Identification of Weak Areas and Worst Served Customers for Power Quality Issues Using Limited Monitoring and Non-Deterministic Data Processing Techniques

A thesis submitted to The University of Manchester for the Degree of

Doctor of Philosophy

in the Faculty of Engineering and Physical Sciences

2012

Mr Nick C Woolley, MEng

Table of Contents

Final word count: 78,718

List of Figures

List of Tables

Abstract

Title: *Identification of Weak Areas and Worst Served customers for Power Quality Issues Using Limited Monitoring and Non-Deterministic Data Processing Techniques*
Mr Nick C Woolley ,The University of Manchester, 2012

The current international trend in distribution networks is towards increased monitoring. This trend is being driven by distribution network operators (DNOs) who hope that through increased monitoring, they will be able to optimise capital and operational expenditure and thus operate a more efficient networks.

One of the key areas of focus relating to the increased interest in distribution network monitoring is power quality. Power quality disturbances affect consumers by interrupting equipment or halting industrial processes and can result in very significant financial losses. DNOs are also financially impacted by power quality issues if they breach regulatory limits or contractual arrangements. To extract value from power quality monitoring, DNOs must process and then interpret data from a variety monitoring devices placed at different locations all potentially measuring different quantities. The challenge of how best to extract useful and practical power quality information from disparate monitoring devices is the subject of this thesis.

This thesis describes and develops monitoring techniques for two power quality phenomena: voltage sags and unbalance. The research presents new techniques which can graphically identify the weakest areas and the worst served customers for voltage sags and unbalance. All the developed techniques utilise non-deterministic methods (such as statistics and artificial intelligence) to deal robustly with network and measurement uncertainties. This thesis can be dissected into four areas: voltage sag monitoring, optimal power quality monitor placement, voltage unbalance monitoring and identification of the weakest areas and worst served customers for both issues.

The first section of this thesis is dedicated to voltage sags. This section introduces a multi-step process to identify and estimate the impacts of voltage sags within networks. The first stage in this process is classification and detection where several different classification methods (including immune inspired techniques) are compared to determine which algorithms work best under the context of limited monitoring. The research then proposes a novel robust method for performing fault location and voltage sag profile estimation using multiple monitors. The method pays particular attention to the errors in measurement inputs and identifies the most likely location for both the fault location and the voltage magnitude using statistical methods. The voltage sag monitoring research concludes by defining the probable impacts of voltage sags on customers, and by introducing a new measure known as the sag trip probability.

The second major section covered by this thesis is optimal monitor placement. This thesis presents a comprehensive methodology which enables network operators to place monitors in locations best suited for voltage sag monitoring based on future likely topological and loading changes.

The third major section covered by this thesis is unbalance monitoring. A three phase distribution system state estimation model is developed which can estimate the location and impact of unbalance within the network, without assuming the loading is balanced.

The final section of this thesis shows how the worst served customers and the weakest areas of the network can be identified presents for both voltage sag and unbalance using limited monitoring and the developed techniques. The results are presented graphically using a series of topological heat maps, and these show visually how the techniques could work to monitor a distribution network.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i) The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii) Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii) The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv) Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy[1], in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations[2] and in The University's policy on Presentation of Theses.

[1] See: http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487
[2] See: http://www.manchester.ac.uk/library/aboutus/regulations

Acknowledgement

I must express my gratitude to my supervisor, Prof Jovica Milanović. It has been a pleasure and a privilege to collaborate and work with him throughout the project. The effort he has applied to review this thesis and the many publications which have arisen as part of this project, has been exceptional.

Special thanks must also go to the Power Networks Research Academy (PNRA) and the Engineering and Physical Research Council (EPSRC) who have jointly sponsored this project. In particular, I would like to thank several members of E.ON New Build and Technology and Western Power Distribution who have helped bring an industrial perspective to this research. My warmest thanks go to Mr Bob Ferris, Miss Kate Grant, Mrs Rachel Stanley and Mr Nigel Johnson.

I am especially appreciative for all of the help, support and guidance of each of the other members of Prof Jovica Milanović's Power Quality and Power Systems Dynamics Group at The University of Manchester. Special thanks go to Mr Robin Preece, Mr Manuel Avendaño-Mora and Mr Muhammad Ali who have selflessly given up countless hours to discuss many of my more fanciful research ideas. A special footnote of gratitude must go to Mr Robin Preece, who has helped me become a more accomplished chess player.

Lastly, but not least, I would like to thank my parents, my wife's parents and my wife, Alice. Alice is an accomplished mathematician, and she has graciously allowed me to talk through many of the complex areas of my research. Without these discussions, this research would not have been possible.

To my wife, parents and dziadkowie

1 Introduction

The motivation for this research is to gain a greater understanding of power quality issues (specifically unbalance and voltage sags) through intelligent monitoring. It is hoped that this will help improve customer service, facilitate the implementation of power quality contracts, reduce maintenance, guide the implementation of mitigation solutions and help obtain visibility of power quality issues throughout the network.

The scope of this research is specifically limited to distribution networks where there is a keen interest in power quality monitoring and a requirement for intelligent monitoring.

The current level of power quality monitoring within distribution networks is such that distribution network operators (DNOs) and customers do not have visibility over power quality issues on many 33kV and 11kV sections of network. An understanding of the level of power quality throughout a network will allow DNOs to pro-actively maintain their networks and mitigate power quality issues where they are not adequate. Through intelligent monitoring, DNOs, regulators and customers should be able to obtain greater visibility of issues to help plan maintenance, implement mitigation solutions and comply with power quality contracts and regulations.

1.1 Background

The term *power quality* in relation to the supply of electricity in a network refers to the compatibility of the power available at a particular busbar of the network with the equipment connected to that busbar[1]. Incompatibilities between the electricity supply and the load can lead to significant negative effects, including tripping of devices, over-heating of machinery, premature aging of electrical assets and significant financial losses for both the end-user and the network operator [1, 2].

Power quality issues can be categorized into a number of different areas including harmonics, voltage sags and swells, voltage unbalance, over voltages and under-voltages, surges, transients and voltage flicker.

This thesis will cover two of these areas: *voltage sags* and *voltage unbalance*.

1.1.1 The Importance of Power Quality

The primary concerns of electricity consumers were traditionally cost and reliability. Over the past twenty years or so, power quality has become more important. Industrial customers with sensitive electromagnetic equipment are concerned by a voltage sag or significant unbalance if it causes a costly interruption to an industrial process.

The compatibility between loads and the network to power quality disturbances can be broadly described using Fig. 1.1 (adapted from [1]).

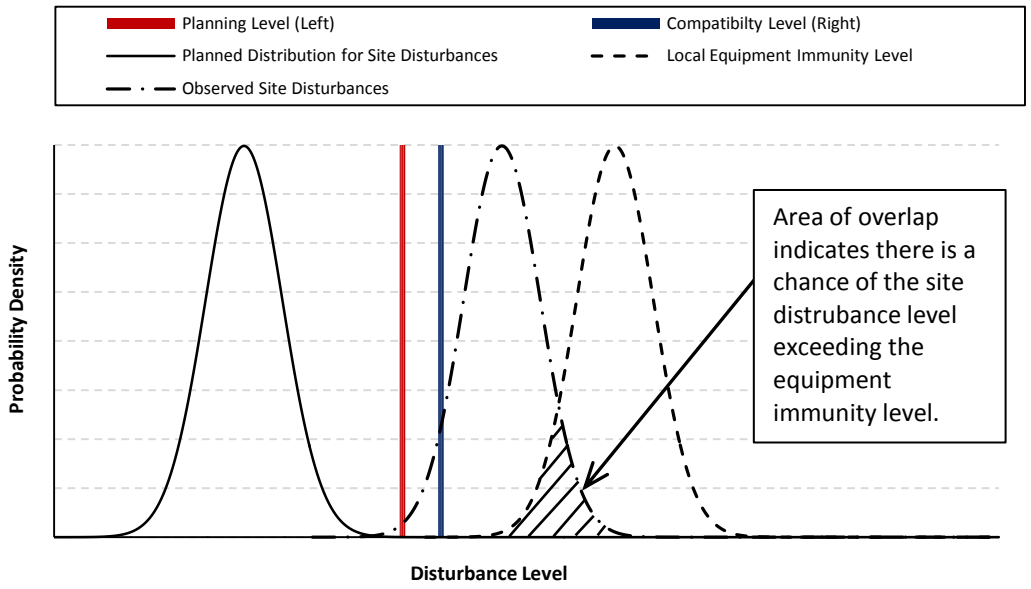


Fig. 1.1. The planning and compatibility levels for power quality events. Adapted from [1].[3]

Fig. 1.1 shows the electro-magnetic coordination between disturbances in a network and locally attached equipment. In theory, devices (such as motors, adjustable speed drives, switched mode power supplies, etc.) are capable of withstanding disturbances up to the electro-magnetic *compatibility level* (shown as the right hand bar in Fig. 1.1). Networks are designed such that disturbances are emitted below the *planning level* (shown as the left hand bar in Fig. 1.1). The designed distribution of site disturbances is shown as the far left hand curve in Fig. 1.1. If the observed distribution of site disturbances is high

[3] Note that there is no reason why the two distributions shown here should follow the Gaussian normal distribution. The statistical distribution will vary depending on network and equipment parameters.

enough (as shown in the middle curve of Fig. 1.1), it can overlap with the distribution of disturbances which can be tolerated by local electrical devices (as shown by the right hand curve of Fig. 1.1). The level of overlap between the two probability distributions indicates probable negative consequences which are often manifested in the form of trips.

The recent relevance and importance of power quality arises from structural and operational changes currently occurring in power networks world-wide:

- Increasing numbers of power electronic loads and generators are being installed which are more sensitive to power quality disturbances than traditional power system loads. These newer load types are more likely to cause power quality disturbances themselves, as is the case with several types of power convertors.
- The regulatory environments are increasing operational pressure on utilities to reduce interruptions and improve the quality of electrical supply. For example, under UK regulation (ofgem's Distribution Price Control Review 5 (DPCR5)), the number and length of all customer interruptions above 3 minutes are routinely monitored, as well as the quality of service provided by call centres.

The objective of power quality monitoring, and this thesis to a certain extent, is to ascertain the probability distribution for site disturbances, to ensure that they match planning and compatibility limits. By using an intelligent monitoring system, this thesis will aim to process standard power quality monitoring information (such as voltages and currents) to obtain probabilistic estimates for the power quality performance of the network.

1.1.2 The Effects of Poor Power Quality

Poor or inadequate power quality impacts on end users and network operators in different ways.

End users will be severely affected if a power quality issue causes an industrial process to be interrupted. The impact of a process interruptions may include lost or wasted work in progress, costs caused by idle staff who are no longer able to work, process slow down, equipment damage, costs associated with restarting the process, potential penalties due to late delivery of finished goods and decreased competiveness [1].

Across the European Union, the total annual costs attributable to power quality related issues were estimated at €150bn [3]. In 2008 / 2009, voltage sags (and faults) caused an

average of 0.73 interruptions per customer and contributed to an average of 76 customer minutes lost (CMLs) over the course of a one year period. These interruptions cause different impacts to different customers. For example, a momentary interruption for a large customer is estimated to cost £216k, whereas a 4 hour interruption on a residential customer is estimated to cost only £4.78 [4]. If these figures are scaled across the UK, the costs of these interruptions are equivalent to economic losses of £3.7bn[4] [2, 5].

Unbalance causes losses in the network, trips (if it is severe enough, and in a critical position), overheating of assets and de-rating of equipment. There were no research surveys available on the exact level of unbalance within the UK network so it is not possible to determine the exact amount of unbalance in the UK network. Unbalance close to the regulatory limit of 2% adds additional losses to network operation, which although negligible, can be quantified [1]. For example, induction motors rated above 100kW and exposed to 2% unbalance generate additional losses of 2.4% [1], synchronous generators rated above 100kW generate additional losses of 4.2% and transformers generate losses between 1% and 4% [1]. Losses generated in transformers are roughly 1/3 of all technical network losses [6] in the UK network. Therefore assuming 2% unbalance is widespread in UK distribution networks, the extra losses caused by unbalance through transformers could be equivalent to between 63GWh and 250GWh[5] for the entire UK [7]. Using a loss incentive of £50 / MWh (as stipulated in DPCR5 [7]), this equates to loss incentives worth between £3.2m and £12.6m.

Distribution network operators (DNOs) are impacted when power quality causes them to breach quality of supply contracts with end users and regulators. DNOs must ensure that their network is maintained to limit power quality disturbances, for example, by cutting tress, maintaining switchgear, washing insulators and maintaining assets. If the level of power quality does not meet the correct regulatory or contractual requirements then DNOs may be required to spend money addressing these requirements, for example by installing FACTS devices [8], adding harmonic filters, rebalancing loads or replacing utility assets. Costs are also incurred as DNOs deal with end user power quality complaints. It should be noted that there is no overall general requirement for a particular level of power quality; although there are often specific requirements which may be set for interruptions, harmonics and the level of unbalance.

[4] Based on the average CML / customer of 1hour / customer, 92% domestic, 6% commercial and 2% industrial customers, 29mln customers. Inflation adjusted between 1996 & 2010.
[5] Total distribution network losses of 18,777GWh in 2009-2010 (available from ofgem loss report).

1.1.3 Why Should a DNO Monitor Power Quality?

Intelligent monitoring could help to improve customer service by reducing complaints and increasing customer satisfaction. Monitoring should also help to facilitate improvements in network efficiency by reducing losses, optimising maintenance and reducing interruptions.

Network wide knowledge of power quality issues also presents a large opportunity for DNOs to potentially make money from new power quality contracts. This will enable DNOs to profit from selling both highly reliable and less-reliable power quality zones [2]. High quality power zones are almost non-existent in the UK [2], but they appeal to international industrial customers where the impact of a power quality issue generates significant financial consequences. In France, distribution and transmission companies offer all customers the ability to be able to contract for extra power quality requirements which are then overseen by the regulator. Networks are also becoming increasingly interested in selling lower quality power as an alternative to network expansion. For example, the Electricity North West Capacity to Customers (C2C) project [9] is seeking to re-use existing network capacity and connect major energy users on less reliable contracts. The results of power quality monitoring will help to quantify the level of unreliability these customers can expect on their new connections.

With intelligent monitoring network operators will also be able to gauge the performance of their network against European regulatory standards such as EN50160 [10] and national regulators. Close monitoring of regulatory limits could even generate extra income for DNOs, as high power quality performance is often incentivised by the regulator.

1.2 Voltage Sags

Voltage sags (voltage dips) are short (typically half a cycle to several seconds) reductions in the RMS voltage at buses in a power system network [11, 12]. They are often caused by faults, large motors starting, transformers energizing and loose and defective wiring. Voltage sags form the most significant component of power quality problems in distribution networks [13] both in terms of gross numbers of events and the high associated costs to end users [11, 14-16]. A study for the European Union estimated that voltage sags and short interruptions contributed to an annual loss of €86bn [3].

Voltage sags are most often characterized by two features: firstly, the voltage sag *magnitude* that defines the magnitude of the retained RMS voltage (expressed in per unit or as a percentage) and secondly the *duration* that defines the length of time that the voltage remains below a specified threshold (typically 0.9 per unit). They are stochastic in nature as they are influenced by the a variety of random factors, including fault type, position and pre-fault voltage [17].

The effects of a voltage sag is also dependent on two other important factors, in addition to magnitude and duration. The first is the *phase angle shift* which is defined by the difference of the pre-sag voltage angle and post sag voltage angle. The second factor relates to the *point on the wave* at which the voltage sag starts. This refers to the instantaneous voltage drop caused when the voltage sag starts [11], which varies dependent on the sag.

A typical three phase voltage profile for a voltage sag is shown in Fig. 1.2:

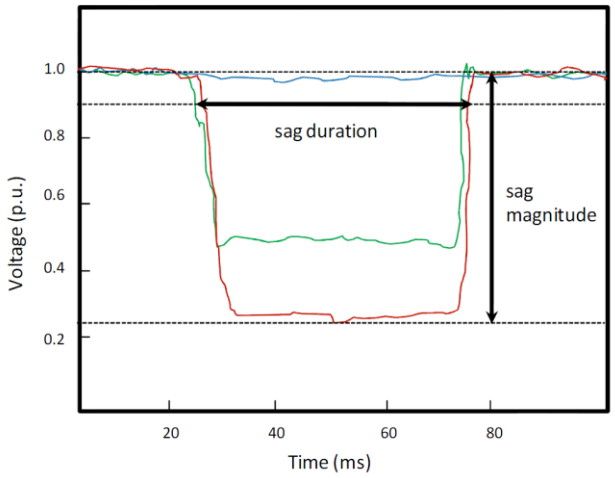


Fig. 1.2. A voltage sag. Adapted from [2].

The impact of voltage sags on end-user equipment is highly dependent on the types of loads connected to the electrical supply. Modern manufacturing processes involve sets of cascaded machinery which can be disrupted if only one of the devices attached to the process is interrupted [18]. Both equipment and processes may be interrupted if the voltage sag is severe enough.

The ultimate consequence of a trip caused by a voltage sag is economic losses. The losses from process interruptions can be extremely high, ranging from €250,000 for the glass industry to €3,800,000 for semiconductor production [19]. Financial losses arise from a variety of factors including lost output revenue, lost opportunities, reduced asset operating life time, product damage, repairs and lost productivity [20].

1.3 Voltage Unbalance

Voltage unbalance is caused by the asymmetric operation of multi-phase electrical power networks. It is regarded as a significant problem which negatively affects both distribution network operators (DNOs) and electrical consumers [21].

IEC 61000 [22] defines voltage unbalance as a condition in multi-phase power systems where the magnitude of phase to earth voltages and angles are not equal. The voltage unbalance factor (VUF) is defined as shown in equation (1.1).

$$VUF = V^{(2)} / V^{(1)} \times 100 \tag{1.1}$$

Where $V^{(2)}$ is the complex negative sequence voltage, and $V^{(1)}$ is the complex positive sequence voltage. The VUF quantifies the level of negative sequence voltage unbalance normalized to the size of the positive sequence voltage.

Voltage unbalance is generated through the emission of zero and negative sequence current at sources throughout the network. Emission sources include asymmetric un-transposed or partially transposed transmission lines [23], single phase and dual-phase loads and unbalanced three phase loads.

Voltage unbalance is a steady-state phenomenon [22], and is separate from the unbalance that arises during asymmetric short circuit faults.

Where there is a path for zero sequence current to flow through the network, it is also possible to observe zero-sequence unbalance. Zero-sequence unbalance is controlled by network system design, rather than emission limits [22]. Zero-sequence unbalance can be problematic where there exists a path for zero-sequence currents to flow up or down voltage levels within a network. However, most networks naturally control zero-sequence current flow through blocks in the current path, such as Δ-Y transformers. From a DNO's perspective, it is important to have a strong understanding of the level of unbalance within the network as this will affect three phase customers, network losses, heating, operational limits (of three phase machines, cables and lines) and influence customer minutes lost (CML). The DNO also has a regulatory requirement to operate their network within statutory unbalance limits (<2% in the UK [10]). It is important for the DNO to understand whether the unbalance is being generated from within its own network, or from unbalanced customer loads.

From a consumer's perspective, the effects of unbalance include three phase equipment trips, heating of industrial machinery (caused by increased negative and zero sequence currents) and torsional oscillations. Unbalance creates additional heating on electrical

equipment which results in a shortened in-service time for equipment installations. This ultimately means the equipment operator may have to replace their equipment earlier than expected; for example, 2.3% earlier for transformers and 9.1% for induction motors both operating at 2% unbalance [1]. The presence of significant unbalance may also necessitate that three phase machinery is de-rated to operate at less than full load.

1.4 Equipment Immunity

Equipment immunity defines the ability of an electrical or electronic device to be able to withstand a power quality disturbance. *Process immunity* defines the ability of a process involving interrelated electrical machinery to withstand power quality disturbances.

It is important to maintain *electromagnetic compatibility* between the equipment immunity of loads in the network and network itself. This includes ensuring that loads do not introduce intolerable disturbances into the network and that they are able to function satisfactorily during power quality events within regulatory limits.

1.4.1 Equipment Immunity to Voltage Sags

The immunity of information technology (IT) and process control equipment to voltage sags is often represented using the Information Technology Industry Council (ITIC) [20] or Computer Business Equipment Manufacturers Association (CBEMA) [20] curve. The ITIC curve is shown in Fig. 1.3.

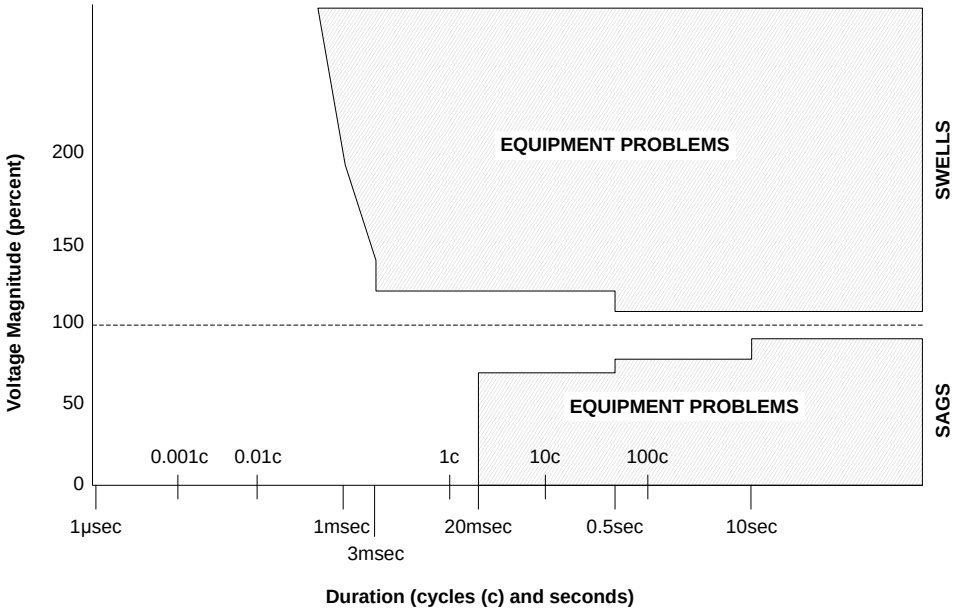


Fig. 1.3. The ITIC voltage sag immunity curve.

The aim of the ITIC curve is to define the operation of equipment after it is exposed to RMS voltage variations with different magnitudes and durations. The curve splits the

operating zone into a safe middle operating zone where no trips are likely to occur, and two upper and lower zones where trips are likely. IT equipment can withstand small voltage sags of a short duration because the internal voltage regulator (principally a capacitor) within many switch mode power supplies can maintain a constant DC voltage for a short length of time across a wide range of input voltages. Another curve which is similar in representation to the ITIC curve is the SEMI F-47 [24] curve. SEMI F-47 defines the voltage sag immunity of semiconductor manufacturing equipment.

Other types of devices such as contactors, induction motors and adjustable speed drives are also sensitive to voltage sags [25]. Contactors and electromechanical relays are used to control loads, such as motors [20]. Contactors have been found to open when the voltage drop is between 50% and 70% of nominal operating voltage [26]. They are also sensitive to the point-on-wave of sag initiation [25]. The response of an induction motor during a voltage sag will depend on the torque speed characteristic of the motor and the amount of mechanical load it is currently serving. A voltage sag of the correct magnitude, duration, phase angle jump and type can cause an induction motor to stall [20]. Adjustable speed drives can be severely affected during voltage sags if the drive protection acts to prevent damage to the device, if the controller fails, or if post-sag over-currents blow fuses protecting power electronics components [25].

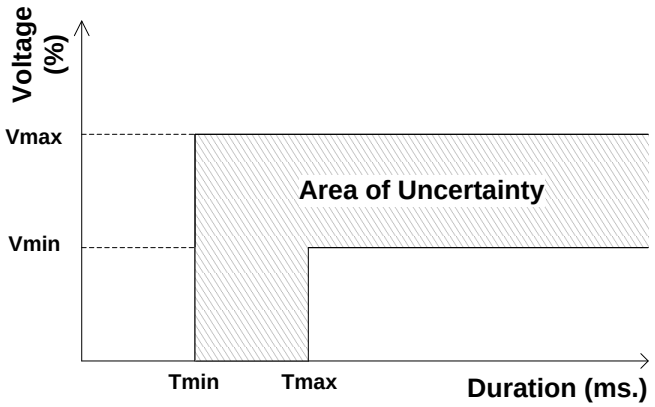


Fig. 1.4. The area of uncertainty associated with voltage sag operation

The immunity of each of these types of equipment to a voltage sag can be summarised using a curve defined in a similar way to Fig. 1.4. Fig. 1.4 describes an area of uncertainty where the operation of the equipment after the sag will be uncertain. This may be caused by factors such as manufacturing tolerances or un-monitored variables such as the point-on-wave of initiation of the sag. The main area of interest in Fig. 1.4 is the region of uncertainty [2], and the core focus of equipment immunity studies is to define this as accurately as possible.

More detailed information on equipment immunity to voltage sags can be found in [20].

1.4.2 Equipment Immunity to Unbalance

The sensitivity of electrical equipment to voltage unbalance varies across different devices. In contrast to voltage sags it is predominantly only three phase electrical equipment that is directly affected by unbalance. The concept of an immunity curve is also not defined for unbalance in the same way as for voltage sags.

Induction motors are affected in three ways by voltage unbalance [1]. The first consequence is that the motor may not be able to operate at full torque. This is caused because the negative sequence voltage applied to the terminals of the induction motor under unbalanced conditions generate a rotating elliptical (rather than circular) magnetic field and applies a braking torque on the motor. Under unbalanced conditions, the motor also suffers from increased mechanical wear on the bearings of the device, because torque components are introduced at double the system frequency. The last consequence is that the stator and the rotor of the motor are heated through the rotating magnetic field of the negative sequence components. If the level of unbalance is large, induction motors must be derated according to the NEMA machine derating curve (as shown in Fig. 1.5).

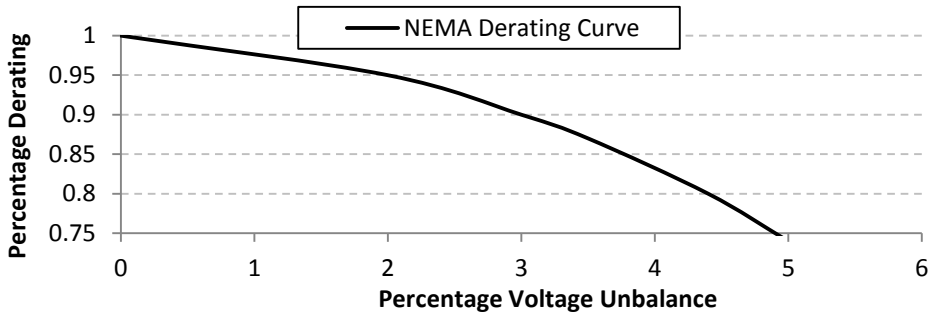


Fig. 1.5. The NEMA derating curve [27] for induction motors.

The main consequence of unbalance applied to synchronous machines is overheating in the windings. Synchronous machines are usually protected from excessive heating by local protection devices, whose settings are specified dependent on the protection device and the machine.

Transformers are affected as unbalance generates negative sequence currents which subsequently cause extra heating and losses through the transformer. Transformers can be additionally affected by unbalance if the transformer has a delta (Δ) winding. A delta

winding traps a circulating zero sequence current and causes extra heating and losses as the current circulates in the transformer windings. If the amount of heating is severe enough, the transformer may need to be derated.

Other loads such as electronic power convertors are also affected by unbalance. The presence of unbalance at the terminals of an electronic power convertor can create harmonic distortion. Network components such as cables, lines and transformers are also affected by unbalance as the flow of worthless negative sequence current generates extra losses and heating on those devices. This limits their capacity to carry useful positive sequence currents.

1.5 Regulatory Limits and Standards for Voltage Sags and Unbalance

European standard EN 50160 [10] specifies that the number of voltage sags in LV[6] and MV networks should be less than 1000 per year. Some countries, notably Norway, impose tighter limits on voltage sags. The Norwegian regulator specifies that the maximum number of voltage sags which can occur in MV and LV networks must not exceed 24 in any 24 hour period [28]. In HV networks, the maximum number of voltage sags cannot exceed 12 in any 24 hour period [28]. The requirements of regulators such as Norway, Hungary and France already enforce more stringent standards than those presented in EN501060 [1].

EN 50160 [10] define that the voltage unbalance factor should be less than 2% for LV and MV networks, and less than 1% for HV network, measured in 10-minite period, with an instantaneous maximum of 4%. IEC 61000-2-2 [29] and IEC 61000-2-12 [30] define that voltage unbalance can be up to 2% in LV and MV networks. The IEC does not define compatibility levels for HV and EHV networks. The Norwegian regulator again implements a stricter standard of unbalance regulation than EN50160, by stipulating that the level of unbalance must not exceed 2% for any 10 min period in LV, MV or HV networks [28], thus removing the 95th percentile requirement.

1.6 Addressing Power Quality Issues

1.6.1 Mitigating Voltage Sags

Voltage sags are most often caused by faults in the network. Their effects can be reduced at the network level by reducing the number of faults in the system, reducing

[6] In this thesis LV is defined phase to phase nominal RMS voltage that does not exceed 1kV, MV concerns phase to phase nominal RMS voltages between 1kV and 35kV and HV concerns voltages above 35kV.

fault duration of faults and reducing the magnitude of sags by re-configuring the network to change its impedance.

Voltage sags can also be mitigated at a customer's site. Equipment immunity of AC drive and power electronic devices can be improved by adding capacitance. The immunity of DC drives to voltage sags can be improved by improving the algorithms which control those drives. Power electronic FACTS devices can also be installed at a customer site to minimise the impact of voltage sags during a fault [8].

1.6.2 Removing Voltage Unbalance

Voltage unbalance is ultimately removed by re-balancing loads or removing asymmetric network conductors [31]. Re-balancing loads can be achieved by manually or automatically reconfiguring the network. Voltage regulators and power electronic devices such as FACTS devices can also be installed to help rebalance three phase circuits.

Transposing overhead cables can help to reduce the effects of unbalance, as un-transposed transmission lines can be a cause of unbalance in a network [31]. Better tower design can also help to limit the effects of electromagnetic induction between conductors [32].

1.7 Power Quality Monitoring

Power quality monitors were historically installed close to customers experiencing power quality problems to troubleshoot specific issues relevant to important customers [33]. The devices were often installed on a temporary basis and removed once the problem had been identified and rectified.

DNOs are placing increasing emphasis on power quality data-collection to obtain greater visibility of power quality events [1] beyond monitoring a limited group of large important customers. Through network wide monitoring, the level of power quality within a network can be quantified in terms of both events and costs and important current problems can be identified. Monitoring allows a DNO to perform reliability benchmarking, monitor power quality contracts and plan predictive maintenance [33]. A DNO's monitoring investment decision is also driven by other factors such as new initiatives like the Smart Grid, changes in the regulatory environment, concerns about customer retention and new competition within the utility sector. The Electric Power Research Institute (EPRI) lists forecasting and short circuit analysis as the two main

reasons for monitoring alongside permanent power quality monitoring [33] for monitoring in future power networks.

The following list provides a summary of the reasons why power quality monitoring devices are being installed in distribution networks [34]:

- Identifying power quality problems, and troubleshooting issues
- Obtaining information on network performance for DNOs and customers
- Speeding up the resolution of new issues by matching similar issues against a history of past events
- Optimising maintenance schedules by guiding operatives to problem areas through localisation of issues
- Helping to guide pro-active maintenance
- Analysing the cause of black outs and other interruptions
- Helping to tweak protection settings (for example, in the case of long voltage sags)
- Monitoring power quality contracts
- Ensuring that regulatory standards are adhered to, and providing a convenient basis for reporting back regulatory information
- Providing equipment manufacturers with network based information (on the numbers and types of events) to formulate standards required for end-user equipment compatibility with the network

It is important to note that information obtained from power quality monitors needs to be intelligently processed by an operator or an algorithm before it can be used by a DNO to fulfil any of the tasks listed above. Different tasks require different levels of processing; for example, issue localisation requires measurements to be algorithmically processed into an event location before it can be used to optimise maintenance. Building an intelligent picture of the network is further complicated (and might need to be estimated) if monitoring is not available at every customer or busbar within the network.

The following section outlines a high level description of how metering information is collected and processed by the network operator.

1.7.1 Power Quality Monitoring Structure

Data for a power quality monitoring system is collected by monitoring instruments which typically measure voltage and current [35]. Power quality monitors are installed

at locations which are selected by the DNO or a partner organisation (such as EPRI for example [33]) and could include important substations or customers. The locations of these monitors are not generally selected based on any rigorous comparative analysis of the quality of different monitoring locations. Typically the DNO will either install monitors at most locations of a certain voltage level within the network (as is the case for [36]) or be guided by engineering judgement and locate monitors in sensible locations, such as the secondary side of 33kV/11kV primary substations.

A distribution management system (DMS) or central database server typically collates monitoring information before it is interpreted by a network operator. Information is sent to the DMS over a communications system, such as the telephone network, mobile network, or utility owned communications infrastructure. Power quality event information is processed by a DMS and published for operators in formats including summary reports, graphs and tables. The broad structure of a power quality monitoring system is described in Fig. 1.6, which is adapted from [33].

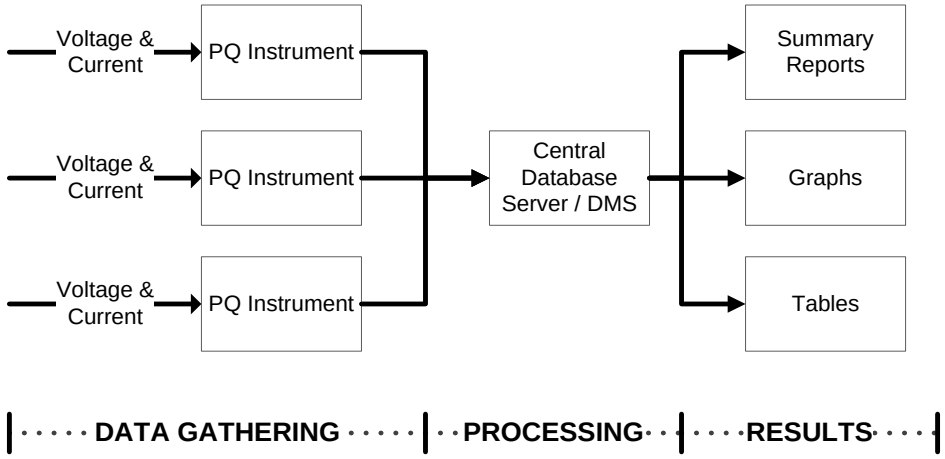


Fig. 1.6. The structure of a typical power quality monitoring system [33]**.**

1.7.2 Distribution Management Systems

Power quality monitoring systems are likely to end up augmenting or fully integrating with DMS systems which are already installed in most utility networks. DMSs originally started as an extension of the supervisory control and data acquisition (SCADA) system [37], and have advanced to cover a variety of applications. Commercial examples include GE's GENe [38] and ABB's Network Manager SCADA / DMS [39].

One of the applications covered by a DMS is fault detection, isolation and service restoration (FDIR) [40]. FDIR algorithms automatically restore service to the maximum number of customers using intelligent optimization restoration algorithms [41-43]. A

DMS can provide approximate information on fault location [40], and pre-empt network issues by highlighting overloaded cables and transformers. FDIR significantly improves the reliability of networks by reducing fault restoration time from several hours to a few minutes [37]. Although not strictly a power quality issue, voltage sags are often the end result of faults, and the results of voltage sag analysis can often be useful for FDIR, and conversely, FDIR can help with the localisation of voltage sag issues.

DMSs also contain functionality to assist with several offline processes which ultimately help DNOs plan network reinforcements and maintenance in a structured and informed manner [40]. Contingency and short circuit analysis evaluates the performance of the network given certain outage conditions [37]. DMSs can also help plan the optimal placement of voltage reinforcement devices by performing offline studies and optimizing the network wide voltage profile for maximal loss reduction. Understanding the power quality performance of circuits will also feed into these studies; voltage sags can be mitigated through the installation of FACTS devices and pro-actively re-balancing circuits with high unbalance will reduce losses.

The current body of research around DMSs suggests that future commercial DMSs will include further functionality to reduce the operational expenditure of network operators and advanced techniques to defer long term infrastructure investment and operate the network closer to stability boundaries. Power quality analysis tools are likely to be added to future commercial DMSs as these types of issues become increasingly important.

For power quality monitoring of steady state issues such as unbalance and potentially harmonics, it is likely that increased monitoring of customers (through the deployment of an Advanced Metering Infrastructure (AMI) [44]) and DNO assets (through increased monitoring of in-accessible units such as pole top transformers) will enable significant advances in the accuracy of a distribution system state estimator [45, 46]. Load estimation and load modelling algorithms are likely to significantly enhance in accuracy, utilising accurate recordings from meters installed at customer premises and the wider network [37].

The overall reliability of the network could also be enhanced through a DMS's advanced FDIR techniques which cover optimization of complicated network topologies [37], potentially utilising advanced algorithmic techniques [47], such as artificial intelligence. Fault and voltage sag localization using short circuit analysis may be available through the DMS using impedance based fault location algorithms [48].

Fault location and classification techniques are already available in the commercially available power quality software PQView [49], and marketed as important in order to reduce the time to restoration after faults.

The increase in power quality network monitoring will allow DNOs to assess system wide power quality issues such as voltage sags [11], unbalance [11], harmonics and flicker, and estimate their resulting impacts on consumers [36]. The core function of the DMS will be to supply accurate historical data which will allow DNOs to analyse the effects of new loads and generation such as PHEVs and distributed generation.

The information architecture of a DMS is likely to incorporate greater use of visualisation incorporating DNO geographical information systems (GIS). It is also likely to integrate securely and more fully into other DNO systems such as outage management systems (OMS) (such as [50]) and meter data management (MDM) systems (for example, [51] or [52]).

1.7.3 Global Pervasiveness of Monitoring Devices

The amount of power quality monitoring conducted by a DNO varies internationally, and is largely dependent on historical interest in power quality monitoring at the national and regulatory level. The number of monitoring devices installed in distribution networks is likely to increase significantly in the future, as new devices are installed to cater for new initiatives such as Smart Grid [53] and tighter power quality regulations come into force along with increased interest in commercial power quality contracts.

In the United States (US), the Electric Power Research Institute (EPRI) has maintained a significant interest in power quality monitoring over the last 15 years. Between 1993 and 1995 the EPRI power quality monitoring project installed 277 monitors in distribution systems at sites between 4.16kV and 34.5 kV. The monitor locations were chosen to represent a wide geographic sample of the US's rural, suburban and urban networks [54]. One-third of the monitors were located at substations just down line from the feeder circuit breaker, while the remaining monitors were randomly placed along three-phase sections of the feeder primary. Over 5961 hours of voltage and current data was collected. The objectives of the study were to understand system performance for power quality issues, characterize specific problems and check contractual arrangements within high quality power zones.

More recent power quality monitoring deployments include the French DNO creating the MAGIQ national power quality database [55] and Enel's in Italy [35, 36].

The French power quality monitoring survey started in 2002 and will complete in 2012 [55]. The French DNO plans to install one power quality monitoring device per transformer in all HV/MV[7] substations as well as a selection of power plants, important industrial consumers, renewable installations and some commercial locations [56]. The instrument transformer attached to each monitoring device was capable of monitoring to an accuracy class of 0.1 (0.1% error at rated current). All of the power quality information is collated in a national database. The information collected from the monitoring deployment includes RMS voltage and currents, harmonics, flicker, active, reactive power, frequency and unbalance. The objective of the study was to verify adherence to the EN50160 power quality standards and to obtain a "macroscopic view of the MV" [55] network.

The Italian power quality survey started at the end of 2005 with monitors being installed on the MV busbars of 360 HV/MV substations and 89 MV/LV substations [35]. This covers 11% of the MV busbars in ENEL's network, which is still some way short of full network monitoring at all MV busbars. The monitor locations were chosen to give a good geographical dispersion of rural and urban locations across the whole of Italy's distribution network. A full set of power quality performance characteristics were recorded including supply voltage variations, voltage sags and swells, short voltage interruptions, voltage harmonics, flicker, voltage unbalance and rapid voltage changes.

In the UK, power quality monitors have been installed at some 33kV/11kV substations where specific power quality issues have been recorded. For example, Western Power Distribution[8] (WPD, a UK DNO) installed 6 power quality monitors at a single geographic location to monitor power voltage unbalance. The monitoring locations were selected based on engineering judgement. The monitors record 3 phase voltage and current waveforms and RMS voltages on all three phases on both the primary and secondary side of the transformer using the Sub.net monitoring device [57]. In this specific example, the power quality monitors cover just over 50% of the 33kV/11kV substations in a 10 substation section of network. The overall density of monitors in the wider WPD distribution network is much less than 50% due to the limited number of power quality monitoring deployments undertaken thus far. This style of monitoring will give good visibility at 33kV, but only covers the start 11kV substations at the beginning of 11kV feeders.

[7] In this thesis LV means phase to phase nominal RMS voltage that does not exceed 1kV, MV concerns phase to phase nominal RMS voltages between 1kV and 35kV and HV concerns voltages above 35kV.
[8] Information obtained through conversations with Western Power Distribution engineers.

In another UK example, SSE[9] has installed power quality monitors which monitor the secondary side of transformers across 30% of the 33kV to 11kV substations in the network. The secondary side of a 33kV to 11kV transformer may feed between 3 and 10 11kV circuits. An added advantage of installing monitors in these substations is that they can be used to perform additional activities such as measuring circuit breaker performance. Scottish Power's future plans are to install a further 700 power quality meters across their network. The installation of power quality monitors is an ongoing process that is currently taking place across many UK distribution networks.

The objectives of all of the power quality monitoring deployments discussed in this section can be generalised and broadly broken down into the following areas:

Performance

- Improving MV distribution network power quality performance
- Characterisation of specific problems
- Troubleshooting specific issues

Research

- Testing hypotheses; for example, the correlation of power quality events with network structure, correlation of power quality events with short circuit power

Contractual

- Testing the feasibility of introducing contractual arrangements for power quality

Regulatory

- Publishing power quality performance of the MV network
- Testing the feasibility of introducing new regulatory limits for power quality
- Adjusting regulatory limits for power quality

1.7.4 Monitoring Summary

Some important conclusions can be drawn from this review of international power quality monitoring deployments which are pertinent for the literature study which follows.

The first point to note is that the numbers of monitors being installed within distribution networks is much less than would be required to make the network fully observable [58]. To achieve observability of all three phase currents and voltages within a radial

[9] Information obtained through conversations with SSE engineers.

fully loaded MV network, every other busbar would have to be monitored. With the exception of some pockets of the UK network, this is not the case for power quality monitoring. Therefore, it is likely that some form of interpolation and / or power quality estimation would be required to obtain the power quality performance of non-monitored busbars. It will also be important to be able to quantify the accuracy of estimation and monitoring deployments.

Beyond data collection, there will be a requirement for the information from power quality monitors to be synthesised to yield results which align to the DNO's objectives for power quality monitoring. It will therefore be important that the DNO has tools available to interpret the information from power quality monitors beyond simple statistics such as voltage, current or the number of interruptions at monitored busbars. Estimation techniques are likely to be study specific and focus on the DNO's objective for the study; for example, fault localisation could be used to improve network maintenance performance.

It is important to note that the locations where monitors have been installed were predominantly selected using engineering judgement. The quality of these monitoring locations may depend on several factors including the power quality issue being monitored (for example, harmonics, sags, unbalance etc.), their location, the accuracy of the monitors, and the objective of the study. As more monitors are placed in the network, it will be important to understand what constitutes a good location for a power quality monitor and whether the location of power quality monitors could be optimised.

The large numbers of power quality monitors being installed now and in the future will also create a significant volume of power quality information. The data will need to be processed and stored before it can be interrogated. There will therefore be a requirement for tools such as data mining, statistics and visualisation tools to help extract knowledge from large datasets. It will also be important to be able to understand the accuracy of research based on power quality datasets, by quantifying the error in processed power quality information.

These conclusions can be summarised into a series of research questions:

- How can measurements from power quality monitors be synthesised to yield a holistic picture of the network for voltage sags and unbalance?
- What algorithms are required to process monitoring from power quality meters into information that aligns with the DNO's monitoring objectives?

- What constitutes a good location for a power quality monitor, and can advanced optimisation techniques yield better results than engineering judgement?
- Are advanced computational techniques (such as statistics and artificial intelligence) required to synthesise data from power quality monitors and draw reliable conclusions?
- How should power quality issues be incorporated and visualised through the DMS to allow operators to interpret the topological variation in power quality?
- What are the quantifiable economic benefits of power quality monitoring?

1.8 Power Quality Performance Indices

It is common practice for utilities to process the information obtained from power quality monitoring into performance information relating to their network. The sources of power quality issues can be considered as *weak areas* and the customers affected can be regarded as the *worst served customers*. The goal of a power quality performance index is to help reveal where the worst served customers or the weakest areas are located.

1.8.1 Worst Served Customers

Identifying the worst served customers for unbalance and voltage sags should summarise the resultant impact of both sags and unbalance at a particular location in the network. This is synonymous with understanding the power quality performance of a specific bus in a network. The most common way of identifying the worst served customers in the network is to record power quality indices for voltage sags and unbalance independently.

The number of voltage sags at a busbar, for example, could be represented by the System Average RMS (Variation) Frequency Index (SARFI) index [11], or a Generalized Sag Table (GST) [59]. The SARFI index was developed as part of the EPRI power quality monitoring survey undertaken in the mid 1990s [33]. The SARFI index defines the historical average number of voltage sags of a specific magnitude at a customer's busbar. The SARFI index is represented with a suffix: for example, $SARFI_{50}$ represents the number of voltage sags recorded where the minimum voltage was less than 50%. The GST [59] exploits the similarity in voltage magnitude of two of the three phases during a voltage sag. The GST divides the magnitude and duration of voltage sags into discrete regions and represents them in a 2-dimensional table showing two

similar phases against the other phase. A greater level of detail of voltage sag performance is available through the GST compared with the SARFI index.

Voltage unbalance is measured as the ratio of the negative to the positive sequence voltages (in %) and called the *voltage unbalance factor*. The voltage unbalance factor is calculated using RMS voltages averaged over 10 minute intervals [10]. DNOs are particularly interested in recording the amount of time that the voltage unbalance factor exceeds a regulatory limit, such as 1% for transmission networks and 2% for distributions networks (in the UK) [60]. 95^{th} percentile readings are also important, as EN50160 stipulates that LV and MV networks must maintain voltage unbalance factors of less than 2% (or 3% in some locations) for 95% of the week. Other useful unbalance metrics include the percentage of nominal losses caused by unbalance [1], and the level of machine derating [27].

A common consequence of both voltage sags and / or unbalance is an equipment trip. Unbalance trips are localized to three phase machinery and defined by local protection settings. Voltage sags affect single phase and three phase loads, and their impacts can spread throughout the network. The resultant impact of both these power quality issues can be summarised by estimating the expected number of customer interruptions (CIs) at a busbar. If information about the duration of the sag or unbalance is available, the total customer minutes lost (CML) can also be found. CIs and CML can also be segmented by customer type, to identify which types of customers are being affected most severely.

The resultant impact of a trip is that electrical energy is not supplied to an area of network for a period of time. The total amount of electrical energy not supplied (EENS) is the total CML multiplied by the size of the load. Using additional knowledge of the types of customers attached to a busbar and therefore the average cost of an interruption for a given customer type, EENS, CML and trips can be combined to obtain the total economic cost of the interruption [2].

Identifying the worst served customers within the network is important from a regulatory and planning perspective. Regulators are interested in power quality performance metrics to ensure customers are being delivered a high quality service. Network planners are keen to understand potential power quality issues before new loads are connected to the network.

1.8.2 Weakest Areas of the Network

The most basic way of assessing the weakest areas of the network for voltage sags and unbalance is to identify the areas within the network with the highest fault rate (or other sag causing disturbances) and the areas with the most unbalanced loads. This analysis can be improved by considering the resultant impact of the voltage sags or unbalance in terms of customer interruptions (CIs), customer minutes lost (CML) or economic costs to either the DNO or the customer. By assigning the impact of the power quality issue to its source, the relative strength of the network can be assessed.

Identifying the weak areas in a network is most relevant to DNOs. It is important for DNOs to be able to identify weak areas so they can plan their re-enforcement strategy. Weak area analyses could also be used to help identify the root cause of specific problems impacting on one or a number of customers in a specific area of a network.

1.9 Overview of Past Research

The process of synthesising power quality monitoring information into knowledge on voltage sag and unbalance performance can be considered as an estimation process. For voltage sags the process of estimating the state of a power system is known as *voltage sag (profile) estimation* [25, 61]. The level of unbalance is defined as the ratio of two voltages, and thus can be determined using *three phase state estimation*.

1.9.1 State Estimation

The traditional way of determining the operating state of a power system is to use single phase *state estimation*. The operating conditions of a power system can be completely determined if voltage magnitude and angle of every busbar is available and if a suitably accurate model for the network is known [62, 63]. If all of the voltages and angles are known at every busbar in the network, the network is said to be observable, as demonstrated in [62] and [64].

