and Styvaktakis (2003) with respect to characterising event waveforms:

1) Each characterised waveform is associated with a fault cause that makes this a suitable dataset for conducting feature selection and classification according to underlying causes.

2) It addresses the limitations faced in many instances with respect to a shortage of data, with a dataset consisting of 2672 waveforms across 220kV, 275kV and 400kV networks. The entire data set is based on measurements from an operational transmission system.

CHAPTER 5

5. CLASSIFYING TRANSMISSION LINE FAULTS

This chapter describes the classification of transmission line faults according to underlying causes. Single nearest neighbour classifiers are built using nested subsets and the classification performance is compared to decision tree, neural network and naïve Bayes classifiers.

5.1 Introduction

Broadly, root cause identification of faults can improve transmission network reliability in two ways:

a) applied in a control centre environment it allows the control centre staff to 1) make appropriate decisions with respect to reclosing breakers, 2) provide field services operators with knowledge about the probable cause of a fault being investigated, which gives them insight into signs to search for when investigating a fault and 3) despatch crews with appropriate equipment to resolve a fault (Xu & Chow, 2006). Appropriate decisionmaking, combined with suitably equipped and forewarned field service operators translate into improved network reliability by reducing the overall time required to respond and repair transmission line faults.

b) accurate identification of fault causes can inform the design and parameter selection of new lines (insulator selection, tower design, footing resistance) as well as identifying lines that perform poorly due to specific causes i.e. identify unknown fault causes. Mitigation methods can then be developed to resolve a particular problem e.g. birdguards for bird streamer problems or improving footing resistance where lightning faults occur more frequently than is deemed acceptable. These

causes may not always be easily identified as field services staff may give vague descriptions or may lack knowledge about the fault mechanisms (Vosloo, 2005). Accurate identification of fault causes leads to improved network reliability by reducing the number of transmission line faults that occur in the long-term.

This work explores the classification of transmission line faults according to underlying cause using pattern recognition techniques. The individual relevance of features is determined with respect to the four major causes and classification is treated as a multiclass problem. Features are ranked according to their relevance in separating fault causes and classifiers are built by nested subsets. Nested subsets are groups of sets where the larger sets contain all the elements of the smaller sets. This concept is illustrated via a Venn diagram in figure 5.1.

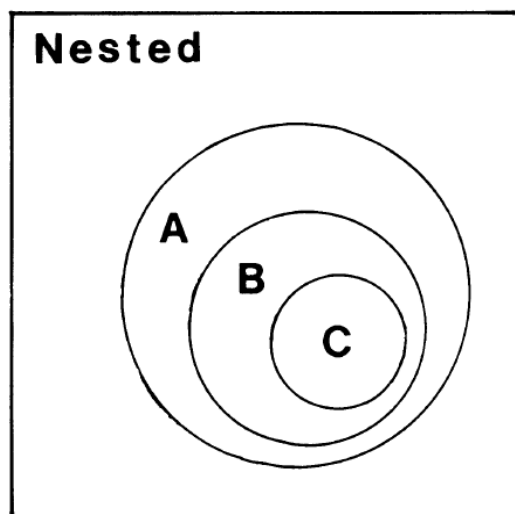


Figure 5.1: Venn diagram of Nested Subset (Patterson, 1987)

Figure 5.1 illustrates a nested subset whereby set B includes all the elements of set C and set A in turn includes all the elements of set B and C.

5.2 Feature Selection and Classification

The full feature set considered here consists of 28 features, made up of 1) the environmental, climatic and diurnal features (these features are referred to as the **contextual features**) discussed in chapter 3 and 2) the twenty-one waveform features (**waveform features**) identified in chapter 4.

The waveform features extracted are:

- Fault type i.e. single-phase-to-ground, phase-to-phase faults
- Current rate of change (positive, negative and zero sequence sequences)
- Maximum negative sequence voltage
- Maximum Zero sequence voltage
- Maximum sequence current, relative to pre-fault currents (positive, negative, zero sequences)
- Current Magnitude, half cycle after fault initiation (positive, negative, zero)
- Current Magnitude, one cycle after fault initiation (positive, negative, zero)
- Fault resistance (1 and 2 cycles after fault initiation)
- Fault initiation phase angle
- Sequence Component Fault Current Time Constant (positive, negative, zero).

A key finding from chapter 2 is that fault frequencies have statistically significant differences with respect to time-of-day, season, and climate (as represented by rainfall area). These differences are not uniform across the voltage levels considered; hence voltage level is also a feature considered.

The contextual feature set also introduces other features considered relevant. These are the geospatial information system (GIS) identifier for the line on which the fault occurred, the geographic region and average lightning ground flash density (Ng).

The contextual features describing fault occurrence are as follows:

- Hour of day
- Month of year
- Rainfall area
- Voltage level
- Line GIS number
- Eskom transmission grid region
- Ground flash density.

In this study, feature selection is used firstly to improve understanding and interpretability of the data and secondly to identify features for building good classifiers for transmission line fault causes. Feature ranking and classification is considered along three scenarios, using:

- a) only the contextual feature set. The earlier analysis has shown statistically significant differences in fault frequencies by time of day, climate and season
- b) only the waveform feature set
- c) combining the waveform and contextual feature sets.

Both feature selection and classification is implemented in the Matlab toolbox PRTOOLS (Duin, et al, 2000).

5.2.1 Feature Ranking

This study starts by selecting features according to the individual relevance to the classification problem. Feature ranking is conducted by using the F-statistic, derived by analysis of variance (ANOVA) which provides a measure of variance due to a feature. This is calculated as the statistical comparison between data sets. This provides a basis for building classifiers as well as giving some insight into the relevance of individual features for identifying fault cause.

Table 5.1: Features ranked by F-statistic

Feature	Type of Feature	F-statistic	Overall Rank	Contextual Rank	Waveform Rank
Hour	Contextual	90.90	1	1	n/a
Region	Contextual	33.09	2	2	n/a
Month	Contextual	32.16	3	3	n/a
Nominal Voltage	Contextual	29.22	4	4	n/a
Average ground flash density (Ng)	Contextual	25.45	5	5	n/a
$I_{2(0.5)}$	Waveform	16.06	6	n/a	1
τ_1	Waveform	11.98	7	n/a	2
Faulted Phases	Waveform	11.58	8	n/a	3
V_{max2}	Waveform	10.44	9	n/a	4
τ_2	Waveform	9.97	10	n/a	5
$I_{2(1)}$	Waveform	8.33	11	n/a	6
V_{max0}	Waveform	8.12	12	n/a	7
ΔI_{max0}	Waveform	6.66	13	n/a	8
I_{2max}	Waveform	6.17	14	n/a	9
ΔI_{max1}	Waveform	6.02	15	n/a	10
ΔI_{max2}	Waveform	5.24	16	n/a	11
τ_0	Waveform	5.14	17	n/a	12
I_{1max}	Waveform	4.97	18	n/a	13
$I_{0(0.5)}$	Waveform	4.60	19	n/a	14
Rfault_onecycle	Waveform	3.59	20	n/a	15
$I_{0(1)}$	Waveform	3.53	21	n/a	16
$I_{1(0.5)}$	Waveform	3.29	22	n/a	17
Rainfall Area	Waveform	2.78	23	6	n/a
Rfault_twocycle	Waveform	2.42	24	n/a	18
$I_{1(1)}$	Waveform	2.24	25	n/a	19
I_{0max}	Waveform	2.12	26	n/a	20
GIS Number	Contextual	1.71	27	7	n/a
FIPA	Waveform	1.57	28	n/a	21

Table 5.1 illustrates the features ranked overall as well as separate ranking lists for contextual and waveform features. It gives a clear picture of the relevance of individual features to separating faults according to cause. The waveform features

with the highest F-statistic scores are maximum negative sequence current (half cycle) and the positive sequence time constant labelled as $I_{2(0.5)}$ and τ_1 respectively in table 5.1. Both these features provide a measure of the dynamic response to an event on a transmission line, which relates to the rate at which sequence currents rise to peak fault current.

Table 5.1 indicates that the individual contextual features related to the time of occurrence and geographic location of the fault, are highly relevant to identifying the underlying cause of the fault. This result ties in with classification efforts on distribution systems (Cai, et al, 2010) that only made use of contextual features to achieve good classification success for animal- and tree-caused faults. The ranking also gives insight into the relative strength of using the contextual versus waveform features to identify fault causes.

5.2.2 Classification by Nested Subsets

The classification problem for transmission line faults is here defined as a multiclass problem with the five classes being bird streamers, fires, lightning, pollution and all other faults. A multiclass classifier is a function $F: X \rightarrow Y$ which maps an instance x into a label $F(x)$ (Duda, et al, 2000). There are 2 common approaches taken to generate F : the first is to construct it through the combination of a number of binary classes e.g. logistic regression or support vector machines; the second approach (used in this study) is to generate F directly e.g. naïve Bayes algorithms or nearest neighbour.