The process of state estimation in power systems can be conducted using single phase [62] or three phase models [45, 46, 65] of the power system. With single phase state estimation, an assumption is made that the network operation is balanced and symmetric. This assumption is clearly not valid for unbalance, where the focus of the study is to estimate the imbalance between all three phases. Three phase state estimation enables the voltages and currents in all three phases to be known throughout the network. For voltage sag estimation studies, the state of the network prior to a fault is

often assumed to be balanced, and thus single phase state estimation models should suffice.

A notable example of three phase state estimation applied to the specific task of estimating the level of unbalance in a power network is [65]. In [65], the authors considered sources of unbalance as either unbalanced loads and unbalanced network topologies and proposed a three phase state estimator able to estimate three phase voltages and currents. [65] uses the weighted least squares optimization procedure. The method of least squares is the same optimization procedure used for most single phase state estimation implementations, including [63] and [62]. The main difference with the three phase approach proposed in [65] and other single phase approaches, is that the Jacobian, state and measurement matrices (see [62]) must cover all three phases, and thus include two additional dimensions.

The outputs from a three phase state estimator are directly applicable to unbalance studies as they can be directly used to derive the voltage unbalance factor. The output of a state estimator is also relevant to voltage sag state estimation algorithms such as [25] and as they require a pre-fault voltage estimate of the system to accurately estimate the post-fault system state. Pre-fault voltages are discussed as an additional cause of uncertainty for the proposed voltage sag estimation algorithm presented in [25], but no method is proposed to help alleviate this problem.

1.9.2 Distribution System State Estimation

The core assumption in traditional state estimation is that the number of monitors available within the network is sufficient to make the network observable [58, 62, 64]. The discussion presented earlier in this chapter demonstrates that the number of power quality monitors deployed in most modern power networks is much less than required to make the network fully observable. This problem has been overcome in past research by using *Distribution System State Estimation* (DSSE) [46].

Using DSSE, missing data at non-monitored busbars can be filled in using *pseudo-measurements* [45, 46, 66] enabling the state of an unobservable system to be estimated. The number of measurements available to estimate the state of the system can then be less than the number of measurements required to fully observe [58] the network. DSSE overcomes observability issues by making use of estimated measurements [46, 67-69].

Like traditional state estimation, DSSE is usually performed under the assumption that the network is balanced and can be represented as a single phase equivalent [67, 70].

[45] and [46] are notable exceptions, where the authors estimate the three phase state of a system using a DSSE formulation.

DSSE is very similar to the process of performing probabilistic load flow. Probabilistic load flow studies are focussed on establishing long term variation in network parameters whilst DSSE aims to establish a system's current state. Both techniques use Newton's method and both aim to estimate the statistical variation of parameters based on uncertainty. Notable works on this topic include [71], where the authors define the structure for three phase load flow and [72] where the authors estimate the variation in network parameters when exposed to uncertain wind generation. [72] is particularly relevant to unbalance studies as the focus of this study was to establish the maximum and 95^{th} percentiles of the unbalance factor.

An important aspect of three-phase DSSE and three phase probabilistic load flow studies for unbalance studies is the correlation between measurement errors. In [46] the authors touched upon the correlative nature of three phase DSSE measurement errors. Correlation in multi-phase networks is covered in more detail in [71-73], where the later reference describes how probabilistic three phase load can be used to estimate the probabilistic distribution of voltage unbalance. Correlation of measurement errors could be incorporated into a three-phase DSSE formulation using a generalised least squares (GLS) approach [74]. References in this area are limited, and the current body of research lacks a rigorous method to incorporate three phase correlated measurement errors into a real-time DSSE.

1.9.2.1 Locating Sources of Unbalance

Locating all of the multiple interacting emission sources is not a straightforward task [75], but it can be achieved using techniques such as [76] which decomposes the level of unbalance made by asymmetrical lines and loads in a section of the Australian network. The authors of [76] identify the relative contributions of lines and loads to each of the unbalances in the network. However, there has been limited research into real-time identification of unbalance sources and their associated effects in distribution networks. There is therefore a gap in the current body of research for a real-time state estimator which can statistically identify sources of unbalance in the network.

1.9.3 Voltage Sag Estimation

The sag performance of a monitored busbar can be obtained by using historical data recorded directly at the busbar. If there is no monitor present, the sag performance of a

busbar must be estimated. The process of estimating the sag performance of a network is known as voltage sag (profile) estimation (VSPE), and this can be accomplished using algorithms such as those described in [17, 25, 61]. These algorithms aim to re-construct the voltage profile at all buses in the network during a voltage sag using a set of monitors much smaller than the number of buses in the network.

The method of fault positions is the most common method used to determine the theoretical sag performance of a network. The fault position method works by defining a series of fault positions across the network, and simulating single line to ground, double line to ground, line to line and three phase faults at each position with an associated probability, as demonstrated in [18, 77, 78] and [79].

Examples of past work which use statistical reliability data to estimate the sag performance of a network include [18, 77, 78] and [79]. The reliability statistics could be obtained from multiple sources, including data recorded in a small portion of the target network, data compiled from power quality surveys (in the same network or other networks) or data compiled from reliability studies and models. Using reliability information is fast (as relatively small amounts of monitoring data are required to study a network), and it also allows the building of robust statistical performance estimates which can be used for network performance and economic analysis (as in [77]).

Reliability information provides an estimate for the expectation of failure or a distribution for the expected failure rate based on historical information. Although component failure rates may tend towards their expected failure rate, the failure rate in a specific section of a network will vary dependent of several conditional factors, including proximity to trees, weather patterns, ground water content (for cables) and cable condition, which may or may not be included in the reliability studies used as the basis of the fault studies.

In contrast, the approach taken by [25] for voltage sags and other sag (fault) location techniques such as [80, 81] and [82] uses data recorded at monitoring sites across the network and does not require reliability statistics to estimate the sag performance of a network. The approach presented in [25] implicitly calculates the reliability of the network through monitored data. The main drawback of performing voltage sag performance estimation using real monitoring information is that the monitors need to be installed for a number of years before they can accurately estimate the sag performance of the network. Reference [11] stipulated that at 1 fault per week, 7 years

of fault information would be required to obtained voltage sag performance information accurate to 10%.

An important work which proposes a solution to the problem of estimating the voltage sag performance of a network using a limited number of monitors is presented in [25]. A limitation of [25]'s technique is that it doesn't quantify how measurement or modelling errors affect the accuracy of the estimation procedure. This is an important consideration for DNOs who will be interpreting the results and drawing conclusions from limited monitoring.

1.9.3.1 Fault Detection and Classification (for Voltage Sag Estimation)

The voltage sag estimation technique proposed in [25], and other fault localisation techniques such as [81], assume that the type of the fault (e.g. single line to ground, line to line etc.) is known before the algorithm is applied. Therefore, there is a requirement for the fault to be detected and classified before voltage sag estimation can be carried out.

Fault classification techniques can be split into two groups: techniques which try to classify the outage-cause of a fault, and those which try to identify its technical characteristics. Fault classification techniques such as [83-85] and [86] are examples of techniques which try to identify the causes of faults in power networks. For example, [86] uses a hybrid artificial immune based neural network to estimate the number of faults caused by animals. The artificial immune recognition system (AIRS) [87] algorithm is used in [84] to classify the causes of faults as caused by either lightning, animals or trees.

In contrast to outage-cause identification, techniques such as [88] attempt to detect and subsequently classify the type of fault (as line to line, single line to ground, double line to ground or three phase). It is shown in [88] that fault classification can be accomplished with nearly 100% accuracy using a random forest algorithm [89] and high frequency measurements taken from one end of a transmission line.

Artificial intelligence (AI) based classification algorithms are ideally suited to fault classification as they are capable of generalising the relationship between a large number of features and a set of discrete classes. AI based classification techniques are common to many fault detection and classification algorithms (such as [86, 88] and [84]). Notable examples of artificial immune based algorithms applied to fault classification and detection are [84] and [86].

It is worth considering that artificial immune system inspired algorithms may be of particular relevance for fault classification and detection in the voltage sag estimation problem, and potentially optimal monitor placement. Artificial immune systems (AIS) are inspired by understanding the concepts involved in the adaptive immune system[90].

The immune system exhibits many important features that are attractive to computational scientists:

- Ability to defend the body from unknown diseases
- Highly parallel and distributed in operation
- Adaptive
- Capable of pattern recognition and classification
- Ability to recognise dangerous invaders and uncreative body cells

These ideas have inspired a number of AIS algorithms, including optimization algorithms , classification algorithms [87] and clustering algorithms [91].

Applying AIS algorithms to tackle power systems problems is largely unexplored in the current body of AIS and power systems research. There are some notable examples such as [84], [92] and [93]. However, the total number of papers applied to the field is still relatively small.

AIS algorithms have been successfully applied to areas such as anomaly detection [85, 94, 95], pattern recognition [96] and data mining [97, 98]. AIS algorithms have been used to solve data clustering problems in the past [91, 99], and [91] demonstrated that for some data sets an AIS algorithm can outperform other clustering algorithms. Similar niches could be exploited and the algorithms successfully applied to power systems research. It is important to stress that the best applications of AIS and other AI techniques are where there was a suitable requirement for their application.

There are several areas where the current body of research on fault detection and classification could be extended for voltage sag profile estimation. Most studies do not consider fault classification and detection using a limited set of monitors, or the impact of erroneous measurement information. It is important to understand how misclassifications will occur with a limited set of monitors and ensure that detection and classification algorithms are robust to monitor error. Although [88] suggests that random forest algorithms can produce accurate results, it would be useful to compare the quality of a variety of techniques to establish the best methods for voltage sag

profile estimation. It would be interesting to compare the performance of random forest against the AIRS algorithm [87] (used in [84]), as this is a popular immune inspired technique [96].

1.9.3.2 Fault Location (for Voltage Sag Estimation)

The task of fault detection is closely related to voltage sag estimation. Voltage sags are most often caused by faults [2] and locating the fault which caused the sag is an intermediary step which is required before the sag performance of the whole network can be established.

The aim of a fault location algorithm is to estimate the source of a fault through measurements taken throughout a power system network. Fault location is often a non-trivial process: algorithms may work well in simulations, but care must be taken to ensure practical implementations are robust to measurement error and unbiased when integrating information from a broad variety of monitoring devices.

There are two main approaches to fault location which have been widely explored in power systems research. The first approach estimates a fault location by using voltage and current measurements taken at the power system's fundamental frequency [81]. The second approach uses high frequency transient components of the voltage and current waveform recorded during a fault [100]. These two approaches can be further sub-divided into single-ended and double-ended methods. Single-ended methods estimate the location of the fault using information from one monitoring device usually placed at the start of a feeder or line [81, 101]. Double-ended methods use measurements from two locations, commonly between two ends of a feeder or line [81].

The main disadvantage of using a high frequency approach is the requirement for high sample rate monitors (in the order of several MHz [102]) to trace the voltage and current waveform. This restricts the information input to a fault locator to a limited number of high frequency monitoring devices. In contrast, steady state methods (such as the impedance based approach proposed in [81]) can use information from lower sample rate monitors. The main drawback of impedance based methods is that they can produce several estimates for a fault location when multiple impedance paths cause the same voltage drop [82, 102]. This is a common problem in distribution networks where the number of monitors is often much less than required to make the network fully observable [62]. This problem was solved for symmetrical faults on feeders with single and two-phase laterals in [82], but no solution was presented for asymmetrical faults. To

a certain extent, this issue can be overcome by taking the intersection of fault location solutions [48]. However, there are no details in [48] on how to deal with solutions which don't overlap because measurement errors or network parameter sensitivities cause the solutions from different monitors to separate.

The main challenge of developing a fault location algorithm for voltage sag profile estimation is how to eliminate errors caused by lack of knowledge of the fault's characteristics. Voltage sags are stochastic in nature as they are influenced by the a variety of random factors including, fault type, position, fault impedance and pre-fault voltage [17]. The method proposed in [25] assumes that fault type, pre-fault voltage, fault type and fault impedance are all known. Other fault location techniques (for example [81]) either eliminate the variables entirely or provide assumptions to cover other variables. There is a lack of guidance in the body of existing research on how these assumptions affect the results of voltage sag profile estimation.

The process by which [81] deals with uncertainty in fault characteristics can be explained by considering each type of fault. For three phase and line to line faults, the fault impedance must be assumed to be entirely real, and pre-fault voltages must be estimated. The assumption of real fault impedance is shown in [103] to be a valid approximation. Reference [81] and [79] approximates pre-fault voltages at the fault location to 1 per unit. This leads to small errors (typically around 2% for the specific test network in [81]) in the estimated fault location [81]. For double line to ground faults a solution to the fault location equations (of [81]) can be found through a single assumption of an entirely real fault impedance. Using the fault location equations shown in [81] for single line to ground faults, no assumptions on pre-fault loading or fault impedance are required to carry out voltage sag profile estimation. In other words, the equations in [81] for single line to ground faults are insensitive to fault impedance and pre-fault loading.

The algorithmic approach to fault location undertaken by a DNO ultimately depends on the number of and quality of the monitoring devices installed. [104] describes how distribution network substations commonly contain a number of monitoring devices including remote terminal units (RTUs) and protective relays which feed information such as voltages, real and reactive power flows, feeder currents and frequency towards a centralized SCADA measurement system. The continued interest in smart grids [104] is encouraging further deployment of monitoring devices such as power quality meters, digital protection relays, smart meters, advanced automation devices, condition

monitors and digital fault recorders. DNOs are therefore interested in synthesising as much information as possible for fault location. This is not typically the focus of fault location algorithms; instead the algorithms tend to focus on locating faults using either one or two monitors.

One of the most important references on fault location using a limited monitoring number of variably accurate monitors is [105]. In the paper, the authors develop a method which is capable of locating faults using measurements taken from any two locations in the power network. They also propose an optimal estimation procedure which is based on similar principles to the method of least squares and state estimation. The method is also able to identify the presence of bad monitoring data, using the chi-squared statistical test.

Reference [105] could be enhanced if the approach was extended so that operators can use measurements from an arbitrarily accurate selection of installed monitoring devices, including relays, power quality meters, disturbance recorders, phasor measurement (PMUs) devices and revenue meters. This would require the method to be altered such that it could cope with more than two monitoring devices. This may yield a higher degree of accuracy and could be more robust than simply calculating the fault location using one or two monitoring devices (as presented in [81]). This extension may also help to overcome issues caused by multiple impedance paths, by reducing the number of fault locations to those that are the most feasible.

Reference [105] could also be extended by formulating the fault location estimate as a statistical distribution whilst taking into account the errors of monitoring devices within the network. This would give network operators a distribution for the area which was likely to contain their fault. This is not possible using the analytical equations (such as those presented in [81] or [105] as these algorithms yield a single deterministic solution.

1.9.4 Power Quality Impact Assessment

The ultimate objective of unbalance or voltage sag estimation is often to estimate the effects of power quality issues on customers.

The effect of a voltage sag can be defined technically in terms of a number of trips [20], or economically in terms a monetary impact on the customer [4]. The technical impact of a voltage sag is assessed by comparing the voltage sag disturbance (using a voltage sag profile estimation such as [25]) against equipment or process immunity (as described in Fig. 1.1). The local equipment immunity level can be defined

deterministically (for example in terms of trip / no trip) or probabilistically using statistical distributions [18, 20, 106]. Assessing the economic impact of voltage sags on customers can be accomplished by combining information on the technical impact of the voltage sag with the associated economic costs of an interruption for that specific customer [4].

One of the most notable developments in equipment immunity modelling is described in [18, 106], where the authors generalised the equipment immunity curve as a probabilistic distribution. The equipment immunity for various types of equipment including programmable logic controllers (PLCs), adjustable speed drives (ASDs), personal computers (PCs) and contactors were all defined as generalised distributions. The authors also described how these generalised probability distributions can be combined to develop a model for process sensitivity. The analysis in [18, 106] was ultimately used to rank different the different network topologies in terms voltage sag performance. This technical approach was extended into the economics of voltage sag events in [4, 18, 77] and [1] to calculate the economic impact of voltage sags to individual customers and complex industrial processes.

An important aspect missing from the current body of research on the impacts of voltage sags is the combination of statistical techniques such as those proposed in [18, 106] with measurement uncertainty. The probabilistic equipment immunity approach taken in [106] could be combined with measurement uncertainty and used to estimate confidence bounded trips and economic results.

There is no analogous concept of an immunity curve for unbalance. Three phase machines, power electronic convertors, three phase drives and single phase lighting loads are discussed in [21] and [1] as sensitive to unbalance. Reference [1] defined the economic impacts resulting from unbalance as a linear summation of incorporating costs caused by additional losses, overheating, pre-mature aging of assets, reactive power compensation and replacement of light sources.

The end result of high levels of unbalance applied to three phase machinery may ultimately result in a customer trip [21], in the same way as defined for voltage sags. However, whether or not a specific device trips will depend on the specific protection settings for the specific three phase machine. The compatibility standards IEC-61000-2-2 [29], IEC-61000-2-12 [30] and EN50160 [10] provide some guidance on the unbalance limits which manufacturers use to design their equipment [31].

It would be useful to gather experimental evidence and survey exact level of equipment immunity to unbalance across a variety of devices including ASDs, power electronic devices, three phase drives, and induction machines. This would define the immunity of equipment to unbalance in the similar way to voltage sag immunity. This would enable the effects of high levels of unbalance to be more easily quantified.

Further research could also be conducted on quantifying the impact of unbalance using statistical techniques. Economic analysis (such as [1]) or technical analysis (NEMA curves [27], for example) could be combined with measurement and equipment uncertainty to generate a statistical picture of the affect of unbalance on consumers.

1.9.5 Optimal Placement Techniques

Optimal monitor placement algorithms are used by DNOs to guide the deployment of monitors in their network. Techniques have been developed to optimally monitor power networks on a wide range of topics including for state estimation [107, 108], power system stability [109, 110] and power quality [17, 111].

Optimal placement algorithms for state estimation largely focus on placing phasor measurement units (PMUs) in transmission networks to obtain full system observability. This objective is usually accomplished using integer programming optimisation techniques [108, 112, 113]. PMUs are not typically installed in distribution networks because the cost of the device is usually too high to warrant their installation.

The methods used to place monitors in the network vary from integer programming [108, 112, 113] to heuristic search methods such as genetic algorithms [114] and immune inspired techniques [115-118]. Heuristic population based search techniques are advantageous because they reduce the time to achieve a near-optimal solution. This is useful when a searching through a large search space, dealing with a wide range of uncertain future scenarios [92], or achieving the best trade-off between competing objectives [114]. [92] exploited this aspect of an immune inspired based search technique to derive a series of sub-optimal solutions for dominance analysis and power system planning.

Most monitor placement techniques focus on placing monitors using a single stage implementation process. A notable example of a multi-stage approach which is [119], which uses mixed integer linear programming to place PMUs to maximise the observability of the system. Uncertainty is often present when considering monitoring deployments over long periods of time because the economic value of a chosen

monitoring solution is not precisely known over all time periods. Placing monitors whilst considering the uncertainty inherent in network topology and loading is an important aspect which is worthy of further research.

For power quality, optimal monitoring techniques have been developed for harmonic state estimation [111], distribution system state estimation [69] and also voltage sag estimation [17, 120]. The techniques developed for distribution system state estimation are most relevant to unbalance estimation.

The methods developed to optimally place monitors for voltage sag estimation in [17, 120] and fault location observability analysis in [48] used a branch and bound integer programming technique to place monitors able to detect events occurring anywhere in the network. The authors of [120] developed a concept known as the monitor reach area (MRA) which defines the area of the network which each power quality monitor covers. The optimal monitoring locations were used in [120] to estimate the voltage sag profile across the whole network. The objective of the optimisation method in [48] was to ensure that faults (or voltage sags) occurring anywhere in the network could be observed by all monitors.

There are several areas where the research presented for optimal monitor placement for voltage sags could be enhanced. Both [48] and [120] assume that the objective of monitor placement is to observe all faults within the network. It may also be useful to develop a method which can best observe faults with a limited set of monitoring, or take account of existing monitoring devices in the network. There is scope to change the objective function used by [120] and [48] to focus on accurately estimating the impacts that voltage sags have on customers, rather than just focussing on detection (through thresholds) [120] or localization [48]. A further enhancement not covered in the existing body of research could be to test the quality of the optimal monitoring solutions across a range of uncertain scenarios. This may require a heuristic search technique such as a genetic algorithm or artificial immune system.

Monitor placement optimisation for unbalance is similar to monitor placement for traditional state estimation. The key differences are that the monitor locations must be selected to ensure three phase (rather than single phase [62]) observability of the system. Once this modification has been made, the linear programming optimisation problem [121] can be formulated and optimised in the same way as single phase state estimation.

Unbalance monitor placement optimisation studies further differ from traditional state estimation techniques if the constraint of complete observability is relaxed. In this scenario, the optimal placement problem focuses on optimally placing monitors for distribution system state estimation. In distribution system state estimation, the relative difference in accuracy between pseudo-measurements and real monitoring devices becomes a critical factor in the resulting accuracy of the state estimate. This problem was formulated and solved for single phase systems in [69] and [122]. The focus of the two papers was subtly different; in [69] the objective of the optimisation was to reduce the overall error at every bus in the network, whereas the objective in [122] was to ensure that the network could be operated and controlled effectively.

The research on meter placement for DSSE could be extended by incorporating all three phases of the power system into the optimisation and also re-focussing the objective function to monitor the network for regulatory breaches at the 2% voltage unbalance factor level. It would also be interesting to research the possibility of placing monitors to optimally estimate the sources of unbalance, rather than estimating the effects.

1.10 Problem Statements

The review of past research frames the current lines of thought in both industry and academia. A series of problem statements can now be defined which describe the key problems which will be addressed in this thesis.

The problem statements are as follows:

- Distribution system state estimation models have been used in the past to create three phase and single phase models of distribution networks. Limited monitoring in future distribution networks require models to be developed which specifically estimate the three phase state of the network for both unbalance (and voltage sags) whilst taking into account the correlated nature of three phase loads.
- Voltage sag estimation algorithms have been developed to estimate sag performance at monitored and non-monitored busbars. There is a requirement for new sag profile estimation algorithms which can quantify the sensitivity of estimates to measurement (and model) error, limited measurements and algorithm choice.
- Existing fault classification algorithms use artificial intelligence methods to classify the cause and type of fault occurring in a network to a varying degree of

accuracy. It is important to understand if these techniques will work under the context of limited monitoring, where misclassifications occur and what are the best classification algorithms.

- Fault location is a core aspect of voltage sag estimation and there are many existing techniques which are able to perform this function by collating information from typically one or two error free measurement devices. Extracting knowledge from increased measurement information available in future networks will require techniques that can synthesise *all* information from an arbitrary collection of monitors whilst addressing measurement error uncertainty.
- Statistical methods have been proposed which generalise the immunity of equipment and processes to voltage sags as statistical distributions. To obtain information on the accuracy of any analysis, there is also a requirement to define both the disturbance level and the electromagnetic compatibility assessment as distributions which incorporate measurement error and all other uncertainties.
- Optimal monitor placement techniques have been proposed which can perform sag estimation and DSSE to estimate the magnitude of voltage sags and unbalance. As more monitors are added to power networks, there will be a further requirement for methods which are able to sequentially and optimally place additional monitors to best estimate the impacts of power quality events on customers and network operators.

1.11 Objective, Hypotheses & Aims of this Research

The main objective of this research is to present new techniques which are capable of identifying the weakest areas and the worst served customers for voltage sags and unbalance of a generic distribution network using a limited set of power quality monitors.

To address the problem statements, the following hypotheses can be asserted.

1. Simple measurements from a limited number of variable accuracy metering devices measuring properties such as voltage magnitude, reactive power and real power can be synthesised to yield an intelligent customer centric picture of the power network in terms of its performance for voltage sags, and unbalance.

2. Performance estimation for voltage sags and unbalance can be quantified to optimise the monitoring locations of any arbitrary number of monitors, perhaps much less than the number of nodes in a network.
3. Artificial immune system heuristic based techniques are required to monitor the state of the power network for voltage sags and unbalance and can help with monitor placement in the networks.

This research aims to address the issues that have not been satisfactorily resolved in past research and answer the problem statements, along with investigating the hypotheses.

The main aims of the research can be summarised as follows:

1. To present a methodology which is capable of classifying faults in distribution networks using a limited number of monitors.
2. To compare the performance of a variety of artificial intelligence fault classification techniques and identify where they miss-classify faults, and which methods work best.
3. To develop a method which is capable of localising faults within distribution networks using a statistical approach which accounts for monitor error and which can work with an arbitrary number of devices.
4. To develop a method which is able to estimate the statistical distribution of a voltage during a voltage sag using all information available from monitors within the network.
5. To develop a new criterion to assess the impact of voltage sags on customers based on the physical behaviour of equipment and the statistical distribution of the disturbance.
6. To demonstrate and experimentally validate that an artificial immune system (AIS) optimization methodology is an appropriate, high quality tool for the optimal monitor placement problem.
7. To assess the suitability of artificial immune systems (AIS) for power systems research.
8. To present a practical and robust methodology capable of optimally locating monitors to observe the effects of voltage sags on customers in a generic distribution network across a range of future uncertain loading scenarios.
9. To provide a framework for performing three phase state estimation in a network where measurement information is incomplete, and must be estimated.

10. To highlight how the results of DSSE can be used to statistically estimate the location, level and impacts of unbalance on a typical distribution network, and thus ultimately help fix unbalance related issues.
11. To demonstrate how the voltage sag and unbalance performance of a network can be combined and presented to identify the worst served and weakest areas of the network for these issues.

1.12 Major Contributions of This Research

The research presented in this thesis has contributed to several areas in the field of power quality monitoring, voltage sags and unbalance. Note that the references used in this section are pre-fixed with a "G" which have been published as part of this research. A full list of journal and conference papers can be found in Appendix G.

The culmination of this research is a comprehensive methodology which is able to simultaneously estimate the impacts of voltage sags and unbalance on customers within a network using a limited amount of monitoring information and realistic set of measurement inputs.

The main contributions of this thesis can be summarized as follows:

- The development of practical methodologies which are able to synthesize information from a number of arbitrarily accurate monitors placed anywhere in a power network and combine this information to statistically estimate the value of network parameters at non-monitored busbars [G2][G3][G8][G11].
- The proposal of a new set of consumer focused statistical frameworks which are able to estimate the impact of sags and unbalance on equipment whilst taking into account both measurement and model uncertainties [G2][G3][G8][G9].
- The design of a new method for improving the performance of both voltage sag monitoring using an immune inspired optimization algorithm and an objective function which focuses on enhancing a DNO's visibility of voltage sag issues [G2][G6][G7][G9].

The following sub-sections review specific contributions of this thesis.

A Method to Detect and Classify Voltage Sags

In contrast to existing techniques, the detection and classification method developed in this research is capable of classifying and detecting the occurrence and the type of fault in a distribution networks and shown to work under the context of limited monitoring

with high levels of accuracy. The method also pays attention to errors which might be present in measurement data.

Development of Probabilistic Fault Location & Voltage Sag Estimation

The probabilistic fault location and voltage sag profile estimation developed in this research and presented (to a certain extent in [G6] and significantly extended) in [G11] has several advantages which haven't been explored by existing techniques. It synthesises all information using statistical distributions. This allows the method to integrate an arbitrary set of monitors, deal effectively with errors in measurements and overcome sensitivity issues with traditional fault location equations. The developed approach can be easily augmented with new monitoring information, uses all the information from power quality monitors in the network and can identify faulty monitoring devices. This approach is different to existing techniques on voltage sag localization, as it focuses on estimating the voltage at non-monitored busbars, rather than just locating the source of a fault.

Development of a Statistical Impact Estimation Framework

A new framework for estimating the impact of a voltage sag in terms of a simple to interpret measure known as the *sag trip probability* which defines the probability that equipment will trip given a voltage sag is proposed in this research [G2][G9]. The general formulation of the *STP* is flexible and configurable and can be used with different equipment sensitivities, different process sensitivities and arbitrary sets of monitors. This measure consolidates existing research on the topic by incorporating the error inherent in the voltage sag profile estimation process (and hence measurement error).

A Four Step Process for Estimating the Impacts of Voltage Sags

A major contribution to the existing research on voltage sags is the development of a four step process which can estimate the impacts of voltage sags using limited monitoring. The process defines how a DNO could practically and effectively monitor their network for voltage sags. This involves methods presented in a variety of research related publications including [G3], [G6], [G7], [G9] and [G10].

A New Optimal Placement Method for Voltage Sag Monitoring

A new method which optimises the placement of monitors with the objective of estimating the likelihood of equipment trips during voltage sags rather than focusing on estimating the voltage sag profile itself is proposed in this research [G2]. The method is

different from existing optimisation techniques as it uses an immune inspired B-cell optimisation algorithm which is capable of developing a strong set of monitoring solutions which are effective over a range of uncertain future network loads and topologies.

A Comparison of the Performance of AIS Based Techniques

Artificial immune system based classification and optimisation methods were used in this research [G1][G5][G10][G2] and compared against a variety of other more traditional methods. It was shown that the artificial immune recognition system (AIRS) provided no benefit over other techniques [G1][G5], whilst immune inspired optimisation for optimal monitor placement may be worthy of future research [G2]. These two conclusions could be used to assist future research on artificial immune systems applied to power systems.

Development of a Method to Identify the Source, Location and Impacts of Unbalance

A new method which can estimate the source, location and impacts of unbalance was proposed in this research [G3][G8]. The research extends existing concepts on distribution system state estimation by building a three phase measurement model which incorporates correlated measurements, dealing with incomplete and erroneous information and also estimating the level, source and location of unbalance.

1.13 Thesis Overview

Chapter 1 provides the background and context to the research. It builds the case for the problems which will be solved in later chapters in the thesis. Particular attention is paid to the current state of power quality monitoring in power networks and a review of current research on identifying the worst served and weakest areas of the network for voltage sags and unbalance.

Chapter 2 describes the fundamental models and the simulation techniques that will be used as the foundation for analysis in later chapters. The chapter begins by discussing monitoring in power systems and then outlining the inputs which will be used for the techniques developed throughout the rest of this thesis. The focus then moves onto model and simulation tool development, with careful attention being applied to modelling assumptions. Chapter 2 is largely theoretical in content and most of the material discussed within this chapter can be found in textbooks on power systems, notably [63] and [11] [123].

Chapter 3 describes some of the intelligent processing techniques which are used as tools throughout this thesis. The following techniques are described in detail:

- Artificial Immune Recognition System (AIRS) for *classification* (immune inspired)
- Support Vector Machine (SVM) for *classification*
- k-Nearest Neighbours (kNN) for *classification*
- Random Forest for *classification*
- Unsupervised Artificial Immune Classifier (UAIC) for *clustering*
- B-Cell Algorithm (BCA) for *optimization* (immune inspired)
- Statistical analysis

Chapter 4 introduces the multi-step process used to build a holistic picture of voltage sags within distribution networks and then subsequently focuses on detection and classification of voltage sags. Chapter 4 is the first of three sequential chapters which all focus on a new overall method which is capable of estimating the impact of voltage sags on customers within the distribution network.

The chapter presents a comparison of four classification algorithms applied to classify faults within a 295 bus generic distribution network. The results are presented as a series of confusion matrices and also visualised on a 295 bus generic distribution network. To test the performance of the algorithms under limited monitoring, the number of monitors in the network is varied from 1 through to 12 monitors.

Chapter 5 focuses on the voltage sag localization and voltage sag profile estimation. In this chapter, a robust method for performing fault location and voltage sag profile estimation using multiple monitors is developed. The method uses statistical analysis to solve the fault location equations and find the most likely location for both the fault location and the best estimate for the voltage profile throughout the whole network during a fault. The method draws on and significantly advances existing research developed in [81].

Chapter 6 follows on from chapters 4 and 5 with a method which estimates the impact of voltage sags on customers within the network. A new criterion to assess the impact of voltage sags on customers based on the physical behaviour of equipment is developed which defines the probability of a trip from a voltage sag as the *sag trip probability* (*STP*). The *STP* is illustrated using simulations of single line to ground faults on the 295 bus generic distribution network.

Chapter 7 presents a methodology which can be used to optimally place power quality monitors into a distribution network to monitor the network for voltage sags. The chapter starts by explaining what makes a good monitoring set, and thus how an optimisation methodology should be configured. The chapter follows on by defining a new way of optimising the location of monitors based on equipment trip probabilities: the sag trip probability (*STP*), from Chapter 6. The proposed optimisation method is an artificial immune system (AIS) based B-Cell algorithm. This technique is tested on a 295 bus generic distribution network with load growth over 15 years and topological variation against a range of other optimisation techniques. The chapter concludes by reviewing the quality of the performance of the B-Cell algorithm compared with other techniques.

Chapter 8 develops a three phase DSSE model which can estimate the location and impact of unbalance within the network, without assuming the loading is balanced. The chapter presents a framework for performing 3 phase state estimation with incomplete measurements. The framework is shown to be capable of identifying the location, level and impacts of unbalance as statistical distributions within the power system network. The DSSE model developed deals with the correlative nature of three phase DSSE measurement errors which are critical to understand how a three phase phenomenon like unbalance spreads throughout the network.

Chapter 9 presents an overview of how the methods presented in this thesis can be synthesised into a illustrative picture of the power network which is capable of identifying the worst served and the weakest areas of the network. The chapter illustrates the techniques developed in the thesis using both a full and a limited set of monitors. The key results of the chapter are presented using a series of topological heat maps. The worst served customers are identified by assessing the percentage of the 24 hour period where unbalance exceeded 2%, the number of trips caused by unbalance and the number of trips caused by voltage sags. The weakest areas of the network are assessed by considering the sources of unbalance, sources of unbalance which caused the most customer interruptions, and the sources of voltage sags which caused the most interruptions. The chapter concludes by globally ranking the worst served and weakest areas of the network.

Chapter 10 outlines the major conclusions of this thesis and suggest areas for future research.

.

2 Power System Modelling

2.1 Introduction

A power system model is a mathematical representation of the physical operation of a power system. A good model will accurately represent reality, but also generalise and simplify the complex inner workings of a system. Once a model is constructed, simulations can be performed on the model to test the system's response to changes in input parameters. In the context of this thesis, the model inputs are measurements at monitoring devices in the power network and the outputs are unbalance and voltage sag performance results.

The goal of this thesis is to present a series of tools which can be used to identify the worst served customers for two power quality phenomena; voltage sags and unbalance. Model inputs, power system models and simulation tools form a tripartite foundation which can be used to validate if this thesis' proposed techniques are theoretically and practical applicable to power networks.

This chapter describes the model inputs and the fundamental models and techniques that will be used for analysis in later chapters. The chapter begins by discussing monitoring in power systems and defining the inputs which will be used for the techniques developed throughout the rest of this thesis. The focus then moves onto model development, with careful attention being applied to modelling assumptions, for both voltage sags and unbalance.

This chapter is largely theoretical in content and most of the material discussed within this chapter can be found in textbooks on power systems, notably [63], [11] and [123].

2.1.1 Power Quality Measurements

The two fundamental components required for power quality monitoring are voltages and currents [34]. Voltages are measured through *voltage transformers* (VTs) and currents are measured through *current transformers* (CTs), both collectively known as instrument transformers. Instrument transformers are often used for multiple applications including protection, automatic voltage control, automation, control, visualisation, and operator analysis.

Most incumbent remote monitoring takes place within distribution networks as part of the supervisory control and data acquisition (SCADA) network and is typically recorded for only one phase. More advanced power quality monitors are being installed, and these devices tend to be installed to monitor all three phases. Power quality monitors and SCADA devices send information back to a central control system for processing.

The following two sub-sections review the availability and accuracy of information from incumbent monitoring devices and newer power quality monitors.

2.1.1.1 Single Phase SCADA Monitoring

The SCADA system is a computer and measurement system which is present within most distribution networks. The information collected by the SCADA system is relevant to some aspects of power quality analysis including providing information on pre-disturbance system state, and providing information on the level and pattern of loading in the network.

Table 2.2 shows the measurements which are routinely collated by the SCADA system of a UK distribution network. The table shows that all feeders from 11kV upwards are monitored using single phase voltage and current measurements. Voltages are measured using VTs which are connected line to line, and current is measured using CTs which are typically connected to the yellow phase. The network's operation is assumed to be balanced. Information on power and reactive power is available on the secondary side of transformers with a secondary winding greater than 11kV. Although 11kV feeders are monitored, 11kV customers and 11kV to LV distribution transformers (and thus LV loads) are not typically monitored using anything more than a maximum demand CT.

Table 2.2 also highlights that more advanced monitoring is being installed into distribution networks across 33kV, 11kV and LV voltage levels. The Schneider PM710[124] is being used to monitor a number of LV customers and the EMS Sub.net [57] is being used for higher voltage levels.

The accuracy of SCADA measurements ultimately depends on the accuracy of the instrument transformer recording the measurement. IEC Standard [125] lists various classes for current transformer ranging from class 0.1 to class 3. The class of a CT translates into a percentage error for the measurement device.

The resolution of SCADA measurements stored centrally in the DNO's control room depends upon the digitisation capabilities at the point of measurement. Some UK SCADA networks have a resolution per byte of approximately 0.12kV across a voltage range of 0 to 12.5kV. This means each measurement will be resolved to within approximately ±0.5V (i.e. 120V / 2^8 = 0.47V).

2.1.1.2 Three Phase Power Quality Monitoring

The current international trend towards power quality monitoring device installation (which was discussed in the introduction) typically involves devices which monitor all three phases and a variety of power system information including (but not limited to) voltage, current, real and reactive power, harmonics, flicker and unbalance. The types of measurements available and the measurement error of a meter will vary depending on the quality of the meter installed.

Table 2.1 highlights the measurement accuracy of three power quality monitors from Siemens [126], Embedded Monitoring Systems (EMS) [57] and General Electric (GE) [127]. The meter from EMS is shown as this is known to be undergoing field trials in a section of the UK distribution network. The Siemens and GE meters are shown as popular comparisons.

TABLE 2.1. THE MEASUREMENT ACCURACY OF VARIOUS POWER QUALITY MONITORS.

Meter	**Symmetrical Voltage**		**Current**	**Voltage**	**Power Demand**		**GPS Synchronisation**
	Magnitude	**Phase**	**Magnitude**	**Magnitude**	**Real**	**Reactive**	
Siemens 9610 [126]	0.2% of full scale	0.2% of 2π (±0.004π at 50Hz)	0.1%	0.1%	±0.2%	±0.2%	±1ms ±0.1π at 50Hz
EMS Sub.net [57]	0.1% of full scale	0.1% of full scale	0.1% of full scale	0.1% of full scale	0.1% of full scale	0.1% of full scale	±1μs $\pm 1\times10^{-4}\pi$ at 50Hz
GE EPM9650 [127]	Not quoted	Not quoted	0.025% of full scale	0.01% of full scale	0.04% of full scale	0.04% of full scale	±1ms ±0.1π at 50Hz

TABLE 2.2. THE MONITORING AND MEASUREMENTS WHICH ARE RECORDED IN A TYPICAL UK DISTRIBUTION NETWORK

Point in the Network	**Pervasiveness of Measurements**	**Types of Measurements**	**Voltage Reference and Number**	**Current Measurements and Numbers**	**No. of Power Measurements**
Primary side of 132kV:XkV Transformer	100%	|S|, |V|, |I|, P* ,Q*	1 mostly line to line (some line to ground)	1 phase (typically yellow)	1 phase
Secondary side of XkV:33kV Transformer	100%	|S|, |V|', |I|, P* ,Q*	1mostly line to line (some line to ground)	1 phase (typically yellow)	1 phase
Secondary side of XkV:11kV Transformer	100%	|S|, |V|', |I|, P* ,Q*	1mostly line to line (some line to ground)	1 phase (typically yellow)	1 phase
Secondary side of XkV:6.6kV Transformer	100%	|S|, |V|', |I|, P* ,Q*	1 mostly line to line (some line to ground)	1 phase (typically yellow)	1 phase
33kV Feeders	100%	|S|, |V|', |I|	1 mostly line to line (some line to ground)	1 phase (typically yellow)	1 phase
11kV Feeders	100%	|S|, |V|', |I|	1mostly line to line (some line to ground)	1 phase (typically yellow)	1 phase
Monitored 33kV and 11kV Feeders	<10% (EMS Sub.net [57])	P, Q, |S|, |V|, |I|	3 line to line or 3 phase to ground	3 phases	3 phases
LV Network	No metering sent back to control centre	Maximum demand indicators are installed on the majority of the LV network transformers. These are periodically read and recorded in central databases by operators.			
Monitored LV Network and Customers Sites	<10% (Ad-hoc Schneider PM710 [124] monitors)	P, Q, |S|, |V|, |I|	3 phase to ground	3 phases	3 phases

Notes:

* The value of P and Q is calculated using phasor line to line measurements taken from all three phases, but the information sent back to the central control room only describes 1 phase.

' Voltage transformers are sometimes not present, thus the voltage measurement is taken to be the nominal voltage at that busbar.

The information shown in Table 2.2 was collected from conversations with engineers working at a UK distribution network in the UK.

Many of the measurement errors listed in Table 2.1 are described as a percentage error of full scale. A signal is said to be at full scale when it has reached its maximum representable value.

In this research, all power quality meters were assumed to follow the error properties of the Siemens 9610 monitor, unless otherwise stated. Voltages were assumed to be accurate to 0.1% of 1 per unit for all line to line and line to ground measurements, and 0.2% of 1 per unit for all symmetrical component magnitudes. Symmetrical voltage

phase angles were assumed to be accurate to $\pm 0.004\pi$. The standard deviation for the symmetrical voltage magnitude and phase measurements was therefore 0.067 per unit and 0.0013π radians respectively [67].

Table 2.1 also describes the accuracy of the GPS synchronisation within the monitoring device. If the voltage angles between two separate locations within the network are to be measured, then measurements must be taken relative to a global time stamp, such as a GPS signal. The accuracy at which the GPS signal is synchronised will affect the accuracy of the relative phase angle measurement between two locations. In the discussions within this thesis, it will be assumed that all measurements can be synchronised to a global GPS time signal.

2.1.2 Measurement Errors

Monitor measurement errors can modelled as a series of independently distributed normal distributions [8, 11]. Given a calibrated percentage error for a monitoring device (*%error*), a monitor's error can be modelled as a normal distribution with a standard deviation (σ_i) as defined in (2.1), given a measurement at *i*th bus z_i.

$$\sigma_i = \frac{z_i \times \%error}{3 \times 100} \tag{2.1}$$

2.1.3 Summary of Monitoring Assumptions

There are several assumptions that will be made throughout this thesis when considering the measurements taken at monitoring devices.

The assumptions relevant to the use of incumbent SCADA measurements within the network are:

- Single phase measurements are available at all substations on the secondary windings of 11kV transformers
- Measurements from the SCADA system are single phase
- The SCADA network measurements implicitly assume that the network's operation is balanced

Other general measurement assumptions include:

- The measurement error of the magnitude and phase of a measurement device are independent
- The error of a device follows a normal distribution as modelled by equation (2.1)

- The grounding of voltage transformers is an accurate representation of the true earth reference
- The error properties of the Siemens 9610 power quality monitor will be used unless otherwise stated

2.2 3-Phase Voltages, Currents and Powers

Voltage measurements are recorded using a VT with reference to another line in the system (line to line) or with reference to earth (line to ground), or a neutral conductor (line to neutral). The reference for the voltage has important implications for both unbalance and voltage sag monitoring. CTs record the value of the current flowing in each phase (line) separately.

The voltage and current waveforms in an alternating current (AC) power systems are made up of a fundamental component at 50Hz, and potentially high frequency transients and harmonics. In power system dynamics and harmonic analysis, it is important to consider the oscillatory nature of the power system. In this thesis, voltage, current and power measurements always concern root mean square (RMS) values at 50Hz. A phasor representation will be used for all voltages, currents and powers.

Throughout this thesis, voltages will be referred to as either line to line or line to ground. Line to line voltages will be described using the notation $V^{(ab)}$, $V^{(bc)}$ and $V^{(ca)}$, with the superscript (i,j) denoting the voltage in phase i relative to phase j. Phase voltages will always be referenced to ground (rather than neutral), and use the notation $V^{(a)}$, $V^{(b)}$ and $V^{(c)}$ to describe phases A, B and C. Line to line sequence voltages will be referenced to phases A and B and notated using $V^{(0LL)}$, $V^{(1LL)}$ and $V^{(2LL)}$ for zero, positive and negative sequences. Line to ground sequence voltages will be referenced to phase A and notated using $V^{(0)}$, $V^{(1)}$ and $V^{(2)}$ for zero, positive and negative sequences. The current flowing in each phase will be denoted as $I^{(a)}$, $I^{(b)}$ and $I^{(c)}$.

The different sets of technical measurements that are typically collected at a substation can be split into the following groups:

1. Line to ground voltages
2. Line to line currents
3. Line currents
4. Real and reactive power consumption

Depending on the availability of monitoring at a substation, some, none, or all of these parameters might be monitored. Other information may also be collected including (but not limited to) temperature and asset condition related measurements.

2.2.1 Line to Ground Voltages

The set of three phase to earth voltages, $V^{(a)}$, $V^{(b)}$ and $V^{(c)}$, can be converted into positive negative and zero sequence components $V^{(0)}$, $V^{(1)}$ and $V^{(2)}$ by using the well known Fortescue [128] transformation. This is shown in equation (2.2):

$$\begin{bmatrix} V^{(0)} \\ V^{(1)} \\ V^{(2)} \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} V^{(a)} \\ V^{(b)} \\ V^{(c)} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} V^{(a)} \\ V^{(b)} \\ V^{(c)} \end{bmatrix} \tag{2.2}$$

where $a = e^{2\pi/3}$.

2.2.2 Line to line voltages and Line Currents

The Fortescue transform [128] can be applied to convert a set of three phase line to line voltages, $V^{(ab)}$, $V^{(bc)}$ and $V^{(ca)}$, into a set of sequence components, $V^{(0LL)}$, $V^{(1LL)}$ and $V^{(2LL)}$. This is shown in equation (2.3):

$$\begin{bmatrix} V^{(0LL)} \\ V^{(1LL)} \\ V^{(2LL)} \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} V^{(ab)} \\ V^{(bc)} \\ V^{(ca)} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} V^{(ab)} \\ V^{(bc)} \\ V^{(ca)} \end{bmatrix} \tag{2.3}$$

Since $V^{(ab)}$, $V^{(bc)}$ and $V^{(ca)}$ form a closed triangle and $V^{(ab)}+V^{(bc)}+V^{(ca)}=0$, the value of $V^{(0LL)}$ is also zero. Line to line voltages therefore have no zero sequence components.

Using $V^{(1LL)}$ and $V^{(2LL)}$ it is possible to determine $V^{(1)}$ and $V^{(2)}$ through equations (2.4) and (2.5). $V^{(0)}$ cannot be determined from line to line voltages alone.

$$V^{(1)} = V^{(1LL)} / 1 - a^2 \tag{2.4}$$

$$V^{(2)} = V^{(2LL)} / 1 - a^2 \tag{2.5}$$

2.2.3 Line Currents

Line currents can be transformed into their symmetrical components using the Fortescue transform[128] as shown in equation (2.6).

$$\begin{bmatrix} I^{(0)} \\ I^{(1)} \\ I^{(2)} \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} I^{(a)} \\ I^{(b)} \\ I^{(c)} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} I^{(a)} \\ I^{(b)} \\ I^{(c)} \end{bmatrix} \tag{2.6}$$

2.2.4 Power Consumption in Three Phase Networks

The total power consumption in a three phase network is defined as shown in equation (2.7).

$$S_{3\phi} = V^{(a)} I^{(a)*} + V^{(b)} I^{(b)*} + V^{(c)} I^{(c)*} \tag{2.7}$$

The power consumption can also be segmented across each of the three lines in the network, as described in equations (2.8), (2.9) and (2.10).

$$S^{(a)} = V^{(a)} I^{(a)*} \tag{2.8}$$
$$S^{(b)} = V^{(b)} I^{(b)*} \tag{2.9}$$
$$S^{(c)} = V^{(c)} I^{(c)*} \tag{2.10}$$

If all line to ground voltages and line currents are known for a location in a network, it is possible to determine the total three phase power flow.

Power consumption can also be defined in the sequence domain, as shown in (2.11) in per unit as:

$$S_{3\phi} = V^{(0)} I^{(0)*} + V^{(1)} I^{(1)*} + V^{(2)} I^{(2)*} \tag{2.11}$$

Again, this can be segmented into each sequence as shown in equations (2.12), (2.13) and (2.14).

$$S^{(0)} = V^{(0)} I^{(0)*} \tag{2.12}$$
$$S^{(1)} = V^{(1)} I^{(1)*} \tag{2.13}$$
$$S^{(2)} = V^{(2)} I^{(2)*} \tag{2.14}$$

A full set of phase to earth voltages fully defines the power flowing in any line and any sequence.

With line to line voltages, the power flowing in each line cannot be determined because the zero sequence component of the voltage cannot be defined. However, positive and negative sequence power (equations (2.13) and (2.14)) can be derived.

2.2.5 Zero Sequence Assumptions

The magnitude of the zero sequence voltage can be determined at certain locations within a network where there is no path for a zero sequence current to flow.

The propagation of zero sequence current is halted by certain types of transformers. Zero sequence currents are halted by Y-Y transformers with one neutral ungrounded, Y-Δ transformers and Δ-Δ transformers. This knowledge can help to define the zero sequence voltage magnitude at certain locations within a network.

Consider the case where an unbalanced (asymmetric) fault occurs and zero sequence currents are injected on the low voltage (LV) side of Y-Δ transformer, as shown in Fig. 2.1. The zero sequence voltage on the high voltage (MV) side of the transformer will be zero, as long as the following assumptions are not violated:

- Any loads connected at MV or above do not inject zero sequence currents into the network, and are therefore balanced.
- Any loads connected to the MV network through Y-Y grounded transformers do not inject zero sequence currents into the network.
- The MV network is uncoupled in the sequence domain, with no mutual coupling between positive and zero sequences or negative and zero sequences.

If these assumptions can be made at the MV transformer, then they can also be made for all busbars connected on the HV side of the transformer.

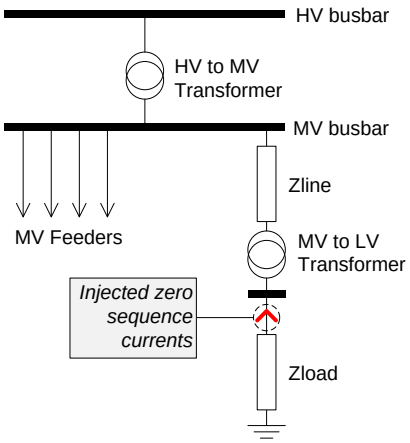


Fig. 2.1. A three phase load connected to an arbitrary power network.

The case of an unbalanced asymmetric fault is a specific case where zero sequence currents are injected into the network. This type of network operation is synonymous with voltage unbalance. In both situations, the key point is that progress of zero sequence currents are halted by certain types of transformers.

2.3 Test Power Systems

There are two power networks which shall be used throughout this thesis. The first is a 295 bus generic distribution which does not physically exist, but was designed to be representative of a typical UK distribution network [8]. The second network is a 24 bus distribution network which is based on a physical section of the UK's distribution network. The physical location of the network not specified for commercial reasons.

2.3.1 295 Bus Generic Distribution Network

The 295 bus distribution network [8] comprises four 275kV transmission in-feeds, 132kV and 33kV sub-transmission networks (which are mostly meshed) and a distribution network which pre-dominantly operates at 11kV, but also has some small sections at 3.3kV (and is mostly radial). There are 295 buses, 276 over-head lines and underground cables and 37 transformers with various winding connections. The average X/R ratio of 11kV, 33kV and 132kV cables and lines are 0.62, 2.58 and 4.13 respectively. All lines vary in length from 50 m to 11 km and resistance from 0.002Ω/km for 132kV underground to 0.66 Ω/km for 11kV overhead lines.

The single line diagram for the network can be found in Chapters 5 and 7. Full details of the network's system parameters and loading can be found in [129].

2.3.2 24 Bus Section of UK Distribution Network

The 24 bus distribution network is a model of a real section of UK distribution network. The network comprises 14 33kV busbars and 10 11kV busbars. The 33kV network is meshed (in an N-1 formation) whilst the 11kV network busbars feed a series of independent feeders. The 11kV feeders are not modelled individually.

A single line diagram for the topology of this network can be found in Chapter 8 and the network's system parameters and loading can be found in Appendix B.

2.4 3-Phase Power System Component Modelling

Three phase modelling of the network is required for both voltage sag and unbalance analysis. A three phase model of the system is required because both voltage sags and unbalance asymmetrically affect network operation, and thus simplified single phase models are not sufficient.

2.4.1 Lines and Cables

Lines and cables were modelled using the equivalent π model. All cables and lines were assumed to be uncoupled in the sequence domain.

Uncoupled transmission lines can be represented as a set of two 3×3 component matrices $\mathbf{Y}_L$, and $\mathbf{Y}_S$ [123].

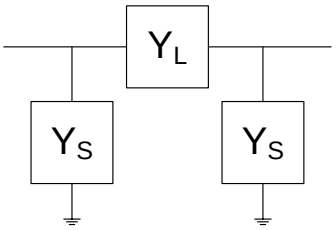


Fig. 2.2. 3 phase component model for transmission line.

Each element of $\mathbf{Y}_L$ and $\mathbf{Y}_S$ can be derived from knowledge of the zero ($\underline{y}^{(0)}$), positive ($y^{(1)}$) and negative ($y^{(2)}$) sequence components. The negative sequence impedance of a line was assumed to be the same as positive sequence impedance, and the zero sequence was assumed to be three times the positive sequence impedance (unless otherwise stated). Equation (2.15) describes the sub-component matrix in the sequence domain for $\mathbf{Y}_L^{seq}$.

$$\mathbf{Y}_L^{seq} = \begin{bmatrix} y^{(0)} & 0 & 0 \\ 0 & y^{(1)} & 0 \\ 0 & 0 & y^{(2)} \end{bmatrix} \tag{2.15}$$

$\mathbf{Y}_S^{seq}$ can be derived in a similar manner from the single phase equivalent of the shunt component of the line. Both $\mathbf{Y}_L^{seq}$ and $\mathbf{Y}_S^{seq}$ can be converted into the phase domain by applying the Fortescue transformation:

$$\begin{aligned} \mathbf{Y}_L &= \mathbf{A}^{-1}\mathbf{Y}_L^{seq}\mathbf{A} \\ \mathbf{Y}_S &= \mathbf{A}^{-1}\mathbf{Y}_S^{seq}\mathbf{A} \end{aligned} \tag{2.16}$$

2.4.2 Generators

All of the networks modelled in this research are distribution networks with a single in-feed from a higher voltage network. The in-feed is modelled as a generator, but its model will vary depending whether voltage sag (fault studies) or unbalance (load flow and state estimation) studies are being conducted.

For three phase load flow and state estimation, generators are modelled using as an ideal voltage source infinite bus. The system is always modelled under steady state conditions and source of the distribution network is assumed to be able to supply as much power as required. The generator terminals were always assumed to supply a balanced three phase set of voltages (unless otherwise stated).

For short circuit studies, the point of common coupling at the in-feed is modelled as an ideal voltage source (E) behind a series sequence resistance (R) and sub-transient sequence reactance (Xd”). This is shown in Fig. 2.3.

The sub-transient reactance was selected as this determines the current in the first few cycles after a fault and the performance of the network in this time frame is relevant to voltage sag studies.

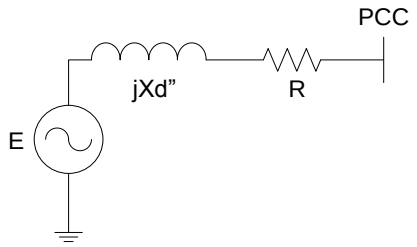


Fig. 2.3. The generator model at the point of common coupling (PCC) for the in-feed of the distribution network.

2.4.3 Loads

Load models are important and can significantly affect the results of dynamic simulations [8]. In this research, a constant impedance load model was used to perform fault analysis for voltage sag studies. This has been shown to be a justifiably accurate approximation for fault studies [129]. Unbalance is a steady state phenomenon, and therefore all loads were assumed to be constant power, when subjected to small perturbations in voltage. This assumption is justified for unbalance because unbalance is always judged over periods of time, typically much greater than 5 minutes. Over long enough periods, it is reasonable to assume that most loads will adapt to changes in voltage by drawing more current to maintain a constant power output.

The equivalent impedance for a constant impedance load was calculated using:

$$R + jX = \frac{V^2}{(P + jQ)^*} \tag{2.17}$$

where P and Q are the active and reactive power demand at the busbar, and V is the complex voltage before the fault.

2.4.4 Transformers

All of the transformers modelled in this thesis are either solidly grounded Y-Y transformers or solidly grounded Y-Δ transformers.

The nodal admittance matrix for a three phase transformer is a 6×6 component block and can be constructed by formulating a connection matrix (**C**) and a primitive admittance matrix ($\mathbf{Y}_{prim}$) and combining them using equation (2.18):

$$\mathbf{Y}_{node} = \mathbf{C}^* \mathbf{Y}_{prim} \mathbf{C} \tag{2.18}$$

where **C*** is the transpose of the connection matrix and $\mathbf{Y}_{node}$ is the transformers nodal admittance matrix.

The connection matrix for a solidly grounded Y-Y and solidly ground Y-Δ transformer can be derived by considering the connections between the primary and the secondary coils [123]. The connection matrices for a Y-Y ($\mathbf{C}^{Y\text{-}Y}$) and Y-Δ ($\mathbf{C}^{Y\text{-}\Delta}$) transformer are shown below in equations (2.19) and (2.20):

$$\mathbf{C}^{Y-Y} = \mathbf{I}_{(6\times6)} \tag{2.19}$$

$$\mathbf{C}^{Y-\Delta} = \begin{bmatrix} \mathbf{I}_{(3\times3)} & \mathbf{0}_{(3\times3)} \\ \mathbf{0}_{(3\times3)} & \mathbf{C}_{\Delta} \end{bmatrix} \tag{2.20}$$

$$\mathbf{C}_{\Delta} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \tag{2.21}$$

where $\mathbf{I}_{(N\times N)}$ is an $N\times N$ identity matrix, $\mathbf{0}_{(N\times N)}$is an $N\times N$ matrix of zeros. The factor of $1/\sqrt{3}$ is necessary to cater for the turns ratio of the delta windings. It should be noted that Δ-Y transformers can be created by swapping the $\mathbf{C}_{\Delta}$ and $\mathbf{I}_{(3\times3)}$ components of the connection matrix.

The primitive admittance matrix ($\mathbf{Y}_{prim}$) for a transformer can be derived by considering the transformer admittance in the sequence domain. The primitive admittance matrix of a transformer with leakage admittance $\tilde{y}$, zero sequence admittance $\tilde{y}_{m0}$ and magnetising admittance $\tilde{y}_m$ [130] is shown in (2.22):

$$\mathbf{Y}_{prim}^{seq} = \begin{bmatrix} \tilde{y}\mathbf{I}_{(3\times3)} & -\tilde{y}\mathbf{I}_{(3\times3)} \\ -\tilde{y}\mathbf{I}_{(3\times3)} & \mathbf{Y}' \end{bmatrix} \tag{2.22}$$

$$\mathbf{Y}' = \begin{bmatrix} \tilde{y} + \tilde{y}_m^{(0)} & 0 & 0 \\ 0 & \tilde{y} + \tilde{y}_m & 0 \\ 0 & 0 & \tilde{y} + \tilde{y}_m \end{bmatrix} \tag{2.23}$$

The magnetising admittance is often neglected and was thus ignored for all of the case studies in this thesis. This is a valid approximation so long as the transformer is not highly saturated [130]. The saturation of transformers becomes important when transformer energizing is the cause of a voltage sag, but this type of sag was not considered in this research.

The primitive admittance matrix in phase coordinates can be obtained by applying the Fortescue transformation to each three by three element of the primitive admittance matrix in the sequence domain as shown in (2.24):

$$\mathbf{Y}_{prim}^{phase} = \begin{bmatrix} \mathbf{A}^{-1}\tilde{y}\mathbf{I}_{(3\times3)}\mathbf{A} & -\mathbf{A}^{-1}\tilde{y}\mathbf{I}_{(3\times3)}\mathbf{A} \\ -\mathbf{A}^{-1}\tilde{y}\mathbf{I}_{(3\times3)}\mathbf{A} & \mathbf{A}^{-1}\mathbf{Y}'\mathbf{A} \end{bmatrix} \tag{2.24}$$

2.4.4.1 Tap Changing Transformers

Transformer taps were modelled by incorporating a primary tap ratio α into the 6×6 nodal admittance matrix $\mathbf{Y}_{node}$. This can be achieved as shown in the following equation:

$$\mathbf{Y}_{node} = \begin{bmatrix} \mathbf{Y}_I / \alpha^2 & \mathbf{Y}_{II} / \alpha \\ \mathbf{Y}_{II} / \alpha & \mathbf{Y}_{III} \end{bmatrix} \tag{2.25}$$

where $\mathbf{Y}_I$, $\mathbf{Y}_{II}$ and $\mathbf{Y}_{III}$ are formed from partitioning the $\mathbf{Y}_{node}$ matrix into four three by three blocks. Transformer tap positions were modelled at discrete intervals dependent on the tap positions of the transformer.

More information on three phase transformer models can be found in [130] and [123] [130].

2.4.5 Admittance Matrix

The component matrices are the building blocks of the three phase admittance matrix. The set of $N \times N$ component matrices can be combined into a single three phase admittance matrix [123] using the same rules required to generate a single phase admittance matrix [63]. The diagonal elements of the matrix are equivalent to the self-admittances and are the sum of all three by three sub-component matrices attached to that node. The off-diagonal elements are the mutual admittances, and these are equivalent to the negated sum of all the $N \times N$ admittance matrices between those two nodes. It should be noted that forming the admittance matrix in this way only works if all of the components have been reduced into sub-components such that there is no mutual coupling between any two sets of sub-component.

Once the three phase admittance matrix for components have been developed, these can be used to perform three phase power flow studies. Further details on the formulation of the three phase admittance matrix can be found in [123] and [65].

2.4.6 Summary of Three Phase Component Modelling Assumptions

Throughout this thesis, the following assumptions were made regarding 3-phase component models:

- All transformers are assumed to be either Y-Y or Y-D and solidly grounded at both star points
- The magnetising impedance of transformers was neglected [130]
- All three phase components were assumed to be uncoupled in the sequence domain
- Negative and positive sequence impedances were assumed to be equal
- Zero sequence impedances were assumed to be 3 times the positive sequence impedance [129]

2.5 Voltage Sag Modelling

The primary purpose of all of the voltage sag research discussed within this thesis is to show that voltage sag performance can be measured using a limited monitor set. To achieve this objective, it is assumed that a voltage sag can be simplified and characterised by two properties: *magnitude* and *duration*.

In reality, the characterisation of a voltage sag into these two parameters is an approximation. The transient evolution of a voltage sag is complex [11], but this approximation has been used extensively in research before [25, 48, 81] and should suffice in this research.

Fig. 2.4 shows how the transient waveform of an arbitrary voltage sag is approximated into magnitude and duration, for each of the three phases in the network.

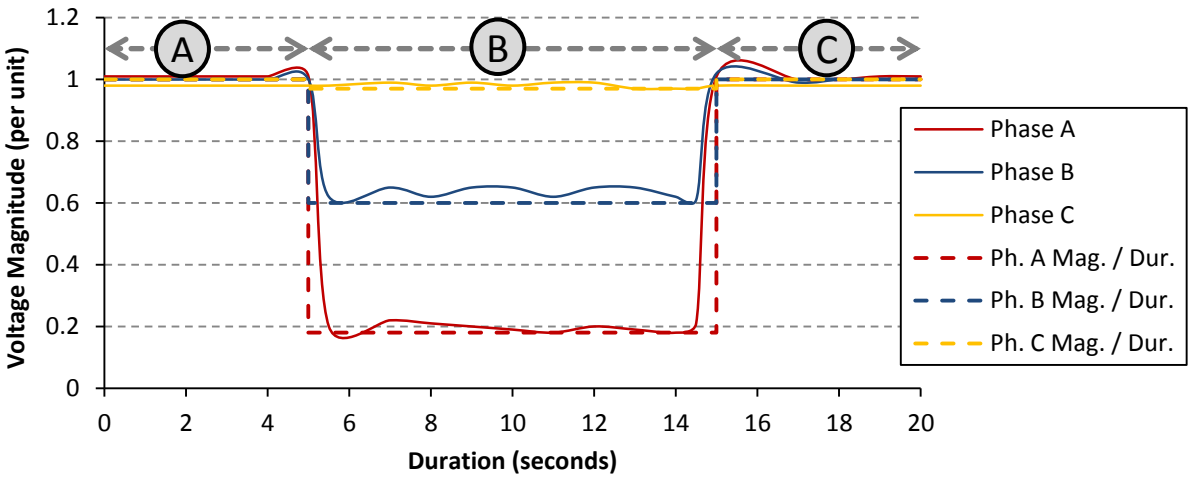


Fig. 2.4. Voltage magnitude per phase for an example voltage sag.

Fig. 2.4 shows how the power system's operation can be split into three regions (A, B and C). All three regions can be modelled by representing voltages as phasors.

Regions A and C are defined by the pre-fault and post-fault loading of the system and can be determined through load flow [63]. If the pre-fault and post-fault loading is assumed to be unchanged, then these two regions are identical.

Region B concerns the state of the system during a voltage sag. The residual voltages in region B can be determined by conducting a fault study.

2.5.1 Stochastic Properties of Faults

Voltage sags are modelled in this thesis using fault studies. Faults are stochastic in nature, and their characteristics are influenced by a number of important factors.

The factors that influence the characteristics of the fault that will be modelled in this thesis include:

- Fault type (for example, three phase, single line to ground, line to line or double-line to ground)
- Fault location
- Fault impedance
- Fault duration

Considering fault type is important since different types of fault will generate different impacts in the network. An asymmetric fault in one phase may only impact on customers in one (or two) of the three phases, whilst a three phase fault affects customers in all three phases. Where the fault is physically located will determine which customers are affected by the fault. Customers very close to the fault will be influenced to a much higher degree than those further away. The impedance of the fault affects the severity of the fault, as a fault with low fault impedance will cause a higher severity sag. The fault's duration is important when considering the resultant impact of in terms of customer trips in a network: a longer duration sag will be more likely to cause customers to trip.

Another factor which affects the characteristic of the fault being studied is the configuration of protection systems which influences both fault clearing time and reclosing time. Protection systems ultimately influence the duration of the fault. The age and wear of network assets is a very important factor as this influences the reliability of components and hence affects the fault rate. External factors such as the weather can also heavily influence the fault rate in the network, as different types of components have different fault rates in different weather conditions. The dynamic response of

equipment such as induction motors, and synchronous machines will also affect the exact depth and duration of a voltage sag.

Fig. 2.5 illustrates the different types of stochastic causes of faults using data recorded in a UK distribution network. The largest numbers of faults are attributed to deterioration due to ageing or wear (excluding corrosion). A significant number of faults are also classified as unknown.

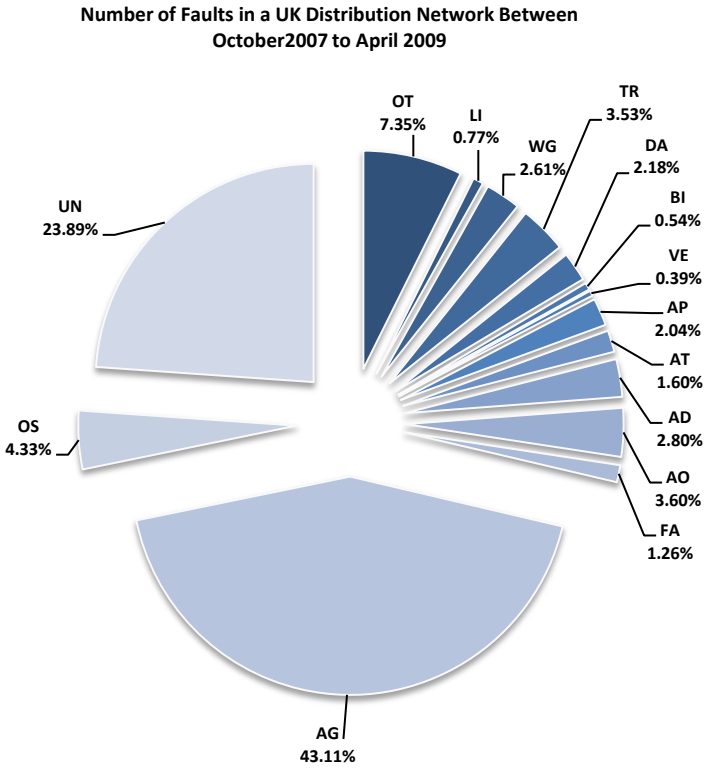


Fig. 2.5. The fault outage causes of faults recorded in a UK distribution network between 2007 and 2009.

TABLE 2.3. FAULT OUTAGE CAUSE KEY.

Acronym	Description
OT	All Other Faults
LI	Lightning Faults
WG	Wind and Gale (excluding windborne material)
TR	Growing or Falling Trees (not felled)
DA	Wilful Damage, Interference or Theft
BI	Birds (including Swans and Geese)
VE	Vermin, Wild Animals and Insects
AP	Accidental Contact, Damage or Interference by private Individuals (excluding 49 and 56)
AT	Accidental Contact, Damage or Interference by unknown Third Parties
AD	Accidental Contact, Damage or Interference by private Developers or their Contractors
AO	Accidental Contact, Damage or Interference by Other Third Parties
FA	Faulty Installation or Construction
AG	Deterioration due to Ageing or Wear (excluding corrosion)
OS	Operational or Safety Restriction
UN	Cause Unknown

2.5.2 Fault Studies

Short circuit fault studies enable the response of the network during short circuit conditions to be analysed. In this thesis, fault studies will be conducted with three phase (LLL), single phase to ground (SLG), line to line (LL) or double line to ground (LLG) faults. Faults can occur anywhere on lines cables or busbars.

The derivation of the fault equations are best illustrated by considering Fig. 2.6. Fig. 2.6 shows a fault occurring on the *l*th line connecting the *p*th and *q*th buses. The pre-fault voltages may be estimated or measured at the *p*th and the *q*th busbars. The fault equations will be derived so that the values of the short circuit currents during a fault can be calculated.

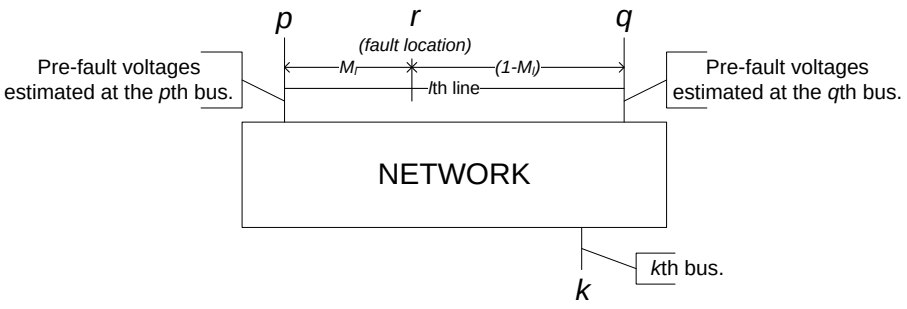


Fig. 2.6. A fault occurring at position *r*, a distance M_l along the *l*th between the *p*th and the *q*th busbar.

In order to calculate the during fault voltages throughout the network, the bus impedance matrix is required. The bus impedance matrix for a general system can be defined as $\boldsymbol{Z'_{bus}}^{(k)}$ [63], where the notation $Z'_{ij}{}^{(k)}$ represents the element contained within the *i*th row and the *j*th column of the *k*th sequence bus impedance matrix ($\boldsymbol{Z'_{bus}}^{(k)}$) [63]. The bus impedance matrix can be derived either from the direct Z bus building [131], or by directly inverting the three phase admittance matrix.

The bus impedance matrix can be readily modified to add a new temporary faulted busbar *r*. Let the bus impedance matrix with the temporary busbar *r* added be $\boldsymbol{Z_{bus}}$, and element $Z_{kr}^{(i)}$ be the element contained in the *k*th row and *r*th column and the *i*th sequence. This matrix can be built using the method of adding an extra busbar to the busbar impedance matrix described in [63]. The $\boldsymbol{Z_{bus}}$ impedance matrix (with the additional *r*th busbar) will be used in the equations that follow.

2.5.3 Fundamental Equations for all Types of Fault

The zero ($V_k^{(0)}$), positive ($V_k^{(1)}$) and negative ($V_k^{(2)}$) sequence voltages during a fault are defined in equations (2.26) to (2.28):

$$V_k^{(0)} = -Z_{kr}^{(0)} I_f^{(0)} \tag{2.26}$$

$$V_k^{(1)} = V_k^{(1)'} - Z_{kr}^{(1)} I_f^{(1)} \tag{2.27}$$

$$V_k^{(2)} = -Z_{kr}^{(2)} I_f^{(2)} \tag{2.28}$$

where $I_f^{(0)}$, $I_f^{(1)}$ and $I_f^{(2)}$ are the zero, positive and negative sequence fault currents at the fault point [63], and $V_k^{(1)'}$ is the positive sequence pre-fault voltage.

2.5.4 3-Phase Faults

For a three phase symmetrical fault, the fault current in each of the three sequences can be defined as follows:

$$I_f^{(1)} = \frac{V_r^{(1)'}}{R_f + Z_{rr}^{(1)}} \tag{2.29}$$

$$I_f^{(0)} = I_f^{(2)} = 0 \tag{2.30}$$

where R_f is the fault impedance, and can be assumed to be entirely real [103]. The zero and negative sequence currents during a three phase faults are assumed to be zero, as the flow of current is assumed to be balanced.

2.5.4.1 Single Line to Ground Faults

For a phase A to ground fault, $I_f^{(0)} = I_f^{(1)} = I_f^{(2)}$.

$$I_f^{(1)} = \frac{V_r^{(1)'}}{3R_f + Z_{rr}^{(0)} + Z_{rr}^{(1)} + Z_{rr}^{(2)}} \tag{2.31}$$

$$I_f^{(0)} = I_f^{(1)} = I_f^{(2)} \tag{2.32}$$

2.5.4.2 Line to Line Faults

For a phase B to C fault, $I_f^{(1)} = -I_f^{(2)}$, $I_f^{(0)} = 0$.

$$I_f^{(1)} = \frac{V_r^{(1)'}}{R_f + Z_{rr}^{(1)} + Z_{rr}^{(2)}} \tag{2.33}$$

$$I_f^{(1)} = -I_f^{(2)} \tag{2.34}$$

$$I_f^{(0)} = 0 \tag{2.35}$$

2.5.4.3 Double Line to Ground Faults

For a double line to ground fault, the fault current equations are as follows [63]:

$$I_f^{(1)} = \frac{V_r^{(1)'}}{Z_{rr}^{(1)} + \dfrac{Z_{rr}^{(0)}(Z_{rr}^{(2)} + R_f)}{Z_{rr}^{(0)} + Z_{rr}^{(1)} + Z_{rr}^{(2)}}} \tag{2.36}$$

$$I_f^{(2)} = -I_f^{(1)} \frac{Z_{rr}^{(0)} + 3R_f}{Z_{rr}^{(0)} + Z_{rr}^{(2)} + 3R_f} \tag{2.37}$$

$$I_f^{(0)} = -I_f^{(1)} \frac{Z_{kk}^{(2)}}{Z_{rr}^{(0)} + Z_{rr}^{(2)} + 3R_f} \tag{2.38}$$

2.5.5 Summary of Assumptions for Fault Studies

Where fault studies are performed in this thesis, the studies are always performed under a series of assumptions. These modelling assumptions can be summarised as follows:

- The system is composed of lines, transformers, generators and buses
- All line, transformer and generator impedances are assumed to be free from error

- Transformers and lines are modelled with no coupling between sequence impedances
- System loading is assumed to be balanced prior to the occurrence of a disturbance
- Transformers are assumed to be solidly grounded, through a negligible impedance
- Fault impedance can be assumed to be entirely real [103]
- Fault type, fault location, fault duration and fault impedance are all modelled for every fault simulation
- The dynamic response of equipment is not considered when simulating faults

2.6 Unbalance Modelling

As described in Chapter 1, voltage unbalance is caused by the asymmetric operation of multi-phase electrical power networks. The voltage unbalance factor is the ratio of the complex negative sequence voltage ($V^{(2)}$) to the complex positive sequence voltage ($V^{(1)}$) as described by (1.1).

Emission sources can include asymmetric un-transposed or partially transposed transmission lines [23], single phase and dual-phase loads and unbalanced three phase loads. The scope of this thesis will focus on identifying the source and affects of unbalanced loads only. Unbalanced network components will be largely ignored, although some of the later case studies will show how the models can be adapted to model the effects of unbalanced network components.

The nature of unbalance will be assumed to be time-invariant, and treated only as a steady-state phenomenon, rather than a transient phenomenon.

2.6.1 Summary of Assumptions for Unbalance Modelling

Where unbalance is modelled in the network, the following assumptions are made:

- The injection of unbalance into the network is assumed to be invariant over time
- Unbalance is assumed to originate from unbalanced loads, rather than unbalanced cables, lines or network assets
- Transient unbalance generated by short circuit asymmetric faults are ignored

2.7 Three Phase Load Flow

The object of a load flow study is to calculate the steady state operating characteristics of a power system network given a certain level of network load and generation. The

principal information obtained from a load flow study is the voltages and angles at all buses in the network. This allows operators to calculate the amount of power flowing throughout the network, losses within the network and the level of power consumed and absorbed at all buses. It also enables operators to subsequently perform further studies based on specific phenomena such as stability, faults or power quality.

The results of load flow are used to calculate the pre-fault and post-fault voltages for all of the research on voltage sags discussed in Chapters 4, 5 and 6. Several components of the three phase load flow formulation are also used for distribution system state estimation, including the Jacobian, the three phase power injection equations and the admittance matrix.

2.7.1 Formulation

A power flow study can be solved by iteratively solving the equation shown in (2.39) [63]:

$$\mathbf{z} = \mathbf{F(x)} \tag{2.39}$$

where **z** is a vector of real and reactive power values, **x** is a vector of voltages and angles and **F**(**x**) is a set of equations which describe all real and reactive power injections throughout the network. The 3-phase power injection equations are shown in equations (2.40) and (2.41). These equations show the real and reactive power injection at the *i*th bus in the *p*th phase.

$$P_i^p = V_i^p \sum_{k=1...N} \sum_{m=\{a,b,c\}} V_k^m [G_{ik}^{pm} \cos(\theta_{ik}^{pm}) + B_{ik}^{pm} \sin(\theta_{ik}^{pm})] \tag{2.40}$$

$$Q_i^p = V_i^p \sum_{k=1...N} \sum_{m=\{a,b,c\}} V_k^m [G_{ik}^{pm} \sin(\theta_{ik}^{pm}) - B_{ik}^{pm} \cos(\theta_{ik}^{pm})] \tag{2.41}$$

where the *p*th or the *m*th phase is either, *a* (1st index), *b* (2nd index) or *c* (3rd index), G_{ij}^{pm} is the real *i,j*th element of the three phase admittance matrix $\mathbf{Y_{bus}}$, and B_{ij}^{pm} is the imaginary *i,j*th element of $\mathbf{Y_{bus}}$. θ_{ij}^{pm} is the angle in radians between the *i*th bus *p*th phase and the *j*th in the *m*th phase.

The load flow equation can be solved iteratively, using the Newton-Raphson method [63]. The update equations are defined in equation (2.42).

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{F}_\mathbf{x}[\mathbf{z} - \mathbf{F}(\mathbf{x}_k)] \tag{2.42}$$

$$\mathbf{F}_\mathbf{x} = \frac{\partial \mathbf{F}(\mathbf{x_k})}{\partial \mathbf{x}} \tag{2.43}$$

where $\mathbf{x}_{k+1}$ is the estimate for the voltages and angles at the (k+1)th iteration. $\mathbf{F_x}$ is the Jacobian matrix whose elements are derived in the same way as for distribution system state estimation.

To build the Jacobian matrix, the derivatives of the load flow equations are evaluated with respect to each state variable of the system. This is achieved through sequential differentiation of equations (2.40) and (2.41). The elements of the Jacobian matrix are fully derived in Appendix A.

2.7.1.1 Slack Bus Model

Both the 295 bus and 24 bus network are fed from a single generation source, modelled as a slack bus. For all load flow (and distribution system state estimation analysis) the slack bus was assumed to be a fixed magnitude and out of phase by exactly 120°. Thus, the slack bus was always assumed to be completely balanced.

2.7.2 Summary of Assumptions for Three Phase Load Flow Analysis

The following modelling assumptions are made when performing load flow analysis:

- The network is operating in a steady state, unaffected by small perturbations in voltage or frequency
- All loads can be modelled as a constant active and reactive power demand
- All line, transformer and generator impedances are assumed to be free from error
- Dynamic concepts of load modelling are not considered

2.8 Distribution System State Estimation

Distribution system state estimation (DSSE) can be considered as an extension of the standard load flow equations which makes them solvable within the context of a practical power system. The objective of DSSE is to establish the current state of the network by minimising the least square error in a set of measurements and estimates taken from a network. The output from a DSSE algorithm is a set of voltages and angles which define the state of the network.

There are several important differences between DSSE and load flow. In DSSE, the number of equations which describe the system may be more than the number of state variables. This is described as an over-determined system, and it has the result of making the Jacobian matrix rectangular. Unlike load flow, the set of equations for

DSSE may also include line flows, voltages, and currents. A DSSE Jacobian matrix therefore also includes derivatives of real and reactive line flow, voltage or current. The DSSE equations can also be weighted by the accuracy of the measurements, whilst load flow implicitly assumes that each power and reactive power measurement is equally accurate.

DSSE algorithms are applied in this research to establish pre-fault voltages in voltage sag profile estimation (Chapter 5) and also to estimate the level of unbalance within a distribution network (Chapter 8).

2.8.1 3-Phase Distribution System State Estimation

The state estimation problem is defined mathematically by equation (2.44).

$$\mathbf{e} = \mathbf{z} - \mathbf{H(x)} \tag{2.44}$$

where **z** is a vector of measurements, **H**(**x**) s a non-linear set of equations that describes the true state of the power system with state variables **x**, and **e** is a vector of errors between the observed measurements and true state of the system. **e**~$N(0,\boldsymbol{R})$, where **R** is the covariance matrix of the measurement errors (**e**).

In this research, the weighted least squares (WLS) and the generalized least squares [74] (GLS) state estimation technique were used. The objective function of the GLS / WLS estimation problem is shown in equation (2.45).

$$\min_{\mathbf{x}}[\mathbf{z} - \mathbf{H}(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{z} - \mathbf{H}(\mathbf{x})] \tag{2.45}$$

The state estimation equations can be solved iteratively, using the Newton-Raphson method. The update equations are defined in equation (2.46).

$$\mathbf{x}_{k+1} = \mathbf{x}_k + (\mathbf{H}_\mathbf{X}\mathbf{R}^{-1}\mathbf{H}_\mathbf{X})^{-1}\mathbf{H}_\mathbf{x}^T\mathbf{R}^{-1}[\mathbf{z} - \mathbf{H}(\mathbf{x_k})] \tag{2.46}$$

$$\mathbf{H}_\mathbf{x} = \frac{\partial \mathbf{H}(\mathbf{x_k})}{\partial \mathbf{x}} \tag{2.47}$$

where $\mathbf{x}_{k+1}$ is the estimate for the state variables at the (k+1)th iteration. $\mathbf{H_x}$ is the Jacobian matrix.

Unlike weighted least squares (WLS) [62], GLS does not assume that the measurements within the covariance matrix are uncorrelated. **R**'s off diagonal elements are non-zero, and incorporate the correlation between pairs of measurement errors. The GLS method is relevant in 3 phase systems where measurement estimates (pseudo-measurements) are often correlated across all 3 phases. Like WLS, each diagonal element of **R** is the reciprocal of the measurement error variances.

The DSSE equations are formulated to solve for the state vector **x**. The state vector **x** fully defines the state of the system [65]. It typically contains a set of voltages for all buses in the network; angles and magnitudes or real and imaginary components. The measurement vector **z** can be configured to include any measurement such as three phase active and reactive power demands, three phase active and reactive power flows, as well as voltage magnitudes, and currents.

2.8.2 Non-Linear Power Flow Equations

The non-linear power flow equations describe the set of equations in **H(x)**.

There are 5 different types of measurements: real $(P_i^{(p)})$ and reactive power $(Q_i^{(p)})$, voltage magnitude $(|V_i^{(p)}|)$, real power flow $(P_{ij}^{(p)})$ and reactive power flow $(Q_{ij}^{(p)})$ in a line. Each of these measurements can be formulated in terms of magnitude and phase of the voltage at all buses within the network.

Equations (2.48) and (2.49) describe the line flow equations, where $G_{p,m}$ and $B_{p,m}$ are elements of 3×6 line admittance matrices $\mathbf{G_L}$ and $\mathbf{B_L}$ relating voltage $(V=[V_i^{(a)},V_i^{(b)},V_i^{(c)},V_j^{(a)},V_j^{(b)},V_j^{(c)}]^T)$ and current $(I=[I_{i,j}^{(a)},I_{i,j}^{(b)},I_{i,j}^{(c)}]^T)$ between buses i and j in the form $I_{ij}=(\mathbf{G}_L+j\mathbf{B}_L)\mathbf{V}_{ij}$, where $\mathbf{I}_{ij}$ is a 3×1 vector representing the current flows in phases *a*, *b* and *c* from bus i to j and $\mathbf{V}_{ij}$ is a 6×1 vector representing the voltages at bus i in all three phases and the voltages in bus j.

$$P_{ij}^{(p)} = V_i^{(p)} \sum_{m=\{a,b,c\}} \begin{matrix} V_i^{(m)} G_{p,m}[\cos(\theta_i^{(p)} - \theta_i^{(m)})] + V_i^{(m)} B_{p,m}[\sin(\theta_i^{(p)} - \theta_i^{(m)})] \\ - V_j^{(m)} G_{p,(m+3)}[\cos(\theta_i^{(p)} - \theta_j^{(m)})] - V_j^{(m)} B_{p,(m+3)}[\sin(\theta_i^{(p)} - \theta_j^{(m)})] \end{matrix} \quad (2.48)$$

$$Q_{ij}^{(p)} = V_i^{(p)} \sum_{m=\{a,b,c\}} \begin{matrix} V_i^{(m)} G_{p,m}[\cos(\theta_i^{(p)} - \theta_i^{(m)})] - V_i^{(m)} B_{p,m}[\sin(\theta_i^{(p)} - \theta_i^{(m)})] \\ - V_j^{(m)} G_{p,(m+3)}[\cos(\theta_i^{(p)} - \theta_j^{(m)})] + V_j^{(m)} B_{p,(m+3)}[\sin(\theta_i^{(p)} - \theta_j^{(m)})] \end{matrix} \quad (2.49)$$

2.8.2.1 Jacobian Matrix

To build the Jacobian matrix (2.47), the derivatives of the measurement equations are evaluated with respect to each state of the system in exactly the same way as for load flow. The derivative of equations (2.40), (2.41), (2.48) and (2.49) are shown in full in Chapter 8.

2.8.3 DSSE Measurements

The accuracy of a DSSE formulation is dependent on the quality of the measurements available in the power system. Both *real measurements* and *estimated measurements*

can be incorporated into a DSSE formulation. Estimated measurements can be sub-divided into two further groups known as *virtual measurements* and *pseudo-measurements*. To work accurately, a DSSE formulation must be fed with accurate estimates of the expected error distributions of real virtual pseudo-measurements.

2.8.3.1 Real Measurements

Real measurements cover all measurements taken by monitoring devices. This can include real and reactive power consumption, voltage and current. The standard deviation of a real measurement is dependent on the accuracy of a monitoring device and can be modelled using equation (2.1).

2.8.3.2 Virtual Measurements

Virtual measurements are used when the value of a measurement is known with almost near certainty. For example, if there are no loads connected to a busbar, then it is possible to say with certainty that the injected power at the busbar is zero. Virtual measurements can be modelled with a very low standard deviation (for example, 2×10^{-7} [69]).

2.8.3.3 Pseudo-Measurements

Pseudo-measurements are estimates for the value of a parameter based on some model of system behaviour. For example, it may also be possible to predict the load at certain busbars by building a model based on historical analysis. The electrical load profile follows a fairly predictable pattern which can be estimated through knowledge of external factors such as ambient temperature, weather conditions, time of year, day of the week, and the types of customers connected to a busbar. Voltages can also be predicted to within certain ranges if voltage control equipment is installed at a busbar. In this research pseudo-measurements are limited to estimates of the real and reactive power injected at a busbar.

A good pseudo-measurement should have a number of properties. A pseudo-measurement should add information to the state estimator to enable full system observability. Pseudo-measurements should have a low standard deviation, and a standard deviation which can be quantified. The error of a pseudo-measurement should (ideally) be uncorrelated with other pseudo-measurement errors, and follow a normal distribution (as the WLS / GLS procedure assumes that measurement errors are

normally distributed). It is also convenient if the pseudo-measurement can be easily estimated using a statistical model, or engineering judgement.

Constructing pseudo-measurements for three phase state estimation is further complicated by the intrinsic correlation in three phase systems. Correlated random variables can be used to take account of this problem [71, 73].

The reported percentage errors for single phase pseudo-measurement models in [69] was between 20% to 50%. This is equivalent to a standard deviation of between 7% and 17%. The multiple linear regression model discussed in [132] was able to estimate loading with an error of between 0.2% and 18.7% for any half hour interval. The knowledge based expert system in [132] predicted loading with an error of between 0.1% and 2.5%.

The pseudo-measurement models developed in [69] and [132] are single phase, and do not take into account the correlation between pseudo-measurements. In Chapter 8, the pseudo-measurement models will be developed in more detail to take into account the error correlation between measurements.

2.8.4 Building the Covariance Matrix R

The covariance matrix is different depending on whether the WLS or GLS estimation procedure is used. The covariance matrix contains the information relating to the accuracy of measurements used in a DSSE formulation.

2.8.4.1 Weighted Least Squares (WLS)

The covariance matrix for WLS is diagonal. The diagonal elements of the matrix **R** are described in equation (2.50).

$$r_{ii} = \sigma_i^2 \tag{2.50}$$

where r_{ii} is the ith diagonal element of **R** and σ_i is the standard deviation of the ith measurement. With WLS, all measurement errors are assumed to be independent.

2.8.4.2 Generalised Least Squares (GLS)

To build the covariance matrix for the GLS formulation, the correlation between pseudo-measurement errors must be taken into account. Real measurements are always assumed to be uncorrelated from one another. The correlative nature of pseudo-measurements can be explored through historical analysis of measurement data.

Full details on the construction of the covariance matrix for GLS can be found in Chapter 8.

2.8.5 Output from Distribution System State Estimation

The output of the distribution system state estimator is a set of voltages and angles at all of the buses in the network.

A distribution for the voltage and angle at a busbar can be obtained if the error properties of the measurements are known. The state estimator can be run repeatedly using a set of measurements which are taken from within the distribution of the measurement errors. This generates a set of solutions to the state estimator which can be plotted as a probability distribution. Care must be taken to ensure that any correlation between pseudo-measurement errors is also taken into account.

2.8.6 Summary of Assumptions for Distribution System State Estimation

The assumptions for DSSE include all of the assumptions discussed in the previous section for three phase load flow. The following additional assumptions must also be made:

- The error in the accuracy of all measurements (both pseudo and real) is assumed to follow a normal distribution
- Measurement error information for both pseudo and real measurements is assumed to be accurate

Additionally, if the WLS technique is used (as opposed to GLS), then the following extra assumptions must also be made:

- All pseudo-measurement and real measurement errors must be assumed to be uncorrelated

2.9 Summary

This chapter presented some of the fundamental modelling techniques which will be used throughout the rest of this thesis.

The chapter started by discussing model inputs and describing the error properties and availability of single and three phase measurements. The modelling concepts of three phase systems were then introduced by discussing technical aspects surrounding three phase voltages, currents and powers as well as introducing three phase models for the

power system. The chapter then discussed the models and assumptions surrounding voltage sags and unbalance modelling.

The chapter also introduced the equations required to conduct fault studies, perform load flow and estimate the three phase state of the system. These fundamentals will be used and built upon in the chapters which follow.

The following chapter expands on the modelling techniques developed in this chapter by introducing some of the advanced computational techniques which will be used in combination with these power system models.

3 Advanced Computational Techniques

3.1 Introduction

This chapter introduces some of the advanced computational techniques which have been used in this thesis to help identify of the worst served and weakest areas of the network. The algorithmic techniques described in this chapter include *classification*, *clustering* and *optimization* algorithms. These types of techniques can be loosely described as *artificial intelligence* (AI) techniques. Statistical methods are also used extensively through this thesis, and they are introduced at the end of this chapter.

Two important AI techniques are *supervised classification* and *unsupervised clustering*. Classification is supervised, as the data is pre-labelled into a series of nominal groups. The aim of a classification algorithm is to build a generalised model (through training) of a dataset. The classifier can then be applied to data points which have not been previously used for training, to test the classifier's performance. Clustering is described as unsupervised as it operates on datasets which are not pre-labelled into nominal groups. The purpose of clustering algorithm is to split a dataset into a set of groups (clusters) by maximising some measure of dissimilarity between data points.

One of the objectives of this research was to assess the suitability of *artificial immune systems* (AIS) for power systems research. AIS are a new computational technique. They are based on the metaphors gained by theoretical and empirical knowledge of the

natural invertebrate immune system. The algorithms selected for use in this thesis therefore include a number of immune inspired techniques.

The following clustering, classification, optimization and other techniques were used in the research presented in this thesis:

- Artificial Immune Recognition System (AIRS) for *classification* (immune inspired)
- Support Vector Machine (SVM) for *classification*
- k-Nearest Neighbours (kNN) for *classification*
- Random Forest for *classification*
- Unsupervised Artificial Immune Classifier (UAIC) for *clustering*
- B-Cell Algorithm (BCA) for *optimization* (immune inspired)
- Statistical analysis

The rest of this chapter describes the details of each of the algorithms and methods.

3.2 Classification Algorithms

The aim of a classification algorithm is to partition observations into a sub-population on the basis of a general set of rules which have been constructed (or learnt) by the classification algorithm upon exposure to a set of training data. The classification algorithm processes a set of training data typically constructed from past observations of important variables related to the specific field of study.

A key advantage of classification algorithms is their ability to generalise and discover important and complicated predictive relationships between variables which may not be immediately obvious to a human operative. Classification algorithms are also able to identify relationships across hundreds of dimensions, making them particularly relevant in power systems.

3.2.1 AIRS with kNN Classification

AIRS is an immune inspired instance creation algorithm [87, 96]. The objective of the AIRS algorithm is to generate a set of points in N dimensional feature space (memory cells) from exposure to a set of training data (antigens). The memory cells describe the dataset in a generalised form. Typically, at the end of training, the memory pool of AIRS contains 50% of the number of cells presented to AIRS at the outset of the algorithm [133].

The immunological inspiration for AIRS is largely based on the principle of the clonal selection that was first described in [134]. Clonal selection describes the process of how B-cells within the immune system become activated, proliferate and produce antibodies that target an invading antigen. During the process B-cells are cloned and then undergo somatic hyper-mutation [135] to produce antibodies which specifically target an invading antigen. Some of the B-cells also develop into long-lived memory cells which are capable of recognising the same antigenic stimulus long after a first exposure. A more detailed overview of the immune system and its uses in immune inspired algorithms can be found in [136] and [137] and specifically for AIRS in [96].

To classify data, AIRS requires the use of a classification algorithm [87]. The classification algorithm used in the original design of the AIRS algorithm (as described in [87]) was the k nearest neighbour classification algorithm. The k nearest neighbour algorithm was also selected as the classification algorithm in this research.

The following section provides a brief overview of the AIRS algorithm. More details on the AIRS algorithm can be found in [87], [138], [139] and [96].

The data that is presented to the AIRS algorithm comprises a set of labelled N dimensional feature vectors. In this research, AIRS is being used for supervised classification; therefore each of the feature vectors also has a class label. Each individual feature vector is modelled as an antigen. The set of all antigens is known as training data.

Along with training data, there are two other pools of cells which are modelled in the AIRS algorithm, namely, the *memory cell pool* (MC) and the *artificial recognition ball* (ARB) *pool*. MC becomes important at the end of the algorithm, as it represents the generalised model for the dataset. The ARB pool on the other hand, is a transient pool which is refined during the algorithm. An ARB itself is an abstract concept that represents a number of similar or identical recognition cells.

AIRS comprises the following 5 key processes. These are:

1) Initialisation
2) Training / learning
3) Competition for resources
4) Memory cell selection
5) Classification

3.2.1.1 Initialisation

Initialisation is a pre-processing step before the core of the AIRS algorithm. During initialization, the training data is normalised in the range [0, 1]. MC and the ARB pool are seeded during initialization by randomly selecting a number of instances from the training data. Lastly, the *affinity threshold* (AT) (shown in equation (3.1)) of the training data is calculated. This is simply the mean affinity within the training data.

$$affinity\ threshold = \frac{\sum_{p=1}^{T}\sum_{q=p+1}^{T} affinity(ag_p, ag_q)}{T(T-1)/2} \tag{3.1}$$

Where T is the total number of antigens in the training set and ag_p represents the pth antigen, and *affinity* represents a function that defines a measure of closeness or similarity between two antibodies or antigens. Small affinity values indicate strong affinity between two objects. Euclidean distance is often selected as the affinity measure in the AIRS algorithm [96].

3.2.1.2 Training / Learning

Training and learning is accomplished using several immune inspired processes, including clonal expansion and somatic hyper-mutation. Training takes place by using a one-shot process: AIRS only passes over the entire training dataset once. During training, each antigen is exposed to the memory cells one at a time, stimulating each cell within the memory pool. Each memory cell is allocated a stimulation value which is inversely proportional to its affinity with the training antigen:

$$stim(ag_i, mc) = 1 - affinity(ag_i, mc) \tag{3.2}$$

where ag_i represents training on the ith antigen, $mc \in MC$ is a memory cell. The memory cell which is stimulated the most is labelled mc_{match}.

Mutated clones of mc_{match} are generated (through the process of clonal expansion and somatic hyper-mutation) in proportion to the stimulation of mc_{match} with the presented antigen. The number of mutated clones ($NumClones_{mc_{match}}$) generated is governed by:

$$NumClones_{mc_{match}} = hyper\_rate \times clonal\_rate \times stim(ag_i, mc_{match}) \tag{3.3}$$

Both the hyper-mutation rate ($hyper\_rate$) and the $clonal\_rate$ are user defined parameters. The mutated clones are then added to the ARB pool.