The use of single nearest neighbour (1-NN) classifier is proposed for identifying the underlying cause of transmission line faults. The 1-NN rule is a suitable benchmark for other classifiers as it 1) requires no user-specified parameters, making its implementation independent and 2) provides reasonable classification performance for the majority of applications (Jain, et al, 2000). A major aim of this work is the identification of suitable features for automatically classifying fault cause, as opposed to primarily identifying an optimal classifier or optimising classifiers for maximum accuracy or robustness. The reasonable performance and simplicity in implementation makes a 1-NN classifier highly suitable in this context.

The single nearest neighbour rule assigns to an unclassified point the classification of the nearest previously classified point (Cover & Hart, 1967).

The nearest-neighbour classifier is commonly based on the Euclidean distance between a test sample and the specified training samples. Let x_i be an input sample with p features $(x_{i1}, x_{i2}, \dots, x_{ip})$, n be the total number of input samples ($i=1, 2, \dots, n$) and p the total number of features ($j=1, 2, \dots, p$). The Euclidean distance between sample x_i and x_l ($l=1, 2, \dots, n$) is defined as:

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \dots + (x_{in} - x_{ln})^2}$$

For the single nearest neighbour rule, the predicted class of test sample x is set equal to the true class ω of its nearest neighbour, where m_i is a nearest neighbour to x if the distance:

$$d(m_i, x) = \min_j \{d(m_j, x)\}$$

Classification is done by building a series of 1-NN classifiers to determine the underlying cause of transmission line faults. The 1-NN classifiers are built comprising 5 classes, one for each of the major fault cause i.e. birds, fire, lightning, pollution and other. The classifier finds the nearest point in the training set to the unclassified point and assigns this to the corresponding label. An advantage of nearest neighbour classification is its conceptual simplicity and ease of implementation, making it capable of dealing with complex problems as long as sufficient training data is available (Garcia, et al, 2008).

Feature selection for the initial classification is done by building nested subsets (as illustrated in figure 5.1) of features of increasing size. Classifiers are built, starting with a subset of one feature (the highest F-statistic) and features are progressively added by decreasing F-statistic. Subsets are constructed according to the rankings illustrated in table 5.1 i.e. the first subset comprises the feature ranked 1, the second subset the features ranked 1 and 2 For each of the classifiers, the data is split using two-thirds for training data and one third for testing. The classifier is trained using the training set and performance is evaluated using the test set. The entire data set is randomly split into two-thirds training and one-third test sets thirty times and the trained classifier evaluated against each test set to reduce variance from the classification results. Evaluation of the classifier is done in this manner so that enough test data is available for the faults causes that occur rarely i.e. other- and pollution- caused faults. Classification in this manner is conducted according to

the three scenarios identified i.e. 1) the contextual features; 2) the waveform features and 3) the combined feature set.

Two combining rules are implemented for classification using all the features. These are:

Combining-rule 1) The waveform and contextual feature sets are combined by relevance using the overall ranking as indicated in table 5.1

Combining-rule 2) The waveform and contextual feature sets are combined by adding (in order of decreasing relevance) a feature from each set, starting with the contextual set. Once the contextual features are all added, the remaining waveform features are then added to form additional subsets.

5.2.3 Assessing Classifier Performance

The common basis for many of the performance measures of classifiers is based on the confusion matrix. For a 2-class classifier (e.g. Yes/No) and a test dataset a two-by-two confusion matrix can be built, shown in table 5.2 (Kubat, et al, 1998).

Table 5.2: Confusion Matrix

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	True Positive (TP)	False Negative (FN)
Actual Negative Class	False Positive (FP)	True Negative (TN)

The most common performance measure calculated from this matrix is Accuracy (Acc), which is the proportion of the total number of predictions that were correct:

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \dots\dots\dots(9)$$

A consideration for many real-world applications is determining classifier performance of imbalanced datasets i.e. when one of the classes represents only a small portion of the total data set. Using Accuracy as a performance measure is insufficient as a classifier will show good performance by simply ignoring the presence of the minority class. The fault performance of the transmission system is imbalanced across the four major fault causes.

Two measures commonly used when classifying unbalanced datasets are Precision (p), the percentage of positive predictions made that are correct and Recall (r), percentage of true positive cases that were correctly identified.

Another measure with its foundation in information retrieval and text classification (Nan, et al, 2012) is the F-measure¹, which is the harmonic mean of Precision and Recall. This provides a balanced measure of the performance of a classifier and will be used to provide insight into balanced classification success for each fault cause.

¹ It should be noted that the F-measure is defined independently from the F-statistic used in ANOVA. F-measure and F-statistic are two completely independent measures defined and used in different ways.

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots\dots\dots (10)$$

The measures used to assess classifier performance are overall classification Accuracy as well as the F-measure for each fault cause.

5.2.4 Overall Classification Accuracy

The overall classification accuracy rates indicate that a classification accuracy of up to 90% is achieved when only using the contextual features. This classification rate is achieved using the five highest ranked contextual features.

Figure 5.1 illustrates overall classification accuracy for contextual, waveform and all features. The features are added according to the respective decreasing rankings shown in table 5.1 for waveform and contextual features and then according to the combining rules defined in section 5.2.2 e.g. features for ACC_All_rule1 are added according to combining-rule 1.

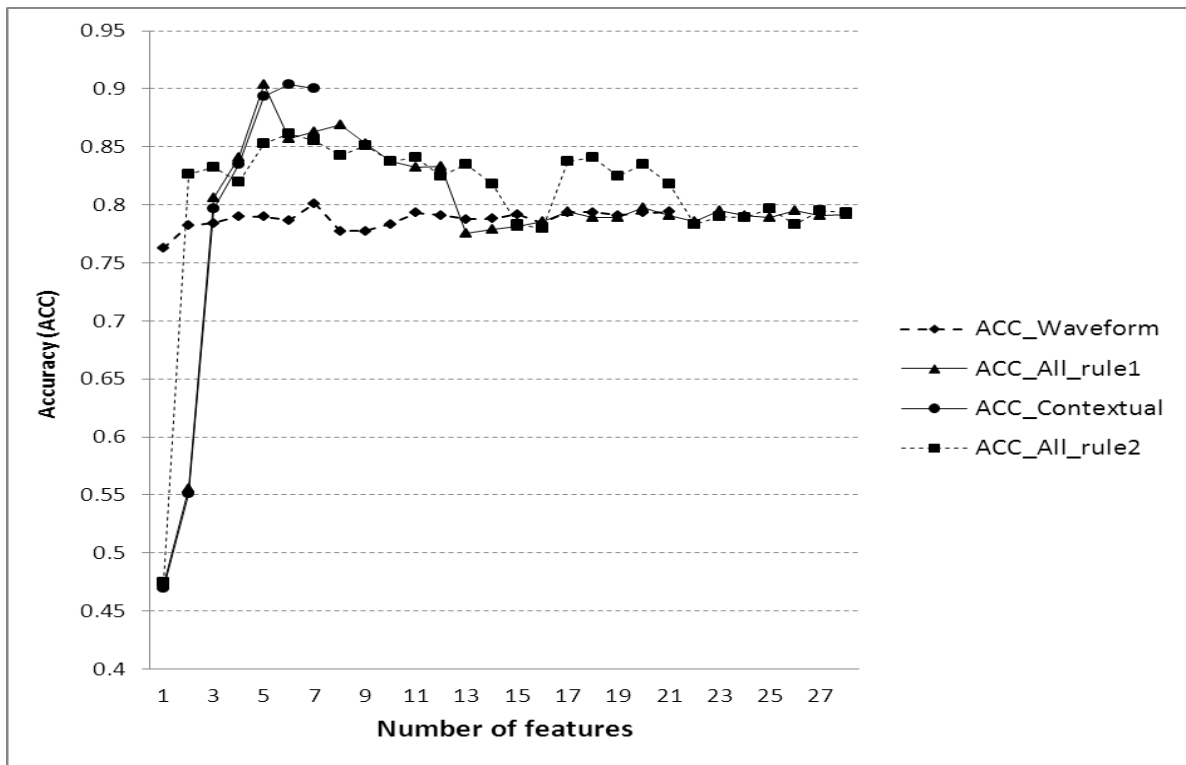


Figure 5.1:Classification Accuracy.

The best classification accuracy rate achieved when only considering the waveform features is 0.801, using the seven highest ranked waveform features. This shows that reasonably good classification performance can be achieved using only the waveform features. This performance does not match that achieved using only contextual features or a combination of contextual and measurement features.

5.2.5 Classification Performance using Waveform Features

The best overall classification accuracy rate is achieved using the best seven waveform features. This shows that reasonably good classification performance can be achieved using only the waveform features.

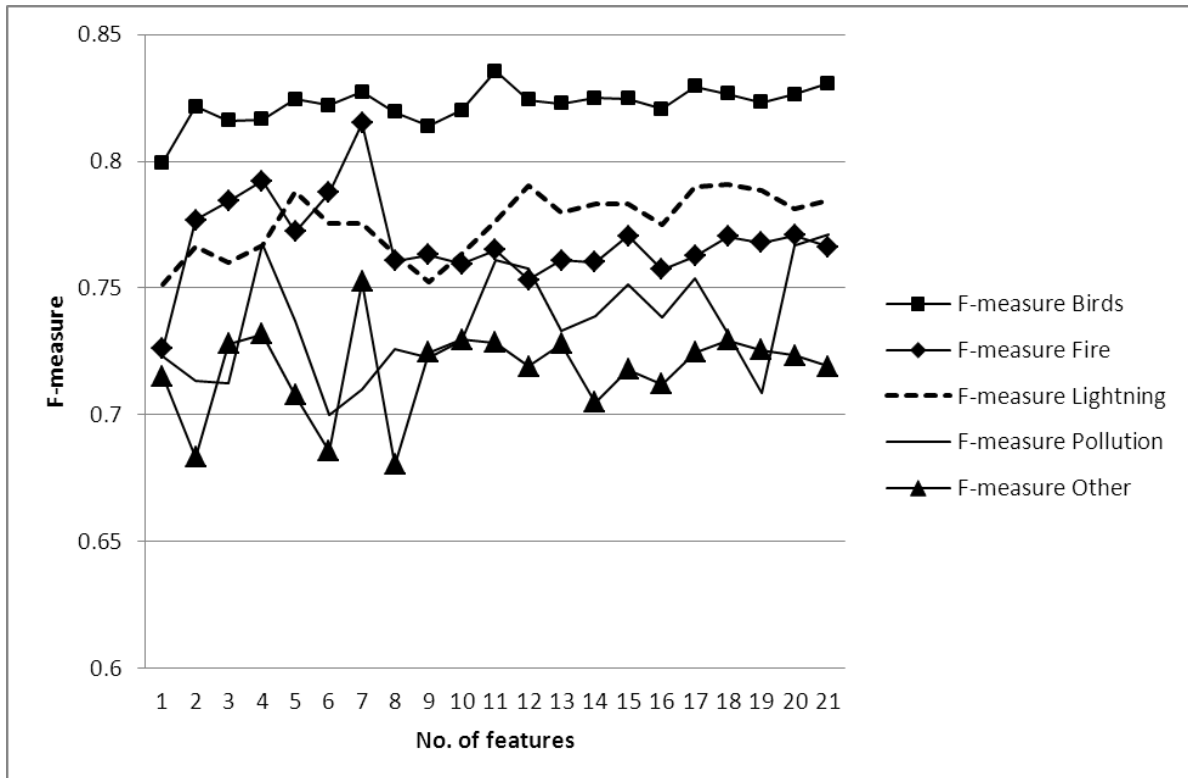


Figure 5.2: F-measure for fault causes

Figure 5.2 (features added by decreasing F-statistic) illustrates the classification success by F-measure for each of the major fault causes. Scores of above 0.75 are achieved using only the two highest ranked features for bird, fire and lightning caused faults. The F-measure for pollution-caused faults is generally lower than the first three classes. However, it needs to be considered that pollution represents a highly imbalanced set.

The performance of the 'other' class of faults is the poorest for most of the feature sets used and is more dependent on the appropriate feature set being selected for achieving reasonable classification performance.

5.2.6 Classification Performance using Contextual Features

The classification performance achieved using contextual features is significantly better once five or more features are used to build the nearest neighbour classifier. F-measure scores above 0.8 are achieved for all classes of faults with bird streamers having the highest score of 0.917.

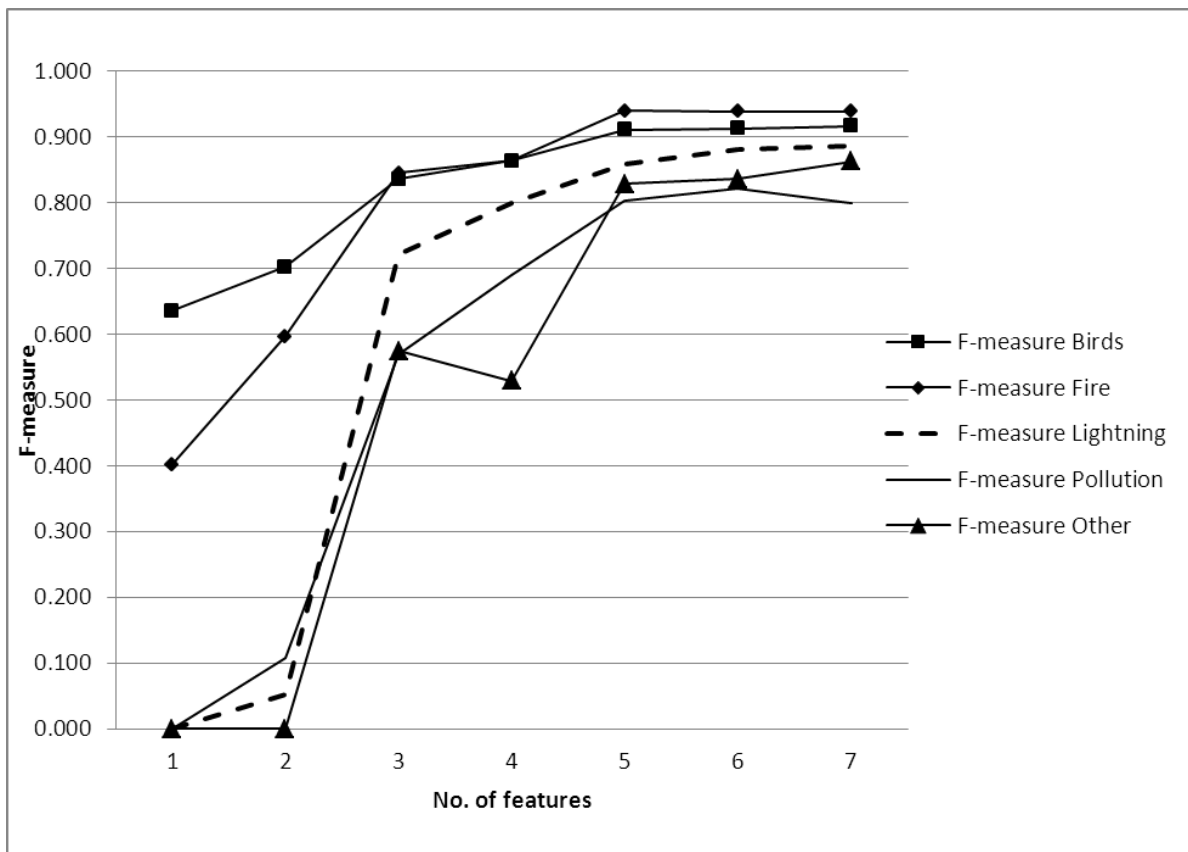


Figure 5.3: F-measure using contextual features.

Figure 5.3 illustrates the F-measure performance achieved with a 1-NN classifier built with contextual features, with features added by decreasing F-statistic.

5.2.7 Classification Performance using Combined Features

The classification performance achieved when combining features indicates that combining the waveform and contextual features does not result in an improved rate of classification performance over that achieved by using only the contextual features. Figure 5.4 illustrates the F-measure performance achieved using combining-rule 1 and demonstrates deterioration in classification performance once waveform features are added to the five highest ranked contextual features. The exception is the balanced performance for pollution-caused faults which shows a slight improvement in classification performance.