3.2.1.3 Competition for Resources

The ARBs within the ARB pool then undergo their own process of clonal expansion and somatic hyper-mutation under a resource limited environment. This process is known as competition for limited resources[87]. During this process, mutated clones are generated in proportion to the stimulation of each $ab \in AB$ to ag_i:

$$NumClones_{ab} = stim(ag_i, ab) \times clonal\_rate \quad (3.4)$$

Resources are also allocated to each of the $ab \in AB$ in proportion to the stimulation of the ab with ag_i. The total amount of resources in the ARB pool is limited by the user defined configuration parameter known as $total\ resources$. After each of the ARBs in the ARB pool have been stimulated by ag_i, the least stimulated ARBs are removed from the ARB pool by allocating resources to the most stimulated ARBs in the pool first until all the resources have been allocated.

The development of the ARB pool is finished when the mean user defined $stimulation\ threshold$ reaches a pre-defined level.

3.2.1.4 Memory Cell Selection

The penultimate step in AIRS is memory cell selection. Memory cell selection selects the best antibodies from the ARB pool to potentially join MC. The most stimulated ARB in the ARB pool is called $mc_{candidate}$. If $mc_{candidate}$ is more stimulated than mc_{match}, a check is made to see if mc_{match} should be removed. mc_{match} will be removed if the $affinity(mc_{match}, mc_{candidate})$ is less than $AT \times ATS$. Otherwise, $mc_{candidate}$ will be added to MC.

3.2.1.5 Classification

The final stage of the AIRS algorithm is to classify the dataset. This is achieved using an instance based classification algorithm such as k nearest neighbour (which is described later in this chapter).

3.2.2 More Information on AIRS

More details on the AIRS algorithm can be found in [87], [138], [139] and [96].

3.2.3 SVM Classification Algorithm

Support vector machines (SVMs) are a classification technique based on statistical analysis. They aim to separate a N dimensional continuous feature space into two

classes. SVMs represent this problem by formulating a separating hyperplane. This hyperplane can then be represented as shown below:

$$\mathbf{0} = \boldsymbol{w}^T\boldsymbol{x} + \boldsymbol{b} \tag{3.5}$$

where $\boldsymbol{w}$ is a N dimensional vector of weights, $\boldsymbol{b}$ is a bias vector and $\boldsymbol{x}$ is a point in N dimensional space. Depending on whether a point lies above or below this plane determines to which class the point belongs.

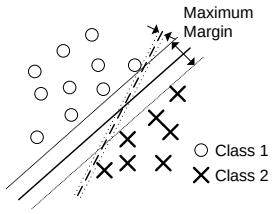


Fig. 3.1 - Two possible linear discriminants to separate the circles from the crosses

Fig. 3.1 shows the construction of two hyperplanes that separate the linearly separable classes: class 1 and class 2. The aim of the SVM is to construct a maximum margin hyperplane that separates the two classes. The construction of the maximum margin hyperplane can be formulated as an optimisation problem.

The SVM can be easily adapted to work for non-linear classification problems. In this instance, the training examples are mapped into a higher dimensional space and a maximum margin linear hyperplane constructed to classify the data. If $\theta(\boldsymbol{x})$ defines a mapping such that $\theta(\boldsymbol{x})\colon \mathbb{R}^n \to \mathbb{R}^{n'}$ where $n' \gg n$ the support vector machine dual optimisation [140] becomes:

$$\begin{cases} \min\limits_{\alpha} \left[\frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{k} \propto_i \propto_j y_i y_j \theta(x_i)\theta(x_j) - \sum_{i=1}^{k} \propto_i\right] \\ s.t \ \sum_{i=1}^{k} y_i \propto_i = 0 \\ \propto_i \geq 0 \ for \ i = 1 \ldots k \end{cases} \tag{3.6}$$

where $\propto_i, \propto_j$ are Lagrange multipliers and $\boldsymbol{x} \in \mathbb{R}^k$.

The inner product between certain sets of two functions can be calculated without ever knowing the individual mapping. This leads to a replacement of $\theta(x_i)\theta(x_j)$ with $K(x_i, x_j)$ where $K(x_i, x_j)$ is a kernel function. When the data is not linearly separable, then a penalization parameter C is introduced. The parameter C represents the trade-off between minimizing the training set error and maximizing the margin.

There are several choices of kernel function available. The polynomial kernel and the radial basis function are two such examples. It has been shown that given sufficient tuning (of its parameters) the polynomial kernel function is a special case of the radial

basis function [141]. The kernel function selected for this study is therefore the radial basis function.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{3.7}$$

The parameter σ affects the above kernel function and also influences the ability of the support vector machine to accurately classify data.

The choice of values for C and σ will affect the classification performance of the classifier. To obtain a high level of classification accuracy, C and σ should be chosen carefully and optimised where possible. The other main variable in the SVM is the kernel function.

Support vector machines are often used for two class classification problems. However, the method can easily be extended to multi class problems by considering a multi-class problem as a set of two class classification problems.

More information on SVM algorithms can be found in [140].

3.2.4 Random Forest Classification Algorithm

A Random Forest is a classifier consisting of a collection of tree classifiers where each tree classifier casts a unit vote for the most popular class given an input **x** [89].

Decision trees are a nonparametric learning technique. They are built by creating sets of tests that split the data between classes. For example [142], a test T may be defined as:

$$T: x_i < t_i \tag{3.8}$$

Where $\boldsymbol{x} \in \mathbb{R}^n$ and n is the number of features in a voltage profile, x_i is the voltage at the ith bus and t_i is the optimal threshold to split the feature x_i into two classes. In the random forest algorithm, n input features are selected for each of the N random tree classifiers. The decision trees are then grown by sampling P cases from a set of P observations in the training set with replacement. Each tree is grown to the largest extent possible with no pruning.

Both the total number of features per tree (n) and the total number of decision tress generated (N) are parameter configuration variables which must be selected carefully to ensure the random forest algorithm works effectively.

The random forest algorithm has several key advantages appropriate for this study. Firstly, the algorithm has been shown to have high classification performance compared with other algorithms. Secondly, they can run efficiently on large datasets. Thirdly, it is possible to interrogate the random forest to identify the most important features of the

classifier and hence identify important monitors and measurements. Lastly, they are robust against over-fitting.

3.2.5 k Nearest Neighbours (kNN) Classification Algorithm

The k nearest neighbours algorithm (kNN) works by performing a majority vote on the k closest training examples to a test vector in the feature space [143]. The class assigned to a test vector is the most prominent class contained within the set of the k nearest neighbours (as defined by some distance metric).

The kNN algorithm was included in this study as a simple comparison to the support vector machine and the AIRS classification algorithm. It is especially relevant as a comparison with AIRS since they both share the same procedure for classification.

3.3 Clustering

The aim of a clustering algorithm is to assign a set of data points into groups (called clusters) so that the data points within each group are more similar to each other than other data points outside of these groups.

The similarity between groups of data points can be customised depending on the target problem. The similarity measure should be defined in such a way so that the distance between similar items is relatively small, and the distance between dissimilar items is relatively large. Appendix F describes the application of a clustering algorithm to a transmission network, and a custom distance measure is used to group similar data points.

3.4 Unsupervised Artificial Immune Classifier (UAIC)

The unsupervised artificial immune classifier (UAIC) algorithm is an unsupervised immune inspired algorithm that is capable of performing clustering on a dataset. The algorithm was proposed in [91] to classify areas of hyper-spectral remote sensing images into areas describing land use: for example road, town, vegetation or sea. It is largely based on ideas first presented in the AIRS algorithm in [87]. UAIC differs from AIRS being an unsupervised rather than supervised algorithm. The full description of the UAIC algorithm can be found in [91].

3.4.1 Immune Components of the UAIC

The UAIC algorithm is based on the principle of immune clonal selection [134]. Like AIRS, the significant immunological inspiration for the UAIC comes from the idea of

clonal selection [134]. The processes of clonal expansion and somatic hyper-mutation [135] are both modelled within the algorithm along with the concept of a long lasting population of memory cells.

The UAIC also exhibits a form of immune system metadynamics [144]. This process is modelled in the algorithm as the continuous changing of the antibody population through antibody proliferation and death. In the UAIC, low affinity antibodies are continuously removed from the population to ensure that only the best selection of antibodies is maintained.

Antigens, antibodies and memory cells are all modelled in UAIC as feature vectors in N dimensional feature space. Antigens are input into the algorithm from a dataset. Antibodies are cloned and mutated during the UAIC algorithm and represent candidate clustering centres. Memory cells are selected as the final locations which most accurately represent the dataset, based on rules defined within the UAIC algorithmic process.

3.4.2 UAIC Algorithmic Process

The clustering procedure of the UAIC algorithm is configured to find a pre-defined number of clusters (N_c) in a dataset.

The UAIC algorithm initializes by generating a memory cell population (*MC*) using the Kaufman initialization procedure [145] which generates memory cells sequentially. The first memory cell is assigned to the location in the centre of the most number of antigens. The subsequent memory cells are placed to minimize the distance between the remaining antigens and the set of memory cells. This process continues until N_c memory cells have been positioned. Each memory cell in the set of cells is assigned a different label corresponding to the cluster it represents. The antibody pool *AB* is also initially seeded with the contents of *MC*.

After initialization, the algorithm enters the processing phase. The first stage of the processing phase is to present the antigens to the algorithm. During this procedure, the order of the antigens is randomized and then presented to the algorithm sequentially. The following steps are then performed on each training antigen:

1) **Antigen Presentation:** Each antigen is assigned to the cluster of the nearest memory cell based on a calculation of affinity with all other memory cells:

$$mc_{match} = arg\min_{mc \in MC} affinity(ag_i, mc) \quad (3.9)^{10}$$

where $mc \in MC$ is a memory cell from the population of memory cells MC, ag_i is the ith presented antigen and $affinity(ag_i, mc)$ is a measure of similarity (or closeness) between the ith antigen and the selected memory cell. $mc_{match,C}$ is the best matching memory cell of the cluster C.

2) **Clonal Expansion:** The antibody population of the antigen's cluster then undergoes a form of clonal expansion. A number of strong affinity antibodies are selected and are cloned rapidly. The number of clones generated depends on the clonal rate and the affinity between the antigen and the antibody.

$$C_l = \sum_{i=1}^{N} round(\frac{Cr}{affinity(ag_{i,C}, ab_C)}) \quad (3.10)^{11}$$

Where C_l is number of clones generated, Cr is the clonal rate defined for the algorithm and $ab_C \in AB_C \subset AB$ is an antibody from the population of antibodies AB labelled with the same cluster C as $ag_{i,C}$.

3) **Affinity Maturation:** The clones are then subjected to affinity maturation. They are mutated at a rate inversely proportional to their affinity with the training antigen. Stronger affinity antibodies are mutated less than those with a weak affinity.

4) **Memory Cell Selection:** Each of the matured antibodies in the pool is then evaluated against the original best classifying memory cell to decide if the new memory cell should be accepted for inclusion into the memory pool. The best new memory cell is found by searching the pool of antibodies:

$$mc_{candidate} = \arg\min_{ab_C \in AB_C} affinity(ag_{i,C}, ab_C) \quad (3.11)$$

Memory cell selection requires the use of the $affinity\ threshold\ scalar$ (ATS) which is a user configured parameter for the UAIC algorithm and the $affinity\ threshold$ (AT) which is defined as described in equation (3.1). If $mc_{candidate}$ is a better match (has a smaller affinity) to the antigen than mc_{match} then $mc_{candidate}$ is considered for inclusion as part of the memory cell population (MC). If this is the case, $mc_{candidate}$ is added to the memory cell population if $AT \times ATS < affinity(mc_{match}, mc_{candidate})$ otherwise, if $AT \times ATS > affinity(mc_{match}, mc_{candidate})$ the memory cell will replace mc_{match}.

[10] In [91], this is a maximization; however it is defined as a minimization here to align with AIRS.
[11] In [91], this is defined as a multiplication, but it is defined as a division here to align with AIRS.

5) **Displace Antibody Pool:** The antibody pool is pruned and regulated by using a displace rate which displaces a percentage of the worst matching antibodies from the pool. D percent of the worst matching antibodies from the antibody pool (AB_C) are replaced with the same number of best matching memory cells (of the antigen's class).

   After training on all the antigens is complete, a termination condition is checked and if it is not satisfied, the algorithm will restart and begin training across the set of antigens once again. The algorithm terminates when the number of antigens changing clusters, from one iteration to the next, reaches a minimum threshold value.

There are several configuration parameters in the UAIC algorithm. The $affinity\ threshold\ scalar$[12] is used to decide whether a newly created antibody will be added to the memory pool (as shown in step 4 below). The clonal rate Cr defines the number of cloned antibodies injected into the antibody pool during each training iteration. The displace rate D defines the percentage of antibodies that will be displaced on each iteration (as shown in step 5 below). The defined $affinity$[13] (or $distance$) function is also a user-defined function and it is often tailored to the specific clustering problem [91].

3.4.3 Similarity between the UAIC Algorithm and AIRS

There is a large degree of similarity between the core algorithmic procedure of UAIC and the AIRS algorithm. AIRS was the first algorithm to be developed, and the UAIC algorithm was developed from ideas present in AIRS and other algorithms.

Both algorithms clone and mutate a set of antibodies which are ultimately selected to become part of a memory cell population using identical procedures. Both algorithms also select antibodies to become part of the memory cell population by considering the $affinity\ threshold$ of the dataset as a whole. The output from both algorithms is also a set of memory cells which generalises the training data.

The two algorithms differ most significantly in initialization and termination. AIRS initializes by randomly seeding a number of memory cells from the dataset while the UAIC uses the Kaufman initialization procedure. AIRS terminates after one run over the dataset whilst UAIC is configured to terminate when the number of antigens

[12] In the original implementation of UAIC, a distance threshold scalar (DTS) is defined, rather than an affinity threshold scalar (ATS). ATS was described in this research to enable a straightforward comparison between UAIC and AIRS algorithm (which uses an ATS).

[13] The original implementation of UAIC referred to affinity as "inversely proportional to distance". AIRS defines affinity in proportion to distance. The description of UAIC in this paper keeps with the definition of affinity as outlined in the AIRS algorithm.

changing clusters between iterations reaches a pre-configured value i.e. after several iterations.

Furthermore, the AIRS algorithm is often applied as a pre-processing algorithm used prior to supervised classification by another algorithm (for example k nearest neighbours). The UAIC algorithm on the other hand, is designed specifically for unsupervised clustering, and cannot be used with supervised data.

TABLE 3.1. COMPARISON OF UAIC ALGORITHM AND AIRS

Process	Difference or Similarity	UAIC	AIRS
Initialization	Different	Kaufman initialization procedure.	Random selection of antigen from training data.
Core Algorithm Process	Similar	Training / learning, resource consolidation, memory cell selection.	
Termination	Different	User configured based on the number of antigens switching cluster per iteration.	One-shot algorithm. Terminates after one run over the dataset.
Data Representation	Similar Ideas	Antigens, memory cells & antibodies (or ARBs in AIRS).	
Mutation	Similar	Immune inspired somatic hyper-mutation.	
Cloning	Similar	Immune inspired clonal expansion.	
Memory Cell Selection	Similar	Use affinity threshold, user configured affinity threshold scalar and candidate and best matched memory cells.	
Typical Usage	Different	Clustering / Unsupervised Classification Only	Classification
Competition for Resources	Different	Antibody are displaced using a displace rate.	Resource limited artificial recognition ball (ARB) pool.
Consolidation of Resources	Similar Ideas	Memory cells with identical session data are merged to limit the memory cell population growth.	The artificial recognition ball (ARB) model to ensure identical antibodies aren't represented twice.

3.5 Optimization

3.5.1 The B-Cell Algorithm

The B-cell algorithm [146] is an immune inspired optimisation algorithm. The algorithm is inspired by the immunological process of clonal selection [134]. The immune system itself is able to track a series of ever changing pathogenic material that enters the body. The B-cell algorithm has been shown to perform strongly when compared with other immune inspired optimisation algorithms on problems that involved dynamic optimisation in [147]. In [148] it was compared against a hybrid genetic algorithm (HGA) and shown to be more effective at tracking fixed points in a changing landscape. The B-cell algorithm was also compared against a HGA in a static

environment in [146], and shown to outperform the HGA in terms of number of evaluations.

The B-cell algorithm is unique in that it operates using a unique mutation operator known as contiguous somatic hyper-mutation. The inspiration for this operation comes from two immunological research sources in [149] and [150].

In the original B-cell algorithm, each B-cell is encoded as an N-dimensional vector of 64–bit strings ($\boldsymbol{v}$). By encoding each solution in this manner, the B-cell algorithm is able to perform continuous function optimization for an objective function with any number of dimensions. Each B-cell belongs to the population of B-cells, P ($\boldsymbol{v} \in P$). The objective function of the chosen optimization problem can be evaluated for any of the B-cells, by computing $f(\boldsymbol{v})$.

An immune inspired optimization algorithm has particular advantages when considering the optimal monitor placement problem. Immune inspired algorithms such as the BCA are able to maintain a diverse range of near optimal solutions [92]. This is not possible with traditional combinatorial optimization techniques which tend to locate one single optimum solution. This will be explored and exploited in more detail in Chapter 7.

There are two key advantages to identifying a range of solutions. Firstly, the optimum solution may be extremely sensitive to small changes in the operating conditions of the monitoring solution. A near optimal solution may be more robust when small changes in operating conditions are considered. The second advantage of a diverse range of solutions is that it adds flexibility into the decision making process. A DNO can choose from a wide range of similarly effective solutions before adopting a monitoring solution. The final selection could be based on additional criteria which may not be directly correlated with sag or unbalance monitoring while still achieving high accuracy of power quality performance assessment.

The B-cell algorithm has been shown to perform strongly when compared with other immune inspired optimization algorithms on problems that involved dynamic optimization in [147]. In [148] it was compared against a hybrid genetic algorithm (HGA) and shown to be more effective at tracking fixed points in a changing landscape. The B-cell algorithm was also compared against a HGA in a static environment in [146], and shown to outperform the HGA in terms of number of evaluations.

The algorithm begins by seeding a random number of B-cells, as defined by the user-configured total population size (P_{max}). The objective function for each of the B-cells is then evaluated. Each of the B-cells is then cloned C_{max} times and a clonal pool C_i is created for the ith B-cell. C_{max} is a user-configured variable and is often set to be equal to P_{max}. To maintain a level of diversity in the candidate solutions (and to mirror the immune process of metadynamics), one B-cell clone in each of the clonal pools is selected and each element is mutated randomly, subject to a probability.

The next process on the clonal pool involves mutation. Each B-cell in the clonal pool is mutated by using contiguous hyper-mutation. This mutation operation works by first selecting a random location (loc_r) on the vector to be mutated. A random length is then chosen between 1 and the length of the vector (len_r). These two values define a region on the B-cell that will undergo the mutation operation as shown in Fig. 3.2.

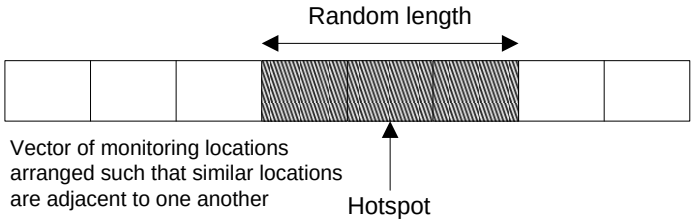


Fig. 3.2. The contiguous somatic hyper-mutation operation.

The B-cell algorithm terminates when a stopping criterion is fulfilled. The stopping criterion used is problem dependent.

3.6 Statistics

Statistical techniques cannot be defined as classification, clustering or optimization algorithms. They are mentioned here as they are used extensively in this thesis to intelligently process monitoring data. They are therefore similar to data mining or machine learning techniques, as they help power system operators extract knowledge from information associated with their network.

The application of statistical techniques often has advantages over more complicated machine learning and artificial intelligence algorithms. Statistical methods are often simpler to implement, and they require no specialist knowledge of the techniques to interpret the results. They are also widely applicable to a variety of problem areas, and backed up by rigorous mathematical research.

Statistical methods are used extensively throughout this thesis. For example, in Chapter 5 they are used to formulate a probabilistic estimate for the location of a voltage sag, in Chapter 8 they are used statistically estimate the distribution of voltage unbalance.

3.7 Algorithm Implementation: Weka Machine Learning Environment

All of the classification algorithms applied in this research were tested using WEKA machine learning environment [151]. The UAIC clustering algorithm and the B-Cell optimisation was implemented using bespoke code developed in MATLAB by following the algorithmic procedure outlined in [91] and [146] respectively. Statistical methods were also implemented in MATLAB using standard toolbox functions.

3.8 Summary

This chapter described the core computational techniques which will be used to help identify the worst served customers and weakest areas of the network for power quality issues.

The selected techniques included classification, clustering, optimization and statistical analysis. Particular emphasis was placed on selecting immune inspired artificial immune systems (AIS). AIS algorithms were selected so that they could be evaluated against other methods.

A summary of the reasons for selecting each of the algorithms is described in Table 3.2.

Supervised classification techniques are used in Chapter 4 to detect and classify voltage sag events in distribution networks and also in Appendix F to classify voltage stability issues and overloaded lines in transmission networks. In both case studies, the performance of the AIS algorithms was compared against other more standard techniques, to try and discover whether they have any particular advantage for power system problems. Appendix F also describes the application of an immune inspired *unsupervised clustering* algorithm. Clustering is a powerful technique which can be applied to data which is unlabelled. By grouping similar modes of failure, the clustering algorithm is able to generalise scenarios which will fail in a similar manner.

Immune inspired *optimisation* algorithms are applied in Chapter 7 to optimally place monitors for voltage sag performance monitoring. In Chapter 7 the immune inspired techniques were compared against other optimisation results and their performance evaluated.

TABLE 3.2. AN OVERVIEW OF THE ADVANCED COMPUTATIONAL TECHNIQUES USED THROUGHOUT THIS THESIS

Name	Application	Description	Reasons for Selection
AIRS	Classification	Immune inspired classifier	• Popular immune inspired algorithm • AIRS has been shown to outperform the kNN classifier [133] • Shown to be an effective classifier across a variety of data sets [96]; outperforms learning vector quantisation in [133] • Limited applications in power systems; exceptions include [83]
SVM	Classification[14]	Popular classifier based on statistics	• Statistical and mathematically based approach • Easy to visualise their operation • Widespread implementations available in environments such as MATLAB & Weka[151] (with LibSVM [152])
kNN	Classification	Popular simple classifier	• Simple to understand and implement • Very popular • Widespread implementation in machine learning environments such as MATLAB & Weka[151] • Same classifier as the AIRS algorithm, so useful for performance comparisons with AIRS
Random Forest Algorithm	Classification	Strongly performing widely implemented classifier based on decision trees	• Run effectively on large datasets • Robust against over-fitting • High classification performance against other algorithms • Widespread implementation in machine learning environments such as MATLAB & Weka[151]
UAIC	Clustering	Immune inspired clustering algorithm based on AIRS	• Similar immune inspiration to AIRS; AIRS equivalent clustering algorithm • Straightforward and documented implementation [145]

B-Cell Algorithm	Optimization		• Population based optimisation able to maintain a set of sub-optimal solutions
			  • Creates a robust set of solutions
			  • Creates flexibility in the decision making process
			• Performs strongly on dynamic optimisation problems [147]
			• Shown to outperform other population based techniques such as a hybrid genetic algorithm [146]

[14] Support vector machines can also be used for linear regression and clustering.

4 Voltage Sags: Detection and Classification

4.1 Introduction

Gaining a holistic view of voltage sags within distribution networks can be considered a multi-stage procedure as shown in Fig. 4.1:

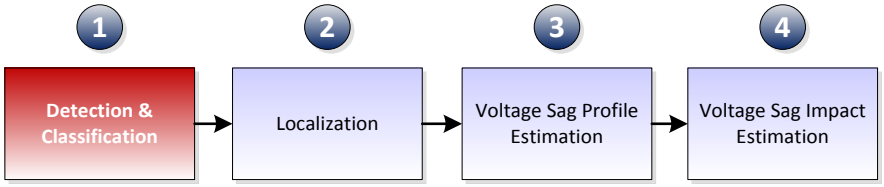


Fig. 4.1. The process of gaining an overview of voltage sags within a power network (stage 1).

Detection and classification is the process of detecting that a voltage sag has occurred, and then subsequently classifying the type of sag which has occurred. Localization is the process of identifying the region where the sag occurred. Voltage sag profile estimation synthesizes the results of the earlier stages into a during fault voltage profile for the whole of the network. Voltage sag impact estimation is the process of identifying the impacts of the voltage sags on each of the consumers in the network.

Algorithms must be developed to deal with each of the stages described in Fig. 4.1. The developed algorithms must be able to synthesize information from a range of arbitrary monitoring devices to produce reliable outputs that can be fed into the next stage of the process.

The following three chapters describe a series of techniques and case studies to tackle each of the stages defined in the process of Fig. 4.1. This chapter covers the detection and classification of voltage sag events using artificial intelligence algorithms to assist with the detection process (as shown as the highlighted box in Fig. 4.1).

4.1.1 Aims

Voltage sag detection, localization and estimation have all been tackled before. One of the most important works on this topic was [25], which tackled the optimal monitoring problem for voltage sag detection, localization and estimation using monitors as threshold monitoring devices. In [25], it was assumed that each power quality monitor is able to register a fault if the voltage magnitude if the worst affected phase descended below a pre-determined threshold level. This led to the development of the concept of a monitor reach area (MRA) for each monitor. Monitoring programs designed in this way assume that power quality meters act autonomously, each covering their particular area of the network. Reference [25] also assumed that the type of fault (e.g. three phase or double line to ground) that had caused the voltage sag event was known a priori. Determining the type of fault is an important requirement of the detection stage and required to achieve accurate localization and estimation.

Classifying the type of fault before performing fault location or voltage sag profile estimation is an important aspect which is absent from the existing body of research relating to voltage sag estimation, but covered as part of broader research on fault classification and detection (in papers such as [88] and [89] which were described in the introduction). Fault classification and detection is a necessary pre-requisite before the voltage sag estimation algorithms which are discussed in Chapters 5 and 6 can be applied to locate faults and estimate their impact on the network.

In contrast to [25], the research developed in this chapter is a centralised technique. It is assumed that data from monitoring devices can be assimilated to a central location, and then processed together to determine the type of fault.

The principle aim of the research described in this chapter is to develop a method capable of detecting and classifying faults using a realistic set of erroneous monitors installed at practical locations whose number is much less than required to observe the whole network. The objective of the voltage sag classification process is to group together faults into similar types; for example, single line to ground (SLG), line to line

(LL), double line to ground (LLG) and three phase (LLL), using a set of measurements available from realistic monitoring devices.

A secondary aim of the research described in this chapter is to test and compare the performance of the developed algorithm whilst varying the internal classification algorithm used to classify the sags.

A high performing algorithm will exhibit a number of qualities. The algorithm should be able to continually discriminate between faulted operating scenarios and normal power system operation whilst minimizing the number of false alarms and missed faults. Both the detection and classification accuracy of the algorithm must be near perfect, to ensure that the localization, estimation and impact estimation phases of the voltage sag monitoring process are not erroneous. The algorithm must have a high immunity to noise, and be able to work with ad-hoc arrangement of monitors installed at different locations potentially measuring different parameters with varying degrees of accuracy.

The research discussed in this chapter contributes and adds to the existing body of research on fault location, classification and voltage sags by providing a viable method based on artificial intelligence algorithms to identify the type of fault which occurred in the network. The research rigorously tests a variety of established artificial intelligence algorithms, and identifies which algorithms are most likely to provide promising results. The developed method can be considered as a pre-processing stage to the techniques discussed in [25], [81] and the algorithms developed later in this thesis. Unlike [25] and [81], the method's discussed in this chapter (and this research) are shown to be capable of handling noisy measurements, which is a key requirement of a practical monitoring solution. It is also hoped that the results presented in this chapter will highlight the suitability of artificial immune systems for the task of voltage sag classification and detection.

4.2 The Aim of a Voltage Sag Profile Classifier

The aim of the voltage sag classifier is to partition measurements recorded by monitors into five separate sub-populations:

1) Normal (N)
2) Single line to ground (SLG) fault
3) Double line to ground (LLG) fault
4) Line to line (LL) fault

5) Three phase (LLL) fault

Four types of classification algorithm were used to build classifiers capable of detecting steady state operating conditions and classifying the type of fault; namely, an artificial immune recognition system [96] (AIRS), a support vector machine[140] (SVM), a random forest (RF) [89] and k-nearest neighbour (kNN) [143].

Artificial immune recognition system (AIRS) was selected for its consistent performance across a wide range of classification problems [96]. A *support vector machine* (SVM) and the *random forest algorithms* were selected due to their success in classifying data in other power system related problems, such as [142], where they were used to identify critical transmission lines and their proper compensation rate with respect to voltage stability. The random forest in particular was selected because its proven success of classifying faults as LL, SLG, LLG and LLL in [88] with near 100% accuracy using measurements taken from one of a transmission line. The *k-nearest neighbour* (kNN) was selected primarily because it is a popular and conceptually simple classifier which often yields comparable results to more complicated methods.

It is particularly important to compare the kNN algorithm's performance with AIRS because AIRS can be considered as a specialisation of the kNN algorithm. AIRS principle function is to pre-process a data set before presenting it to the kNN algorithm for classification. If AIRS is enhancing the classification process, the classification performance of AIRS should be greater than that of kNN alone.

4.3 Method

The aim of the study is to develop and test the performance of an algorithm capable of detecting and classifying voltage sags. The test methodology is summarised in Fig. 4.2.

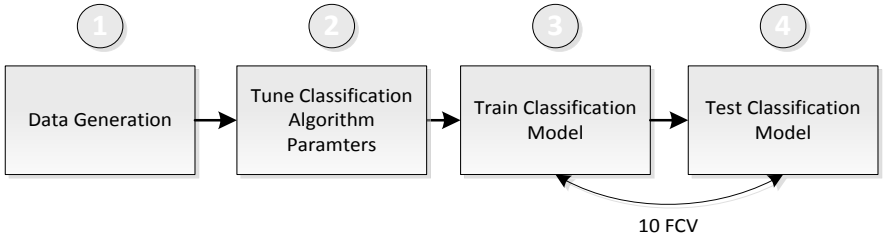


Fig. 4.2. Methodology for testing detection and classification performance.

The first step of the methodology is data generation. During the data generation phase, a dataset was constructed which covers a realistic set of inputs and therefore incorporates variation in pre-fault voltage, fault impedance, fault location, measurement accuracy and number of measurements.

In the second step of the methodology, the classification algorithms are tuned. Each of the algorithms has a variety of configurable parameters and these must be selected in such a way to yield the best possible performance.

The third step of the testing methodology is to build a classification model. The classification model was built by feeding in a training data set into the classification algorithm.

The last step is to classify the test data set into the trained classification algorithm and record the results. The classification algorithm was tested using a technique known as ten-fold cross validation [154].

4.3.1 Data Generation

The input for the classification algorithms was prepared during the data generation phase by generating a fault dataset of faulted and non faulted scenarios [8]. The output of the data generation phase is a *tuning dataset* and *testing dataset* which will be used to tune and then subsequently train and test the classifiers.

Each dataset must incorporate variation in pre-fault voltage, fault location, fault impedance, measurement accuracy, measurement location and the number of measurements. A series of fault studies were conducted to incorporate the variation in each of these parameters.

Pre-fault voltages were varied by creating by creating 6 pre-fault loading scenarios derived from loading the 295 bus network at 50%, 60%, 70%, 80%, 90% and 100% of the maximum expected half hourly load. The load was raised and lowered evenly across all buses ensuring that the power factor remained constant. For each fault study, the load was picked from one of these 6 scenarios.

Fault location was varied by simulating faults at 4 fault positions along each of the 278 lines in the network. Each fault was simulated at a per unit distance of either 0.01, 0.337 0.663 or 0.99 along each line. Fault location was configured as a fixed series of locations (rather than random locations) to ensure that the results evenly covered all areas of the network.

Fault impedance was modelled as a uniformly distributed random variable between 0Ω and 100Ω. For each fault simulation, the fault resistance was selected from this distribution.

The accuracy of the installed measurement devices depends on the actual devices installed within the network. At monitored buses, voltages are measured through a voltage transformer. The accuracy of the measurement will vary depending on the quality of the monitoring device[15].

Measurement errors were modelled using the expression presented in Chapter 2, equation (2.1). In this research, an error of 0.2% of the true value was assumed for all symmetrical component magnitudes and 0.2% of 2π for phase measurements. The standard deviation for the voltage magnitude and phase was therefore $\sigma_{|V|}=0.067\%$ and $\sigma_{\phi}=0.0013\pi$ radians respectively.

Both the locations for the monitors and the numbers of monitors were selected by considering a mixture of previous research defining optimal monitoring locations for voltage sag monitors [25] in a distribution network and engineering judgement. Left unconstrained, the number of possible locations for monitoring devices in the network is a combinatorial problem with a high number of possible solutions (for example, 5.4×10^{80} ways of placing 100 monitors into the 295 bus network). It is therefore important to constrain monitors to positions representative of locations where monitoring devices are likely to be installed.

The optimal monitoring research discussed in [25] presents a method for placing monitors within a network based on a pre-defined voltage threshold. In this research, a set of 12 monitors known as the MRA monitor set was found by applying the method presented in [25] to the 295 bus network and setting the voltage threshold to 0.9 per unit. This is the minimum number of monitors required to detect the presence of all faults in the 295 bus network with 100% accuracy (assuming perfect error free measurements) using the threshold method. Fault detection is accomplished using the method presented in [25] by simply noting whether the voltage in the worst affected phase dips below 0.9 per unit. The 12 monitor MRA solution is shown in Fig. 4.3 with monitors installed at buses 289, 288, 287, 286, 285, 247, 63, 179, 240, 38, 121 and 137. The ordering of the MRA list of monitors ranks buses in ascendance of the number of faults that they can each independently observe. Twelve different sets of monitoring locations were selected from the MRA set, each consisting of the last *N* monitors in the list.

[15] The Siemens 9610 Power Quality meter for example, quotes an accuracy of ±0.2% for symmetrical component magnitude and phase at full scale of 0 to 106volts and $-\pi$ to π.

Although 12 monitors are required to detect all SLG, LL, LLG and LLL faults, only 4 monitors at buses 137, 121, 38 and 240 are required to detect LLL faults alone. This is an important point which will be drawn upon in the analysis of results section which follows later.

In the case where a distribution network operator does not utilise an optimal monitor placement algorithm, monitors are likely to be installed at accessible locations based on the engineering judgement of operators. It would be reasonable to assume that monitors would be installed at the start of 11kV and 33kV feeders on the secondary side of a transformer, as existing SCADA single phase monitors tend to be installed in these locations. In the 295 bus network, there are 10 33kV and 11kV feeders. The results that follow also include a comparison of the classification and detection performance of the algorithms using these 10 monitors installed at buses.

The full *testing dataset* contains 10008 voltage magnitudes and angles for each type of fault, giving a total of 50040 fault voltage profiles. The 10008 voltage profile consists of 9 randomly selected combinations of pre-fault loading, fault resistance simulated at 4 locations across all 278 lines. The *testing dataset* was further reduced to 5004 profiles, by randomizing the data set and selecting 10% of the data. The dataset was reduced to ~5000 instances to ensure that training the classification algorithms within the Weka machine learning environment could be carried out within a feasibly short length of time (<1 hour on a Pentium 4 2.8GHz PC).

The *tuning dataset* contains 100 randomly selected voltage profiles for each type of fault, giving a total of 500 fault voltage profiles.

Each of the instances contained within the training and tuning dataset were labelled with SLG, LL, LLG, LLL or N, depending on the type of fault simulated.

4.3.2 Optimising the Parameters of the Classification Algorithms

Each of the classification algorithms have parameters which can assume a variety of values. These parameters influence the accuracy of the classification algorithm's performance, and it is therefore important to tune the algorithms to ensure that the selected parameter values are optimal, before building the classification model across the whole dataset. The algorithmic parameters were tuned by running the algorithm across the *tuning dataset* comprising 500 points, with 100 points selected randomly from each of the 5 fault classes.

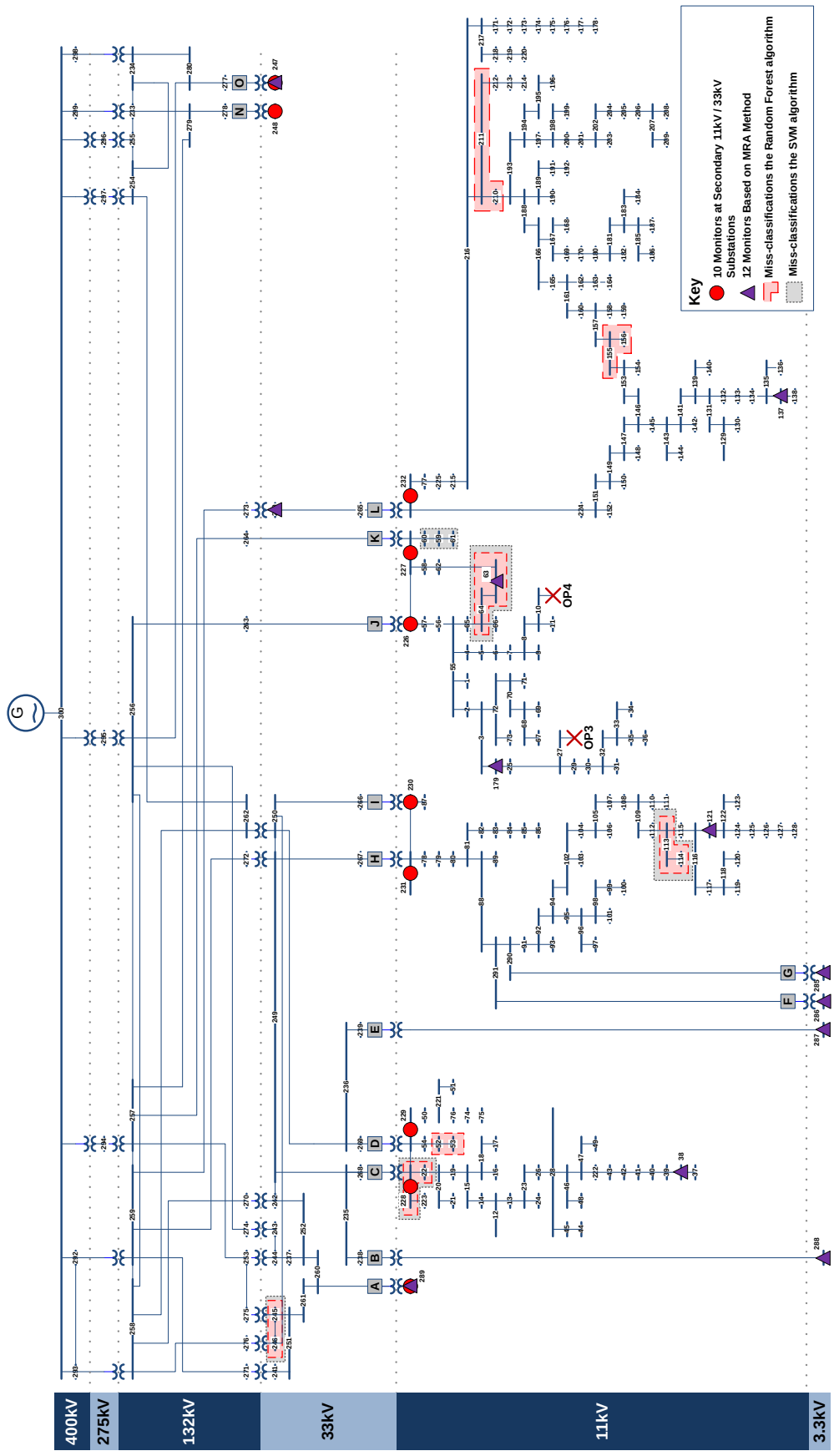


Fig. 4.3. The 295 bus network showing the monitor locations which will be used to test the classification and detection performance of the algorithms. The two highlighted zones represent regions where faults were miss-classified by the random forest and the SVM classification algorithms.

The optimal parameter values for each algorithm were determined by defining a series of ranges and intervals for each parameter, and varying all parameters until every possible combination of parameter had been searched. For each set of parameter values, a classification model was built and tested using 10 fold cross validation. The best set of parameter values was then noted, ready to be used for classification on the *testing dataset*.

The following sections describe the parameter selection process for each of the 5 algorithms.

4.3.2.1 Optimising the Parameters of the SVM Algorithm

The parameters *C* and σ featuring in the SVM algorithm were optimised using cross validation and a grid search algorithm. *C* was chosen to range from 2^{-3} to 2^{15} whilst σ was chosen to range from 2^{-15} to 2^{3}. The step interval for the exponent of the parameters was set to 0.1. Selecting the parameters in the ranges describes above describes a grid of parameters where the SVM algorithm can be tuned [155] for the specific classification problem. All the tests were run using LibSVM [152].

The best value of C was found to be C=17559 ($2^{14.1}$), and the best value of σ was found to be σ=0.04419 ($2^{-4.5}$).

4.3.2.2 Optimising the Parameters of the AIRS Algorithm

There are several user configurable parameters available for AIRS. These include:

- Clonal rate (CR)
- Hyper-mutation rate (HR)
- Total resources (TR)
- Stimulation threshold (ST)
- k value for the k nearest neighbour algorithm (k)

Parameter tuning for the AIRS algorithm was accomplished taking each parameter in turn and finding the best value for that parameter alone within its specified range. The next parameter was then selected and tuned by varying this parameter within its specified range. This method does not search all the possible combinations of parameters, but it does ensure that a set of locally optimised set of parameters were discovered.

The parameters ranges were selected based on the suggested values discussed in [87] and [139]. Table 4.1 shows the optimum parameter values selected for the AIRS algorithm.

TABLE 4.1. PARAMETER OPTIMISATION OF AIRS ALGORITHM FOR FAULT CLASSIFICATION.

Parameter Name	Start	Min	Max	Step	Optimum
Clonal Rate	10	5	15	1	7
Hyper-mutation Rate	2	1	3	0.5	1.5
Total Resources	150	100	500	50	150
Stimulation Threshold	0.9	0.7	0.98	0.2	0.98
kNN	3	1	30	1	1

It is worth explaining the low value of k=1, for the kNN classifier. This value can be explained by considering that the AIRS algorithm generalises the 500 point dataset into a set of 81 classification memory cells. The number of cells in each of the classes varies from 27 representative memory cells for double line to ground faults, to 1 representative cell for normal operating conditions. With only 1 memory cell representing normal operating conditions, using a k value greater than 1 will cause this entire subset to be miss-classified, and limit the classification accuracy to a maximum of 80%. Setting k=1 ensures that normal operating conditions have a chance of being classified correctly.

4.3.2.3 Optimising the Parameter of the Random Forest Algorithm

The two configuration parameters for the random forest algorithm are the number of features per decision tree (n), and the number of trees to generate (N). A grid search was performed to find the optimum parameters with n varied between 1 and 30 in intervals of 2 and N varied between 2 and 30 in intervals of 2.

The optimum values of the configuration parameters was found to be n=9 and N=26.

4.3.2.4 Optimising the Parameter of the k Nearest Neighbours Algorithm

The only variable parameter in the k nearest neighbour classifier is the value of k. Values of k were tested in the range from 1 to 30. 30 was the maximum tested value of k, since this represents a large percentage (6%) of the dataset being used for the majority voting classification procedure.

The highest performing value of k was found to be k=3.

4.3.3 Training the Classification Model

The classification model was built using Weka [151] from 10 randomly selected sets of 90% of the full *testing dataset* and setting the parameters as described during the tuning process. The algorithms were then tested using the remaining 10% of the *testing dataset*. This technique is known as 10 fold cross validation.

4.3.4 Testing the Classification Performance of the Classification Algorithms

The goal of the voltage sag detection and classification algorithm is to successful detect when a fault occurs and then subsequently classify operating conditions into five classes: SLG (single line to ground faults), LL (line to line faults), LLG (double line to ground faults), LLL (three phase faults), and N (normal operation).

The detection accuracy (D) of a classification algorithm can be calculated by considering the total number of faults of type SLG, LL, LLG and LLL correctly classified as either SLG, LL, LLG or LLL (N_{CF}), over the total number of faults (N_F) of type SLG, LL, LLG or LLL. The distinction between of the type of fault is arbitrary, since detection performance is related to splitting the dataset into two classes only: one containing all faults (SLG, LL, LLG or LLL) and one containing normal operating conditions (N). Detection accuracy is summarised in equation (4.1):

$$D = n_{CF} / n_F \times 100 \tag{4.1}$$

The classification accuracy (C) of a classification algorithm can be calculated from the number of correctly classified instances of all types ($n_C = n_{LLLCLLL} + n_{SLGCSLG} + n_{LLGCLLG} + n_{LLCLL} + n_{NCN}$, where $n_{LLLCLLL}$ is the denotes the number of "LLL" instances classified as "LLL") over the total number of instances of all types ($n_I = n_{LLL} + n_{SLG} + n_{LLG} + n_{LL} + n_N$, where n_N is the total number of instances in that class). This is defined in equation (4.2).

$$C = n_C / n_I \times 100 \tag{4.2}$$

The detection and classification accuracy of the classifier can be further validated by considering the false alarm (FA) rate and the missed fault (MF) rate of each classifier. False alarms are triggered when the classification algorithm incorrectly assigns a normal operating condition (N) to one of the four fault types (SLG, LL, LLG and LLL). Missed faults are caused when the classification algorithm fails to detect faults in the system and classifies a fault as a standard operating point (N). Both the FA rate and the MF rate are defined as percentages, and can be considered as the rate at which the algorithm misclassifies faults and normal operating conditions. They are defined in equations (4.3) and (4.4).

$$FA = (n_{NCLLL} + n_{NCSLG} + n_{NCLL} + n_{NCLLG}) / n_I \times 100 \tag{4.3}$$

$$MF = (n_{LLLCN} + n_{SLGCN} + n_{LLNCN} + n_{LLGCN}) / n_I \times 100 \tag{4.4}$$

4.3.4.1 Classification Performance Required for a Practical Monitoring Deployment

It is worth considering the detection, classification, false alarm and missed fault rates which would be acceptable for a practical monitoring deployment.

It was noted previously that the MRA method [25] can perform the process of fault detection, and with 12 monitors this method can detect 100% of faults anywhere in the network using a threshold of 0.9 per unit on the worst affected phase. A detection rate of close to 100% is therefore the target for the proposed classification algorithm in this study.

Classification performance, false alarm and missed fault targets can be obtained by reviewing the expected fault rates for each type of fault (shown in Table 4.2).

TABLE 4.2. THE NUMBER OF FAULTS PER YEAR FOR DIFFERENT TYPES OF FAULTS WITHIN THE 295 BUS GENERIC DISTRIBUTION NETWORK [131].

LLL	SLG	LLG	LL	Type	Voltage Level
0.0032	0.0584	0.0136	0.0048	Bus	All
0.348	6.35	1.479	0.522	Overhead Line	11kV
0.1482	2.7012	0.6288	0.222	Overhead Line	33kV
0.0240	0.438	0.102	0.036	Overhead Line	132kV
0.1962	3.5772	0.8328	0.294	Cable	11kV
0.1482	2.7012	0.6288	0.222	Cable	33kV
0.024	0.438	0.102	0.036	Cable	132kV

During a typical year, it is expected that that there will be 1287 SLG faults, 106 LL faults, 300 LLG faults and 4 three phase faults. The number of normal operating points (N) will depend on the number of events presented to the classification algorithm. It was assumed that in a typical year, 50% of the events logged by the monitoring system represent normal operation and 50% were faults, leaving a total of 1697 normal operating points.

It can be concluded from these fault rates that high levels of detection performance can be achieved by discriminating between two groups only:

Asymmetric Faults: SLG, LL and LLG faults

Symmetric Faults: LLL faults and N (normal operation)

This conclusion asserts that three phase faults (LLL) are largely irrelevant, since they are very unlikely to occur. Indeed, if an algorithm were to be able to discriminate 100% of the time between asymmetric and symmetric whilst consistently misclassifying LLL faults, a detection performance of 99.7% could be achieved. However, LLL faults are important since the small ratio of LLL to N will disproportionately detrimentally affect the false alarm rate.

The key factor of an algorithm's classification performance will be correct classifications within the asymmetric group and symmetric group. The classification performance required ultimately depends on the objective of the study being performed. For example, if the classification results are subsequently used for fault location and then given to a maintenance team to investigate, it is critical that the correct type of fault is identified. It would be reasonable to assume a classification accuracy of around 1 error in 100 faults, or 99%. If however, the fault classification results are being used for historical analyses to obtain a general (rather than a specific) understanding of network voltage sag performance, then classification performance is less important. An accuracy of around 1 error in 20 faults (or 95%) would provide a reasonable guide to network wide voltage sag performance.

Similar arguments can be put forward for both the false alarm (FA) and missed fault (MF) rate. For fault location, a FA and MF rate of less than 1% would be desired and less than 5% for more general studies.

4.4 Results

Detection accuracy, classification accuracy, FA rate, and MF rate were all recorded for each of the 4 algorithms. Inter-class misclassifications were analysed using a confusion matrix. All the results were obtained using the testing dataset for all twelve combinations of monitors in the MRA set and also the ten monitors in the ENG monitor set.