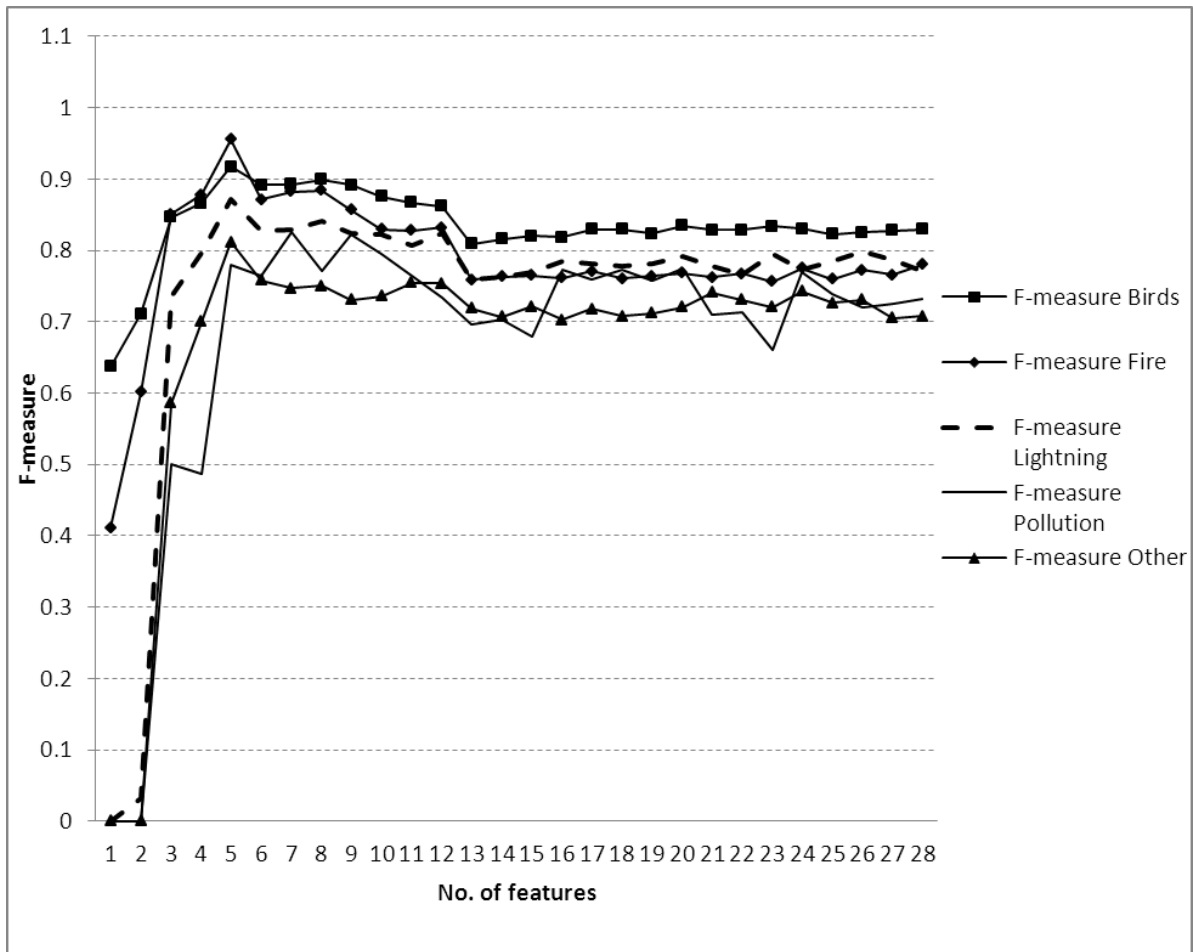


Figure 5.4: F-measure using waveform and contextual features combined by rule 1.

Applying combining-rule 2 does not improve the balanced performance for any of the fault causes, however good classification performance is achieved using as few as one feature from each feature set, as illustrated in figure 5.5.

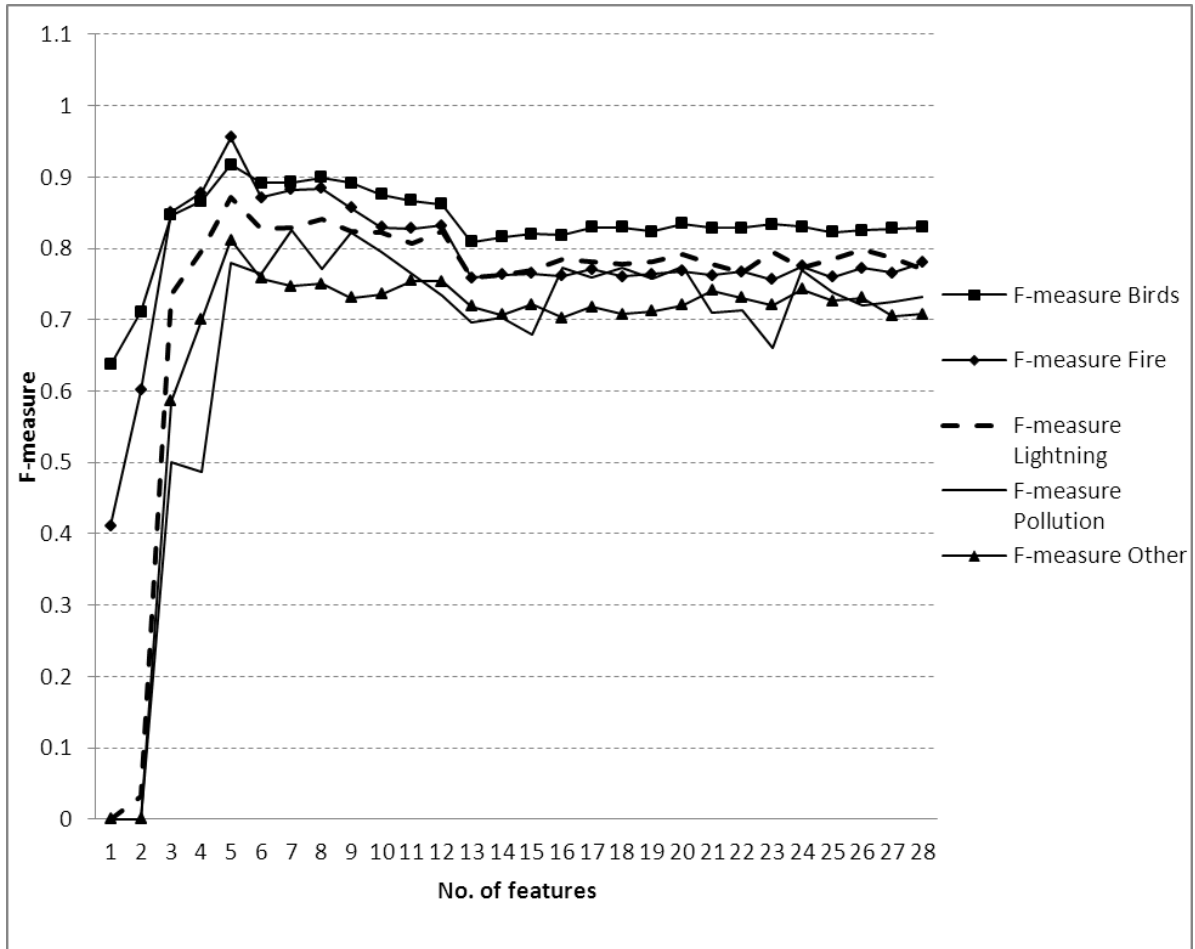


Figure 5.5: F- measure using waveform and contextual features combined by rule 2

The best classification accuracy is achieved using six features (the top three ranked from the waveform and contextual sets). These features are: **Hour, Negative sequence current (half cycle after fault initiation), Region, Positive sequence fault current time constant, Month and Faulted Phases.**

5.3 Comparing Classifier Performance

The best performance achieved with the 1-NN classifier using nested subsets is compared with the classification performance of several common classifiers with feature selection conducted by sequential forward (SFS) and sequential backwards (SBS) wrapper selection based on best performance for a particular classifier. The classifiers used for the comparison are radial basis neural network, decision tree and naïve Bayes classifiers. The implementation of these classifiers is done utilising the existing models within the PRTools toolbox (Duin, et al., 2000). Overall accuracy is compared as well as F-measure for individual fault causes.

SFS and SBS are implemented for an individual classifier based on the 28 waveform and contextual features. The features identified in the selection process are used to train and test the respective classifiers. For each classifier, the data is split using two-thirds for training data and one third for testing. The classifier is trained using the training set and performance is evaluated using the test set. The entire data set is randomly split into two-thirds training and one-third test sets thirty times and the trained classifier evaluated against each test set to reduce variance from the classification results. Evaluation of the classifiers is done in this manner so that enough test data is available for the faults causes that occur rarely i.e. other- and pollution- caused faults.

5.3.1 Results

The overall accuracy and F-measure for each classifier per fault type is shown in table 5.3.

Table 5.3: Classification performance using wrapper selection methods

CLASSIFIER	Feature Selection	ACC	F-measure Birds	F-measure Fire	F-measure Lightning	F-measure Pollution	F-measure Other
Decision tree	SFS	0.7366	0.7795	0.8221	0.6668	0.3089	0.1665
Decision tree	SBS	0.7258	0.7711	0.8129	0.647	0	0.201
Neural Network	SFS	0.7378	0.7832	0.8075	0.6657	0.2959	0.1556
Neural Network	SBS	0.7397	0.7847	0.8104	0.731	0.2805	0.1502
Naïve Bayes classifier	SFS	0.7403	0.7995	0.7765	0.6776	0.2857	0.14
Naïve Bayes classifier	SBS	0.7378	0.7778	0.7984	0.6767	0.3383	0.1501
1-NN (combining rule 2)	Nested subsets	0.8612	0.8896	0.8778	0.8418	0.7493	0.7372

The classification results shown in table 5.3 demonstrate that the best performance achieved with the 1-NN classifier is between 12-14% higher than achieved with any of the other classifiers using either forward or backward selection. The 1-NN classifier is also shown to have considerably better classification performance when considering the minority causes (pollution and other causes together make up less than 10% of all faults).