4.4.1 The Performance of the Algorithms using the MRA Monitor Set

Fig. 4.4 to Fig. 4.15 show graphs of classification and detection performance as well as the confusion matrices for each of the four algorithms using the MRA monitor set.

Fig. 4.4 shows the classification performance of the SVM algorithm. Between 4 and 12 monitors, the algorithm detects 99.9% of faults, misclassifies 1.5% of the operating scenarios with a missed fault rate of almost ~0.05% and a false alarm rate of 3% or less. For less than 4 monitors, the classification and detection performance of the algorithm deteriorates, finishing at a classification rate of almost 88% with 1 monitor installed.

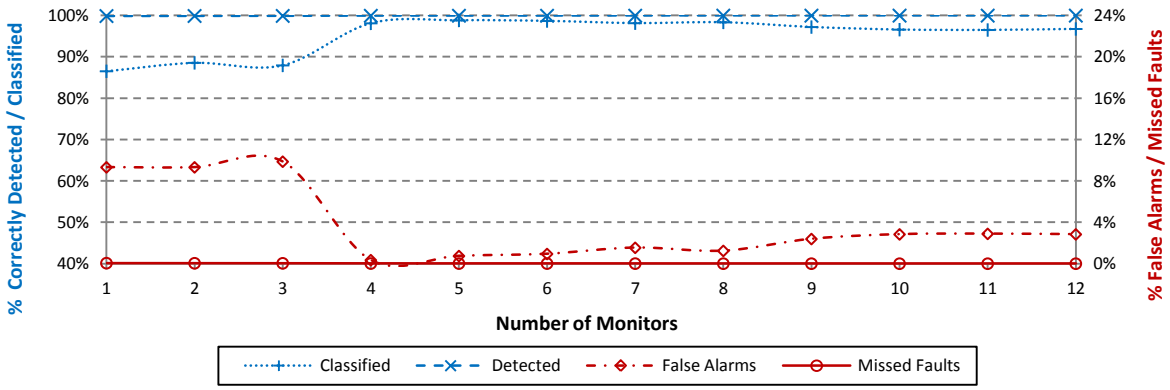


Fig. 4.4. Percentage detection, classification, false alarm (FA) and missed fault (MF) rates for the SVM algorithm.

SLG	LL	LLG	LLL	N	Pred. / True
99.59	0.41	-	-	-	**SLG**
1.13	**96.33**	2.42	-	-	**LL**
-	1.53	**98.55**	-	-	**LLG**
-	-	-	**91.05**	9.07	**LLL**
-	-	-	5.60	**94.32**	**N**

Fig. 4.5. Confusion matrix (in %) for the SVM algorithm using 12 monitors.

SLG	LL	LLG	LLL	N	Pred. / True
93.52	6.42	0	-	-	**SLG**
12.65	**84.71**	2.62	-	-	**LL**
-	13.46	**87.00**	0.20	-	**LLG**
-	-	-	**48.97**	51.74	**LLL**
-	-	-	18.39	**81.36**	**N**

Fig. 4.6. Confusion matrix (in %) for the SVM algorithm using 1 monitor.

The confusion matrices (Fig. 4.5 and Fig. 4.6) for the SVM algorithm show classification errors between SLG compared with LL, LLG compared with LL and LLL compared with N. Fig. 4.6 shows that with one monitor, the SVM algorithm additionally starts to miss-classify between LLL compared with LLG, albeit at low rates. For 12 monitors, Fig. 4.5 shows the highest levels of misclassification are caused where LL are misclassified as LLG and LLL are misclassified as N. Fig. 4.6 shows that with one monitor, the highest levels of misclassification are caused when LLL are misclassified as N and vice versa

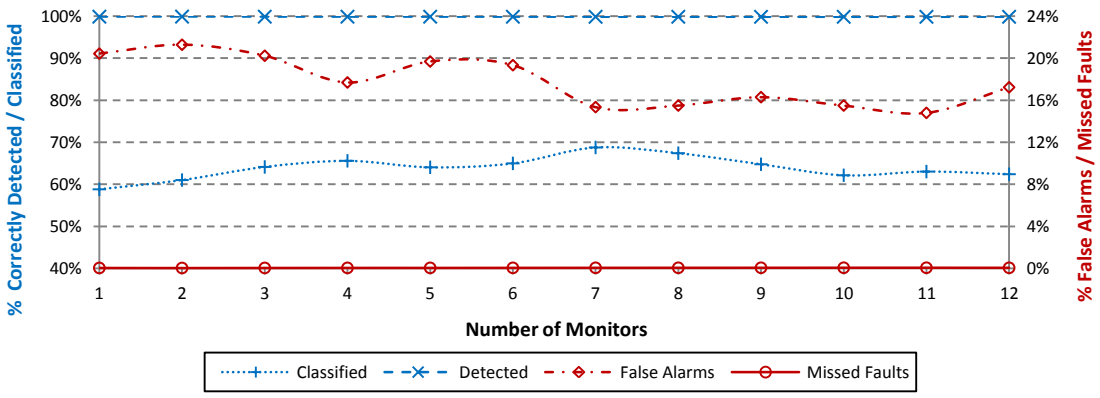


Fig. 4.7. Percentage detection, classification, false alarm (FA) and missed fault (MF) rates for the AIRS algorithm.

Fig. 4.7 shows that the detection performance of the AIRS algorithm is almost 100% as the number of monitors is moved between 1 and 12. The classification performance of the algorithm varies between 60%-70%, and is invariant to the number of monitors. The missed fault rate is close to zero due to good discrimination between asymmetric and symmetric faults, whilst the false alarm rate ranges between 15% and 22%.

SLG	LL	LLG	LLL	N	Pred. \ True
53.76	42.94	0.51	-	-	**SLG**
41.38	**57.46**	3.79	-	-	**LL**
1.38	14.12	**84.14**	-	-	**LLG**
-	-	-	**43.34**	55.68	**LLL**
-	-	-	35.08	**65.52**	**N**

Fig. 4.8. Confusion matrix (in %) for the AIRS algorithm using 12 monitors.

SLG	LL	LLG	LLL	N	Pred. \ True
56.69	8.70	14.45	-	-	**SLG**
22.22	**86.27**	23.18	-	-	**LL**
13.99	19.47	**68.28**	0.04	-	**LLG**
-	-	-	**54.47**	46.16	**LLL**
-	-	-	40.31	**59.12**	**N**

Fig. 4.9. Confusion matrix (in %) for the AIRS algorithm using 1 monitor.

The confusion matrices of Fig. 4.8 and Fig. 4.9 shows high levels of misclassification between SLG, LL and LLG faults, as well as confusion between LLL faults and N.

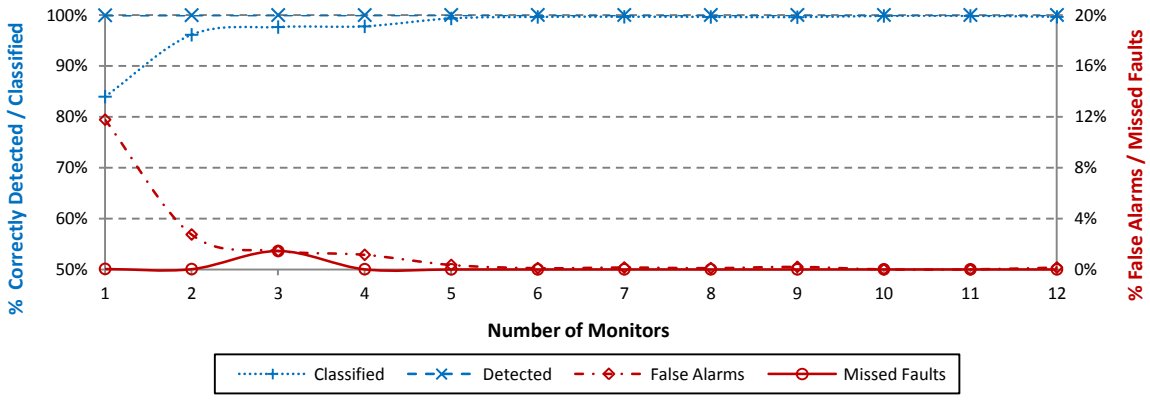


Fig. 4.10. Percentage detection, classification, false alarm (FA) and missed fault (MF) rates for the random forest algorithm.

Fig. 4.10 shows that with 5-12 monitors the detection and classification performance of the Random Forest algorithm remains above 99%, with a false alarm rate always virtually zero, and a missed fault rate of less than 0.5%. With 4 monitors and less, the algorithm performance declines, finishing with a classification performance around 80% when one monitor is used to classify all faults.

Fig. 4.11 shows that there is greatest misclassification between where LL faults are miss-classified as LLG (3.69% of LLG faults), and LLL faults are misclassified as N (4.49% of LLL faults). As the number of monitors is reduced, Fig. 4.12 shows that the level of misclassification grows significantly between LLL and N faults (to almost 35% of all LLL and N faults), and also between LL, LLG, and SLG faults. It is interesting to

note that there are no misclassifications between SLG and LLG faults using the Random Forest algorithm and 12 monitors.

SLG	LL	LLG	LLL	N	Pred. / True
99.79	0.20	-	-	-	**SLG**
1.03	**95.11**	3.69	-	-	**LL**
-	0.10	**99.90**	-	-	**LLG**
-	-	-	**95.58**	4.49	**LLL**
-	-	-	0.29	**99.70**	**N**

Fig. 4.11. Confusion matrix (in %) for the random forest algorithm (12 monitors).

SLG	LL	LLG	LLL	N	Pred. / True
90.33	9.38	0.19	-	-	**SLG**
12.65	**85.22**	3.98	-	-	**LL**
0.21	1.43	**98.45**	-	-	**LLG**
-	-	0.10	**66.76**	33.60	**LLL**
-	-	-	23.21	**76.47**	**N**

Fig. 4.12. Confusion matrix (in %) for the random forest algorithm (1 monitor).

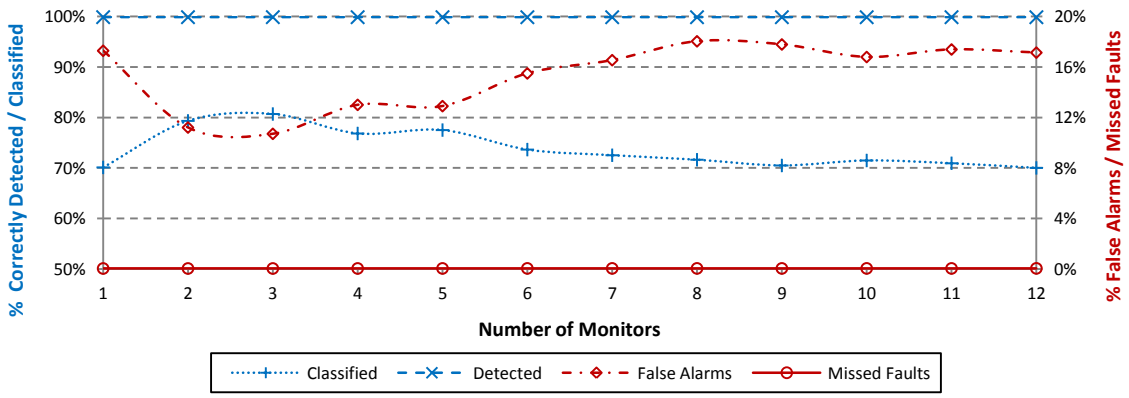


Fig. 4.13. Percentage detection, classification, false alarm (FA) and missed fault (MF) rates for the kNN algorithm.

Fig. 4.13 shows that the kNN algorithm's performance is largely invariant as the number of monitors is changed from 1 to 12 monitors. Its detection performance is always almost 100%, its classification performance varies between 70% and 80%, with virtually no missed faults and a false alarm and missed fault rates of between 12% and 18%.

SLG	LL	LLG	LLL	N	Pred. / True
69.86	29.87	-	-	-	**SLG**
35.19	**61.16**	3.78	-	-	**LL**
-	1.22	**98.84**	-	-	**LLG**
-	-	-	**49.46**	51.25	**LLL**
-	-	-	33.82	**65.70**	**N**

Fig. 4.14. Confusion matrix (in %) for the kNN algorithm using 12 monitors.

SLG	LL	LLG	LLL	N	Pred. / True
74.49	24.06	1.16	-	-	**SLG**
29.73	**63.61**	6.60	-	-	**LL**
10.70	9.58	**80.80**	0.20	-	**LLG**
-	-	-	**56.24**	44.37	**LLL**
-	-	-	34.12	**65.40**	**N**

Fig. 4.15. Confusion matrix (in %) for the kNN algorithm using 1 monitor.

The confusion matrices shown in Fig. 4.14 and Fig. 4.15 show misclassifications between SLG, LL, LLG faults, as well as misclassifications between LLL faults and N

(normal operating points). Fig. 4.14 shows that there are no misclassifications between LLG and SLG faults when 12 monitors are used to classify faults, however, misclassifications are present between LLG, SLG and LL as well as LLL and N.

4.4.2 The Performance of the Algorithms using the ENG Monitor Set

As well as comparing the performance of the algorithms using the MRA monitor set and varying the number of monitors, the algorithms were also compared using the ENG monitor set (as shown in Fig. 4.3).

Table 4.3 shows the variation in performance across each of the four algorithms using the ENG monitor set. The random forest algorithm yields the highest level of performance, although the SVM algorithm is very close in performance to the random forest.

TABLE 4.3. DETECTION AND CLASSIFICATION PERFORMANCE AND FALSE ALARM AND MISSED FAULT RATES FOR THE SVM, AIRS, RANDOM FOREST AND KNN CLASSIFICATION ALGORITHMS USING THE ENG MONITOR SET.

Algorithm	Detection Performance	Classification Performance	False Alarm Rate	Missed Fault Rate
SVM	99.95%	91.66%	7.53%	0.03%
AIRS	99.90%	70.14%	17.30%	0.05%
Random Forest	**99.99%**	**99.78%**	**0.00%**	**0.01%**
kNN	99.88%	67.98%	19.19%	0.06%

Fig. 4.16 to Fig. 4.19 shows the confusion matrices for each of the four classification algorithms. All of the algorithms miss-classify the faults in two distinct groups consisting of asymmetric network conditions (SLG, LL and LLG) and symmetric network operation (LLL and N).

The best performing algorithms were the SVM algorithm (Fig. 4.16) and the random forest algorithm (Fig. 4.18). The most significant difference between these algorithms' confusion matrices is the higher level of misclassifications that occur between LLL faults and N for the SVM algorithm. For the random forest, the average classification performance between LLL and N is 93%, whilst the classification performance of the SVM algorithm between these two groups is 81%.

SLG	LL	LLG	LLL	N	Pred. / True
99.38	0.61	-	-	-	**SLG**
1.44	**94.80**	3.59	-	-	**LL**
-	4.69	**95.54**	-	-	**LLG**
-	-	-	**77.97**	22.33	**LLL**
-	-	-	14.85	**84.95**	**N**

Fig. 4.16. Confusion matrix (in %) for the SVM algorithm using 10 monitors installed at 33kV and 11kV sub-stations.

SLG	LL	LLG	LLL	N	Pred. / True
74.49	24.06	1.16	-	-	**SLG**
29.73	**63.61**	6.60	-	-	**LL**
10.70	9.58	**80.80**	-	-	**LLG**
-	-	-	**56.24**	44.37	**LLL**
-	-	-	34.12	**65.40**	**N**

Fig. 4.17. Confusion matrix (in %) for the AIRS algorithm using 10 monitors installed at 33kV and 11kV sub-stations.

SLG	LL	LLG	LLL	N	Pred. / True
99.79	0.10	0.10	-	-	**SLG**
0.51	**95.11**	3.78	-	-	**LL**
-	0.61	**99.42**	-	-	**LLG**
-	-	0.10	**87.71**	12.36	**LLL**
-	-	-	1.28	**98.70**	**N**

Fig. 4.18. Confusion matrix (in %) for the Random Forest algorithm using 10 monitors installed at 33kV and 11kV sub-stations.

SLG	LL	LLG	LLL	N	Pred. / True
70.47	29.26	-	-	-	**SLG**
40.74	**54.23**	5.14	-	-	**LL**
-	1.73	**98.35**	0.04	-	**LLG**
-	-	-	**50.84**	49.85	**LLL**
-	-	-	37.86	**61.62**	**N**

Fig. 4.19. Confusion matrix (in %) for the kNN algorithm using 10 monitors installed at 33kV and 11kV sub-stations.

4.5 Discussion of Results

The results highlight a wide variation in the classification and detection accuracy of each of the tested algorithms. The best performing algorithm is the Random Forest, whilst the worst performance is obtained with either the kNN or AIRS algorithms.

It is anticipated that both the SVM algorithm and the random forest algorithm could be used in practical monitoring deployment to detect and classify faults. With 12 monitors, both algorithms correctly detect almost 100% of faults, and can further classify over 99% of faults into one of the 5 types. With only 5 monitors in a 295 bus network, the classification accuracy of the random forest is still 99.3% and classification accuracy of the SVM is 98.8%. The false alarm rate and missed fault rate of both algorithms increases sharply as the number of monitors is reduced below four.

If the classification results are weighted by the expected fault rate for each type of fault, and considered in absolute terms, the results for the random forest algorithm are even more compelling. Fig. 4.20 and Fig. 4.21 show confusion matrices for the random forest algorithm whilst classifying 1697 faults and 1697 normal operating points; the same numbers expected during a typical year. The algorithm detects all 1697 faults in both cases, and the algorithm only misclassifies 13 events with 12 monitors and 24 events

with 5 monitors. There are no missed faults, and there are only 5 false alarms with 12 monitors and 5 false alarms with 5 monitors. All of these results are well within the level of targeted level of accuracy for a practical monitoring deployment.

SLG	LL	LLG	LLL	N	Pred. / True
1284	3	-	-	-	**SLG**
1	**101**	4	-	-	**LL**
-	-	**300**	-	-	**LLG**
-	-	-	**4**	-	**LLL**
-	-	-	5	**1692**	**N**

Fig. 4.20. Confusion matrix (absolute numbers of faults / year) for the random forest algorithm for 12 monitors.

SLG	LL	LLG	LLL	N	Pred. / True
1282	5	-	-	-	**SLG**
1	**101**	4	-	-	**LL**
-	2	**298**	-	-	**LLG**
-	-	-	**4**	-	**LLL**
-	-	-	12	**1658**	**N**

Fig. 4.21. Confusion matrix (absolute numbers of faults / year) for the random forest algorithm for 5 monitors.

Classification errors for both the random forest and SVM algorithm can be visualised using the network diagram shown in Fig. 4.3. Fig. 4.3 shows several zones where simulated faults (of type SLG, LL, LLG or LLL) were misclassified by either the SVM or the random forest algorithm. The zones for both algorithms correlate strongly, indicating that both algorithms miss-classify faults when they are simulated in certain locations of the network. Most of the zones are contained within the 11kV network (although there is one zone in the 33kV network), and they are generally fairly close to the secondary side of 33kV to 11kV transformers. It is interesting to note that some misclassifications overlap with monitored busbars, which would not intuitively be expected. These misclassifications are caused between LLG and LL faults, where the during fault voltage profile at the busbars are very similar.

All of the algorithms are strong at separating asymmetric faults from symmetrical network operation. This is reflected in the near 100% detection performance for all of the algorithms. For example, even with just one monitor, Fig. 4.15 shows that the poorly performing kNN algorithm only misclassifies 2 out of a 5004 operating conditions (0.03%) as symmetric, where the true classification was asymmetric.

Within the asymmetric and symmetric groups, there are also some common trends across all of the algorithms. The highest degree of misclassification occurs between three phase faults (LLL) and normal operating conditions. These misclassifications severely impact on detection accuracy, false alarm rate and missed fault rate, because high detection performance requires accurate discrimination between the LLL and N groups. This is shown Fig. 4.10 and Fig. 4.12 for the random forest algorithm; as the number of monitors is decreased, the detection performance decreases whilst the false

alarm rate and missed fault rate both rise due to misclassifications between LLL faults and N.

Misclassifications between the LLL and N group could be overcome by utilising the threshold detection method of the MRA algorithm. As all of the algorithms separate asymmetric and symmetric faults with near 100% accuracy, therefore a threshold detection algorithm could be used to subsequently classify symmetrical operating points as either LLL or N. If an operating point is symmetric and the voltage in the worse affected phase at one monitor or more is below 0.9 per unit, then logically, there must have been a LLL fault. Using this extension, the random forest algorithm's performance would further improve and it would be expected to misclassify a total of 8 faults per year (classification accuracy of 99.8%) with 12 monitors and 12 faults per year (classification accuracy of 99.6%) with 5 monitors, with no missed faults, false alarms and a detection accuracy of 100%. Note that this procedure could only be applied if monitors were installed such that the monitor reach area of each monitor covers the whole network for three phase faults [25]. All MRA monitor sets with 4 or more monitors and the ENG monitor set satisfy this criterion.

The worst performing algorithms are the kNN and AIRS classifiers. Fig. 4.7, Fig. 4.13, Fig. 4.4, Fig. 4.7 and Table 4.3 show that kNN and AIRS perform consistently poorly regardless of the monitor location or number. The detection accuracy for either algorithm never exceeds 90%, and the average classification accuracy languishes around 75%. An average false alarm rate of around 7% and an average missed fault rate of close to 10% would cause an impractically high number of errors if the algorithm was to be implemented in a practical monitoring deployment.

The effect of monitor location can be considered by comparing the results using the 10 monitor MRA set with the ENG monitor set. Table 4.4 highlights the salient differences in terms of detection and classification and detection performance for all of the algorithms. Both the SVM and the random forest algorithms perform marginally better using the MRA monitor set, whilst the AIRS and the kNN algorithm perform better using the ENG monitor set. For the SVM and random forest algorithms, these results indicate that using an optimal monitoring algorithm to place monitors is important, and can yield higher performing results. The random forest algorithm would be a practical algorithm to use with either the MRA or ENG monitor set, as it exceeds the target performance of 99% classification accuracy in both situations, whilst the SVM would only be suitable if the MRA monitor locations were used.

TABLE 4.4. A COMPARISON OF THE CLASSIFICATION AND DETECTION PERFORMANCE OF ALL ALGORITHMS USING THE ENG AND THE MRA MONITOR SETS.

	10 Monitors MRA Monitor Set		10 Monitors ENG Monitor Set	
Algorithm	**Detection Accuracy**	**Classification Accuracy**	**Detection Accuracy**	**Classification Accuracy**
SVM	99.98%	99.59%	99.95%	91.66%
AIRS	99.86%	62.17%	99.90%	70.14%
Random Forest	99.99%	99.78%	99.97%	99.07%
kNN	99.88%	71.52%	99.88%	67.98%

4.5.1.1 Comparison of the Algorithms Performance

Although the algorithms yield different levels of detection and classification performance, it is important to establish whether or not the results are statistically and scientifically significant. This can be achieved by attempting to reject a null hypothesis which can in general be described as follows:

"Algorithm A and Algorithm B yield the same distribution of classification performance when classifying faults using the methodology described in this study on the 295 bus power system network."

This hypothesis can be statistically analyzed by applying the Mann-Whitney-Wilcoxon non-parametric rank-sum test [156] which tests whether or not one of two samples of independent observations tends to have larger values than the other. The aim of the test is to ascertain rejection of the null hypothesis; that each of the algorithms has identical distributions and medians of performance. In this research, a significance level of 5% was used and a *p*-value of <5% shows that the null hypothesis may be rejected.

Scientific significance attempts to measure the magnitude of the differences in classification performance between the tested algorithms. Given a large enough sample size it is often possible to show a statistical significance. A non-parametric measure such as the Vargha-Delany A statistic[157] assesses whether or not the effect is scientifically significant. The A value ranges between 0 and 1. 0.5 indicates a non-existent effect whilst values closer to 0 or 1 indicate an increasingly large effect size. An effect size of <0.36 or >0.64 is taken as a guide from [157] to indicate scientific significance.

The input datasets to the statistical hypothesis tests were four datasets consisting of each of the algorithm's classification performance across the 13 sets of monitoring locations.

Table 4.5 shows the results of the attempted rejection of the 6 null hypotheses. It also shows an overview of whether or not the size of any statistical significance is scientifically significant.

TABLE 4.5. THE STATISTICAL (STAT.) AND SCIENTIFIC (SCI.) SIGNIFICANCE OF CLASSIFICATION PERFORMANCE WHEN COMPARING THE SVM, AIRS, RANDOM FOREST AND KNN ALGORITHMS.

Hypothesis Number	**Comparison**	**Rank-sum *p* Value**	**Vargha-Delany [157] A Statistic**	**Statistical Significance**	**Scientific Significance**
Hypothesis 1	SVM to AIRS	1.65×10^{-5}	1	Yes	Yes
Hypothesis 2	SVM to Random Forest	0.0120	0.7929	Yes	Yes
Hypothesis 3	SVM to kNN	1.65×10^{-5}	1	Yes	Yes
Hypothesis 4	AIRS to Random Forest	1.65×10^{-5}	1	Yes	Yes
Hypothesis 5	AIRS to kNN	2.32×10^{-5}	0.99	Yes	Yes
Hypothesis 6	Random Forest to kNN	1.65×10^{-5}	1	Yes	Yes

The results show that the difference in classification performance for each of the 6 hypotheses and algorithmic comparisons are both scientifically and statistically significant. The effect size is smallest for the comparison between the SVM and the Random Forest algorithm, although it is still classed as a scientifically significant result since $A>0.64$.

Although both the AIRS and the kNN algorithm deliver similarly poor results in terms of detection and classification performance, the statistical and scientific significance tests show rejection of the hypothesis that both algorithms produce the same distribution of classification performance at the 5% significance level. This suggests that the AIRS and kNN classification processes are distinct, even though they are algorithmically very similar as they share the same k-nearest neighbour classification process.

4.6 Summary

The first aim of this research was to develop a method capable of detecting and classifying faults using a realistic set of erroneous monitors installed at practical locations whose number is much less than required to observe the whole network. The method developed in this chapter was shown to be capable of detecting voltage sags with a nearly 100% level of accuracy and subsequently discriminating sags into two distinct groups: asymmetrical operating conditions and symmetrical operating conditions. The best performing classification algorithms were shown to be able to correctly classify over 99% of three phase, single phase to ground, double line to ground and line to line faults expected during the normal course of distribution network

operation. The developed method is practical, and the results of all the algorithms showed resilience to measurement noise, fault impedance variation, pre-fault load variation and as well as working with monitors installed at different locations within the network.

The second aim of the research was to test and compare the performance of the developed method using a range of internal classification algorithms. Although all the algorithms were capable of detecting faults, it is anticipated that only support vector machine (SVM) and the random forest algorithms could yield high enough classification rates (in excess of 99%) to be used for fault location or voltage sag profile estimation. Both of these algorithms should perform comparably in terms of performance and also when considering where in the network faults were most likely to be misclassified. It is anticipated that the random forest algorithm would be the best choice of fault location algorithm, since this yielded consistent performance when the number of monitors was reduced or the location of monitors was altered. This corroborates with the findings of [88], which also found that a random forest can be used to accurately classify faults using measurements at one end of a feeder.

The research discussed in this chapter adds to the existing body of research on the voltage sag monitoring within distribution networks. Some of the advantages of this research include:

- Classification of faults using very few monitors (5 in a 295 bus network; note that this ratio of monitors to busbars might be specifc to this network, and other topological variations may change the results)
- Classification of faults with a high performance (>99% faults correctly classified)
- Ability to confidently use erroneous measurement data as input to the algorithm
- Robustly classify faults using monitors installed in a range of locations with varying fault impedances and pre-fault loading conditions
- Flexible centralised technique which can utilise measurements from all monitoring devices located anywhere in a distribution network

The research also quantifies the performance of several popular artificial intelligence techniques, and assesses their viability for future studies involving fault classification. The results support the conclusion that using the artificial immune recognition (AIRS) system is not the best choice for this style of classification problem.

The next two chapters describe how the results from classification and detection can be used to perform fault location, voltage sag profile estimation and then subsequently estimate the impact of voltage sags on consumer equipment. The classification and detection stage is a fundamental pre-requisite for all of the research that follows.

5 Voltage Sags: Localisation & Estimation of Voltage Magnitude

5.1 Introduction

Classification and detection of voltage sags gives a power system operator visibility on the numbers and types of voltage sags occurring in a network. Voltage sag (and fault) *Localization* uses the classification results to identify the source of the fault, and *Voltage Sag Profile Estimation* subsequently estimates how other buses and customers in the network were affected by the fault. These two processes are highlighted in Fig. 5.1.

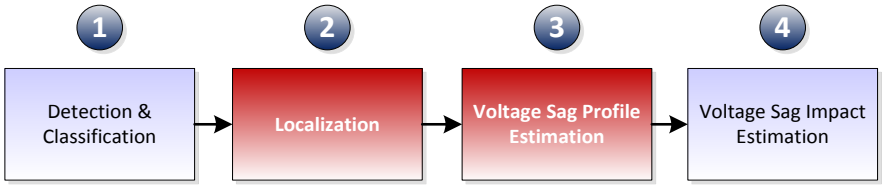


Fig. 5.1. The process of analysing voltage sags within a power network (stages 2 and 3).

The aim of a voltage sag localisation (or a fault location) algorithm is to estimate the source of a fault through measurements taken throughout a power system network. Fault location is often a non-trivial process: algorithms may work well in simulations, but care must be taken to ensure practical implementations are robust to measurement error

and unbiased when integrating information from a broad variety of monitoring devices. Voltage sag profile estimation (VSPE) involves taking the fault location and using this information to estimate the voltage sag depth (and hence the technical impact) of the sag at non-monitored.

Existing works on fault location [81] and voltage sag profile estimation [25] often assume that error free measurements are available from monitoring devices. This is never practically the case, as small errors will be introduced from monitors when readings are taken. If the equations (such as those presented in [81]) are highly sensitive to errors, it is feasible that their practical application will significantly diminish. It is therefore important to ascertain the sensitivity of the fault location equations (in [81]) to small changes in inputs parameters and also consider how the equations can be modified to cope with errors to return the best available estimate for the voltage sag profile across the network.

The most notable existing research on the impacts of measurement error on fault location estimation is [105]. In [105] the authors proposed an approach which was capable of minimising the impact measurement errors in the fault location process for a double circuit transmission line and subsequently identifying the presence of a faulty monitoring device. However, [105] did not consider how information from multiple monitoring devices could be combined to yield a distribution (rather than a point estimate) for the most likely location for a fault. The example presented in [105] also did not consider the situation where the number of monitors was much less than required to obtain a single unique fault location estimate.

In this chapter, a robust method for performing fault location and voltage sag profile estimation using multiple monitors is developed. The method uses statistical analysis to solve the fault location equations and find the most likely location for both the fault location and the best estimate for the voltage profile throughout the whole network during a fault. It comprehensively tackles both measurement error sensitivity and collating information from a range of disparate devices. The method draws on and significantly adds to existing research developed in [81] and [105].

The developed method has several advantages over existing techniques. Firstly, it allows power system operators to formulate a robust statistical estimation for both the fault location and voltage sag profile whilst taking into account the errors of monitoring devices within the network. This would not normally be possible if the analytical equations (such as those presented in [81]) were used alone.

Secondly, the method is independent of the accuracy of the monitoring device, so operators can utilise measurements from an arbitrarily accurate selection of installed monitoring devices, including, relays, power quality meters, disturbance recorders, phasor measurement (PMUs) devices or revenue meters.

Lastly, the method synthesises a fault location using all available information from an arbitrary number of devices. By using information from all available monitors, both the fault location and voltage profile estimate are more accurate than the single or double ended approaches presented in [81]. There is also an added advantage of being able to isolate bad data [63] at a single monitoring device. It also goes some way to overcoming issues caused by multiple impedance paths, by reducing the number of fault locations to those that are the most feasible.

5.2 Impedance Based Voltage Sag Localisation and Profile Estimation

Impedance based fault location algorithms utilize the impedance of the network and the observed voltage drop to arrive at an estimated fault location. Impedance based methods may be transient, or steady state, and single-ended or double-ended. This research focuses and advances the steady state single-ended impedance based algorithms developed in [81] by extended their application so they can be applied with any number of monitoring devices.

Although the algorithm developed in this research builds on the single-ended equations from [81], it should not be strictly considered either single-ended or double-ended. The proposed approach can be thought as arbitrarily-ended since it utilizes information from all available monitors in the network, regardless of their locations. The approach is also independent of the number monitoring devices, since any number from one upwards can be used to establish a fault location.

The single-ended fault location equations developed in [81] calculate the fault location using measured data from one or two monitoring devices. The equations require pre-fault voltage measurements at the monitored bus ($V_k^{(i)'}$) and at the ends of the faulted line ($V_p^{(i)'}$ and $V_q^{(i)'}$), as well as during fault voltage measurements at the monitored busbar ($V_k^{(i)'}$). The equations can be shown to be independent of fault resistance (R_f), if the fault resistance is assumed to be entirely real.

5.2.1 The Aim of the Voltage Sag Localisation Equations

The aim of the voltage sag localisation equations can be illustrated by returning to Fig. 2.6 which was first introduced in Chapter 2. Fig. 5.2 extends Fig. 2.6 by showing an arbitrary un-monitored bus *i*. The objective of the equations is to locate the fault at position *r* along the *l*th line and then subsequently perform voltage sag profile estimation at a non-monitored bus *i*.

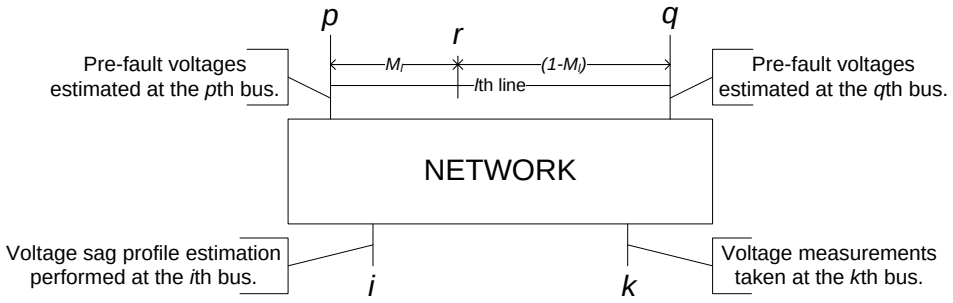


Fig. 5.2. A representation of the estimated and monitored buses required to perform fault location and voltage sag profile estimation in an arbitrary power network.

Assuming the network's impedance parameters can be derived accurately, the accuracy of the fault location equations will depend upon the values of the measured pre-fault voltages at the *k*th busbar and the estimated pre-fault voltage at the ends of the *l*th line. The accuracy of the measurement at the *k*th busbar is dependent on the installed monitor's accuracy, and this may vary depending on the device that is installed. Unless there happens to be a monitor at either the *p*th or the *q*th busbar, these voltages must be estimated through distribution system state estimation (DSSE) (as described in Chapter 2). The accuracy of the DSSE voltage estimates will directly depend on the accuracy of the pseudo-measurements used to estimate the load throughout the network.

The descriptions of the fault location equations that follow focuses on identifying which of the variables will be subject to errors, which variables are estimated and how the statistical properties of both errors and estimates can be determined.

5.2.2 Fundamental Equations for all Types of Fault

The equations developed in this section build upon the equations (2.26) to (2.35) developed in Chapter 2 and [81]. Without shunt capacitances, the following equations can be derived [81]:

$$Z_{kr}^{(i)} = B_k^{(i)} + C_k^{(i)} M_l \tag{5.1}$$

$$Z_{rr}^{(i)} = A_0^{(i)} + A_1^{(i)} M_l + A_2^{(i)} {M_l}^2 \tag{5.2}$$

where $Z_{kr}^{(i)}$ and $Z_{rr}^{(i)}$ are the *k,r*th and *r,r*th elements respectively of the $\boldsymbol{Z'_{bus}}$ impedance matrix which includes the *r*th busbar. The parameters $B_k^{(i)}$, $C_k^{(i)}$, $A_0^{(i)}$, $A_1^{(i)}$and $A_2^{(i)}$ are network dependent and are derived from elements in the $\boldsymbol{Z_{bus}}$ impedance matrix (which excludes the *r*th busbar), measured at the monitored bus (*k*) (details of their derivation are shown in [81] and in the following section). M_l is the distance in per unit along the faulted line *l*. Each of the network parameter (B_k, C_k, A_0, A_1 and A_2) are assumed to be known and free from error.

It is often convenient to formulate the voltage at the fault point ($V_r^{(i)}$) in terms of the voltages at each end of the faulted lines. This is described in equation (5.3) (again, ignoring shunt capacitance):

$$V_r^{(i)} = (V_p^{(i)'}(1-M_l) + M_l V_q^{(i)'}) \tag{5.3}$$

where $V_p^{(i)'}$ and $V_q^{(i)'}$ are the pre-fault voltages at each ends of the faulted line (buses *p* and *q*).

5.2.2.1 Derivation of $B_k^{(i)}$ and $C_k^{(i)}$

The constants $B_k^{(i)}$ and $C_k^{(i)}$ are the same as those described in [81]. If $\boldsymbol{Z'^{(i)}_{bus}}$ is the well documented bus impedance matrix [8] for the *i*th sequence (without an added extra busbar for the fault location *r*), then:

$$B_k^{(i)} = z'^{(i)}_{pk} \tag{5.4}$$

$$C_k^{(i)} = -(z'^{(i)}_{pk} - z'^{(i)}_{qk}) \tag{5.5}$$

where $z'^{(i)}_{pk}$ is the *p*th, *k*th element of the $\boldsymbol{Z'^{(i)}_{bus}}$ impedance matrix, with *p*, *k* and *q* as shown in Fig. 2.6.

5.2.2.2 Derivation of $A_0^{(i)}$, $A_1^{(i)}$ and $A_2^{(i)}$

$A_0^{(i)}$, $A_1^{(i)}$ and $A_2^{(i)}$ were first defined [81]. If $\boldsymbol{Z'^{(i)}_{bus}}$ is the well documented bus impedance matrix [8] for the *i*th sequence (without an added extra busbar for the fault location *r*), then:

$$A_0^{(i)} = [z_{1,pp}^{(i)} z_{1,qq}^{(i)} + z_L^{(i)} z_{1,pp}^{(i)} - (z_{1,pq}^{(i)})^2]/D \tag{5.6}$$

$$A_1^{(i)} = [(z_L^{(i)})^2 + z_L^{(i)}(z_{1,qq}^{(i)} - z_{1,pp}^{(i)})]/D \tag{5.7}$$

$$A_2^{(i)} = -(z_L^{(i)})^2/D \tag{5.8}$$

$$D = z_{1,pp}^{(i)} + z_{1,qq}^{(i)} - 2z_{1,pq}^{(i)} + z_L^{(i)} \tag{5.9}$$

$$z_{1,pp}^{(i)} = z'^{(i)}_{pp} - (z'^{(i)}_{pp} - z'^{(i)}_{pq})^2/D_0 \tag{5.10}$$

$$z_{1,qq}^{(i)} = z'^{(i)}_{qq} - (z'^{(i)}_{pq} - z'^{(i)}_{qq})^2/D_0 \tag{5.11}$$

$$z_{1,pq}^{(i)} = z'^{(i)}_{pq} - (z'^{(i)}_{pp} - z'^{(i)}_{pq})(z'^{(i)}_{pq} - z'^{(i)}_{qq})/D_0 \tag{5.12}$$

$$D_0 = z'^{(i)}_{pp} + z'^{(i)}_{qq} - 2z'^{(i)}_{pq} - z^{(i)}_L \tag{5.13}$$

where $z_L^{(i)}$ is the impedance of the line, between buses *p* and *q*.

5.2.3 Three Phase Symmetrical Faults

For three phase faults, combining (2.29) with (2.27) yields:

$$V_k^{(1)} = V_k^{(1)'} - Z_{kr}^{(1)} \frac{V_r^{(1)'}}{R_f + Z_{rr}} \tag{5.14}$$

Substituting (5.1), (5.2) and (5.3) into (5.14) yields the following fault location equation:

$$V_k^{(1)} = V_k^{(1)'} - \frac{(B_k^{(1)} - C_k^{(1)} M_l)(V_p^{(1)'}(1 - M_l) + M_l V_q^{(1)'})}{R_f + A_0^{(1)} + A_1^{(1)} M_l + A_2^{(1)} M_l^{\,2}} \tag{5.15}$$

where $V_k^{(l)'}$ is the pre-fault voltage at the monitored bus k, and $V_p^{(l)'}$ and $V_q^{(l)'}$ are the pre-fault voltages at each end of the line. The fault location M_l on the *l*th line can be found by splitting (5.15) into real and imaginary parts [81], and solving the imaginary side of the equation for M_l. The fault impedance R_f can be found through back substitution into the real side of equation (5.15), using the solved value of M_l.

Solving the quadratic equation (5.15) results in two possible values for M_l, which may be complex numbers. Values of M_l between 0 and 1 are feasible, and indicate that the fault occurred on the *l*th line.

The accuracy of the solved values of M_l and R_f will be dependent the accuracy of the state estimation estimates $V_q^{(l)'}$ and $V_p^{(l)'}$ and also the accuracy of the measured value $V_k^{(l)'}$. The accuracy of M_l is also dependent on the validity of the assumption that the R_f is entirely real.

5.2.4 Single Line to Ground Faults

Noting that $I_f^{(0)} = I_f^{(1)} = I_f^{(2)}$ for a phase A to ground fault, the ratio of equations (2.28) and (2.26) can be taken and combined with (5.1) to find the fault location. This is shown in equation (5.16).

$$M_l = \frac{B_k^{(2)} - G_k B_k^{(0)}}{G_k C_k^{(0)} - C_k^{(2)}} \tag{5.16}$$

where G_k is the ratio of the negative sequence voltage to the zero sequence voltage ($V_k^{(2)}/V_k^{(0)}$) as measured at bus *k*.

By considering the function shown in (5.16), it is possible to gain an understanding of how small errors in voltage measurement propagate to errors in the fault location M_l. Equation (5.17) shows G_k as a function of M_l.

$$G_k = \frac{B_k^{(2)} + C_k^{(2)} M_l}{B_k^{(0)} + C_k^{(0)} M_l} \tag{5.17}$$

For single line to ground faults, M_l can be solved independently of R_f and pre-fault voltage estimates. The accuracy of M_l is dependent only on the accuracy of measurements at the monitored bus k ($V_k^{(2)}$ and $V_k^{(0)}$).

5.2.5 Line to Line Faults

For phase B to C line to line faults, $I_f^{(1)}$=-$I_f^{(2)}$ and $I_f^{(0)}$=0. An equation for positive sequence fault current during a fault was shown in (2.33). Equation (2.33) can be combined with (5.3) to yield (5.18) [81], and then this can be combined with (2.27) to yield (5.19).

$$I_f^{(1)} = \frac{(V_p^{(1)'}(1-M_l) + M_l V_q^{(1)'})}{R_f + Z_{rr}^{(1)} + Z_{rr}^{(2)}} \tag{5.18}$$

$$V_k^{(1)} = V_k^{(1)'} - Z_{kr}^{(1)} \frac{(V_p^{(1)'}(1-M_l) + M_l V_q^{(1)'})}{R_f + Z_{rr}^{(1)} + Z_{rr}^{(2)}} \tag{5.19}$$

Through the substitution of equations (5.1) and (5.2) into equation (5.19), the fault location for a line to line fault can be found by solving the following quadratic fault location equation:

$$V_k^{(1)} = V_k^{(1)'} - \frac{(B_k^{(1)} - C_k^{(1)} M_l)(V_p^{(1)'}(1-M_l) + M_l V_q^{(1)'})}{R_f + (A_0^{(1)} + A_0^{(2)}) + (A_1^{(1)} + A_1^{(2)}) M_l + (A_2^{(1)} + A_2^{(2)}) M_l^{\,2}} \tag{5.20}$$

Equation (5.20) is solved by splitting the equation into real and imaginary parts and solving the resultant quadratic equation in the same way as for three phase faults [81].

The solved values of R_f and M_l are dependent on the accuracy of the monitor at the kth busbar, and the pre-fault voltages at buses p and q obtained through DSSE. M_l will also be dependent on the validity of the assumption that the fault impedance can be assumed to be entirely real.

5.2.6 Double Line to Line to Ground Faults

For phase B to C line to line to ground faults, the solution for the fault location M_l involves solving a quartic polynomial.

For a line to line to ground fault, the ratio of equations (2.37) and (2.38) yields:

$$\frac{I_f^{(2)}}{I_f^{(0)}} = \frac{(Z_{rr}^{(0)} + 3R_f)}{Z_{rr}^{(2)}} \tag{5.21}$$

In a similar way to single line to ground faults, the fault location equations are solved by considering the ratio of the negative to the zero sequence voltages as a complex quantity G_k as measured at the *k*th busbar.

Equation (5.22) can be derived by combing (2.26) and (2.28) into (5.21) and then (5.1) and (5.2) to yield:

$$G_k = \frac{(B_k^{(2)} - C_k^{(2)} M_l)(A_0^{(0)} + A_1^{(0)} M_l + A_2^{(0)} {M_l}^2 + 3R_f)}{(B_k^{(0)} - C_k^{(0)} M_l)(A_0^{(2)} + A_1^{(2)} M_l + A_2^{(2)} {M_l}^2)} \tag{5.22}$$

R_f can be eliminated from the equation by noting that the equation can be split into real and imaginary parts.

The accuracy of the estimate for the fault location M_l is dependent on the accuracy of the measured values of the negative and zero sequence voltages at the *k*th busbar. Like three phase and line to line faults, the accuracy of M_l is also dependent on the validity of the assumption that the fault impedance is entirely real. If this assumption cannot be made then equation (5.22) cannot be solved. In this situation, two linearly independent fault location equations can be combined using two monitoring locations to eliminate R_f and then subsequently solve for M_l.

5.2.7 Calculating Voltage Sag Depth at Non-Monitored Busbars

The calculated values of the fault location M_l can be used to calculate the during fault voltage drops at the *i*th unmonitored busbar. The during fault voltage drops at bus *i* can be considered by re-writing equations (2.26) to (2.28) as functions of the fault location and the measured voltage at bus *k* only. The during fault voltage drops in the sequence domain are expressed in equations (5.23), (5.24) and (5.25):

$$V_i^{(0)(k)} = \frac{Z_{ir}^{(0)} V_k^{(0)}}{Z_{kr}^{(0)}} = V_k^{(0)} \frac{B_i^{(2)} + C_i^{(2)} M_l}{B_k^{(2)} + C_k^{(2)} M_l} \tag{5.23}$$

$$V_i^{(1)(k)} - V_i^{(1)'} = -\frac{Z_{ir}^{(2)} (V_k^{(1)'} - V_k^{(1)})}{Z_{kr}^{(1)}} = (V_k^{(1)'} - V_k^{(1)}) \frac{B_i^{(1)} + C_i^{(1)} M_l}{B_k^{(1)} + C_k^{(1)} M_l} \tag{5.24}$$

$$V_i^{(2)(k)} = \frac{Z_{ir}^{(2)} V_k^{(2)}}{Z_{kr}^{(2)}} = V_k^{(2)} \frac{B_i^{(2)} + C_i^{(2)} M_l}{B_k^{(2)} + C_k^{(2)} M_l} \tag{5.25}$$

Note that the second superscript (k) in $V_i^{(s)(k)}$ (equations (5.23), (5.24) and (5.25)) denotes that $V_i^{(s)}$ is calculated from measurements taken from the *k*th monitor (where *s* is an arbitrary sequence). Equations (5.23), (5.24) and (5.25) are independent of pre-

fault voltage estimates and fault impedance. However, M_l may be dependent on pre-fault voltage estimates and / or and fault impedance, depending on the type of fault (as discussed in the previous sections).

5.2.8 Uncertain Quantities in the Voltage Sag Localisation Equations

5.2.8.1 Errors in Measured Voltages

The voltage sag localisation equations use measurements from monitoring devices which may be subject to errors to synthesise a fault location. At monitored buses, voltages are measured through a voltage transformer. The accuracy of the measurement will vary depending on the quality of the monitoring device [16] [126].

Measurement errors were modelled as a series of independently distributed normal distributions [8, 11] as described in Chapter 2 and equation (2.1). An error of 0.2% of the true value was assumed for all symmetrical component magnitudes and 0.2% of 2π for phase measurements. The standard deviation for the voltage magnitude and phase was therefore $\sigma_{|V|}=0.067\%$ and $\sigma_{\phi}=0.0013\pi$ radians respectively [67].

To define the distribution of the voltage completely, the mean of the voltage magnitude (|V|) and phase (ϕ) is required. Since each measurement is taken at a unique point in time, there is only one measurement available to estimate the mean. The best estimate for the mean at the *k*th busbar is therefore the measured value itself, $|\overline{V}_k^{(s)}|$ and $\overline{\phi}_k^{(s)}$, in an arbitrary sequence *s*. Equations (5.26) and (5.27) define the probability distributions of a measurement at the *k*th bus in an arbitrary sequence *s*. Both $|V_k^{(s)}|$ and $\phi_k^{(s)}$ are assumed to be independent.

$$|V_k^{(s)}| \sim N(|\overline{V}_k^{(s)}|, \sigma_{|V_k|}) \tag{5.26}$$

$$\phi_k^{(s)} \sim N(\overline{\phi}_k^{(s)}, \sigma_{\phi_k}) \tag{5.27}$$

5.2.8.2 Errors in Pre-Fault Voltage Estimates

Pre-fault voltages can either be obtained by direct measurement, or alternatively by estimating the voltage using DSSE. If the pre-fault voltages are obtained using a monitor, the voltages can be assumed to be distributed as defined by equations (5.26) and (5.27).

DSSE was used to estimate the pre-fault voltages at non-monitored busbars. DSSE estimates voltages by minimising the summed squared differences between a set of

[16] The Siemens 9610 Power Quality meter for example, quotes an accuracy of ±0.2% for symmetrical component magnitude and phase at full scale of 0 to 106volts and -π to π.

measurements when applied to a non-linear model which represents the power system. The measurement inputs into the DSSE model are comprised of both measurements from physical monitoring devices and also pseudo-measurements.

Physical measurement inputs to the distribution system state estimator were modelled using the measurement model described in equations (5.26) and (5.27), assuming independence between monitoring devices. Pseudo-measurements are estimates for both real and reactive power at a non-monitored bus based on some knowledge of the customer type and load profile at that bus. Pseudo-measurements were modelled by assuming that the value of the real and reactive power injection (load) at each busbar can be estimated based on historical knowledge of the load profile. A percentage error of 20% was selected for all load estimation pseudo measurements (as in [69], and mentioned in Chapter 8) resulting in a pseudo-measurement standard deviation equivalent to 7% of the true value. The output of the distribution system state estimator is a set of pre-fault voltages.

Running DSSE to generate a set of correlated output voltages is a computationally intensive task as it involves the solution of many sets of load flow equations. To overcome this computational issue, the output voltages were fitted to a correlated multivariate normal distribution. Using this distributional approximation, tens of thousands of random numbers can be quickly generated to simulate the results of performing DSSE. Equation (5.28) describes the multivariate distribution of a random variable vector $\mathbf{X}$ whose elements are represented by the magnitudes and angles of the voltages calculated by DSSE.

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5.28}$$

where $\boldsymbol{\mu}$ is a set of mean voltages and angles and $\boldsymbol{\Sigma}$ is the covariance matrix obtained through fitting a multivariate distribution to the results of distribution system state estimation.

The accuracy of the pre-fault voltage estimates obtained using DSSE can be illustrated using a simple example. The standard deviation for pseudo-measurements was set to 7% (as described above) and the standard deviation of real measurements was set to 0.2% (as described above), with both assumed to follow a normal distribution. Using the three monitors as shown in Fig. 5.3, the maximum standard deviation for the voltage magnitude of a non-monitored bus ranges between 1.5% (or ~0.015 per unit) and 4.3% (or ~0.043) of the true voltage magnitude. (Note that the use of standard deviation here assumes that the output voltages accurately approximate a normal distribution. For

small variations in input measurements (<10%), this was found to be an accurate approximation.) This is as at least 3% more accurate than the pseudo-measurement inputs (which have a standard deviation of 7%) but far less accurate than measuring the voltage using physical monitors (where the standard deviation is 0.2%).

5.2.9 Difficulties Using the Voltage Sag Localisation Equations

The voltage sag localisation equations are extremely powerful, but they introduce some difficulties when used in a practical monitoring deployment.

A general difficulty with fault location and voltage profile estimation arises when the set of monitors being used to calculate a fault location is not able to unambiguously observe all the voltages within the network. In this scenario, multiple fault location solutions can occur where the same sets of voltages and currents can be observed for multiple locations within the network [158]. Reference [48] tackled this issue by estimating the fault location as the intersection of several monitors' fault location estimates. There is no guidance however for how close fault locations should be before they are considered to overlap.

A further difficulty with the fault location equations involves faulty monitoring devices. As there is no redundancy in the fault location estimate provided with one or two monitors in the methodology proposed in [48], there is no way of identifying a faulty monitoring device.

The equations also involve complex numbers and therefore if any of the measurements are slightly erroneous, the fault location may be two dimensional. The application in [81] of the equations proposed in did not provide guidance on how to deal with complex solutions, which are often possible if measurements are slightly in error.

Perhaps the most important issue with the fault location equations is their sensitivity to small changes in input measurement changes. Even small errors in inputs to the equations may result in large changes in the estimated output, and solutions to the equations which may diverge into the complex plane.

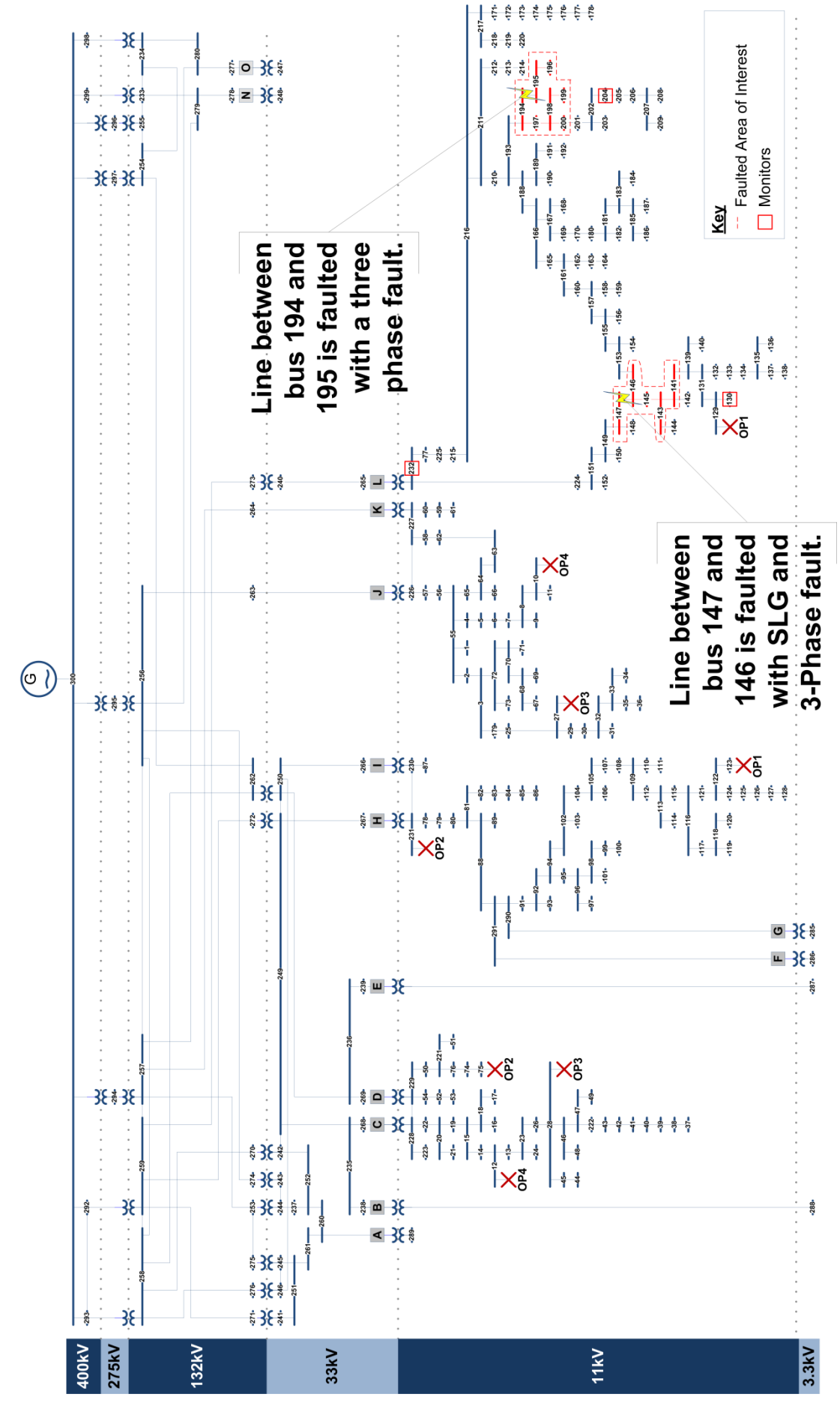


Fig. 5.3. The 100 buses connected to feeder L on the 295 bus network.

5.2.9.1 Sensitivity of the Equations

The sensitivity of the equations can be demonstrated by considering an example. Fig. 5.3 shows a generic 295 bus distribution network {zhang}, with feeder L highlighted. Consider the situation where a single monitor is installed at bus 232 which observes a single phase to ground fault in phase A between buses 147 to 146 (with zero fault resistance). During the fault, the voltages at bus 232 drop to 0.62, 0.94 and 0.98 per unit in phases A, B and C respectively.

Fig. 5.4 shows contour lines representative of introducing a constant percentage error in the zero or negative sequence voltage magnitude only. The figure also shows valid values for G_{232} (where the value of M_l is between 0 and 1 – and not complex) along lines connecting buses 147 to 146 and 146 to145.

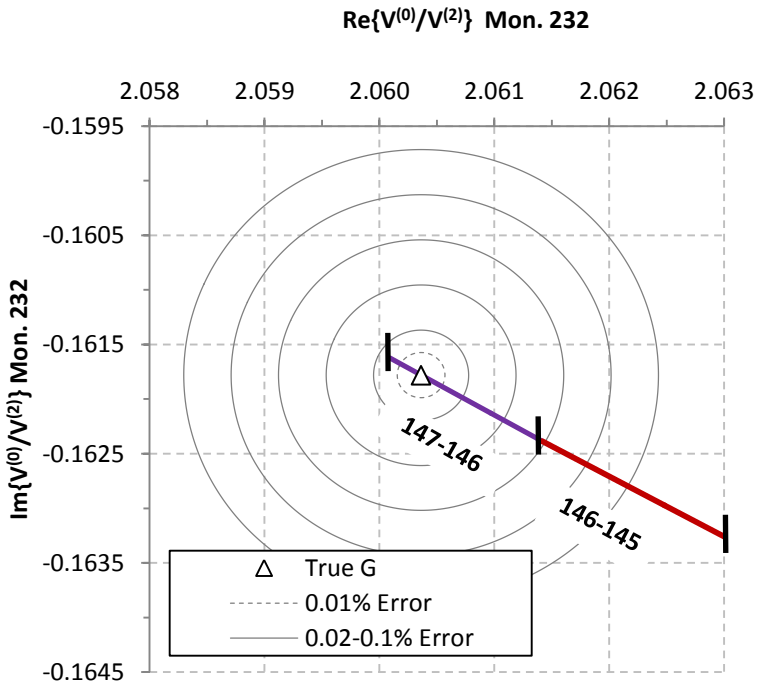


Fig. 5.4. Contours of constant G_{232} for errors in voltage magnitude in either zero or negative sequence.

Fig. 5.5 highlights that even a small error of 0.1% in a magnitude measurement could lead to the fault location solution being assigned to the incorrect line. Indeed, an error of this magnitude could place the fault anywhere on the line between buses 147-146 and up to half the distance along the line connecting buses 146-145. To localise this fault within 10% of the length of the line, a monitor at 232 would need to be accurate to within 0.01% of the true voltage magnitude. This is a very high level of precision for a distribution network monitoring device.

Fig. 5.5 shows how these errors map into solutions for $M_{147\text{-}146}$ using equation (5.16). The diamond sign in Fig. 5.5 illustrates the difficulty of formulating a solution for $M_{147\text{-}146}$ based on erroneous measured voltages. The example solution (M_e) is only in error by 0.1%, but the solution to the fault location equations is now complex. A basic estimate for the fault location could be achieved by either taking the real part of M_e ($\mathrm{Re}\{M_e\}$, as used in {liao EPSR paper}) or the magnitude of M_e ($|M_e|$). Both of these methods result in solutions which are above 1 (1.5 for $\mathrm{Re}\{M_e\}$ and 1.58 for $|M_e|$). Thus it would be concluded incorrectly that the fault did not occur on the line. A more advanced approach would be to consider the possibility that the measurement was erroneous, and predict a fault location given a known measurement error distribution. In Fig. 5.5 a dotted line (labelled "closest") shows most likely position for the fault (M_e=1) which maps M_e to the closest point on the valid line of solutions for $M_{147\text{-}146}$.

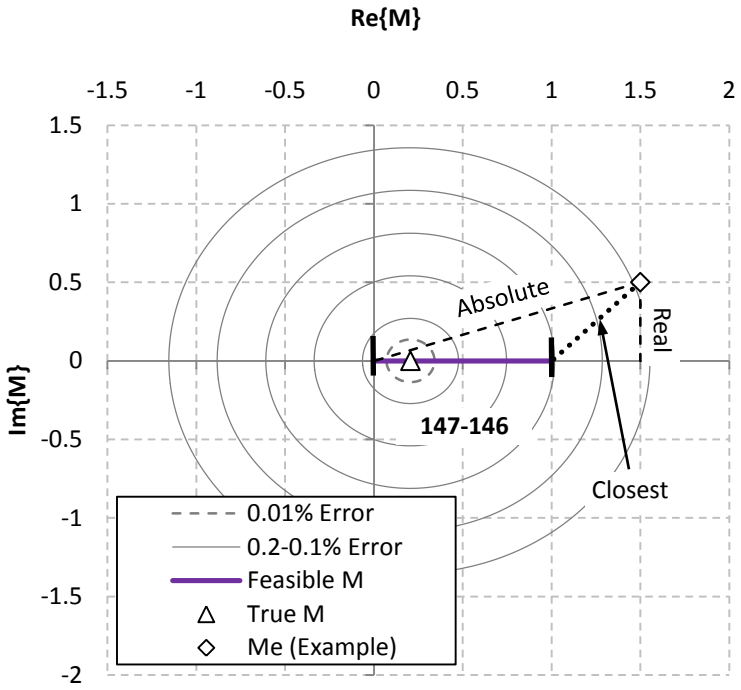


Fig. 5.5. The range of true and erroneous solutions for $M_{147\text{-}146}$ given a constant error in zero or positive sequence voltage magnitude.

There are two problems which are shown by this brief analysis, which are not covered in the existing body of research. Firstly, it is clear that the fault location equations are highly sensitive to even small measurement errors. If measurement error cannot be robustly eliminated, this will limit the usefulness of an impedance based approach using these equations in a practical monitoring deployment. Secondly, when the fault location equations are subjected to small errors they generate complex solutions. Although the only physically meaningful solutions for equation (5.16) are real, it is also important to consider the likelihood of a complex solution given some measurement error.

It should be noted that the examples in Fig. 5.4 and Fig. 5.5 are purposely simplistic to aid illustration; the analysis only considers variation in one dimension, namely zero or negative voltage magnitude. In practice, magnitude and / or phase of all voltages may be in error, thus creating a more complicated four dimensional error space. The shape of this space will be explored briefly in the later case studies.

5.2.10 Summary

The impacts of the problems with the voltage sag localisation equations on voltage sag profile estimation accuracy can be successfully minimised by extending the fault location equations to incorporate the statistical properties of erroneous measurements and bad data. Using similar ideas to state estimation [63], the rest of this chapter will propose a statistical maximum likelihood method of determining the probability distribution for the fault location and voltage sag profile estimate. The proposed technique incorporates readings from all monitors to establish a more accurate estimate than could be achieved with one monitor alone. With redundant estimates, bad data can be eliminated by performing statistical tests on the data. Fault location aliasing will never be eliminated entirely, unless the monitor set is fully observable. A probabilistic approach however, means that fault locations are converted from discrete deterministic points into distributions, therefore identifying the *most likely* fault location and therefore the *best estimate* for the voltage sag profile.

5.3 A Probabilistic Approach to Voltage Sag Localisation & Estimation

5.3.1 Probabilistic Fault Location

The aim of probabilistic fault location is to find a probability density function (f_F) for the fault location F, conditional that a fault occurred on a line within the network. To estimate a fault location for a given type of fault, the fault location equations are solved for each of the lines in the network. Solutions to the fault location equations at customer sites are not considered.

Solving the fault location equations yields a solution for M_l: the location of the fault along the lth line. The values of M_l which are physically meaningful are located on the real axis, between 0 and 1. Errors in the fault location equation's variables may cause M_l to move off the real axis and into the complex plane. Note that a fault at a busbar is equivalent to a value of M_l equal to 0 or 1, and is thus covered by considering fault location solutions in the range [0,1].

It should be noted that M_l is a random variable because it is a function of erroneous voltage measurements and estimates. At an arbitrary point m_l along the lth line, for a specific monitor k, the kth monitor's measurement errors (and the errors of estimates performed using DSSE) can be used define the probability distribution of M_l as $f_{M_l^{(k)}}(m_l^{(k)})$, where the superscript (k) indicates that the distribution of M_l *is* formulated from measurements at the kth monitor only.

Since each monitor measurement and pre-fault estimate can be considered as independent, the value of a multivariate probability density function can be calculated as shown by (5.29).

$$f_{M_l^{(1)}M_l^{(2)}\ldots M_l^{(N)}}(m_l^{(1)}\cdot m_l^{(2)}\cdot\ldots\cdot m_l^{(3)}) = f_{M_l^{(1)}}(m_l^{(1)})\cdot f_{M_l^{(2)}}(m_l^{(2)})\cdot\ldots\cdot f_{M_l^{(N)}}(m_l^{(N)}) = f_{M_l}(m_l) \tag{5.29}$$

If it is noted that $f_{M_{l_j}}(m_{l_j})$ is only of interest where m_{lj} is real and between 0 and 1, the distribution of M_{lj} can be re-written as a conditional distribution as shown in (5.30).

$$f_{M_l}^{[0,1]}(m_l) = f_{M_l}(m_l \mid M_l \in [0,1]) = \frac{f_{M_l}(m_l)}{\int_{\Re_{[0,1]}} f_{M_l}(m_l)dm_l} \forall m_l \in [0,1] \tag{5.30}$$

$\Re_{[0,1]}$ is a line in the complex plane along the real axis between 0 and 1. Equation (5.30) describes the probability distribution for the fault location along the lth line conditional that the fault occurred along the lth line in the network.

Equation (5.30) can be extended to a multi-line system to find the probability density function for the fault location M_l on a specific line l contained within a set of all lines L ($l \in$ L).

$$f_{M_{l_j}}^{[0,1]}(m_{l_j}) = \frac{f_{M_{l_j}}(m_l)}{\sum_{l_i \in L}\int_{\Re_{[0,1]}} f_{M_{l_i}}(m_{l_i})dm_{li}} \forall m_{l_j} \in [0,1] \tag{5.31}$$

$$\sum_{l_j \in L}\int_0^1 f_{M_{l_j}}^{[0,1]}(m_{l_j}) = 1 \tag{5.32}$$

$$f_F = \left\{ f_{M_{l_1}}^{[0,1]}(m_{l_1}),\ f_{M_{l_2}}^{[0,1]}(m_{l_2}),\cdots,\ f_{M_{l_N}}^{[0,1]}(m_{l_N}) \right\} \tag{5.33}$$

$$f_F^{(k)} = \left\{ f_{M_{l_1}^{(k)}}^{[0,1]}(m_{l_1}^{(k)}),\ f_{M_{l_2}^{(k)}}^{[0,1]}(m_{l_2}^{(k)}),\cdots,\ f_{M_{l_N}^{(k)}}^{[0,1]}(m_{l_N}^{(k)}) \right\} \tag{5.34}$$

Equation (5.31) describes the probability of the fault location being found on the l_jth line in the network. Equation (5.32) shows that if each function is summed and integrated over all fault locations in all lines, the resulting integral is equal to unity. The probability density function which describes the fault location, F, is given by (5.33), as

a set of continuous functions for each of the lines in the network. In (5.34) the same probability density function is shown for only one specific monitor *k*.

5.3.2 Probabilistic Voltage Sag Profile Estimation

Voltage sag profile estimation involves estimating the resulting voltage sag profile from a limited number of monitors. This is accomplished by calculating the voltage at a non-monitored bus *i* using equations (5.23) to (5.25).

The voltage sag profile estimation equations calculate sequence voltages ($V_i^{(0)(k)}$, $V_i^{(1)(k)}$ and $V_i^{(2)(k)}$) which can be transformed into the phase voltages by applying the Fortescue transformation. The aim is to calculate the distribution of the magnitude of the phase voltages ($|V_i^{(a)(k)}|$, $|V_i^{(b)(k)}|$ and $|V_i^{(c)(k)}|$), $f_{|V_i^{(j)(k)}|}\left(\left|v_i^{(j)(k)}\right|\right)$, where $j \in \{a,b,c\}$.

Since each monitor is independent, a multivariate probability distribution can be formed for the voltage magnitude at the *i*th bus as shown in (5.35).

$$f_{|V_i^{(s)(1)}|\cdots|V_i^{(s)(N)}|}\left(\left|v_i^{(s)(1)}\right|,\cdots,\left|v_i^{(s)(N)}\right|\right) = f_{|V_i^{(s)(1)}|}\left(\left|v_i^{(s)(1)}\right|\right)\times\cdots\times f_{|V_i^{(s)(1)}|}\left(\left|v_i^{(s)(N)}\right|\right) = f_{|V_i^{(s)}|}\left(\left|v_i^{(s)}\right|\right) \quad (5.35)$$

The PDF $f_{|V_i^{(s)}|}\left(\left|v_i^{(s)}\right|\right)$ describes the voltage magnitude at the *i*th bus in the *s*th sequence estimated using information from all of the monitors within the network.

5.3.3 Estimating the PDFs for Fault Location and Voltage Sag Profile

The PDF distribution of f_F and $f_{|V_i^{(j)}|}\left(\left|v_i^{(j)}\right|\right)$ can be estimated by performing a Monte Carlo simulation with non-parametric kernel density estimation [12]. For each monitor, the Monte Carlo simulation is run by drawing zero, positive and negative sequence voltages from the distributions shown in equations (5.26) and (5.27). For each busbar where the voltage was estimated using DSSE, voltages are drawn from the distribution shown in equation (5.28). It should be noted that the pre-fault voltages found from DSSE may be correlated.

5.3.3.1 PDFs for Fault Location

Each time the voltages are drawn, the fault location equations are solved. The simulation generates a set of fault location estimates for each of the monitors in the network. Although the fault location solution should always be between 0 and 1, it is mathematically possible (and indeed likely) that the solution may be complex, exceed the bounds of 0 to 1 or be found in multiple positions on the same or other line. In the case where a monitor's set of solutions are not entirely real, the probability density

function can be estimated using bi-variate kernel density estimation [12]. If the solutions are entirely real, kernel density estimation in a single dimension will suffice. The total PDF for the fault location, f_F, can be calculated by multiplying together the kernel density estimated PDFs for all of the monitors.

5.3.3.2 PDFs for Voltage Sag Profile Estimation

The probability density function of the voltage magnitude at the *i*th bus, in the *j*th phase, measured by the *k*th monitor, across all lines ($f_{\left|V_p^{(j)(k)}\right|}(\left|V_p^{(j)(k)}\right|)$) can be found by using the results of fault location. Firstly, the probability density function of the fault location, $f_{M_L^{(k)}}^{[0,1]}(m_l^{(k)})$, is used to generate values of $M_l^{(k)}$ by selecting random values from this distribution. Next, during fault voltages $V_k^{(s)(k)}$ and pre-fault voltages $V_k^{(s)(k)'}$ are selected using the distributions shown in equation (5.26) and (5.27). Thirdly, the selected values are used to compute a set of during fault voltage estimates using equations (5.23) to (5.25). Lastly, a probability density function can be estimated by using bivariate kernel density estimation. The combined PDF for the voltage sag profile, $f_{\left|V_i^{(j)}\right|}(\left|v_i^{(j)}\right|)$, can be calculated by multiplying together the kernel density estimated PDFs for all of the monitors.

5.3.4 Identifying a Faulty Monitoring Device

If one of the monitors is faulty, the fault location and voltage sag profile estimates could be detrimentally affected. The location of a faulty monitoring device can be identified by considering the statistical properties of the errors.

A single faulty monitor can be identified through a multi-step process. The first stage is to run a test to establish the presence of bad data in a set of monitor readings. This test is initially carried out using all information available from all monitors in the network. If this test fails, the test is run again using N_m-1 measurements from N_m sets of monitors where the members of the N_m sets each include a different combination of monitors. If one of these tests passes, then it can be deduced that the faulty monitor is the device absent from this test.

These two tests are outlined in the following flow chart:

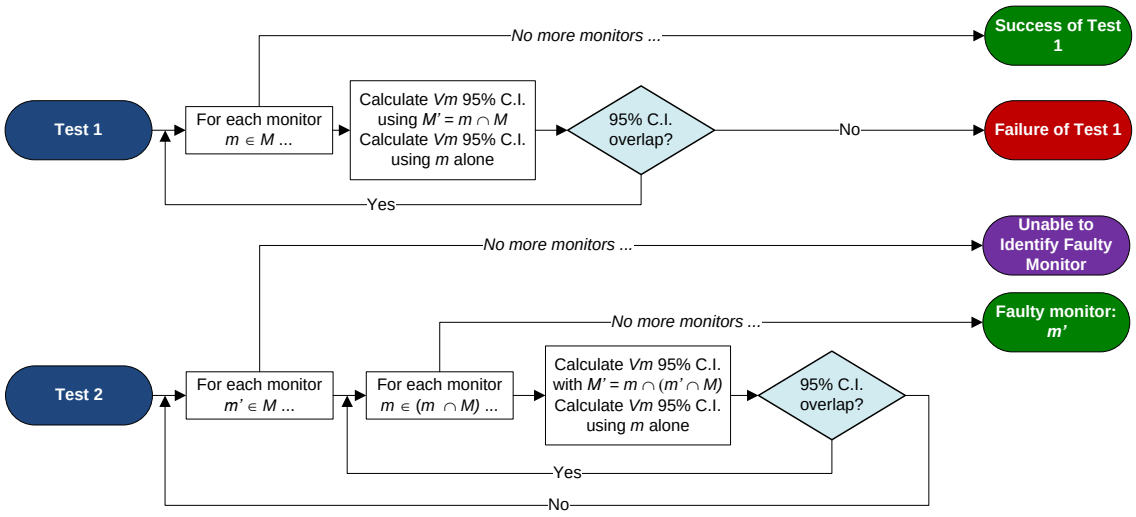


Fig. 5.6. The two tests required to identify a faulty monitor.

In both tests, a statistical test is performed which looks for overlap between the 95% confidence interval of a measured voltage magnitude, $\left|\hat{V}_m^{(s)(m)}\right|$ and an estimated voltage, $\left|\hat{V}_m^{(s)(M')}\right|$, obtained through probabilistic voltage sag estimation without using the measurements from the *m*th monitoring device. If the confidence intervals don't overlap, this implies that either the true value of the voltage, $\left|V_m^{(s)}\right|$, falls outside the 95% confidence interval of one or both of $\left|\hat{V}_m^{(s)(m)}\right|$ and $\left|\hat{V}_m^{(s)(M')}\right|$ or alternatively, that there was some other bias or error introduced into the measured or estimated voltages.

A set of overlapping confidence intervals does not necessarily imply that $\left|V_m^{(s)}\right|$ definitely resides within the 95% confidence interval of both estimates. With a large sample of two independently identically distributed normal random variables, two 95% confidence intervals will overlap 99.4% of the time [159], and not overlap 0.6% of the time. It the case of VSPE, the measured voltage is normally distributed, but the estimated voltage typically follows a non-parametric distribution. To provide an indicative guide for the likely overlap rate when performing VSPE, test number one was run using a set of 1000 measurements picked from a distribution which follows the expected error of each of the monitors shown in Fig. 5.3. This test was found to fail 16 times, indicating that the test is likely to provide false alarms roughly 1.5% of the time. Higher failure rates should indicate that there may be a problem with one of the monitors.

The aim of the second test is to identify which of the monitors is most likely to be faulty. The second test assumes that only one of the monitors is faulty. If more than one monitor is faulty, the test is likely to end in an inconclusive result (shown in purple in

Fig. 5.6). It also assumes that only one set of monitors will pass the overlap test; which is a significant assumption since it is possible to pass the overlap test without the $\left|V_m^{(s)}\right|$ being within 95% confidence interval of $\left|\hat{V}_m^{(s)(M')}\right|$ or $\left|\hat{V}_m^{(s)(m)}\right|$. Despite these two assumptions, the test works, and it will be shown to correctly identify a faulty monitor in the results which follow.

5.4 Case Studies

A section of the 295 bus distribution network can be used to illustrate the concept of probabilistic fault location and probabilistic voltage sag profile estimation. Fig. 5.3 shows the section of network chosen. It covers a total of 100 busbars connected to feeder L.

Three monitors were assumed to be installed in the network. The monitoring locations chosen corresponded to a subset of the optimal monitoring locations found through monitor placement optimisation (which are derived in Chapter 7). It should be noted that the positions of monitors for this study are not important; the probabilistic fault location method could be applied to any arbitrary set of monitors. All the monitors were assumed to measure three phase voltage, real power and reactive power.

Three fault case studies were chosen; one for a single line to ground fault between buses 147 and 146, a second for a three phase fault between buses 147 and 146 and a third with a three phase fault between buses 194 and 195. All three faults were simulated at m=0.206 along the line. Fig. 5.3 shows the location of all three case studies.

The aim of the first two case studies is to predict the voltages at buses 141, 143, 145, 146 and 147 during the fault. Table 5.1 shows the during fault voltages for both case studies for each of these 5 buses. For the first case study, the results will also highlight how bad data from a faulty monitoring device can be identified.

The aim of the third case study is to show how the proposed probabilistic approach to voltage sag localisation helps to eliminate multiple fault location estimates.

TABLE 5.1. DURING FAULT VOLTAGES FOR CASE STUDY 1 & 2.

Bus	Case Study 1: SLG Fault Between Bus 147 & 146			Case Study 2: 3 Phase Fault Between Bus 147 & 146		
	Phase A	Phase B	Phase C	Phase A	Phase B	Phase C
Bus 147	0.117	0.956	1.293	0.0122	0.0122	0.0122
Bus 146	0.129	0.956	1.305	0.0010	0.0010	0.0010
Bus 145	0.130	0.956	1.304	0.0006	0.0006	0.0006
Bus 143	0.130	0.956	1.303	0.0004	0.0004	0.0004
Bus 141	0.131	0.954	1.302	0.0002	0.0002	0.0002

5.5 Results

The following two case studies illustrate the probabilistic fault location and voltage sag profile estimation algorithms in the 295 bus generic distribution network.

5.5.1 Case Study 1: Single Line to Ground Fault between Bus 147 & 146

For the single line to ground fault, the key parameter in equation (5.17) is the ratio of the zero sequence voltage measurement to the negative sequence voltage measurement, G. Fig. 5.7 shows a plot of G_{232} (equation (5.17)) which shows how the measured value of G_{232}, the true value (representing M_{232}=0.206), and the expected variation in G_{232} around the measured point. The contour lines represent the lines of constant probability. The variation of M_l in the complex plan is represented by the sections of lines shown for each of four power lines of interest.

The probability distribution for the fault location $f_l^{(232)}$ is shown in Fig. 5.8. This is a 2-dimensional representation of the height of the contours in Fig. 5.7 as they cut through the G plane with different values of M_l along each of the 4 lines.

Fig. 5.8 highlights that a monitor at bus 232 cannot accurately locate the fault between bus 147 and 146. Indeed, $f_l^{(232)}$does not vary significantly across all of the lines, and the peak between bus 146 and 145 does not correspond to the true fault location. This evidence suggests that a monitor at bus 232 is poor at locating faults within this section of the power network.

Fig. 5.9 and Fig. 5.10 show G_{130} and $f_l^{(130)}$ for a monitor placed at bus 130 whilst Fig. 5.11 and Fig. 5.12 show G_{204} and $f_l^{(204)}$ for a monitor placed at bus 204. The distribution of $f_l^{(130)}$ and $f_l^{(204)}$ are now very tight, and the fault is accurately localised at m_{232}=0.206 along the line between buses 146 and 147. The variation of the error in G_{130} and G_{204} relative to the solutions for G along the line connecting buses 147 to 146 is relatively small, and markedly different to the relative variation for bus 232. The error distribution for both G_{130} and G_{204} cuts the line at almost exactly m=0.2.

There is a significant difference between the solutions of G_{130} and G_{204} along the lines connecting buses 146 to 141. The monitor at bus 204 sees the lengths of these connections as negligibly short distances when compared to the length of the line connecting bus 147 to 146. In practical terms, this means a monitor placed at bus 204 will not be able to accurately localise faults between buses 146 to 141, or near to the end of the line connecting buses 147 to 146.

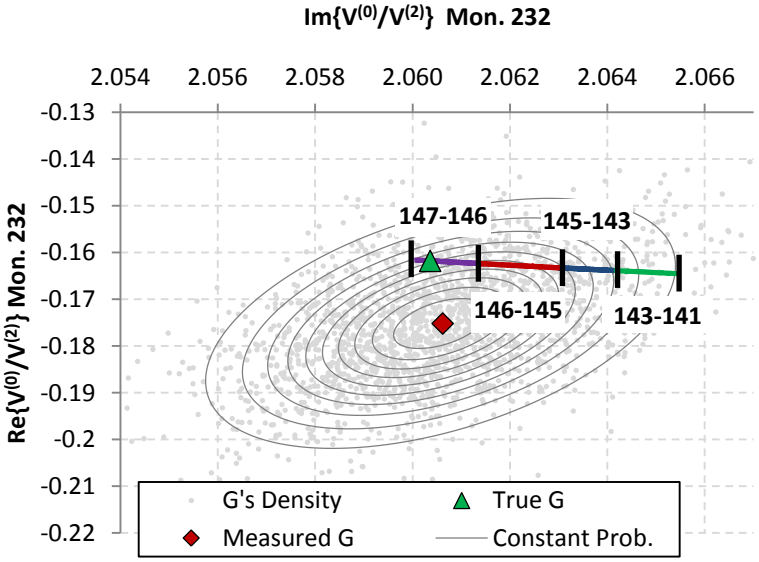


Fig. 5.7. The range of expected values of G for V(0)/V(2) for a monitor placed at bus 232.

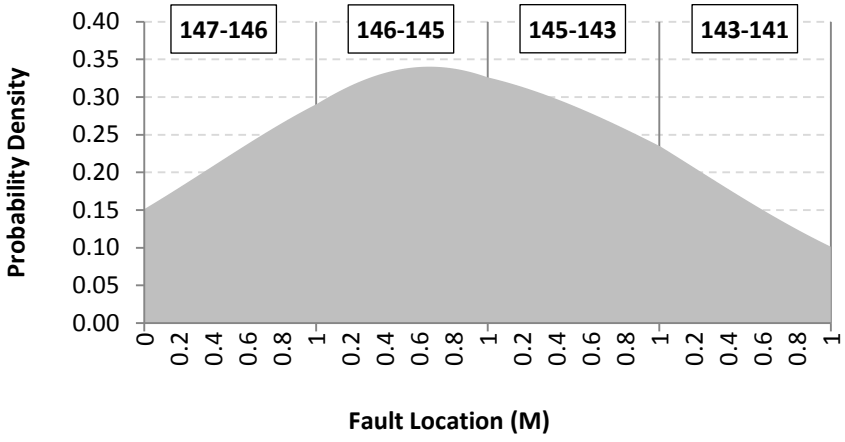


Fig. 5.8. The probability distribution of the fault location when using a monitor at bus 232 to locate the fault.

As well as estimating the location of a fault using a single monitor, the proposed method is able to combine multiple monitoring solutions into a single estimate for the fault location. Table 5.2 highlights medians, 50% and 95% intervals for the fault location using each monitor and subsequently combing the estimates from every monitor. A 95% interval correctly spans the true fault location for all three monitors

whereas the 50% interval only spans the correct location for a monitor placed at bus 204. The combined estimate for the fault location is much closer to the fault point than any of the other monitors, and the intervals are approximately 45% narrower than for the best performing individual monitor. (Note that taking intervals across the dataset is only valid if the lines are adjacent to one another. If the considered lines are not adjacent, then intervals should be taken on a line by line basis because there is no reason to believe that the PDF will be a continuous function.)

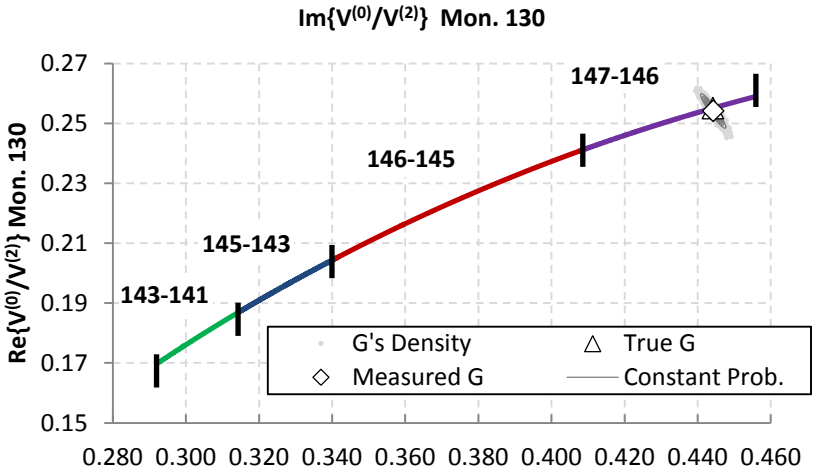


Fig. 5.9. The range of expected values of G for $V^{(0)}/V^{(2)}$ for a monitor placed at bus 130.

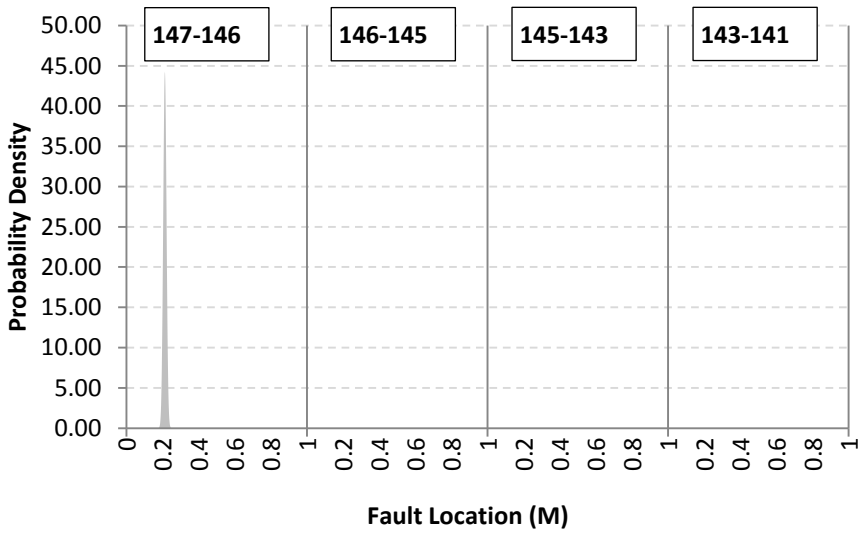


Fig. 5.10. The probability distribution of the fault location when using a monitor at bus 130 to locate the fault.

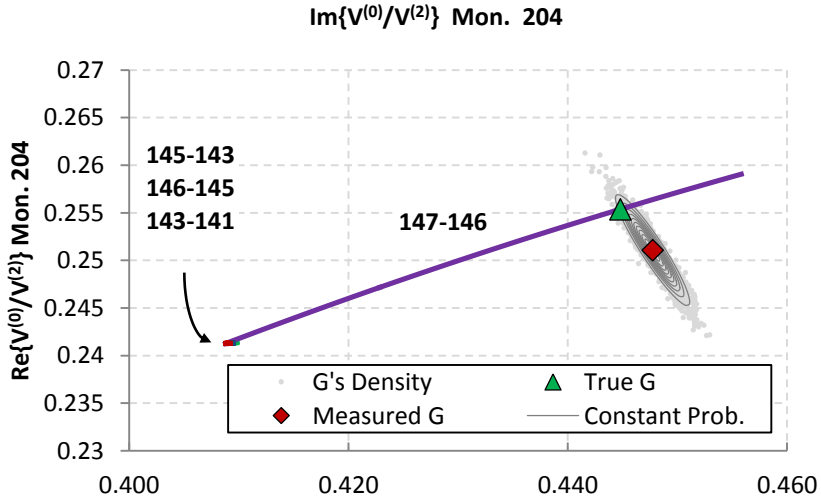


Fig. 5.11. The range of expected values of G for $V^{(0)}/V^{(2)}$ for a monitor placed at bus 204.

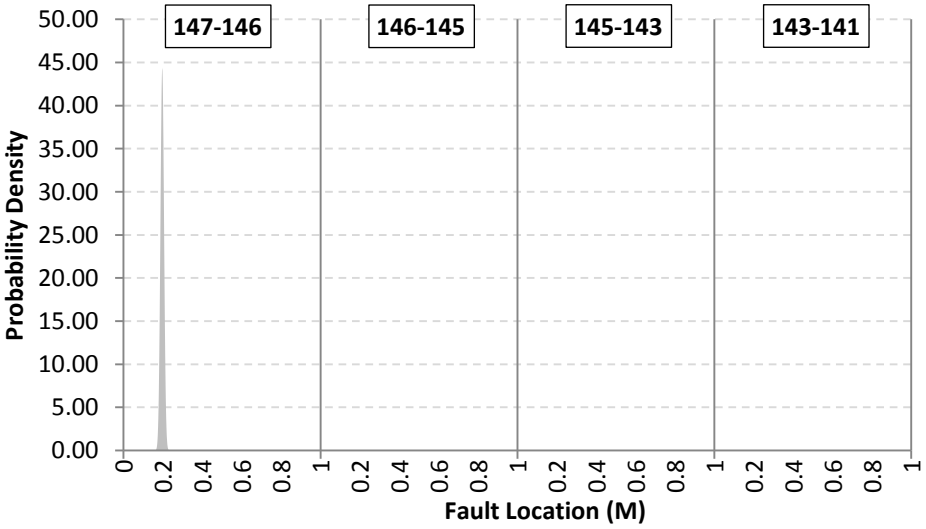


Fig. 5.12. The probability distribution of the fault location when using a monitor at bus 204 to locate the fault.

TABLE 5.2. VARIATION IN THE VALUES OF M FOR DIFFERENT MONITORS FOR A FAULT BETWEEN BUS 146 AND 145.

Monitor	Median M	50% of the Data			95% of the Data		
		Upper M	Lower M	Range M	Upper M	Lower M	Range M
Bus 232	1.235 (Line 146 to 145)	2.337	0.833	1.504	3.653	0.101	3.552
Bus 130	0.213 (Line 147 to 146)	0.219	0.207	0.012	0.230	0.196	0.034
Bus 204	0.197 (Line 147 to 146)	0.203	0.191	0.012	0.214	0.180	0.034
All Monitors	0.208 (Line 147 to 146)	0.211	0.205	0.006	0.2180	0.199	0.0190

The values of M in the table are shown between 0 and 1, 1 to 2, 2 to 3 and 3 to 4 for lines connecting buses 147 to 146, 146 to 145, 145 to 143 and 143 to 141 respectively.

The accuracy of the fault location estimate directly impacts on the accuracy of a voltage sag profile estimate for a non monitored bus (as described in equations (5.23) to (5.25)).

The probability distributions for the fault location were used to build an estimated probability distribution for the voltage at each of the five monitored buses. Fig. 5.13 illustrates the shape of the estimated probability density functions for the voltage in phase A at bus 146 using each of the three monitors individually ($f_{|V_{146}^{(a)(130)}|}$, $f_{|V_{146}^{(a)(204)}|}$ and $f_{|V_{146}^{(a)(232)}|}$, equation (5.35))and the estimated probability density function using all the monitors combined ($f_{|V_{146}^{(a)}|}(|v_k^{(a)}|)$). The density function for all monitors is clearly more accurate than either of the individual monitors, narrowing in on the true value of 0.129 per unit. The inaccuracy in fault location using only a monitor at bus 232 is apparent in $f_{|V_{146}^{(a)(232)}|}$, whose distribution peaks at around 0.19 per unit, 0.06 per unit from the true during fault voltage.

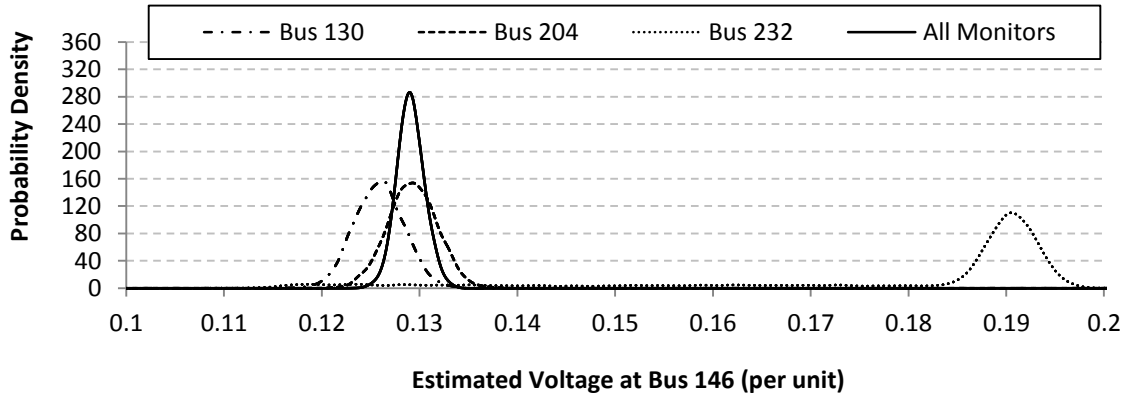


Fig. 5.13. The estimated probability distribution of the voltage magnitude in phase A at bus 146 using measurements from bus 130, 204, 232 and all buses combined for a single phase to ground fault.

Fig. 5.14, Fig. 5.15 and Fig. 5.16 show the distribution of the estimated phase voltage for each of the five buses. Placing a monitor at bus 232 yields the most inaccurate estimate for the during fault voltages at each of the 5 buses. For example, the 95% confidence interval for the voltage magnitude at bus 141 is almost 0.6 per unit using a monitor at bus 232 and approximately 0.02 per unit using a monitor at bus 130 or 204. This large range of inaccuracy in voltage magnitude estimation will have a significant impact if the data was to be used to estimate whether or not equipment attached to that bus tripped during the sag.

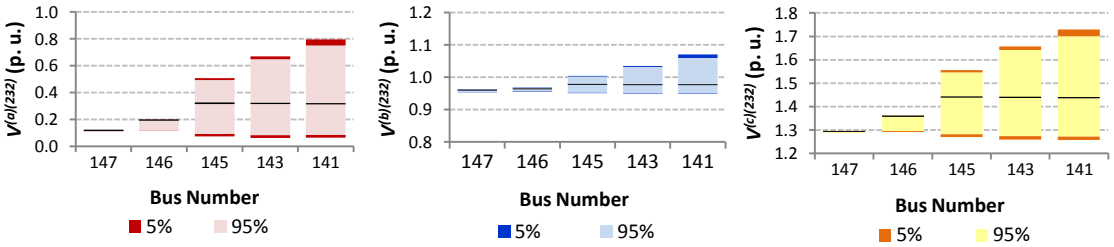


Fig. 5.14. The probability distribution of the voltage magnitude in phase A (left), phase B (middle) and phase C for buses 147, 146, 145, 143 and 141, estimated using a monitor placed at bus 232.

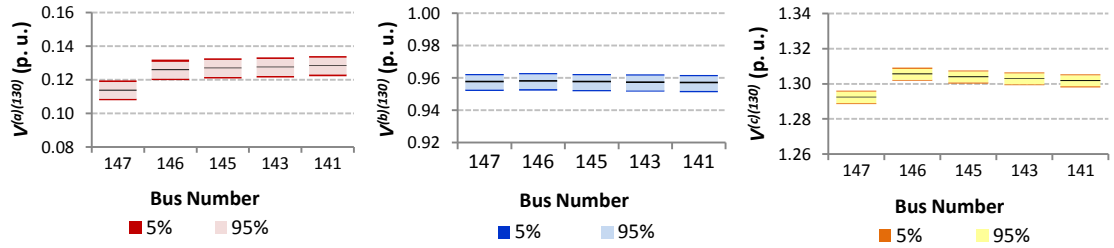


Fig. 5.15. The probability distribution of the voltage magnitude in phase A (left), phase B (middle) and phase C for buses 147, 146, 145, 143 and 141, estimated using a monitor placed at bus 130.

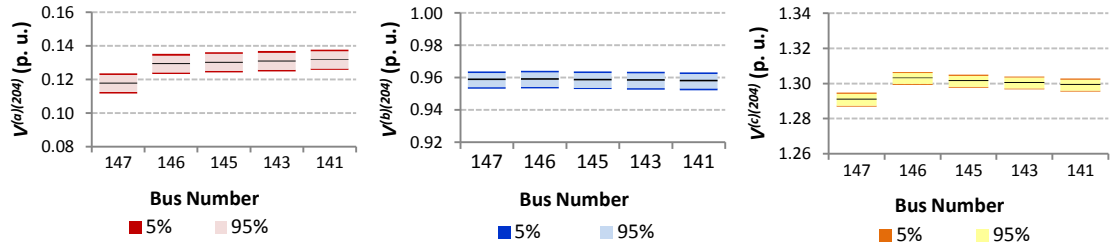


Fig. 5.16. The probability distribution of the voltage magnitude in phase A (left), phase B (middle) and phase C for buses 147, 146, 145, 143 and 141, estimated using a monitor placed at bus 204.

Fig. 5.17 shows the estimated distribution of the voltage when all 3 monitors are used to predict the voltage magnitude at each of the 5 buses. As for fault location, the accuracy improvement is significant: for phase A at bus 147, the width of the 95% confidence interval is 50% smaller the best performing single monitor.