5.4 Discussion and Conclusions

The use of a single nearest neighbour classifier has been proposed for the classification of transmission line faults according to underlying cause.

The results achieved by building a series of 1-NN classifiers by nested subsets of decreasing relevance has shown that the 1-NN classifier is a suitable classifier for the identification of transmission line fault causes.

Classifiers developed using only the contextual features demonstrate the highest level of balanced classification performance (90%). The relevance of this is that in many instances where no physical evidence of the fault is present, operators assign fault causes based on contextual information e.g. a fault during a thunderstorm may not necessarily be due to lightning but is often classified as such (Barrera et al. 2012). This practice has relevance when considering the success achieved with contextual feature set as these features overlap directly with the evidence available to operators. It is also relevant when considering the lower classification accuracies achieved when considering waveform features as the waveform evidence may point to a different result from the contextual evidence.

It has been found that using only waveform features to build a classifier produces reasonable success levels, even with only the single most relevant feature. While this is an important finding with respect to identifying and using fault waveform features for fault identification, the classification performance achieved may not be adequate for practical applications. A guide for an acceptable Accuracy rate can be

set at 80% as a minimum threshold for practical classification applications. This level of accuracy is achieved using a) only contextual features and b) combining contextual and waveform features.

The concern with classifying faults only considering the contextual features is that actual measurements (or an observation) of the event is not considered when identifying faults. In practice, physical observations are used as far as possible to confirm fault cause e.g. flashover markings on towers.

Classification has been conducted on faults occurring on the South African transmission network. These results demonstrate the classification of faults at transmission level voltages and can be extended to other networks where 1) accurate and extensive fault records and waveform measurements are available, 2) fault types are well documented and understood and 3) climatic factors are considered.

For automatic classification of transmission line faults according to underlying cause it is recommended that measured fault waveform features be used to construct fault cause classifiers, where measurements are available.

The classification performance of 1-NN classifier has been compared with radial basis neural networks, decision trees and naïve Bayes classifiers and it has been shown to have superior classification performance with respect to overall accuracy and F-measure for the minority fault causes.

CHAPTER 6

6. CONCLUSIONS

This chapter presents an assessment of the research hypothesis, highlights the contributions made and discusses the significance of the results and future work.

This thesis has focussed on the analysis and classification of transmission line faults. Emphasis has been placed on improving the understanding of the impact of the primary causes of line faults on the transmission network to assist in decision-making for planning and design, network maintenance and operation of the network.

The first focus area of this thesis analyzed faults to develop fault frequency statistics that present more information on the network performance than average annual frequency.

The second focus area investigated measured fault waveforms for characteristics linked to the fault cause and building a suitable feature set classifying faults using statistical pattern recognition techniques.

The overall aims of the research was to 1) improve understanding of the impact that the climate and environment has on the causes and frequency of faults on the South African transmission network; 2) identify electrical fault waveform characteristics relevant to identifying fault causes and 3) ultimately automate the

classification of transmission line faults using statistical pattern recognition techniques.

6.1 Assessing the hypothesis

The basic hypothesis that this research addressed is:

Transmission line faults can be automatically classified according to underlying event cause using pattern recognition techniques. However, this requires knowledge of the external environment influencing the event.

The original hypothesis is assessed by reviewing the original research questions, these are:

- 1) *What are the primary causes of faults on transmission lines on the South African transmission network and how do they impact the fault frequency performance?*

The primary causes of transmission line faults on the South African transmission network have been identified as bird streamers, lightning, fire and pollution. Results have been presented for the impact of these causes across the network voltage levels and with respect to the external environment such time-of-day, time-of-year and rainfall area.

- 2) *Can significant variables related to interruption performance be identified?*

Fault statistics presented in terms of a 4 by 4 matrix, linked to rainfall area has shown that time-of-day, time-of-year and climate (indicated by rainfall area) represent variables which affect the interruption performance

3) *Can event characteristics be identified that are relevant features for automatically classifying transmission line faults according to underlying cause? If so, which characteristics are these?*

It has been shown that both contextual and waveform features can be identified which have a varying degree of relevance to performing classification of events. The accuracy of classification using these features has been demonstrated by building classifiers using nested subsets. Best classification is achieved using only contextual features; however classification in this manner does not make use of any measurements. Taking waveform features based on measurements into account classification, accuracy of 86% is achieved using the following set of contextual and waveform features: Hour, Negative sequence current (half cycle after fault initiation), Region, Positive sequence fault current time constant, Month and Faulted Phases.

4) *Can faults be classified using only electrical waveform characteristics?*

The highest classification accuracy achieved using only waveform features is 80%. This indicates that faults can be classified for underlying cause using only waveform features with a reasonable level of success. However this performance is inferior to classification accuracies achieved when using waveform features in combination with contextual features.

The classification performance achieved here is based on measurements taken on a transmission system. These results would need to be tested on data from distribution system voltages, which are typically 132kV and lower, to verify classification accuracy and choice of features at these voltage levels.

5) *What classification performance is achieved?*

Classification results from this study demonstrate that the best classification accuracy is achieved when using contextual features (which are representative of the external environment) either independently or in combination with features derived from fault waveforms. The best classification accuracy using contextual features is 90%, against 86% for classification using a combination of waveform and contextual features.

The possible misclassification of fault causes by human operators makes it impossible to determine with certainty whether the higher classification success achieved using contextual features would still be retained if these errors could be addressed.

The hypothesis can thus not be validated in its present form and is revised as follows:

Transmission line faults can be automatically classified according to underlying event cause using pattern recognition techniques. However, this requires knowledge of the external environment influencing the event for best classification performance.

Considering the research hypothesis in this form it is shown that the best classification results are achieved when external features are considered.

6.2 Contributions

The contributions and novel elements of this thesis are:

- An analysis method is presented that relates the frequency of faults on overhead lines to the climate of the area in which the line is located and the causes of power system faults
- An approach to analysis of power system faults that presents a statistically significant set of fault data by cause, climate, season and time of day
- The statistical significance of the differences between mean fault frequencies for fault causes, climate, time of day and season is established
- Establishing a basis for the inclusion of non-electrical characteristics of faults in reliability analysis and root cause identification of transmission line faults
- The development of waveform features to characterise faults: 1) across multiple voltage levels, 2) at specific time intervals from fault initiation and 3) according to the influence of the fault flashover mechanism
- The ranking and selection of contextual and waveform features for identifying the causes of transmission line waveforms
- The classification of transmission line faults by Single Nearest Neighbour classification has been proposed and demonstrated
- It has been shown that transmission line faults can be classified using only contextual features as well as a combination of contextual and waveform features.

6.3 Significance of Results

The information derived from the collection and analysis of fault frequency data leads to two applications.

One will be better modelling of the whole system's reliability, using interruption duration and network loading data similarly classified by a 4 by 4 matrix, with substantial implications for both planning and system operations. A second application is the design and selection of parameters for specific lines and identifying lines with relatively poor performance needing to be improved.

The research work undertaken has touched on several questions raised with respect to the classification of power system events and has developed results that can be utilised to address these, including:

- 1) Manual analysis of faults is a labour intensive exercise that consumes considerable man hours. Automated techniques with high accuracy levels have the potential to provide significant man hour savings to utilities.
- 2) The identification of faults is a complex problem that requires significant knowledge and experience to undertake successfully. These skills are often in short supply, with the result that inexperienced staff will make significant errors in diagnosing fault causes.
- 3) Many faults do not leave physical evidence (e.g. flashover marks or animal carcasses) of its cause. Operators may often assign speculative or generic causes to these faults. An expected error is hence built into any set of fault analysis due to this uncertainty. Automatic classification techniques using both contextual and

waveform features provides a means to determine causes based on both the external environment and the physical evidence of the waveform.

The classification of faults plays a role in the design and management of networks. Incorrect classification leads to uncertainty in diagnosing appropriate solutions to improving network reliability performance and can also lead to wasteful financial expenditure due to implementing solutions that are inappropriate to mitigating the real cause of faults. The design of new networks within close proximity of existing networks is often done on the basis that these networks will be exposed to similar conditions and these will then be designed and tailored to minimise the impact of prevalent fault causes e.g. birdguards will be included at the outset or clearance will be increased to minimise bird streamer faults. The result of incorrect fault cause identification is a) wasteful expenditure and b) higher fault frequencies as the causes of faults are not addressed.

A second major impact of automatic fault identification relates to the immediate operational response to faults. In a typical scenario, once a fault has occurred an operational crew is dispatched to patrol the line, identify the fault and conduct corrective work to bring the line back into operation. Automatic classification can provide information for dispatchers to ensure that teams are appropriately equipped to resolve a fault and bring the line back into operation. Another scenario is that lines are often not brought back into service immediately until a complete line inspection is conducted due to uncertainty of the fault cause and concern that the equipment may be damaged. An early identification of the fault would allow a line to be brought back into service sooner.