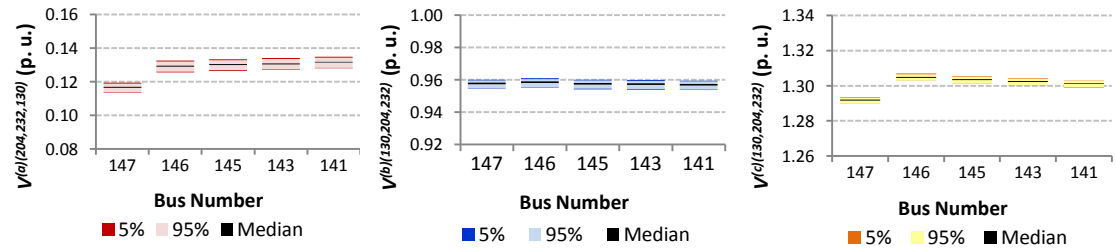


Fig. 5.17. The probability distribution of the voltage magnitude in phase A (left), phase B (middle) and phase C for buses 147, 146, 145, 143 and 141, estimated using all monitors to estimate the voltage.

5.5.2 Case Study 2: Three Phase Fault between Bus 147 & 146

Fig. 5.18 to Fig. 5.21 shows the PDFs for the fault location (f_l) using each monitor individually and all monitors together to locate the fault. The most accurate estimate for the fault location is obtained using data from a monitor placed at bus 232 whilst the least accurate monitor estimate is formulated using data from a monitor at bus 130.

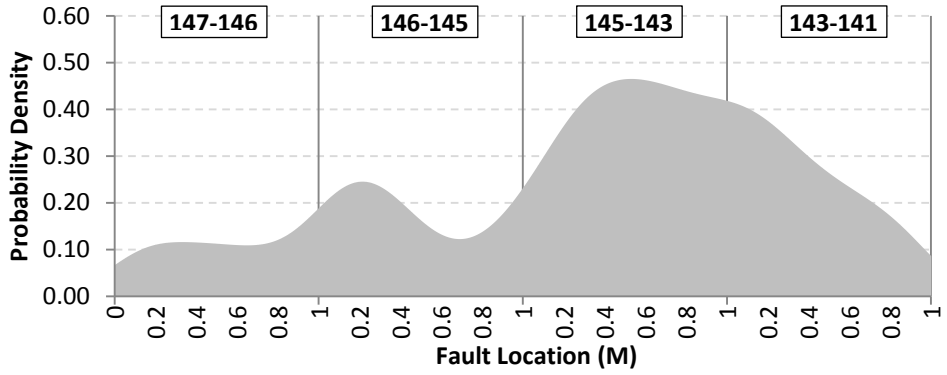


Fig. 5.18. The PDF of the fault location when using a monitor at bus 130 to locate the fault.

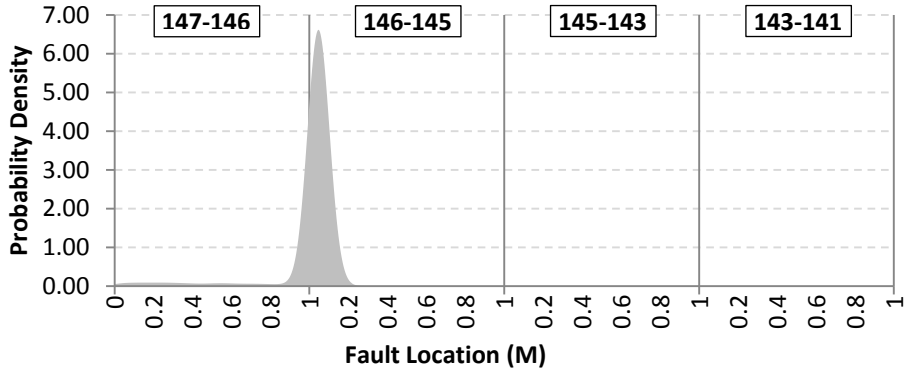


Fig. 5.19. The PDF of the fault location when using a monitor at bus 204 to locate the fault.

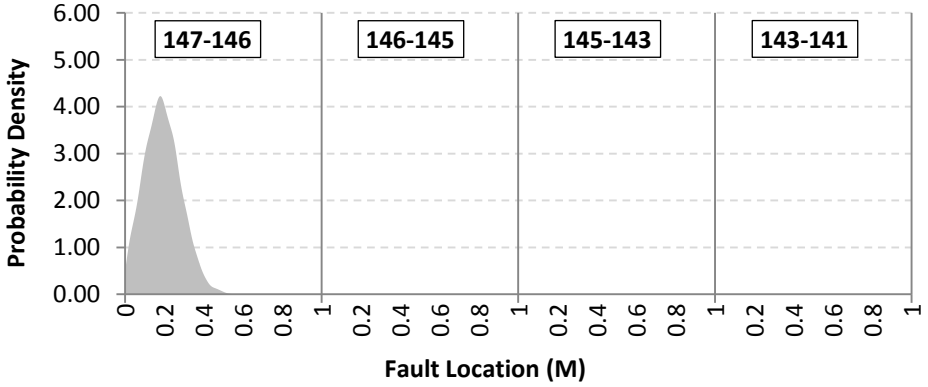


Fig. 5.20. The PDF of the fault location when using a monitor at bus 232 to locate the fault.

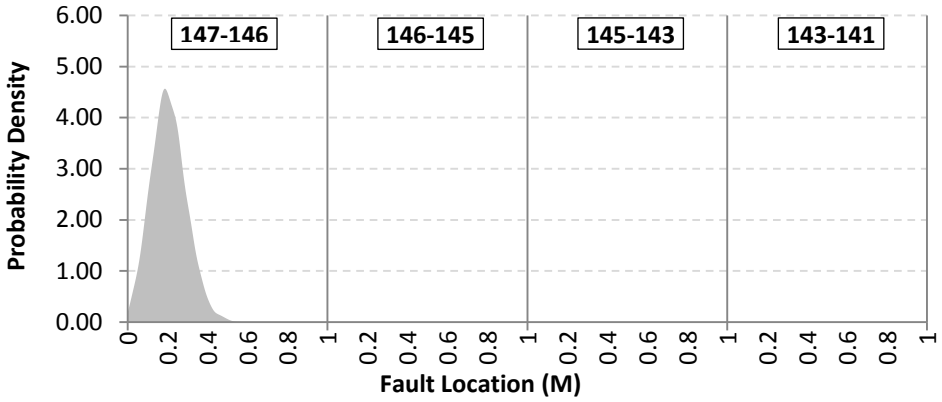


Fig. 5.21. The PDF of the fault location when using all monitors to locate the fault.

The shapes for each of the PDFs shown in Fig. 5.18 to Fig. 5.21 can be explained qualitatively by considering the network topology and monitor's location with respect to the fault.

A three phase fault with zero fault resistance between bus 146 and bus 147 causes all phase voltages to drop to approximately zero downstream of the fault location. When solving for M_l using a monitor at bus 130 (Fig. 5.18), the sensitivity of equation (5.15) is increased and the fault location estimate is consequently uncertain.

Fig. 5.19 shows that the solutions for m using a monitor at bus 204 yield solutions on both the line connecting buses 146 to 147 and the line connecting buses 146 to 145. This can be explained by noting bus 204 nearest connection to the network section of interest is via bus 146. Solving equation (5.15) therefore yields two solutions, which both represent identical impedance paths between bus 204 and the fault location. The line connecting buses 146 to 145 has over twice the impedance of the line connecting buses 146 to 147. This causes the $f_l^{(204)}$to peak in Fig. 5.19 at approximately m=0.1 along the line connecting bus 146 to 147. The PDF still correctly overlaps the true fault location on the line connecting bus 146 to 145 at m=0.206, but the equation which governs the line connecting bus 146 to 145 is non-linear and sensitive such that equation (5.15) yields solutions outside of the length of the line, thus reducing the overall area of the PDF on the line.

Fig. 5.20 shows the results of using a monitor at bus 232 to locate the fault. The PDF peaks at 0.2, and there are no aliased fault locations on other lines. Bus 232 is connected to the primary in-feed for this area of the network, and is thus a relatively strong bus; it experiences a drop in voltage magnitude of 64% compared with the near 100% drop in voltage magnitude at bus 130. Equation (5.15) is therefore relatively insensitive to errors in monitoring and an accurate estimate for the fault location is calculated.

Fig. 5.22 shows the results of voltage sag profile estimation when using each of the monitors to estimate the voltage magnitude at bus 146. Using information from a monitor placed at bus 232 yields the best performance, with a monitor at bus 204 also correctly pin-pointing the voltage within the 0 to 0.02 per unit range. Although the voltage profile predicted by a monitor at bus 130 covers a large range between 0 and 0.3 per unit, any of these voltages would definitely cause serious disruption to equipment attached to bus 146 and the error is therefore insignificant in practical terms.

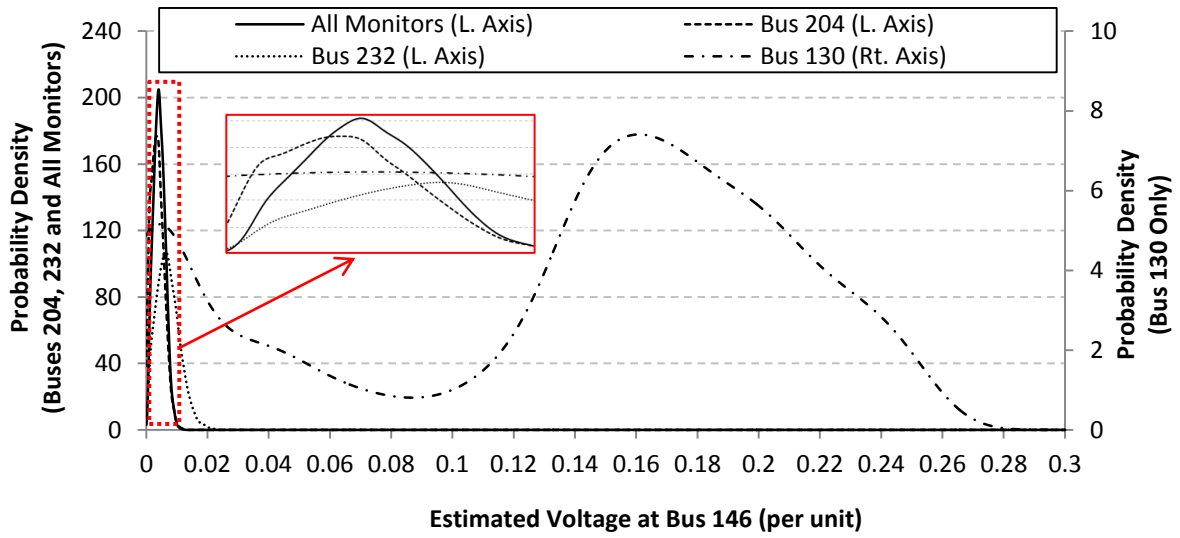


Fig. 5.22. The estimated probability distribution of the voltage magnitude in all three phases at bus 146 using measurements from bus 130, 204, 232 and all buses combined for a three phase fault. The inset shows a zoomed diagram of voltages less than 0.01 per unit.

Fig. 5.23 shows the 5%, 95% and median of the voltages for each of buses147, 146, 145, 143 and 141 in phases A, B or C. It is clear from Fig. 5.23 (C) and Fig. 5.23 (D) that estimating the voltage sag profile using a monitor placed at bus 232 or 204 yields good results, because the range distribution of the voltage magnitudes is contained within a tight range. Fig. 5.23 (A) shows that using monitoring information from bus 130 provides the worst results, whilst the results which synthesise information from all monitors are the best.

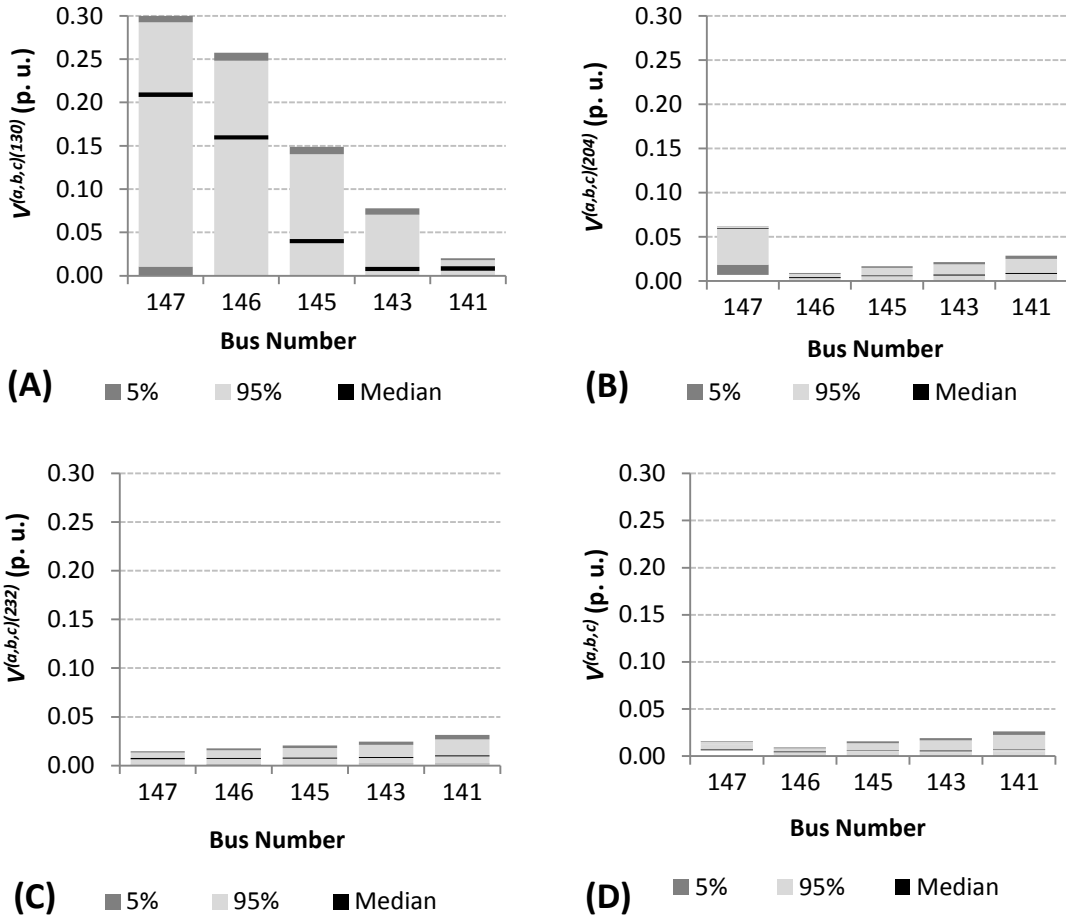


Fig. 5.23. The probability distribution of the voltage magnitude in all phases for a three phase fault for buses 147, 146, 145, 143 and 141, estimated using a monitor placed at (from top left to bottom right) bus 130, 204, 232 and all of these monitors.

It is interesting to note that the best monitor at predicting the during fault voltages for this specific single line to ground faults was a monitor at bus 130 and the worst was a monitor at bus 232. The results would also be the same if the single line to ground fault was in phase B or C. For three phase faults, the converse is true. This could be due to the asymmetry of single phase to ground faults which means that higher during fault voltages will be registered in some phases of bus 130 during a single phase to ground fault compared with a three phase fault. It indicates that the best location to detect single line to ground faults is not necessarily the same as the best location for three phase faults.

5.5.3 Case Study 3: Eliminating Multiple Voltage Sag Location Estimates

The benefits of using the proposed method to identify a unique solution for the source of the voltage sag can be explored by considering a three phase fault simulated between bus 194 and 195 (as shown in Fig. 5.3). This section of network is shown in more detail in Fig. 5.24.

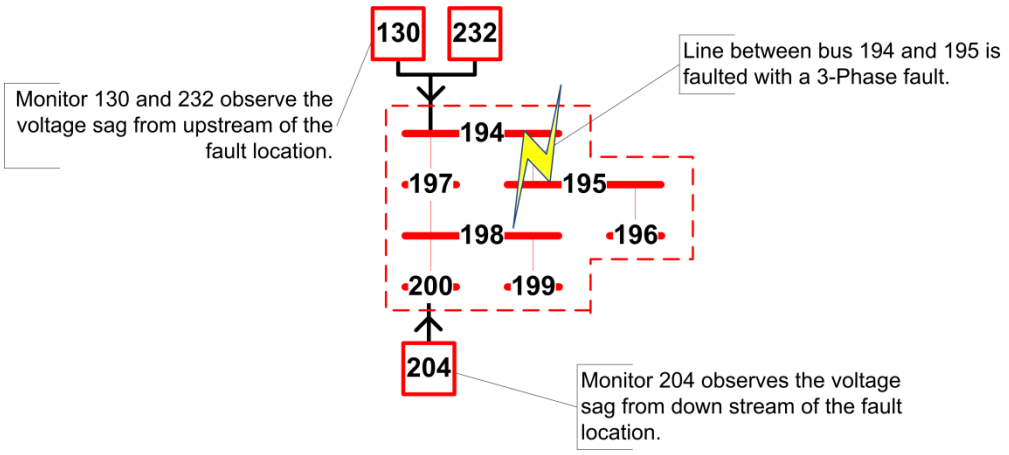


Fig. 5.24. A three phase fault simulated between buses 194 and 195. Monitors are buses 204, 130 and 232 observe the voltage sag.

Solving the three phase fault location equations for each monitor independently yields a probabilistic solution for the fault location, as observed by each monitor alone. The probability distribution for each of the three monitors and the combined distribution of all three monitors is shown in Fig. 5.25.

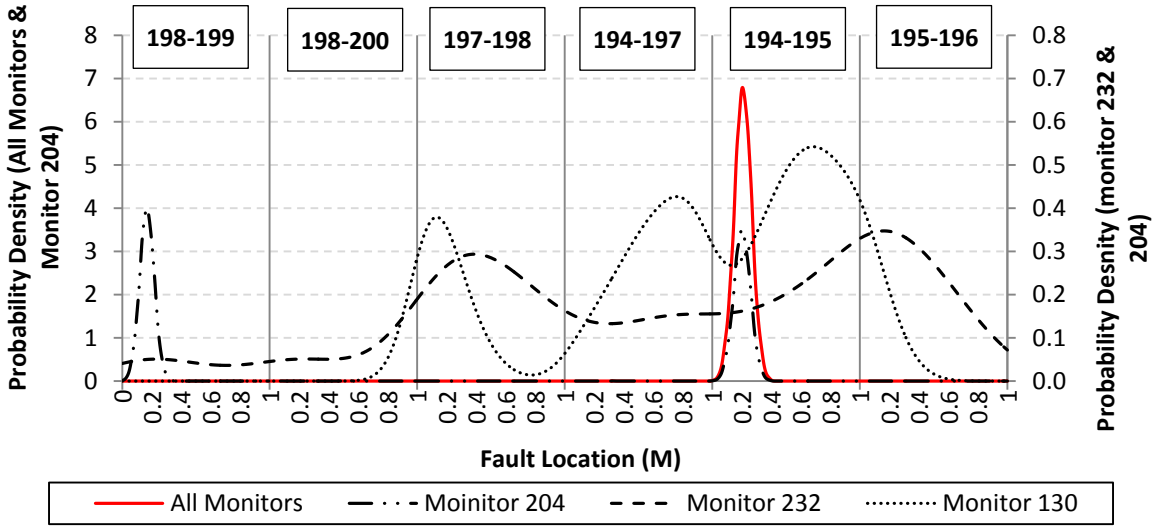


Fig. 5.25. The probability density across 6 lines for a three phase fault at 0.2 per unit along the line connecting bus 194 to bus 195.

Fig. 5.25 shows that each of the three monitors predict that the fault could have occurred at multiple locations within the network. For example, the density function for the monitor installed at bus 130 shows three peaks in the PDF for the fault location on the lines connecting buses 197 to 198, buses 194 to 197 and buses 194 to 195. The monitor at bus 130 observes the network upstream of the fault and sees two impedance paths beyond bus 194; one to bus 197 and one to bus 195 (see Fig. 5.24). The monitor at bus 130 can't distinguish between these two paths, and this ambiguity is reflected in the multiple peaks of the probability distribution shown in Fig. 5.24. The PDF for the

monitor at bus 232 also shows ambiguity in the fault location because it also observes multiple impedance paths from upstream of the fault location. Fig. 5.24 also shows multiple peaks in the PDF for the monitor at bus 204. This is caused because the monitor sees two similar impedance paths from bus 198 to bus 196 and from bus 198 to bus 199.

The strength of the proposed technique at helping to eliminate multiple voltage sag location solutions is highlighted in the combined graph for all monitors shown in Fig. 5.25. By using the information available from all monitors, the PDF shows only one peak at the true fault location.

5.5.4 Identifying a Faulty Monitoring Device

The methodology for finding a faulty monitoring device can be demonstrated through the results of the first case study. Suppose that the monitor at bus 204 measures the voltages in the positive, negative and zero sequence with a fixed offset error of 0.005 per unit. Simulating the same single line to ground fault between bus 147 and 146, with the added error in monitor 204 voltage measurements causes the estimated fault location to shift by approximately 0.1 from m=0.2 to m=0.1.

Running the first test for the presence of bad data generates the 95% confidence intervals shown in Fig. 5.26. Although the mean difference in the estimated and measured voltages is always small, the 95% intervals don't overlap which indicates there could be a problem.

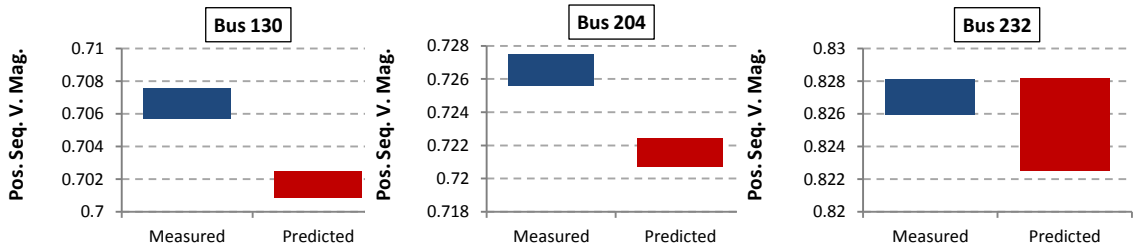


Fig. 5.26. The 95% confidence intervals for bus 130, 204 and 232 (left to right) when predicted using their own measurements independently (left hand series) and measurements at buses 204 & 232, 130 & 232 and 130 and 204 respectively whilst introducing at 0.005 per unit offset error at bus 204.

Upon the failure of the first test for a faulty monitor, the second test highlights which monitor is faulty. Fig. 5.27 shows the 95% confidence interval for each of the three sets of Nm-1 monitors selected in the first loop of the second test. The first four graphs Fig. 5.27 all involve measurements from bus 204, and each of these tests has at least one set

of positive sequence voltages that do not overlap. The right hand two graphs of Fig. 5.27 show the 95% confidence intervals overlapping. This evidence suggests that there is a fault with the measurements from bus 204.

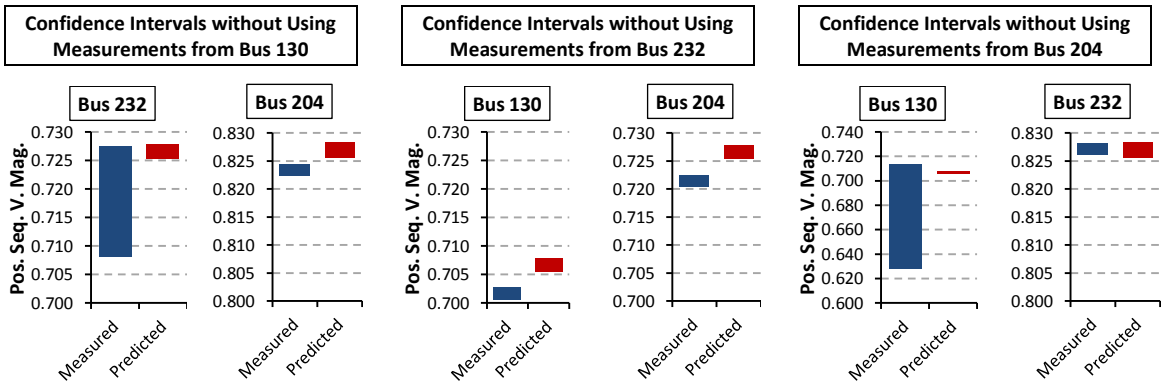


Fig. 5.27. The 95% confidence intervals for buses 130, 204 and 232 and the results of the second test which identifies a faulty monitor in the network.

5.6 Summary

This chapter presented a comprehensive method to identify a statistical distribution for both the fault location and the during sag voltage profile using only a small selection of arbitrarily accurate monitors. The method takes into account the potential variability of device measurements, and integrates relevant information from all monitoring devices in a network into a single statistical estimate. The method can be applied to three phase, single line to ground, line to line and double line to ground faults. Unlike existing techniques for fault location and voltage sag profile estimation, the method estimates probability distributions for both the fault location and voltage profile.

It was also hoped that the method would be able to work with less monitors than required for full observability. The results validated this aspiration, showing that the method was capable of estimating the probability distribution for the fault using only 3 monitors in a 94 bus section of a large 295 bus distribution network.

The research presented has several advantages over existing techniques. Some of the advantages of the method include:

- The ability to use information from one or more monitoring devices to produce a statistical estimate for the location or depth of a voltage sag
- The ability to integrate the accuracy of monitoring devices into a statistical estimate
- The flexibility to add newer and more accurate monitoring devices

- The capability to deal with the sensitivity of the fault location equations
- Increased accuracy versus traditional single or double ended fault equations techniques by synthesising information from more than two monitors
- The ability to eliminate multiple fault location estimates
- The ability to identify faulty monitors using statistical tests

The research presented in this chapter focuses on estimating the technical impact of a voltage. From a customer perspective, it is important to be able to identify how customers are being affected by voltage sags, and which customers are the worst served within the network. The next chapter discusses how the statistical estimation techniques for a voltage sag voltage profile estimation discussed in this chapter can be used to estimate the impact of voltage sags on customers.

6 Voltage Sags: Estimation of Impact on End Users

6.1 Introduction

In Chapter 4 and 5, two methods were developed which firstly classify and detect voltage sags and secondly locate and estimate their technical impact within a network. The next natural step for this process is to estimate the impact of voltage sags on customers within the network. This process is defined as *Voltage Sag Impact Estimation* as shown in Fig. 6.1.

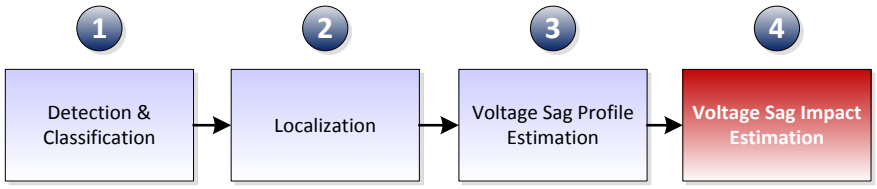


Fig. 6.1. The process of gaining an overview of voltage sags within a power network (stage 4: Voltage Sag Impact Estimation).

The overall theme of all the research presented in this thesis is the intelligent monitoring using a limited monitoring set. In the previous chapter, the discussion focussed on building a statistical model for both the location of a fault and its voltage sag magnitude profile across the network, using an arbitrary set of monitors. The statistical distribution of the depth of a voltage sag is a purely technical representation of the sag and it only presents a partial interpretation of the sag performance of a network. In order to present a full picture, the impact of sags on customers and their effects on consumer equipment

must be considered, whilst taking into account the inaccuracies of using data from an arbitrary set of monitors.

The principal aim of the research presented in this chapter is to introduce a new criterion to assess the impact of voltage sags on customers based on the physical behaviour of equipment. To accomplish this aim, the method must satisfy a number of requirements including:

1. Working with an arbitrary set of monitors
2. Dealing with uncertainty in attached equipment
3. Dealing with uncertainty in voltage sag characteristics

The research will build on the probabilistic approach for VSPE discussed in the previous chapter, to arrive at an overall statistical measure for the impact of voltage sags on customer equipment. The approach will be tested on the 295 bus network to validate its effectiveness.

6.2 Problem Definition

To consider how best to estimate the impact of a voltage sag, it is worth re-considering its characterisation. As described in the introduction, voltage sags are most often (if not exclusively in current industrial practice) characterized by two parameters: firstly, the voltage sag *magnitude* that defines the magnitude of the retained RMS voltage (expressed in per unit or as a percentage) and secondly the *duration* that defines the length of time that the voltage remains below a specified threshold (typically 0.9 per unit). Other voltage sag characteristics (previously discussed in the introduction) are excluded from this analysis as they typically do not feature in current industrial practice. Voltage sags themselves are stochastic in nature as they are influenced by a variety of random factors including, fault type, position and pre-fault voltage [17]. Though voltage sags can be caused also by motor starting, load variation and transformer energizing these were considered to be much less frequent causes of voltage sags and are excluded from the analysis in this research.

To incorporate the impact of voltage sags on consumers, the model developed in chapter 5 must be extended to include voltage sag duration to fully characterise a voltage sag. With duration added, the problem of estimating the impact of a voltage sag on customer equipment can be tackled by understanding the tolerance of equipment to voltage sags using equipment immunity curves.

6.3 Equipment Voltage Sag Tolerance Curves

The Information Technology Industry Council (ITIC) curve [160] is the most frequently used graphical metric that highlights the ability of connected equipment to withstand voltage sags. The ITIC curve defines acceptable and unacceptable power quality operating regions in terms of sag duration and sag magnitude (Fig. 6.2).

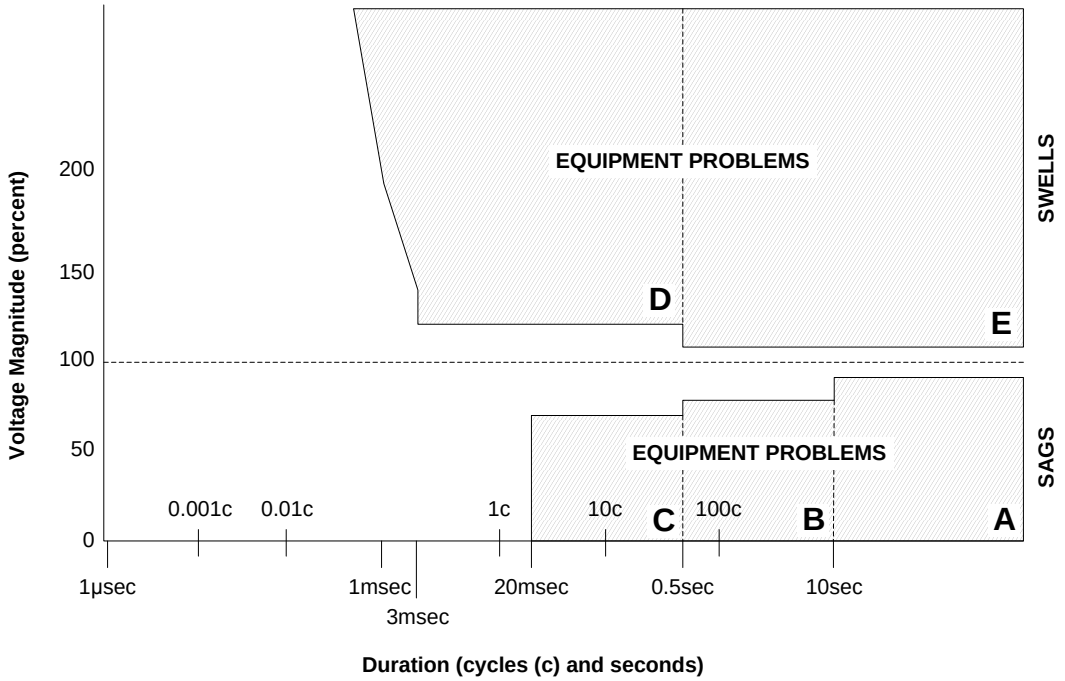


Fig. 6.2. ITIC power acceptability curve.

The ITIC voltage sag tolerance curve provides a valuable guide of equipment tolerance to voltage sags. Although powerful, the ITIC curve also has some important limitations and assumptions which should be considered here as they are particularly relevant to power quality studies. The curve assumes that equipment tolerance to voltage sags can be crisply divided into acceptable and unacceptable regions based on the historical and experimental performance of information technology equipment when exposed to sags of different magnitudes and durations. However, not all loads follow this precise acceptability curve [18, 161, 162] (also known as an immunity curve). Adjustable speed drives (ASDs), personal computers (PCs) and programmable logic controllers (PLCs) all have different responses when considering their immunity to voltage sags [106]. Differences in equipment, applications, years of use and operating conditions will also affect a device's voltage sensitivity [162]. A further limitation of the curve is that it is only strictly applicable to single phase loads.

Fig. 6.3 shows the general region of uncertainty for PCs, ASDs and programmable logic controllers (PLCs) [18]. The graph indicates that a sag deeper than V_{min} with a longer

duration than T_{max} will cause equipment to trip. However, there is a region (shaded area in Fig. 3) where it is uncertain as to whether the equipment will trip or not. The uncertainty in Fig. 6.3 is caused because the model is built from two types of equipment each with different immunity curves. This technique can be extended to combine all types of loads connected to a busbar to form a single immunity curve.

Fig. 6.3 also shows a shaded circle defining the expected range of duration and magnitude of a single sag event. The sag event is shown as a fuzzy region, since the exact depth of the sag is uncertain due to measurement error. The duration of the sag will also be uncertain. The bound for the duration will be defined predominantly by the protection devices located within the network section experiencing a voltage sag, though load dynamics (e.g. the presence of large numbers of induction motors) can also contribute to varying sag duration.

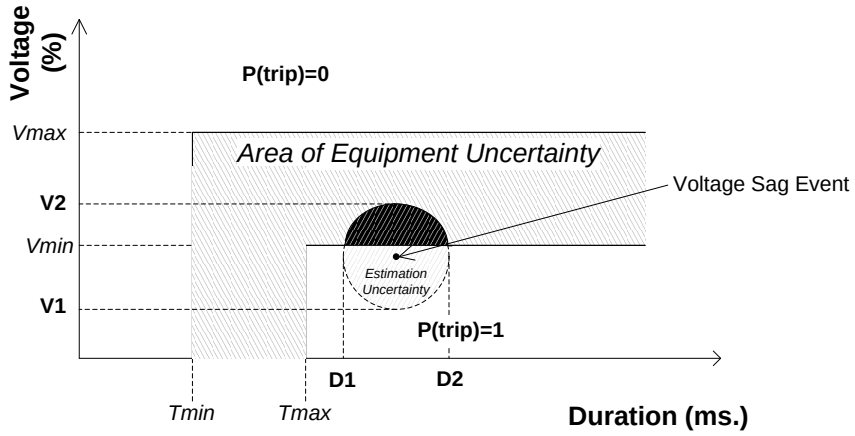


Fig. 6.3. The region of uncertainty as defined for sensitivity curves of PCs, PLC and ASDs[18]. The shaded region represents where it is uncertain whether equipment will trip.

At the network level, the expected number of trips caused by voltage sags is complicated by a number of factors. The distribution of expected sag durations at individual busbars in the network will vary, depending on the voltage of network section being studied (e.g. medium voltage (MV, 11kV in the UK[17]), high voltage (HV, 33kV in the UK[17]), extra high voltage (EHV, 132kV in the UK[17])) [11]. The total maximum expected sag duration at each location in the network will be typically defined by protection system setting. Each bus bar will have a diverse range of attached equipment which implies a different power acceptability curve [161], or even an uncertain combined acceptability curve [106]. The point on the wave at which a sag is applied as well as the phase angle jump during the sags are also important (though to a

[17] Voltages are line to line and indicative of UK distribution network standards.

lesser extent than sag magnitude and duration), and will affect the number of trips registered at busbar [106]. The depth of the voltage sag will ultimately be affected by the strength of the network surrounding the fault, the location of the fault, and the type of fault which caused the sag.

6.4 A New Metric for VSPE Accuracy: Sag Trip Probability

The uncertainty in equipment immunity, sag depth and sag duration can be all combined into a single measure which can be used to assess the impact of voltage sags on a network. It is proposed that sag trip probability (*STP*) can be defined as *the probability that equipment attached to a busbar will trip given a sag*. *STP* aims to incorporate all of the observations on equipment and network intricacies highlighted previously. It is a physically meaningful measure and not simply a relative index that is hard to interpret.

STP can be derived by considering sag immunity curves and utilizing knowledge of sag depth and duration. It can also be generalised over multiple events, by taking into account the likely durations and magnitudes of voltage sags across a population of events. The mathematical definition of the *STP* will now be given.

Let the sag duration be defined as a random variable D and the duration at which equipment will trip as a random variable D_{Trip} (D_T for short). Let the voltage magnitude at a busbar (in per unit) be defined as a random variable V and the voltage magnitude at which the equipment will trip as a random variable V_{Trip} (V_T for short).

The STP is defined mathematically as follows:

$$STP = \Pr(D > D_{Trip} \cap V > V_{Trip}) \tag{6.1}$$

which mathematically describes the probability that the random variable D is greater than D_{Trip} and V is greater than V_{Trip}. Physically, equation (6.1) describes the chance that equipment will trip at a busbar during a sag.

The immunity of a group of devices to voltage sags can be defined through an immunity curve, or more generally as a probability distribution. Let $f_{D_T V_T}(d,v)$ define the joint probability density function of both the sag depth and duration at which the equipment will trip. For the function $f_{D_T V_T}(d,v)$, d is generally not independent from v, although they may be independent for defined regions of the immunity curve. If $f_{D_T V_T}(d,v)$ is the joint probability density function, then $F_{D_T V_T}(d,v)$ can be defined as the joint cumulative distribution function, equal to $Pr(D_T \leq d, V_T \leq v)$. Defined in words, $F_{D_T V_T}(d,v)$ describes

the probability that the random variables V_T and D_T take on a values less than v and d respectively. $F_{D_T V_T}(d,v)$ is shown in equation (6.2):

$$F_{D_T V_T}(d,v) = \Pr(D_T \le d, V_T \le v) = \int_0^{v_{\max}} \int_0^{d_{\max}} f_{D_T V_T}(d,v)\,dd\,dv \tag{6.2}$$

where d_{max} and v_{max} are suitable upper limits for the integration which can be taken as 2 per unit for voltage magnitude and >20 seconds for the duration. Over these limits the probability of a trip will be almost certain.

The CDF $F_{D_T V_T}(d,v)$ is a function of attached load (type of equipment), the point on the wave at which the voltage sag hit, phase angle jump during the sag and the overall uncertainty in the immunity of equipment to any of the above.

Whether or not a device will trip will depend on the severity of the voltage sag event (or events). Let $f_D(d)$ describe the expected range of durations for the voltage sag(s) and let $f_V(v)$ describe the expected range of voltage magnitude drop(s) for the voltage sag(s). By considering duration and magnitude as two separate distributions implies that event duration is independent of magnitude. This seems like a reasonable assumption, since duration and magnitude are influenced by different factors: duration is strongly dependent on local protection settings and magnitude depends on fault type, proximity and network impedances. Both $f_D(d)$ and $f_V(v)$ are functions of the voltage level at the attached bus, the protection system covering the section of the network, and the maximum time a sag can remain in the network.

By combining the immunity model and event characteristics, the *STP* can be defined more specifically:

$$STP = \Pr(D > D_{Trip} \cap V > V_{Trip}) = \int_0^{v_{\max}} \int_0^{d_{\max}} F_{D_T V_T}(d,v) f_D(d) f_V(v)\,dd\,dv \tag{6.3}$$

Equation (6.3) can be visualised as quantifying the amount of the thick shaded area shown in Fig. 6.3 which overlaps the area of uncertainty defined by the equipment's immunity curve.

6.4.1 Process Trip Probability

Industrial processes use many pieces of sensitive equipment. Whether or not a process is interrupted by a voltage sag is dependent on both the mutual connections between

equipment within the process [18], as well as the equipment sensitivities of the individual devices. The overall probability of a process trip can be written as:

$$F_{D_T V_T(i,j)}(d,v) = \int_0^{v_{\max}} \int_0^{d_{\max}} f_{D_T V_T(i,j)}(d,v)\,ddv \tag{6.4}$$

$$F_{D_{PT} V_{PT}}(d,v) = \left[1 - \prod_{i=1}^{m} \left(1 - \prod_{j=1}^{n} F_{D_T V_T(i,j)}(d,v) \right) \right] \tag{6.5}$$

where $F_{D_T V_T(i,j)}(d,v)$ is a cumulative probability density function (CDF) of the *j*th equipment of the *i*th serially connected equipment group. *m* is the number of series connected pieces of equipment and *n* is the number of parallel pieces equipment in the *i*th equipment group. $f_{D_{PT} V_{PT}(i,j)}(d,v)$ is the PDF that describes the duration at which the *j*th equipment in the *i*th serially connected group will trip. $f_{D_{PT} V_{PT}(i,j)}(d,v)$ describes the CDF of a process trip given a sag of duration *d* and magnitude *v*. The random variable D_{PT} and V_{PT} is similar to D_T and V_T respectively; however they now describe the duration at which a process, rather than an individual piece of equipment, will trip.

Equations (6.4) and (6.5) can be combined with Fig. 6.3 to form an overall equation for process *STP* (*PSTP*) where there is knowledge about the underlying process involved as shown in equation (6.6). In a similar way to the *STP*, the *PSTP* is defined as *the probability that a process at a busbar will trip (or be interrupted), given a sag.*

$$PSTP = P(D > D_{PT} \cap V > V_{VT}) = \int_0^{v_{\max}} \int_0^{d_{\max}} F_{D_{PT} V_{PT}}(d,v) f_D(d) f_V(v)\,ddv \tag{6.6}$$

Note: The overall process sensitivity to voltage sags, however, depends on Process Immunity Time (PIT) as defined in [30]. If the PIT is rather short, typically less than 4 seconds, the overall process sensitivity depends on the combined sensitivity of the equipment that controls it.

6.4.2 The CDF for the ITIC Immunity Curve

By considering the ITIC immunity curve model as a crisp boundary, the cumulative probability density function, $F(d_T,v_T)$, for the both, the upper and lower portion of the ITIC curve can be defined as follows:

$$F_{D_T V_T}(d_T, v_T) = 1 \forall \begin{cases} v < 0.9 \text{ and } y > 10s & \text{(Region A)} \\ v < 0.8 \text{ and } 0.5\text{s} < y < 10s & \text{(Region B)} \\ v < 0.7 \text{ and } 20\text{ms} < y < 0.5s & \text{(Region C)} \\ v > 1.1 \text{ and } y > 0.5s & \text{(Region D)} \\ v > 1.2 \text{ and } 0.5 > y > 3ms & \text{(Region E)} \end{cases} \tag{6.7}$$

$$otherwise, 0 \qquad \text{(Safe, No Trip Region)}$$

where the values for the limits of the inequalities are determined from the ITIC curve, and approximating region D into a rectangle, rather than a trapezium. Regions A, B and C are defined in the lower curve of Fig. 6.2 and regions D and E are defined in the upper half of the curve.

The STP can therefore be defined by summing the probability that the voltage sag occurred in any of the five regions (note that the trip regions D and E would be effectively caused by voltage swell or short overvoltage):

$$STP = \Pr(D > D_{Trip} \cap V > V_{Trip}) = \Pr(\text{Region A}) + \Pr(\text{Region B}) + \Pr(\text{Region C}) + \Pr(\text{Region D}) + \Pr(\text{Region E}) \tag{6.8}$$

$$\Pr(\text{Region A}) = \int_{0}^{0.95} \int_{10}^{d_{\max}} f_D(d)\, f_V(v)\, dd dv \tag{6.9}$$

$$\Pr(\text{Region B}) = \int_{0}^{0.8} \int_{0.5}^{10} f_D(d)\, f_V(v)\, dd dv \tag{6.10}$$

$$\Pr(\text{Region C}) = \int_{0}^{0.7} \int_{20\times10^{-3}}^{0.5} f_D(d)\, f_V(v)\, dd dv \tag{6.11}$$

$$\Pr(\text{Region D}) = \int_{1.1}^{v_{\max}} \int_{0.5}^{d_{\max}} f_D(d)\, f_V(v)\, dd dv \tag{6.12}$$

$$\Pr(\text{Region E}) = \int_{1.2}^{v_{\max}} \int_{0.3\times10^{-3}}^{d_{\max}} f_D(d)\, f_V(v)\, dd dv \tag{6.13}$$

6.4.2.1 Assumptions Required when Using the ITIC STP Analysis

In the examples that follow, the ITIC curve will be used to calculate *STP* and *PSTP* (described in (6.3) and (6.6) respectively). Using the ITIC curve at all buses assumes that all equipment attached is voltage sensitive as described by the ITIC curve [161]. This is often not the case, as the curve only describes the immunity of computer equipment, and it is unlikely that all the equipment can be de divided into precisely acceptable and unacceptable regions. The ITIC curve is also only relevant to loads connected to one phase only. Three phase machine immunity is likely to be different to the ITIC curve.

6.4.3 Sag Duration Models

The probability distribution used to model the duration of a voltage sag event, $f_D(d)$, can be defined for a single event or a group of events.

For a single event, with a defined event duration the simplest model for the $f_D(d)$ is:

$$f_D(d) = \delta(t_{event}) \tag{6.14}$$

where t_{event} is the duration of the voltage sag event.

Where there is no information about the specific duration of an event or group of events, an approximate model for $f_D(d)$ can be built by considering empirical evidence. Reference [34] highlights the results of a study which show the durations of voltage sags from sites in the EHV, HV and MV networks.

The evidence from [34] was combined into a single probability distribution by fitting a lognormal distribution to the data by minimizing the squared sum of the residuals. The resulting model was estimated as a lognormal distribution with parameters μ = -2.4 and σ=0.75. This distribution was assumed to be constant throughout all of the network's voltage levels. This distribution is shown in Fig. 6.4. For example, 80% of the sags will last less than 0.18 seconds and almost all sags will last less than 0.5 seconds (as shown highlighted on Fig. 6.4).

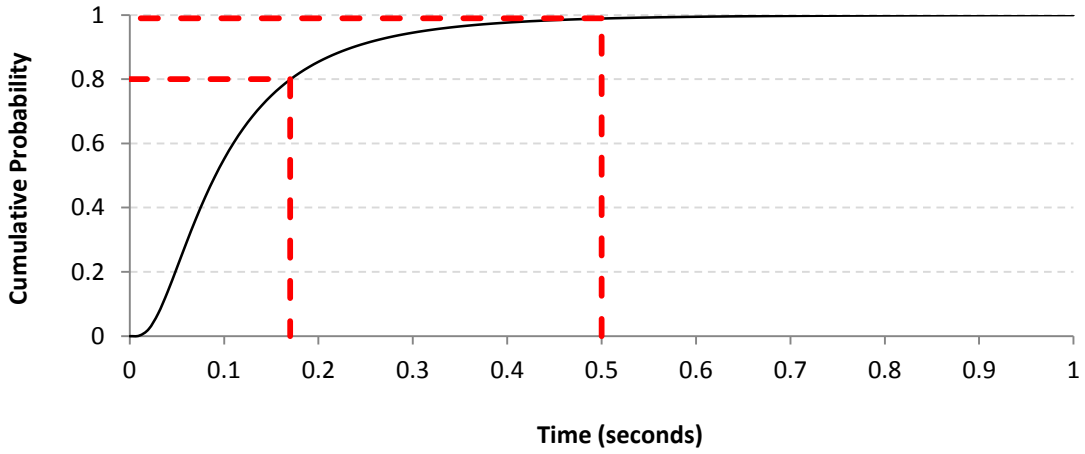


Fig. 6.4. The cumulative probability distribution of sag durations.

The mathematical equation for $f_D(d)$ is therefore:

$$f_D(d;\mu,\sigma) = \frac{1}{d\sigma\sqrt{2\pi}} e^{-\frac{\ln(d-\mu)^2}{2\sigma^2}} \tag{6.15}$$

6.4.4 Number of Trips from the STP

Upon calculating the STP at a busbar, the number of trips across a series of events can be estimated. If the *k*th busbar experiences a number of independent sag events S_1 to S_N, each with STP A_1 to A_N, the expected number of trips is defined as:

$$\text{E(number of trips)} = \sum_{i=1}^{N} A_i \tag{6.16}$$

6.5 The STP for a Specific Event

Both *STP* and *PSTP* can be used for VSPE. In the descriptions that follow, the equations use *STP* only. *PSTP* could, if required, be interchanged for *STP* where there is knowledge about the underlying process at a busbar.

The *STP* can be illustrated by considering the case of the single phase fault discussed in the previous chapter. Consider again a SLG fault between bus 147 and 146 as shown in Fig. 6.5.

Fig. 6.5 shows three monitors observe the fault at buses 130, 204 and 232 (as shown highlighted with thick purple squares in Fig. 6.5). The network is loaded according to the base loading as described in [129]. It was assumed that the duration of the fault follows the distribution described in equation (6.15).

The aim is to estimate the impact of the fault at buses 138, 225 and 174 (as shown with highlighted using thick black circles in Fig. 6.5) by calculating the sag trip probability (*STP*) at each bus. To illustrate how the *STP* can be calculated using limited information, it was first calculated using each monitor independently, and then using information from all the monitors combined.