Both the abovementioned scenarios would lead to improvement in network reliability due to decreased interruption durations.

Finally, line faults are a major cause of voltage dips. The management of voltage dips and their impact on customers is an important aspect of any utilities management of power quality. Fast and accurate identification of the causes of dips enables a utility to manage the impact of voltage dips on customers.

6.4 Future Work

The main conclusions drawn from this work are based on the analysis of measurements from an operational transmission network, each associated with a cause. Measurements from lower voltage levels more typical of distribution systems ($\leq 132\text{kV}$) are required for testing of automatic classification using waveform features at these levels.

A classification accuracy of 90% is achieved using only the identified contextual features. This result indicates that this automatic classification approach can be applied to determine fault causes where voltage and current measurements are not available.

Looking at the operational application of these results, an opportunity exists for integrating classification into the operational measurement and analysis systems of utilities. The questions that arise are whether these systems collect and store data in a manner that facilitates the process of automatic classification.

REFERENCES

- Alvehag, L., & Soder , L. (2011). A reliability model for distribution systems incorporating seasonal variations in severe weather. *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 910-919.
- Barrera , V., Melendez, J., Kulkarni, S., & Santoso , S. (2012). Feature analysis and automatic classification of short circuit faults resulting from external causes. *European Transactions on Electrical Power*, doi 10.1002/etep.674.
- Bernadi, M., Borghetti, A., Nucci, C., & Paolone, M. (2006). A statistical approach for estimating the correlation between lightning and faults in power distribution systems. *2006 International Conference on Probabilistic Methods Applied to Power Systems*. Stockholm.
- Bollen , M., Goossens, P., & Robert, A. (2003). Assessment of voltages in HV-networks: deduction of complex voltages from measured rms voltages. *IEEE Transactions on Power Delivery*, Vol. 18, No.3, pp. 783-790.
- Bollen, M. (2003). Algorithms for characterizing measured three-phase unbalanced voltage dips. *IEEE Transactions on Power Delivery*, vol. 18, no. 3, pp.937-944.
- Bollen, M., Gu, I., & Styvaktakis, E. (2007). Classification of Underlying Causes of Power Quality Disturbances: Deterministic versus Statistical Methods. *EURASIP Journal on Advances in Signal Processsing*, vol. 1, pp. 1-17.

Bollen, M., Gu, I., Santoso, S., Mcgranaghan , M., Crossley, P., Ribeiro, M., et al. (2009). Bridging the gap between signal and power. *IEEE Signal Processing Magazine*, vol. 26, no. 4, pp. 12-31.

Burger, A., & Sarduski, K. (1995). Experimental investigation of bird initiated ac flashover mechanisms. . *Cigré Study Committee 33-95, Workgroup 07 Seminar*. Mabula, South Africa.

Burnham, J. (1995). Bird streamer flashovers on FPL transmission lines. *IEEE Transactions on Power Delivery*, vol.10, no. 2, pp. 970-977.

Cai, Y., & Chow, M. (2009). Exploratory analysis of massive data for distribution fault diagnosis in smart grids. *IEEE PES General Meeting*. Calgary.

Cai, Y., Chow, M., Lu, W., & Li, L. (2010). Statistical feature selection from massive data in distribution fault diagnosis. *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp.642-648.

Chow, M., Yee, S., & Taylor, L. (1993). Recognizing animal-caused faults in power systems using artificial neural networks. *IEEE Transactions on Power Delivery*, vol. 8, no. 3, pp. 1268-1273.

Chowdhury, A., & Koval , D. (2000). Development of transmission reliability performance benchmarks. *IEEE Transactions on Industrial Applications*, vol. 36, no. 3, pp. 899-903.

Chowdhury, A., & Koval, A. (2008). High voltage transmission equipment forced outage statistics including different fault types. *Proceedings of the 10th international*

conference on probabilistic methods applied to power systems , PMAPS '08.
Rincon.

Cigre. (2011). *Economic Framework for Power Quality, JWG Cigre-CIRED C4.107.*

Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, vol. IT-13(1), PP.21-27.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, vol. 1, pp.131-156.

Davies, D., Vosloo , F., & Vannan, S. (2008). Near real-time fire alert system in South Africa: from desktop to mobile service. *ACM conference on Designing Interactive Systems 2008*. Cape Town.

Duda, R., Hart, P., & Stork, D. (2000). *Pattern Classification*. Wiley-Interscience.

Duin, R., Juszczak, P., Paclik, P., Pekalska, E., De Ridder, D., Tax, D., et al. (2000). *PRTools 4, A Matlab Toolbox for Pattern Recognition*. Delft University of Technology.

Duin, R., Roli, F., & De Ridder, D. (2002). A note on core research issues for statistical pattern recognition. *Pattern Recognition Letters*, vol. 23(no.1), pp. 493-499.

Duncan Glover, J., & Sharma , M. (2002). *Power System Analysis and Design*. Pacific Grove: Brooks/Cole.

Edimu, M., Alvehag, K., Gaunt , C., & Herman , R. (2013). Analyzing the performance of a time-dependent probabilistic approach for bulk network reliability assessment, vol. 104, no.1. *Electric Power Systems Research*, pp. 156-153.

Edimu, M., Gaunt, C., & Herman, R. (2011). Using probability distribution functions in reliability analyses. *Electric Power Systems Research*, vol.81, no. 4, pp. 915-921.

Eskom. (2009). *Annual Report, 2009*. Sandton.

Eskom. (2010). *Ground Flash Density Map of South Africa from 1 March 2006 to 1 March 2010 (20km resolution)*. Sandton.

Evert, R., & Schulze , G. (2005). Impact of a new lightning detection and location system in South Africa. . *Inaugural IEEE PES Conference and Exposition*. Durban.

Fernandez, R., & Rojas, H. (2002). An overview of wavelet transform in application in power systems. *Proceedings of the 14th Power Systems Computation Conference*. Sevilla.

Fortescue, C. (1918). Method of symmetrical coordinates applied to the solution of polyphase networks. *Transactions of the AIEE*, vol. 37, no.2, pp. 1027-1140.

Garcia, V., Mollineda, R., & Sanchez, J. (2008). On the k-nn performance in a challenging scenario of imbalance and overlapping. *Pattern Analysis and Applications*, Vol. 11, pp. 269-280.

Gargoom, A., Ertugul, N., & Soong, N. (2005). A comparative study in effective signal processing tools for power quality monitoring. *Proceedings of the 11th European Conference on Power Electronics and Applications*. Dresden.

Glover, J., & Sarma, M. (1994). *Power System Analysis and Design*. USA: PWS Publishing Company.

Gu, I., & Styvaktakis, E. (2003). Bridge the gap: Signal processing for power quality applications. *Electric Power Systems Research*, vol. 66, no. 1, pp. 83-96.

Guikema, S., Davidson, R., & Liu, H. (2000). Statistical models of the effects of tree trimming on power system outages. *IEEE Transactions on Power Delivery*, vol. 21, no. 3, pp. 1549-1557.

Guyon, I. (2008). Practical feature selection: from correlation to causality. *Mining Massive Datasets for Security*.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable feature selection. *Journal of Machine Learning Research*, vol. 3, pp.1157-pp.1182.

Guyon, I., Gunn, S., Ben Hur, A., & Dror, G. (2004). Result analysis of the NIPS2003 feature selection challenge. *Proceedings of the NIPS2004 Conference*. Vancouver.

Herman , R., & Gaunt, C. (2010). Probabilistic interpretation of customer interruption cost (CIC) applied to South African systems. *PMAAPS Conference*. Singapore.

IEC. (2008). *Electromagnetic compatibility (EMC) - Part 4-30: Testing and measurement techniques - Power quality measurement methods*. Geneva: IEC.

Iliceto, F., Babalioglu, M., & Dabanli, F. (1981). Report of failures due to ice, wind and large birds experienced on the 420kV lines of Turkey. *Paper 111-15, Cigre Symposium 22-81*. Stockholm.

Jain, A., Duin, R., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37.

Jandrell, I., Blumenthal, R., Anderson, R., & Trengrove, E. (2009). Recent lightning research in South Africa with a special focus on kerunopathology. *16th International Symposium on High Voltage Engineering*. Cape Town.

Jonker, P., Duin, R., & De Ridder, D. (2003). Pattern recognition for metal defect detection. *Steel Grips*, vol. 1(no.1), pp. 20-23.

Koch, R., Botha, A., Johnson, P., McCurrach, R., & Ragoonanthun, R. (2001). Developments in voltage dip (sag) characterisation and the application to compatibility engineering of utility networks and industrial plants. *Proceedings of the 3rd Southern African Power Quality Conference*. Livingston.

Kohavi, R., & John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, vol. 97, pp.273-324.

Kubat, M., Holte, R., & Matwin, S. (1998). Machine learning for the detection of oil spills in radar images. *Machine Learning*, vol. 30, no. 1, pp. 195-215.

Lyon, W. (1954). *Transient analysis of alternating current machinery*. New York: J. Wiley and Sons.

Macey, R., Vosloo, W., & De Turreil, C. (2006). Chapter 4: Environmental Considerations. In *The practical guide to outdoor high voltage insulators*, Eskom Power Series Volume 3. Crown Publications.

Mcferren, G., & Frost, P. (2009). The Southern African fire information system. *6th International ISCRAM International Conference*. Gothenburg.

Meher, S., & Pradhan, A. (2010). Fuzzy classifiers for power quality event analysis. *Electric Power systems Research*, vol. 80, no. 1, pp. 71-76.

Michener, H. (1928). Where engineer and ornithologist meet: transmission line troubles caused by birds. *The Condor*.

Minnaar, U., Gaunt, C., & Nicolls, F. (2010). Signal processing tools for voltage dip analysis. *Proceedings of the Southern African Universities Power Engineering Conference (SAUPEC2010)*. Johannesburg.

Minnaar, U., Gaunt, C., & Nicolls, F. (2012). Characterisation of power system faults on South African transmission power lines. *Electric Power Systems Research*, vol. 88, no. 1, pp. 25-32.

Mora-Florez, J., Cormane-Angarita, J., & Ordonez-Plata, G. (2009). K-means algorithm for power quality applications. *Electric Power Systems Research*, vol. 79, no. 5, pp. 714-721.

Nan, Y., Chai, K., Lee, W., & Chai, H. (2012). Optimizing f-measure: a tale of two approaches. *29th International Conference on Machine Learning*. Edinburgh.

NERSA. (2007). *NRS 048-2:2007, Electric Supply - Quality of Supply Part 2: voltage characteristics, compatibility levels, limits and assessment methods.*

Nunez, V., Kulkarni, S., Santoso, S., & Melendez, J. (2010). SVM-based classification for overhead distribution fault events. *14th Conference on Harmonics and Quality of Power.* Bergamo.

Pahwa, A., Hopkins, M., & Gaunt, C. (2007). Evaluation of outages in overhead distribution systems of South Africa and Manhattan, Kansas, USA. *7th International Conference of Power System Operations and Planning.* Cape Town.

Patterson, B. (1987). The principle of nested subsets and its implications for biological conservation. *Conservation Biology*, vol.1(no.4), pp. 323-334.

Perez, E., & Barros, J. (2006). Voltage event detection and characterization methods: A comparative study. *Proceedings of the Transmission & Distribution Conference and Exposition: Latin America.* Caracas.

Peter, L., & Vajeth, R. (2005). Strategies for bating environmental influences on overhead lines to improve quality of supply. *IEEE PES Conference and Exposition in Africa.* Durban.

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

Styvaktakis, E., Bollen , M., & Gu, I. (2002). Expert system for classification and analysis of power system events. *IEEE Transactions on Power Delivery*, vol. 17, no.2, pp. 423-428.

Sukhnandan , A., & Hoch, D. (2002). Fire induced flashovers of transmission lines: theoretical models. *Proceedings of the 6th IEEE Africon Conference*. George.

Teive, R., Coelho, J., Charles, P., Lange, T., & Cimino, L. (2011). A bayesian network approach to fault diagnosis and prognosis in power transmission systems. *Proceedings of the 16th International Conference on Intelligent System Application to Power System (ISAP)*. Crete.

Van Rooyen, C., Vosloo, H., & Harness, R. (2003). Watch the birdie: eliminating bird streamers as a cause of faulting on transmission lines. *IEEE Industry Applications Magazine*, pp. 55-60.

Vosloo, H. (2005). *The need for and contents of a life cycle management plan for Eskom transmission line servitudes*. University of Johannesburg.

Vosloo, H. F., Britten, A. C., Burger, A. A., & Muftic, D. (2009). The susceptibility of 400kV transmission lines to bird streamers and bush fires: a definitive case study. *Proceedings of the Cigré 6th Southern African Regional Conference*. Somerset West.

West, H., Brown, J., & Kinyon , A. (1971). Simulation of EHV transmission line flashovers initiated by bird excretion. *IEEE PES Winter meeting, Paper 71, TP 145-WR*. New York.

Xu, L., & Chow, M. (2008). Distribution fault diagnosis using a hybrid algorithm of fuzzy classification and artificial immune system. *Proceedings of IEEE PES General Meeting - Conversion and Delivery of Electrical Energy*. Raleigh.

Xu, L., & Chow, M. (2006). A classification approach for power distribution systems fault cause identification. *IEEE Transactions on Power Systems*, vol. 21, no.1, pp. 53-60.

Xu, L., Chow, M., & Gao, X. (2005). Comparisons of Logistic Regression and Artificial Neural Network on Power Distribution Systems Fault Cause Identification. *2005 IEEE mid-summer workshop on soft computing in industrial applications*. Espoo.

Ziolowski, V., da Silva, I., & Flauzino, R. (2007). Automatic identification of faults in power systems using control technique. *Proceedings of the 16th International Conference on Control Applications*. Singapore.

APPENDIX A: FAULT FREQUENCY DATA FOR SOUTH AFRICAN TRANSMISSION LINES

Fault data is tabulated per cause for the relevant network voltages and rainfall areas.

220 kV Lines

Bird streamer

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
220 kV				
<i>Season 1: Jan-Mar</i>	0.3915; 0.2065	0.0242; 0.0634	0.0097; 0.0399	0.1160; 0.1009
<i>Season 2: Apr-Jun</i>	0.3238; 0.2333	0.1160; 0.1089	0.0048; 0.0199	0.1450; 0.1496
<i>Season 3: Jul-Sep</i>	0.1788; 0.1544	0.1257; 0.1770	0.1015; 0.1284	0; 0
<i>Season 4: Oct-Dec</i>	0.4882; 0.2386	0.0290; 0.0405	0.0242; 0.0634	0.1257; 0.0657

Lightning

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
220 kV				
<i>Season 1: Jan-Mar</i>	0.1469; 0.0929	0.1202; 0.0273	0.1849; 0.0359	0.2907; 0.1001
<i>Season 2: Apr-Jun</i>	0.0647; 0.0273	0.1202; 0.0273	0.1667; 0	0.2450; 0.0483
<i>Season 3: Jul-Sep</i>	0.0693; 0.0323	0.1202; 0.0273	0.1712; 0.199	0.2314; 0.0273
<i>Season 4: Oct-Dec</i>	0.1058; 0.0577	0.1157; 0.0199	0.2214; 0.0634	0.2450; 0.0483

Fire

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
220 kV				
<i>Season 1: Jan-Mar</i>	0;0	0;0	0;0	0;0
<i>Season 2: Apr-Jun</i>	0;0	0;0	0;0	0;0
<i>Season 3: Jul-Sep</i>	0;0	0.0048; 0.0199	0.0242; 0.0698	0;0
<i>Season 4: Oct-Dec</i>	0;0	0;0	0;0	0.0145; 0.0598

Pollution

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
220 kV				
<i>Season 1: Jan-Mar</i>	0.0145; 0.0323	0; 0	0; 0	0.0097; 0.0273
<i>Season 2: Apr-Jun</i>	0.0097; 0.0273	0; 0	0;0	0; 0
<i>Season 3: Jul-Sep</i>	0.0338; 0.0825	0; 0	0; 0	0.0097; 0.0399
<i>Season 4: Oct-Dec</i>	0.0145; 0.0434	0; 0	0; 0	0.0048; 0.0199

Other

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
220 kV				
<i>Season 1: Jan-Mar</i>	0.0338; 0.0508	0; 0	0.0145; 0.0323	0.0242; 0.0634
<i>Season 2: Apr-Jun</i>	0.0290; 0.0819	0.0435; 0.0423	0;0	0.0193; 0.0359
<i>Season 3: Jul-Sep</i>	0.0097; 0.0273	0.0387; 0.1052	0.0338; 0.0825	0.0145; 0.0598
<i>Season 4: Oct-Dec</i>	0.0532; 0.1616	0.0145; 0.0323	0.0338; 0.0585	0.0242; 0.0483

275 kV Lines

Bird streamer

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
275kV				
Season 1: Jan-Mar	0.2018; 0.0743	0.0410; 0.0338	0.0131; 0.0186	0.1268; 0.0790
Season 2: Apr-Jun	0.1964; 0.0576	0.0479; 0.0312	0.0131; 0.0147	0.1399; 0.0644
Season 3: Jul-Sep	0.0874; 0.0380	0.0379; 0.0282	0.0147; 0.0153	0.0820; 0.0414
Season 4: Oct-Dec	0.1013; 0.0457	0.0394; 0.0354	0.0070; 0.0094	0.0657; 0.0335

Lightning

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
275kV				
Season 1: Jan-Mar	0.1395; 0.0578	0.1352; 0.0292	0.3872; 0.1063	0.4040; 0.0676
Season 2: Apr-Jun	0.0738; 0.0155	0.1206; 0.0128	0.1966; 0.0161	0.2551; 0.0335
Season 3: Jul-Sep	0.0672; 0.0215	0.1148; 0.0101	0.1813; 0.0187	0.2492; 0.0260
Season 4: Oct-Dec	0.1381; 0.0443	0.1454; 0.0425	0.4164; 0.1116	0.4420; 0.0814

Fire

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
275kV				
Season 1: Jan-Mar	0; 0	0.0023; 0.0052	0.0178; 0.0502	0.0008; 0.0032
Season 2: Apr-Jun	0.0116; 0.0160	0.0417; 0.0403	0.1701; 0.1732	0.0232; 0.0296
Season 3: Jul-Sep	0.0170; 0.0179	0.0990; 0.0592	0.2861; 0.1377	0.0348; 0.0267
Season 4: Oct-Dec	0.0046; 0.0103	0.0170; 0.0190	0.0603; 0.0544	0.0085; 0.0131

Pollution

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
275kV				
<i>Season 1: Jan-Mar</i>	0.0046; 0.0092	0.0008; 0.0032	0.0015; 0.0044	0.0015; 0.0044
<i>Season 2: Apr-Jun</i>	0.0124; 0.0150	0.0031; 0.0074	0.0023; 0.0069	0.0139; 0.0320
<i>Season 3: Jul-Sep</i>	0.0108; 0.0243	0.0031; 0.0087	0.008; 0.0032	0.0116; 0.0290
<i>Season 4: Oct-Dec</i>	0.0054; 0.0114	0.0015; 0.0064	0.0015; 0.0064	0.0023; 0.0052

Other

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
275kV				
<i>Season 1: Jan-Mar</i>	0.0240; 0.187	0.0247; 0.0236	0.0325; 0.0259	0.0325; 0.0291
<i>Season 2: Apr-Jun</i>	0.0077; 0.0067	0.0209; 0.0228	0.0263; 0.0301	0.0170; 0.0179
<i>Season 3: Jul-Sep</i>	0.0101; 0.0136	0.0193; 0.0208	0.0216; 0.0282	0.0077; 0.0081
<i>Season 4: Oct-Dec</i>	0.0216; 0.0191	0.0209; 0.0242	0.0464; 0.0384	0.0410; 0.0297

400 kV Lines Thunderstorm Rainfall Area

Bird streamer

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Thunderstorm				
<i>Season 1: Jan-Mar</i>	0.1226; 0.0586	0.0314; 0.0243	0.0035; 0.0078	0.0506; 0.0349
<i>Season 2: Apr-Jun</i>	0.1261; 0.0926	0.0593; 0.0254	0.0035; 0.0069	0.0721; 0.0505
<i>Season 3: Jul-Sep</i>	0.0564; 0.0329	0.0244; 0.0232	0.0076; 0.0124	0.0384; 0.0234
<i>Season 4: Oct-Dec</i>	0.0732; 0.0443	0.0110; 0.0130	0.0064; 0.0092	0.0296; 0.0207

Lightning

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Thunderstorm				
<i>Season 1: Jan-Mar</i>	0.0477; 0.0301	0.0052; 0.0062	0.0732; 0.0571	0.0726; 0.0401
<i>Season 2: Apr-Jun</i>	0.0070; 0.0103	0.0041; 0.0061	0.0052; 0.0086	0.0128; 0.0103
<i>Season 3: Jul-Sep</i>	0.0035; 0.0060	0.0012; 0.0033	0.0145; 0.0161	0.0035; 0.0049
<i>Season 4: Oct-Dec</i>	0.0337; 0.0257	0.0099; 0.0144	0.1005; 0.0381	0.0878; 0.0516

Fire

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Thunderstorm				
<i>Season 1: Jan-Mar</i>	0; 0	0; 0	0.0058; 0.0105	0.0012; 0.0048
<i>Season 2: Apr-Jun</i>	0.0052; 0.0093	0.0302; 0.0243	0.1197; 0.1300	0.0081; 0.0132
<i>Season 3: Jul-Sep</i>	0.0145; 0.0175	0.0808; 0.0548	0.4103; 0.2085	0.0296; 0.0229
<i>Season 4: Oct-Dec</i>	0.0029; 0.0058	0.00128; 0.0109	0.0465; 0.0345	0.0058; 0.0121

Pollution

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Thunderstorm				
<i>Season 1: Jan-Mar</i>	0.0070; 0.0125	0; 0	0; 0	0.0006; 0.0024
<i>Season 2: Apr-Jun</i>	0.0029; 0.0058	0.0029; 0.0058	0; 0	0.0006; 0.0024
<i>Season 3: Jul-Sep</i>	0.0064; 0.00218	0.0122; 0.0199	0.006; 0.0024	0; 0
<i>Season 4: Oct-Dec</i>	0; 0	0.0006; 0.0024	0; 0	0; 0

Other

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Thunderstorm				
<i>Season 1: Jan-Mar</i>	0.0099; 0.0105	0.0110; 0.0160	0.0163; 0.0215	0.0105; 0.0102
<i>Season 2: Apr-Jun</i>	0.0093; 0.0262	0.0139; 0.0213	0.0128; 0.0143	0.0064; 0.0060
<i>Season 3: Jul-Sep</i>	0.0174; 0.0283	0.0163; 0.0135	0.0244; 0.0333	0.0093; 0.0183
<i>Season 4: Oct-Dec</i>	0.0110; 0.0114	0.0087; 0.0092	0.0291; 0.0274	0.0192; 0.0208

400 kV Lines Frontal Rainstorm Area

Bird streamer

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Frontal				
<i>Season 1: Jan-Mar</i>	0.0863; 0.0322	0.0211; 0.0264	0.0037; 0.0117	0.0266; 0.0245
<i>Season 2: Apr-Jun</i>	0.0964; 0.0341	0.0569; 0.0321	0.0046; 0.0073	0.0496; 0.0341
<i>Season 3: Jul-Sep</i>	0.0844; 0.0353	0.0477; 0.0151	0.0018; 0.0052	0.0385; 0.0313
<i>Season 4: Oct-Dec</i>	0.0955; 0.0409	0.0073; 0.0097	0.0046; 0.0092	0.0257; 0.0165

Lightning

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Frontal				
<i>Season 1: Jan-Mar</i>	0.0477; 0.0301	0.0052; 0.0062	0.0732; 0.0571	0.0726; 0.0401
<i>Season 2: Apr-Jun</i>	0.0070; 0.0103	0.0041; 0.0061	0.0052; 0.0086	0.0128; 0.0103
<i>Season 3: Jul-Sep</i>	0.0035; 0.0060	0.0012; 0.0033	0.00145; 0.0161	0.0035; 0.0049
<i>Season 4: Oct-Dec</i>	0.0337; 0.0257	0.0099; 0.0144	0.1005; 0.0381	0.0878; 0.0516

Fire

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Frontal				
<i>Season 1: Jan-Mar</i>	0.0009; 0.0038	0.0064; 0.00157	0.0413; 0.0514	0.0110; 0.0205
<i>Season 2: Apr-Jun</i>	0.0028; 0.0082	0.0018; 0.0052	0.0193; 0.0300	0.0009; 0.0038
<i>Season 3: Jul-Sep</i>	0.0028; 0.0061	0.0046; 0.00120	0.0119; 0.0161	0.0028; 0.0061
<i>Season 4: Oct-Dec</i>	0.0018; 0.0076	0.0064; 0.0124	0.0395; 0.0512	0.0009; 0.0038

Pollution

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Frontal				
<i>Season 1: Jan-Mar</i>	0.0762; 0.2275	0.0220; 0.0394	0.0110; 0.0257	0.0138; 0.0270
<i>Season 2: Apr-Jun</i>	0.0092; 0.0157	0.0073; 0.0167	0.0009; 0.0038	0.0046; 0.0092
<i>Season 3: Jul-Sep</i>	0.0037; 0.0088	0.0018; 0.0076	0; 0	0.0055; 0.0165
<i>Season 4: Oct-Dec</i>	0.0046; 0.0107	0.0009; 0.0038	0; 0	0.0009; 0.0038

Other

S/I	00:00-05:59	06:00-11:59	12:00-17:59	18:00-23:59
400kV Frontal				
<i>Season 1: Jan-Mar</i>	0.0092; 0.0085	0.0092; 0.0085	0.0102; 0.0115	0.0116; 0.0114
<i>Season 2: Apr-Jun</i>	0.0116; 0.0254	0.0116; 0.0254	0.0130; 0.0170	0.0089; 0.0148
<i>Season 3: Jul-Sep</i>	0.0123; 0.0166	0.0123; 0.0166	0.0160; 0.0147	0.0078; 0.0106
<i>Season 4: Oct-Dec</i>	0.0106; 0.0164	0.0106; 0.0164	0.0102; 0.0052	0.0153; 0.0145