To illustrate the impact of uncertainty in voltage magnitude and duration, 1000 samples of voltage magnitude (in each phase) and duration were sampled from the distribution from a distribution of voltage magnitude and duration. These points were then overlaid on the ITIC curve. The PDF for the voltage magnitude was calculated using the techniques discussed in the previous chapter. The PDF for sag duration was modelled using equation (6.15).

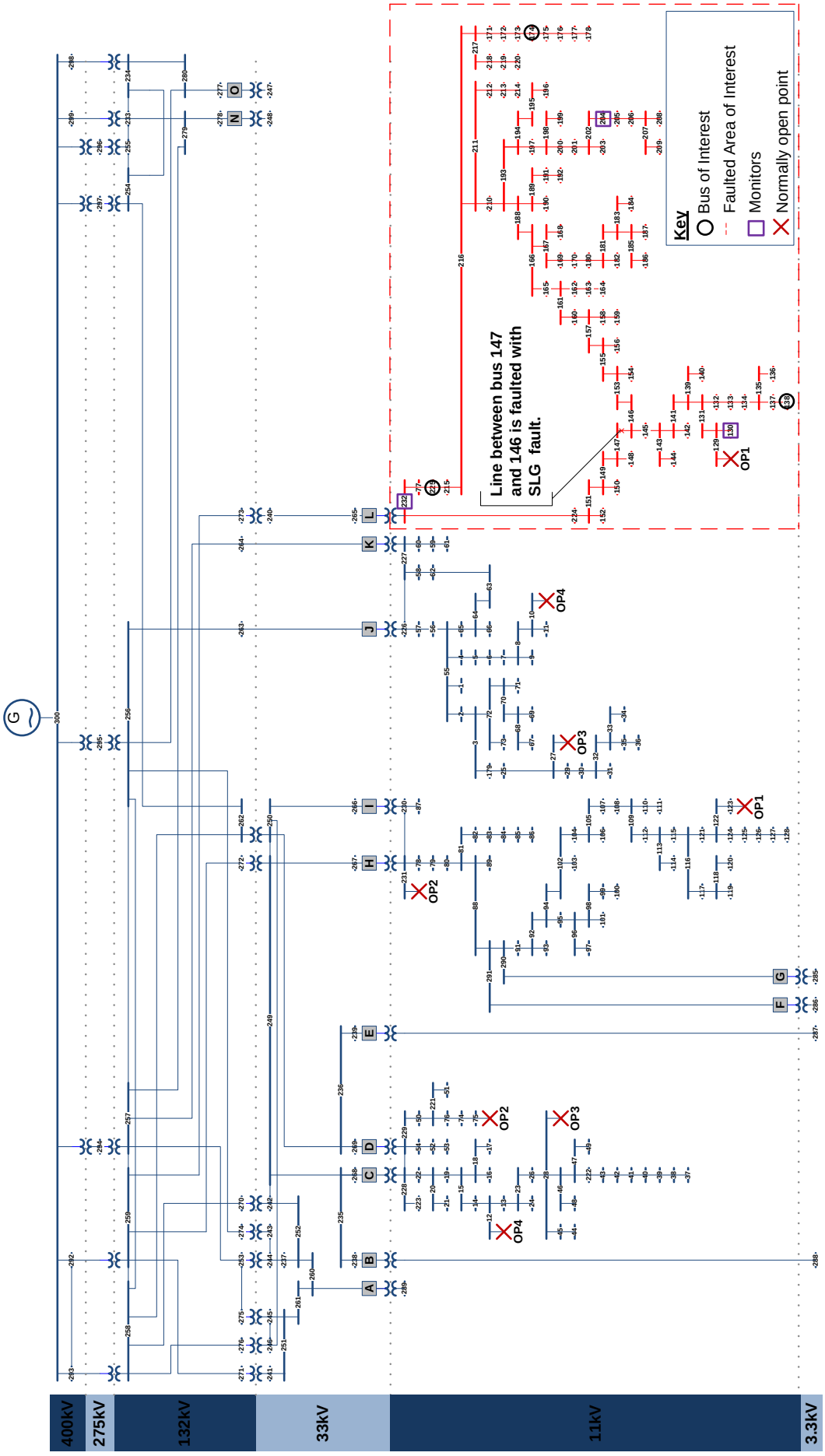


Fig. 6.5. The 96 buses connected to feeder L on the 295 bus network.

6.6 Results

Fig. 6.6 to Fig. 6.14 shows the distribution of each phase for the SLG fault using each of the monitors independently estimating the voltage and duration of the fault.

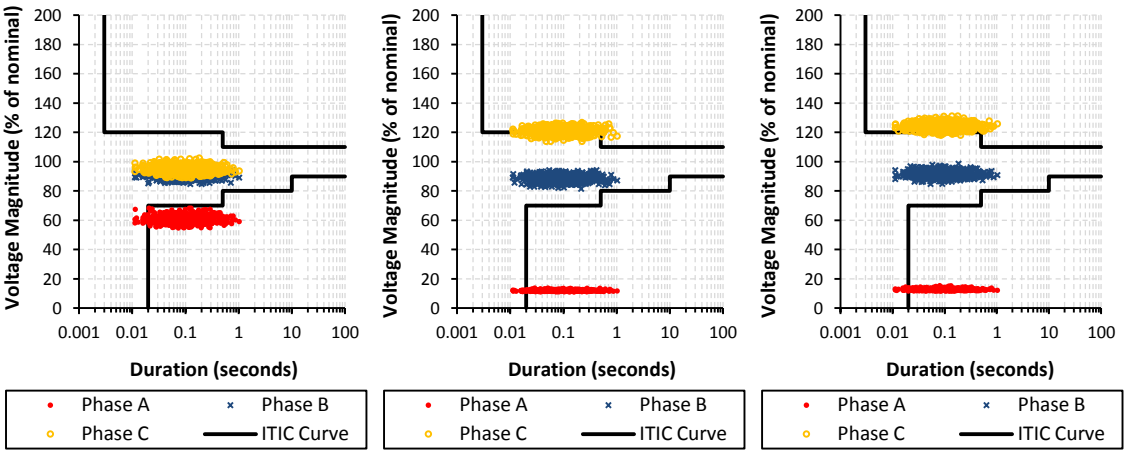


Fig. 6.6 (left), Fig. 6.7 (middle) and Fig. 6.8 (right). The ITIC curve and distribution of $|V^{(a)}|$, $|V^{(b)}|$, $|V^{(c)}|$ against duration at bus 225 (left), 138 (middle) and 174 (right) for a SLG fault between bus 147 and 146, overlaid on ITIC using a monitor at bus 130.

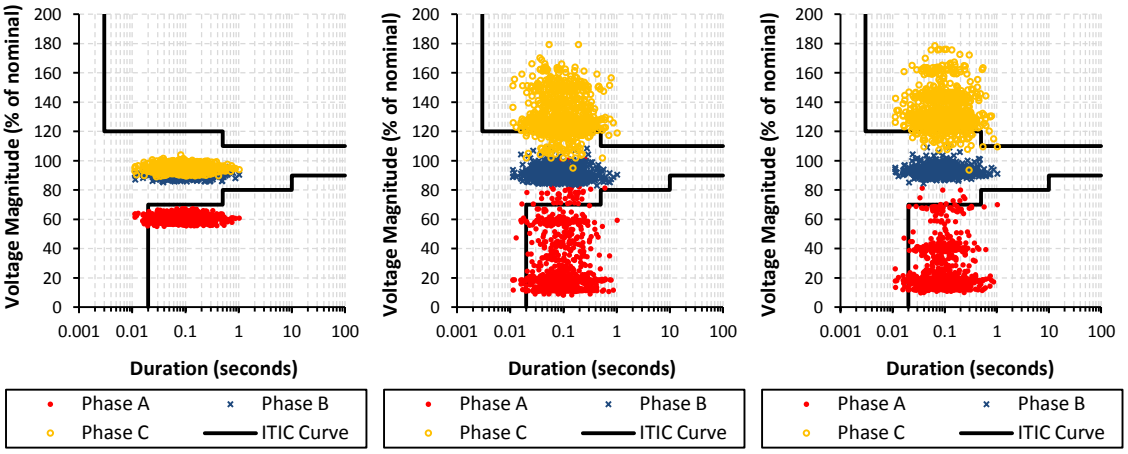


Fig. 6.9 (left), Fig. 6.10 (middle) and Fig. 6.11 (right). The ITIC curve and distribution of $|V^{(a)}|$, $|V^{(b)}|$, $|V^{(c)}|$ against duration at bus 225 (left), 138 (middle) and 174 (right) for a SLG fault between bus 147 and 146, overlaid on ITIC using a monitor at bus 232.

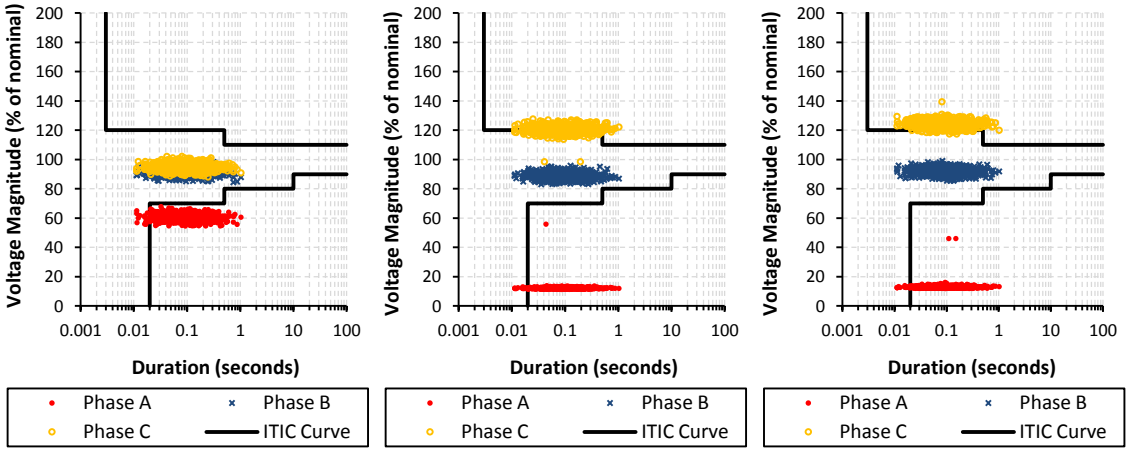


Fig. 6.12 (left), Fig. 6.13 (middle) and Fig. 6.14 (right). The ITIC curve and distribution of $|V^{(a)}|$, $|V^{(b)}|$, $|V^{(c)}|$ against duration at bus 225 (left), 138 (middle) and 174 (right) for a SLG fault between bus 147 and 146, overlaid on ITIC using a monitor at bus 204.

It is possible to conclude that monitor 130 provides the most information about the voltage sag depth, closely followed by a monitor at bus 204. This conclusion can be formulated by reviewing the vertical size of each of the distributions shown in Fig. 6.6 to Fig. 6.14, which are much smaller for bus 130 and 204 compared with bus 232. This corroborates with the results in the previous chapter, which found that the monitors at bus 130 and 204 were the best at estimating voltages in other areas of the 295 bus network.

The distribution of the blue crosses is always within the safe section of ITIC curve, indicating any of the monitors would predict a *STP* of 0 for all locations in phase B. This is an interesting point, as it shows that any of the monitors are equally good at estimating the impact of voltage sags in phase B, even though monitor 130 and 204 are both technically better at voltage sag profile estimation.

Fig. 6.10 and Fig. 6.11 show the results of bus 232 predicting the voltage in all three phases at buses 138 and 174 respectively. The spread of the distribution of the voltage in phase A and C is large when compared with bus 130 (Fig. 6.7 and Fig. 6.8) or bus 204 (Fig. 6.13 and Fig. 6.14). However, large uncertainties in the voltage may or may not impact on the *STP* calculation since *STP* is a function of the expected number of points in the safe and non-safe regions of the ITIC curve. This is highlighted in Table 6.1 which shows that the *STP*s estimated in phase A using a monitor bus 232. Table 6.1 shows that the STPs estimated using a monitor at bus 232 were 0.983 at bus 225, 0.934 at bus 138, and 0.963 at bus 174; all very similar to the result for all monitors which were 0.983, 0.934 and 0.963 for buses 225, 138 and 174 respectively. Thus it could be

concluded that although bus 232 is a poor estimator of voltage sag magnitude, it is still a high quality estimator of the probability of a trip (STP). Using information from either a monitor at bus 232 or information from all monitors in the network results in the same conclusion that equipment attached to phase A at buses 225, 174 and 138 were very likely to trip.

TABLE 6.1. THE SAG TRIP PROBABILITIES FOR BUSES 225, 138 AND 174 OBSERVED FROM BUSES 130, 232 AND 204

Monitor	**Bus 225**			**Bus 138**			**Bus 174**		
	Phase A	**Phase B**	**Phase C**	**Phase A**	**Phase B**	**Phase C**	**Phase A**	**Phase B**	**Phase C**
Bus 130 STP	0.983	0	0	0.983	0	0.691	0.983	0	0.985
Bus 232 STP	0.983	0	0	0.934	0	0.862	0.963	0	0.94
Bus 204 STP	0.983	0	0	0.983	0	0.632	0.983	0	1
All Mon. STP	0.983	0	0	0.983	0	0.771	1	0	1

The smallest probability of a trip in Table 6.1 is recorded for phase C of bus 138. Using all monitors, the probability of a trip at bus 138 is 0.771. If the estimate taken using all monitors is considered a best estimate for the true probability of a trip, then the information from bus 232 overestimates the likelihood of a trip, and bus 204 and 130 underestimates the likelihood of a trip at this busbar.

Fig. 6.15 to Fig. 6.17 show the distribution of the voltage magnitude and duration when all of the monitors are used together to predict the voltage magnitude. All of the plots show a much narrower spread for the voltage magnitude, as the information from all devices eliminates some of the uncertainty. Table 6.1 shows that the reduced variation has a minimal impact on the *STP* as the *STP* for all monitors is very similar to the *STP* for monitors 130 and 204.

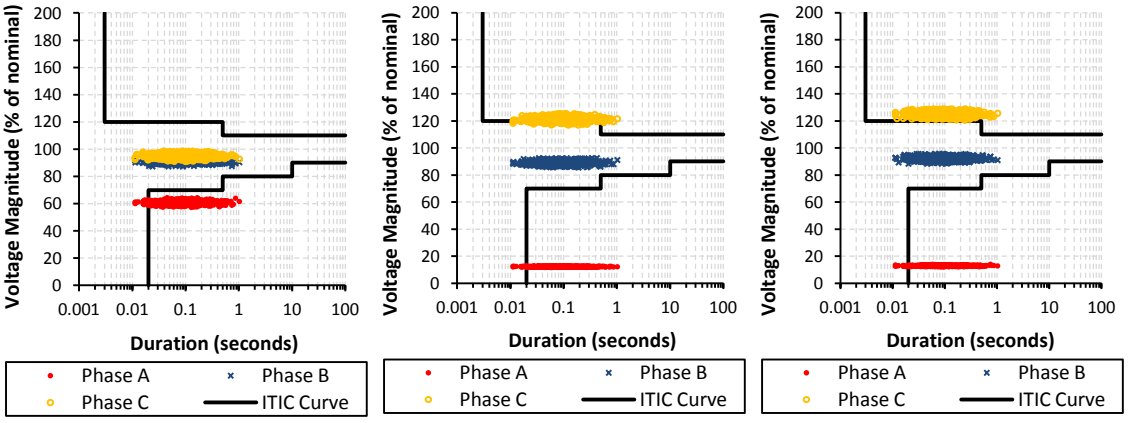


Fig. 6.15 (left), Fig. 6.16 (middle) and Fig. 6.17 (right). The ITIC curve and distribution of $|V^{(a)}|$, $|V^{(b)}|$, $|V^{(c)}|$ against duration at bus 225 (left), 138 (middle) and 174 (right) for a SLG fault between bus 147 and 146, overlaid on ITIC using monitors at bus 130, 232 and 204.

6.7 Summary

The research presented in this chapter introduces a new method for estimating the probability of a trip from a voltage sag, namely, the sag trip probability (*STP*). The general formulation of the *STP* is can be configured to be used with different equipment sensitivities, different process sensitivities and arbitrary sets of monitors. The *STP* is formulated using statistics which allows uncertainties in equipment and voltage sag characteristics to be robustly accommodated.

As well as formulating the *STP*, a generalised derivation for a process trip probability (*PSTP*) was also presented. This allows network operators to gauge the potential impact of voltage sags on important industrial processes within their network.

The presented method was validated by illustrating how the *STP* can be calculated for a single phase to ground fault within the 295 bus network. The results illustrated that estimating the impact of a voltage sag on a consumer using the *STP* does not always require a precise estimate for voltage sag depth. Indeed the results that the *STP* can be estimated with a high level of precision, even if there is large uncertainty in voltage sag depth. It could therefore be concluded that estimating the *STP* to a defined relative level of accuracy is more straightforward than estimating the voltage sag magnitude to the same relative level of accuracy. In the following chapter, this point will be used to guide the placement of monitoring devices to best estimate the *STP* and hence most accurately estimate the resultant impact of voltage sags on consumers in the network.

Using the *STP* approach, this chapter also described a method whereby the number of trips caused by voltage sags could be estimated over a series of events during a defined time period. This aspect of the *STP* will be re-visited in Chapter 9 when the number of trips will be estimated over a period of time.

7 Optimal Placement of Monitors for Voltage Sag Monitoring

7.1 Introduction

The previous three chapters have focussed on techniques for voltage sag monitoring with a limited number of monitors. As well as monitoring a network using pre-existing monitors, it is also important to consider how best to optimise their locations to enhance the observability of a network for voltage sag issues.

The purpose of the following chapter is to introduce a methodology which can be used to optimise the placement of monitors for voltage sag monitoring in a distribution network. The developed methodology allows distribution network planners to specify budgetary requirements and future loading forecasts and determine a robust set of monitoring locations for voltage sag performance monitoring.

Existing research on optimal monitoring for voltage sags focuses on placing monitors throughout the network to either detect or localize faults that cause voltage sags. In [25], the author defines the concept of a monitor reach area that defines an area in a network where a monitor can detect all sags of a defined magnitude. In [48], the authors place monitors to localize faults within the network.

An added complexity of the optimal monitor placement problem which has not been considered in existing research is that DNOs must deal with a large degree of uncertainty when deciding where to install monitors in their network. This uncertainty is present because the economic value of a chosen monitoring solution is not precisely known over all time periods. Load growth forecasts and cyclical demand levels are two factors that influence the estimated loading on the network.

The focus of the research presented in this chapter is different to existing optimal placement methodologies in two key ways. The first difference is the objective function of the optimization algorithm. The optimization uses the STP (described in the previous chapter) to assess the suitability of monitoring positions. This is an important contribution, as it ensures monitors are placed where they are best able to estimate the impacts of voltage sags, rather than simply estimating voltage magnitudes.

The second important difference is that the optimization process is based on an immune inspired (AIS) optimization algorithm [146], known as the B-Cell Algorithm (BCA) which was first introduced in Chapter 3. This allows a diverse range of near-optimal monitoring solutions to be developed based on minimization of the error in expected equipment trips at a busbars throughout the network. This increases the robustness of solutions and gives power system operators' choice over monitor locations.

There are two aims for the research presented in this chapter:

- To present a practical and robust methodology capable of optimally locating monitors to observe the effects of voltage sags on customers in a generic distribution network across a range of future uncertain loading scenarios.
- To demonstrate and experimentally validate that an artificial immune system (AIS) optimization methodology is an appropriate, high quality tool for the optimal monitor placement problem.

These aims will be accomplished by testing the optimization methodology against different monitoring sets selected from existing optimal placement research and engineering judgement. The monitoring sets will be evaluated and compared to identify which monitoring sets can most accurately monitor the network for trips.

The monitor sets which are developed in this chapter are used throughout this thesis to test the methods developed for unbalance and voltage sag monitoring. It should be noted that the optimization techniques developed in this chapter are aimed at sag

monitoring only, but this does not exclude the data from power quality voltage sag monitors being used for unbalance or any other power systems monitoring.

7.2 What Makes a Good Monitoring Set?

A high quality voltage sag monitoring set must be able to perform voltage sag performance estimation as outlined in Chapters 4 to 6. The ultimate objective of voltage sag monitoring is to estimate the resultant impact of voltage sags on customers.

The complexities of locating monitors which can accurately estimate the number of trips at a busbar can be explained by considering Fig. 7.1. Fig. 7.1 shows a busbar which is exposed to two voltage sag events. Both of these events occur in the region of the immunity curve where an equipment trip is certain ($P(trip)=1$).

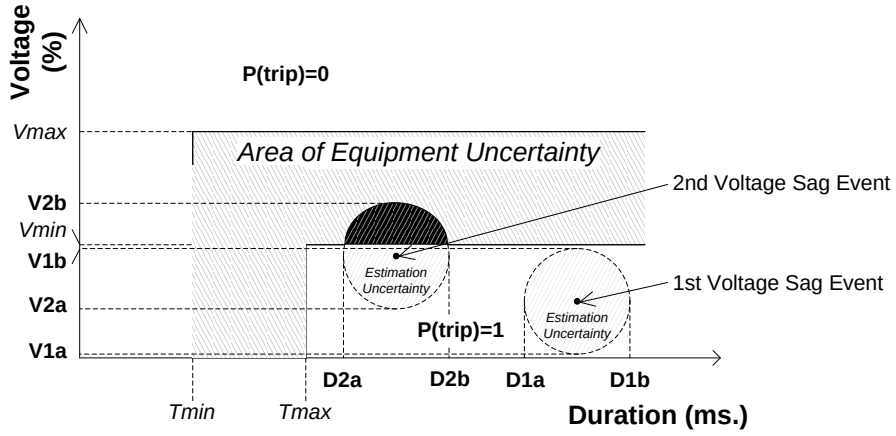


Fig. 7.1. The uncertainty in both equipment immunity and measurements.

Fig. 7.1 also shows the area of voltage sag profile estimation uncertainty described by variation in both voltage magnitude and duration for the two voltage sags. The error in the estimation process for both of the sags is identical (as shown by equal diameter circles). For the 1st voltage sag, the estimation uncertainty will not impact on the estimated probability of a trip since the entire region of uncertainty is contained within the region with $P(trip)=1$. However, for the 2nd voltage sag, the same amount of estimation uncertainty causes $P(trip)\neq 1$. Therefore, accurate estimation is preferential near the immunity curve boundary and less important elsewhere.

The accuracy of the estimation process is heavily influenced by the location of power quality monitors in the network. Existing optimal monitor placement methodologies (such as [25] and [48]) focus on accurately estimating the during sag voltage magnitude at non monitored busbars rather than considering the resulting impact of voltage sags on customer equipment. These techniques optimise monitor locations by minimising the

three phase error in voltage magnitude across all monitored and un-monitored buses in the network. Another similar approach could involve minimising the error in generalized sag table [59] (GST) estimation for all monitored and un-monitored busbars. If the GST could be predicted accurately, the voltage sag performance of a busbar in terms of customer trips could also be estimated.

Both the voltage magnitude and the GST approach have a common deficiency: their objective functions optimise monitor locations in proportion to absolute errors in voltage magnitude estimation whilst ignoring the resultant impact of the sag on end user equipment. A correct assessment of voltage sags needs to consider magnitudes, expected durations and equipment immunity. A 40% voltage sag is somewhat similar to a 20% voltage sag (of similar duration); both cause serious disruption to equipment connected to those buses in the network. However, a 90% voltage sag is very different from a 70% voltage sag (of the same duration). The 90% sag may not cause equipment to trip. Therefore, when estimating trips, it is more important to estimate the difference between a 70% and 90% voltage sag than the difference between a 20% and 40% voltage sag.

The rest of this chapter will focus on a method capable of positioning monitors in such a way to maximise their accuracy of estimation on the equipment immunity curve boundary.

7.2.1 STPs for Optimal Monitor Placement

The *STP* defines the probability of equipment tripping at the busbar given a sag. The *STP* is a probability, and as such it will always range between 0 and 1. The *STP* is made up from uncertainty in equipment immunity and uncertainty in voltage sag profile estimation. Uncertainty in the *STP* can be reduced by improving the voltage sag estimation procedure by placing the monitors in high quality locations.

Let STP_{true} represent the value for the *STP* which would be obtained if the during fault voltage magnitude could be estimated with full monitoring and no error, but whilst still considering uncertainty in equipment immunity curves. Let STP_{est} represent the value for the *STP* obtained when using voltage sag profile estimation (VSPE, as described in Chapter 5). Both STP_{est} and STP_{true} are calculated using (6.3). Equation (7.1) defines the quality of a VSPE estimate based on the STP_{est} and STP_{true}.

$$u(v_{true}, v_{est}) = STP_{true} - STP_{est} \tag{7.1}$$

$u(v_{true},v_{est})$ defines the absolute difference in STPs for the estimated voltage profile, v_{est} and true voltage profile, v_{true} (obtained assuming no estimation or measurement error). In other words, $u(v_{true},v_{est})$ this is the difference in the probable number of equipment trips estimated using a limited monitoring set, and the probable number estimated using a full monitoring set.

A false alarm (*FA*) can be defined where the algorithm estimates a high STP (STP_{est}), but the real STP (STP_{true}) is low. A missed trip (*MT*) is where the algorithm estimates a low *STP*, but the real *STP* is high. For example, if the estimated *STP* is zero, and the real *STP* is 1, then we can be certain we have missed a trip. Indeed, if the difference between these two probabilities is high (say 0.95), then we can be almost certain that a false alarm or missed trip has occurred. The two equations that define false alarms and missed trips across all *N* buses in a network for a fault *f* are shown in (7.2) and (7.3).

$$FA_f = \sum_{i=1}^{N} \sum_{p\in\{a,b,c\}} 1 \;\forall\; STP_{est_{f,i}} - STP_{true_{f,i}} > 0.95 \tag{7.2}$$

$$MT_f = \sum_{i=1}^{N} \sum_{p\in\{a,b,c\}} 1 \;\forall\; STP_{true_{f,i}} - STP_{est_{f,i}} > 0.95 \tag{7.3}$$

Where $STP_{true,f,i}$ is the true STP at bus *i* for an arbitrary fault *f* and $STP_{est,f,i}$ is the estimated STP at bus *i* for the same arbitrary fault *f*.

The definition of *MTs* and *FAs* will be used as the basis of an optimal monitor placement algorithm in the methodology which follows. *MTs* and *FAs* will be integrated into an objective function which will optimize the position of monitors using a B-cell algorithm (as described in Chapter 3) to estimate sag performance over a range of uncertain loading scenarios.

7.3 Methodology

The overall methodology was designed to satisfy both of this chapter's aims. The methodology is broadly split into 4 stages as shown in Fig. 7.2:

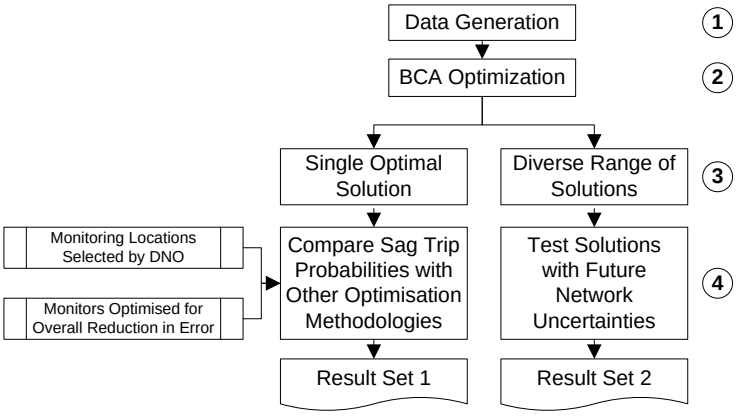


Fig. 7.2. The 4 stages of the methodology used in this paper.

The first set of results will be used to compare the *STP* optimization methodology with other optimal monitoring techniques. The second set of results will evaluate the robustness of the BCA optimal monitoring solutions over a range of uncertain future operating scenarios.

7.3.1 Data Generation

Data was generated for the study based on a proposed load growth model for the next 15 years. In this scenario, load is expected to average between 60% and 270% of the original system loading (l_b), as shown in Fig. 7.3. Simulated data is required to firstly find through optimization a wide variety of near optimal monitoring locations and secondly to test the quality of these solutions when projected into an uncertain future. The actual load growth scenarios, however, do not influence the methodology qualitatively, thought quantitative results may be different for different assumed future load growth patterns.

7.3.1.1 Modelling Future Network Uncertainties

The future loading of the network was modelled using a Monte Carlo simulation [92]. The oscillatory nature of the power system loading was modelled using a sinusoidal oscillation that spans a time period of 365 days. Random variation was added into the model by adding Gaussian noise at each time interval.

Equations (7.4) and (7.5) describe the long term load growth model used in this research:

$$p(i) = r(i) + \delta_{pi} \tag{7.4}$$

$$r(i) = 1 + \frac{\cos(\frac{2\pi i}{365})}{8} \tag{7.5}$$

p(i) is a vector which describes the power system demand at every node in the power network at the *i*th time interval. *r(i)* is a function which describes the long term seasonal oscillation in demand of the power system. *p(i)* depends on both *p(i)* and δ_{pi} where the later is a Gaussian random variable normally distributed with a mean of *0* and standard deviation of σ_{pi}. At the *i*th time interval, σ_{pi} was set to 0.2% of the total power demand, scaled by a factor of $r(i)/r(0)$.

It is important to note that sinusoidal oscillations were selected due to the ease of application without any loss of generality. Over a 365 day period, the electrical load profile in the UK broadly correlates with the number of hours of daylight (due to limited air conditioning load in the summer). Therefore, a model which follows a sinusoidal pattern with additional random variation in the form of Gaussian noise reasonably covers the range of expected variation in electrical demand. Different functional forms *r(i)* could be modelled without affecting the proposed methodology.

Fig. 7.3 shows 10 Monte Carlo simulations generated with a load growth of 5% per year over the next 15 years.

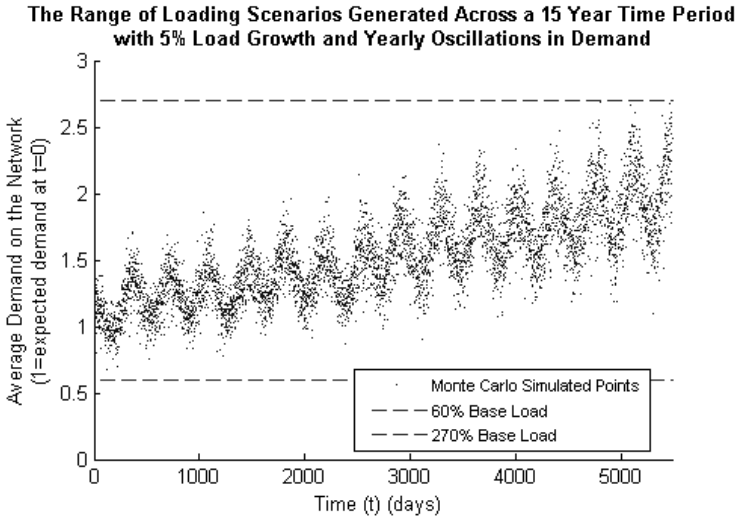


Fig. 7.3. 10 runs to generate Monte Carlo samples across a period of 15 years based on a load growth model of 5% per year and oscillations in the yearly loading of the network.

The developed load growth model used in this research assumes that the power factor remains constant at the base power factor. The base power factor is the set of power factors specified by the loads in the generic distribution network at the default (or base) operating load. At the base operating load, the base power factor ranges between 0.95 (lagging) and unity, as one would expect to be the case at 11kV and 33kV buses in modern distribution network. The average base power factor across all buses in the

network is 0.98. In this research, power factors were assumed to be largely invariant over the 15 year study period and therefore kept constant at the base power factor.

Varying the power factor significantly will create variation in pre-fault voltages throughout the network, which will in turn affect the accuracy of the fault location estimate (M_l) derived using equations (5.15) and (5.20) (for three phase and line to line faults respectively). The equations for single line to ground and double line to ground faults (equations (5.17) and (5.22) respectively) are not functions of pre-fault voltage, and are therefore insensitive to changes in the power factor. The during fault voltage drops for all faults are ultimately calculated using equations (5.23) to (5.25), which are independent of pre-fault voltages at non-monitored busbars.

Voltage support devices such as capacitor banks, and tap-changing transformers were not modelled in the data sets which were generated in this study, as they were not present in the test 295 bus network. Capacitor banks change the circuit impedance and consequently alter the voltage magnitude at buses where they are connected as well as to a lesser extent at surrounding busbars. If these devices were present, they would add more operating points to the set of generated operating conditions generated by load variation and variations in network topology. Variation in network topology can also be included in the uncertainty analysis. Fig. 7.5 shows the location of 4 open points (positioned based on [163]) that connect various feeders in the 295 bus distribution network. The network in Fig. 7.5 is assumed to operate for 96% of the time with the open points (OP1, OP2, OP3 and OP4 in Fig. 7) open. However, for 1% of the time, it is expected that each of the 4 open points will be closed independently. Thus, there are 5 different topologies on which the network could operate. It should be noted that there are often many permutations of topologies on which a network can operate. In this study, only the 5 most common topologies are considered to illustrate the methodology.

7.3.1.2 Datasets for the Optimization Procedure

For the optimization section of this study 42 datasets (Opt_D) were generated. To thoroughly test the robustness of the proposed methodology, variations in fault position, loading conditions, topology, fault type and fault impedance must all be considered.

Each dataset represents the results of short circuit analysis on the power system network with load divided equally between 60% and 270% of the base system load in 5% intervals on the base network topology. The load variation at each of the buses was kept consistent with the base load, and the power factor kept as the base power factor. 1668

fault positions (6 faults for each of 278 lines) were selected and a single line to ground (asymmetrical) fault simulated at each of the locations.

To optimise the monitor locations, 5475 datasets ($Test_D$) were generated; one dataset for each day over a 15 year period (365×15). Variation in network topology was incorporated by randomly selecting 55 datasets (1% of all datasets) for each of the four new topologies. Each new topology is created by closing one of the four normally open points. The remaining 96% of the $Test_D$ datasets operate using the default network topology. All the datasets consist of the results of simulating 1668 single line to ground faults (6 fault locations on each of the 278 lines). The pre-fault voltages for the fault studies were assigned randomly according to a Monte Carlo simulation which models 5% annual growth and stochastic variation in daily load, for the next 15 years.

Single line to ground faults are by far the most common type of fault, accounting for perhaps as much as 80% of all recorded faults [164]. Because of their prevalence, the monitor placement optimization performed in this chapter used data generated from single line to ground fault simulations only.

7.3.2 BCA Optimization

7.3.2.1 Optimization Problem

Determining the location of monitors for VSPE can be defined as an optimization problem. For a given VSPE algorithm A, a network topology n, a network loading $l \in L$, a selection of monitors $\mathbf{m} \in M$, and a set of fault positions F, an objective function can be formulated as:

$$U(\mathbf{m},n,l,F) = \frac{\left[3NN_F - \sum_{f\in F}\begin{cases} 1 \forall \frac{\lceil FA_f + MT_f \rceil}{3N} > 0.01 \\ 0 \forall \frac{\lceil FA_f + MT_f \rceil}{3N} < 0.01 \end{cases}\right]}{3NN_F} \times 100$$

$$C = \max_{\mathbf{m}\in M, n\in N, l\in L} U(\mathbf{m},n,l,F) \tag{7.6}$$

$$\text{Subject to:} \begin{cases} g_1(\mathbf{m}) \le a_1 m_1 + \ldots + a_i m_i + \ldots + a_N m_N \\ g_2(\mathbf{m},i) \le \left\lceil \sum_{f\in F} FA_{f,i} + MT_{f,i} \right\rceil \times 100 \end{cases} \tag{7.7}$$

where C is a function that describes the computation of U for the given combination of $\boldsymbol{m}$,A,n,l and F and g_1 is a function that represents the cost of installing a selected monitoring solution. A simple way of describing U is the percentage of the network where missed trips and false alarms occur at less than 1% of the buses in the network. a_i

represents the cost of installing a monitor at bus i and $m_i \in \{1,0\}$ is 1 if a monitor is installed at the ith bus. N is the number of buses in the network (295 in this study), and N_F is the total number of simulated faults (1668 in this study).

g_2 is a function that ensures that the number of false alarms and missed trips at the ith bus is kept below a certain threshold level. This is useful when there are important industrial processes operating in a network which must be monitored.

In this research, the cost associated with installation and maintenance of a monitor at a given location is assumed to be constant ($a_i=1 \ \forall \ i=[1...N]$ unit per monitor). Practically this would incorporate the cost of the monitoring device, communications infrastructure costs, cyber-security costs, and other bespoke installation requirements. Information on the pre-fault voltages in the network is also assumed to be available from load flow study.

The 295 bus generic distribution system (as shown in Fig. 7.5) has 10 primary substations at 11kV and 33kV. It was therefore decided that g_1 should be constrained to 10 monitoring locations. This allows the solutions generated by the optimization algorithm to be directly compared with an engineered solution (ENG) which represents installing a monitor at every 33kV and 11kV primary sub-station. g_2 was left unset, except when optimizing to find the STP_2 solution where it was set to 99.9% on the secondary side of feeder A (Fig. 7.5) to simulate an important 11kV customer.

7.3.3 BCA Algorithm Details

The optimization algorithm problem is tackled using the B-cell algorithm (BCA) and the Opt_D dataset. The BCA is described in detail in Chapter 3. For each of the 42 datasets, the BCA is used to find a set $S_d \in S$ (d=[1,42], d=1:60% load, d=8:100% load, etc.) of 10 near-optimal monitoring solutions.

7.3.3.1 Data Representation

The BCA used in this paper was slightly adapted from the original work in [146] (and Chapter 3) to work with discrete function optimization rather than continuous function optimization using 64-bit strings. Instead of representing each B-cell as a series of 64-bit strings, a different representation was employed in this research to represent each B-cell. Each B-cell was represented as an N-bit binary string. The N bits represents the N possible sites for monitors in the network. A 1 indicates that a site has a monitor installed, and a 0 indicates that the site has no monitor installed.

7.3.3.2 Configuration of the B-Cell Algorithm

The BCA has two configurable parameters. Firstly, the total number of B-cells in the population and secondly the number of clones for each B-cell produced on each iteration of the algorithm. The aim of the BCA in this research is to generate a wide variety of near-optimal monitoring solutions.

The BCA was setup to maintain a population of 10 B-cells. The algorithm is often configured to run with the number of clones configured as the same as the population [146]. In this research, the number of clones was configured in a similar manner, with each B-cell producing 10 clones.

7.3.3.3 Termination

The BCA was configured to terminate when the best 20% of the B-cells (2 in this case) had reached a steady value for the last 10 iterations. Only 20% of the B-cell population was tested for convergence to ensure that the whole population did not simply converge to a single global optimum.

7.3.3.4 Output of the BCA

The output of the BCA was 42 sets of monitoring solutions, one for each network loading scenario. The best solution from S_8 (the set of solutions at base load) was extracted for comparisons against the other optimization methods (as in the following). This solution is known as the STP_1 solution.

The set S was consolidated to remove any duplicated solutions found from each of the 42 optimization runs. This formed a new set, known as S'. These solutions are required to test the robustness of the algorithm to network uncertainties.

7.3.3.5 Output of the BCA

A second STP solution known as STP_2 was also generated from the BCA which is the result of optimizing monitoring locations with the constraint g_2 activated at 99.9% on feeder A.

7.3.4 Voltage Sag Impact Estimation

Voltage sag profile estimation was performed by locating each fault using equation (5.16) (described in Chapter 5 and [81]) and then subsequently estimating the voltage during the fault at non-monitored busbars using equations (5.23) to (5.25). To reduce

the computational time required to run the optimization process, all simulations were assumed to be free from monitoring error.

The localization accuracy of the fault location equations may be affected by fault impedance, fault type and pre-fault loading. The equations developed in Chapter 5 for single line to ground (SLG) faults either eliminate these variables entirely, or provide robust assumptions which limit the inaccuracy fault localization and subsequent voltage sag profile estimation.

To estimate the voltage profile at non-monitored busbars, each monitor was tested in turn to determine a set of fault locations for each fault by solving equation (5.16) for all lines in the network. The monitor with the lowest number of locations was then selected. The voltage sag profile at all non-monitored busbars was then estimated using the selected monitor, each fault location solution and equations (5.23) to (5.25). These solutions were then averaged to obtain an estimate for the voltage at all non-monitored busbars. The reader is referred to equations to Chapter 5 for more information on the fault location and voltage sag profile estimation equations. Note that this method of VSPE is slightly different to the statistical approach incorporating monitoring errors outlined in Chapter 5. The key difference is that a single estimate for the during sag voltage magnitude is obtained rather than a statistical estimate, beacuse of the assumption that all monitors were free from measurement error.

STP_{est} was calculated using (6.1), the estimated during sag voltage profile estimate and assuming a sag duration model as shown in Fig. 6.4. STP_{true} was calculated using a perfect during sag voltage profile and the same sag duration model.

Using the equations in Chapter 5 (and [81]) for single line to ground faults, no assumptions on pre-fault loading or fault impedance are required to carry out fault localization and VSPE. In other words, the equations for single line to ground fault location are insensitive to fault impedance and pre-fault loading. Other types of faults such as three phase, double line to ground and line to line faults were not simulated in the results presented of this paper. Chapter 5 describes (and [81]) formulates equations for all types of faults, and any of these equations can be incorporated into the optimal placement methodology developed in this chapter.

To use the equations for other types of faults, assumptions on fault impedance and pre-fault voltages must be considered. For three phase and line to line faults, the fault impedance must be assumed to be entirely real, and pre-fault voltages must be estimated [165]. The assumption of real fault impedance is shown in [103] to be a valid

approximation. Pre-fault voltages can be estimated using distribution system state estimation as shown in Chapter 5. For double line to ground faults a solution to the fault location equations can be found through a single assumption of an entirely real fault impedance. The reader is referred to the equations in Chapter 5 for more details on the assumptions required to estimate during sag voltages at non-monitored busbars.

It should be noted that the impact estimation equations described in Chapter 5 and 6 (or the algorithm described in [17] with correctly configured thresholds) are able to estimate a sag location with fewer monitors than required to make a network fully observable [166] (from a state estimation observability perspective). This is an important quality, as full observability requires more monitors and therefore necessitates higher associated deployment costs. A key advantage of the approach proposed in this research is its ability to estimate sag trip performance using much fewer monitors than required to make a network fully observable.

7.3.5 Comparison of STP Optimization with Other Techniques

The quality of the both STP_1 and STP_2 solutions were compared against three other monitor placement methodologies. Firstly, a solution (labelled *LOC*) was produced which is the result of running the BCA optimization algorithm with the objective function changed to minimizing the sum of absolute deviation in per unit voltage magnitude at all buses in the network. Secondly, a solution was created based on the monitor reach area placement method presented in [17] (labelled *MRA*). Lastly, a solution (known as *ENG*) was created where monitors were placed at primary sub-stations on the secondary side of each 11kV and 33kV feeder. The *ENG* solution is representative of how DNOs are likely to place monitors in the network without the assistance of an optimal placement algorithm, and is thus un-optimized.

Each of the optimizations utilized a constraint to limit the number of monitors installed. To provide some context for the results, an unconstrained optimization was also performed to determine the minimum number of monitors required to achieve full observability of all voltages in the network. This was accomplished using [167]. Full observability allows any voltage and hence any fault to be localized anywhere in the network.

7.3.6 Testing Solutions with Future Network Uncertainties

It is not feasible to exhaustively search every monitoring positioning, on every network, for all loadings and all fault positions. There are simply too many combinations to

search. However, the quality of a proposed monitoring solution can be assessed using two figure of merit (FOM) functions. For a monitoring solution $s' \in S'$, the following FOM functions are defined:

1. Across all scenarios, the total percentage area (sum of U) covered should be as high as possible.
2. Across all scenarios, the range of U should be as small as possible.

The FOM functions are shown algebraically in the following formulae:

$$f_1(s') = \frac{\left\lfloor \sum_{U \in \mathbf{U}(Test_D)} U \right\rfloor}{5475} \tag{7.8}$$

$$f_2(s') = range(\mathbf{U}(Test_D)) \tag{7.9}$$

Where $\mathbf{U}(Test_D)$ is 5475 long vector with each element representing the computation of U for the different operating conditions in $Test_D$.

f_1 and f_2 can be analyzed by discovering the Pareto set of non-dominated solutions. In terms of a maximization, a solution $s'^* \in S'$ is said to be non-dominated if there exists no other solution $s' \in S'$ where the value of $f_i(s'^*) \geq f_i(s') \forall i$ and for at least one i, $f_i(s'^*) > f_i(s')$. The Pareto optimal set of solutions S'^*, form a set of solutions which can be used to make the final engineering decision about the location of the monitors within the network.

7.4 Results

7.4.1 STP Monitor Placement Methodology

The *STP*, *LOC* and *ENG* solutions were all compared to demonstrate the viability of the STP optimization methodology. Each of the solutions compared are shown in Fig. 7.5, for the specific case of 10 monitors. The *LOC*, $STP_1$and STP_2 solutions are all placed mainly in the 11kV network. These three solutions are also placed in an even density across each of the 11kV networks.

The performance of each of the solutions was compared in terms of their ability to correctly estimate the STP. This can be measured in terms of false alarms (FAs) and missed trips (MT).

Table 7.2 shows the mean number of missed trips and false alarms for a fault for each of the three monitoring solutions. It also shows the percentage of the network where faults can occur and the monitoring solution produces FAs and MTs at less than 1% of the

buses (*U*). A higher percentage value indicates a larger area of the network were the STP can be reliably estimated. This area is shown in Fig. 7.5 for the STP_1 monitoring solution at base load.

As a comparative benchmark, the number of monitors required to achieve full observability of the network were computed using [167]. It was found that 80 monitors were required to achieve full observability of all voltages in the network, and hence a value for *U* of 100%. The busbar monitor locations for the full observability solution is shown in Table 7.1.

TABLE 7.1. OPTIMAL MONITORING SOLUTIONS

Monitor Set	**Voltage Level**	**Monitoring Locations**
Monitoring Solution for Full Observability	11kV	5, 8, 11, 14, 18, 20, 23, 29, 31, 34, 36, 39, 42, 44, 47, 48, 52, 55, 56, 61, 64, 69, 71, 73, 74, 77, 79, 83, 85, 89, 91, 95, 99, 100, 101, 105, 107, 109, 112, 117, 124, 127, 129, 131, 134, 139, 141, 142, 146, 149, 150, 153, 155, 158, 160, 163, 166, 173, 175, 180, 183, 184, 187, 188, 191, 194, 197, 201, 206, 212, 218, 226, 228
	33kV	242, 247, 250, 266, 269
	132kV	272
	275kV	294
STP_1	11kV	130 , 204, 217, 60, 27, 128, 291, 53, 221
	33kV	261
STP_2	11kV	289, 50, 122, 135, 149, 61, 5, 187, 206, 176
LOC	11kV	28, 121, 68, 87, 60, 137, 204, 176
	33kV	237, 266
ENG	11kV	289, 228, 229, 231, 230, 226, 227, 232
	33kV	248, 247

TABLE 7.2. FAS AND MTS FOR ALL PLACEMENT METHODS

Placement Method	**Mean FAs**	**Mean MTs**	**Mean MTs + Mean FAs**	**Percent Network <1% FAs & MTs (*U*)**
STP1	6.61	1.97	8.59	78.06%
STP2	4.35	5.44	12.05	76.80%
LOC	7.19	1.98	9.16	75.90%
MRA	3.40	11.10	14.50	51.74%
ENG	9.38	10.37	19.75	53.0%

Table 7.2 shows that the STP_1 solution has the lowest number of missed trips and false alarms. It also shows the STP_1 solution to be the best in terms of the amount of network where missed trips and false alarms can be accurately estimated.

TABLE 7.3. FAS & MTS FOR 10 TO 20 MONITORS USING STP1 & LOC PLACEMENT

Placement Method	Percent Network <1% FAs & MTs (*U*) for Each No. of Monitors					
	10	**12**	**14**	**16**	**18**	**20**
STP1	78.6%	80.8%	80.2%	80.9%	83.9%	84.5%
LOC	77.0%	79.1%	80.4%	82.3%	83.9%	83.5%

Table 7.3 shows how U varies as the number of monitors is varied between 10 and 20 monitors in the network. The table shows that for each increase in monitors, the performance (U) of the monitoring program increases. From an observability perspective, the number of buses where the voltage is observable for 2 to 20 monitors increases from 8% to 37% and 7% to 38% of busbars for STP_1 and *LOC* respectively.

Fig. 7.5 shows that the STP1 algorithm misses trips in the 132kV network, whilst false alarms occur at all voltage levels. This is probably understandable, given that the STP_1 monitoring has 9 monitors at 11kV, 1 monitor at 33kV and no monitors at 132kV, and there is no preference in the optimization algorithm as to which voltage level to place monitors.

Fig. 7.5 also shows that the optimization algorithm has located a monitor on feeder A for the STP_2 solution to accurately monitor the STP at the important customer.

It is also worth comparing the accuracy of all the monitor placement solutions with regards to estimating voltages across the network. It is expected that the *LOC* solution should perform best on this comparison since the objective of this placement method is to minimize the error in estimating voltages at non-monitored buses. Fig. 7.4 shows the per unit error in estimating voltages between 0 and 1.5 per unit.

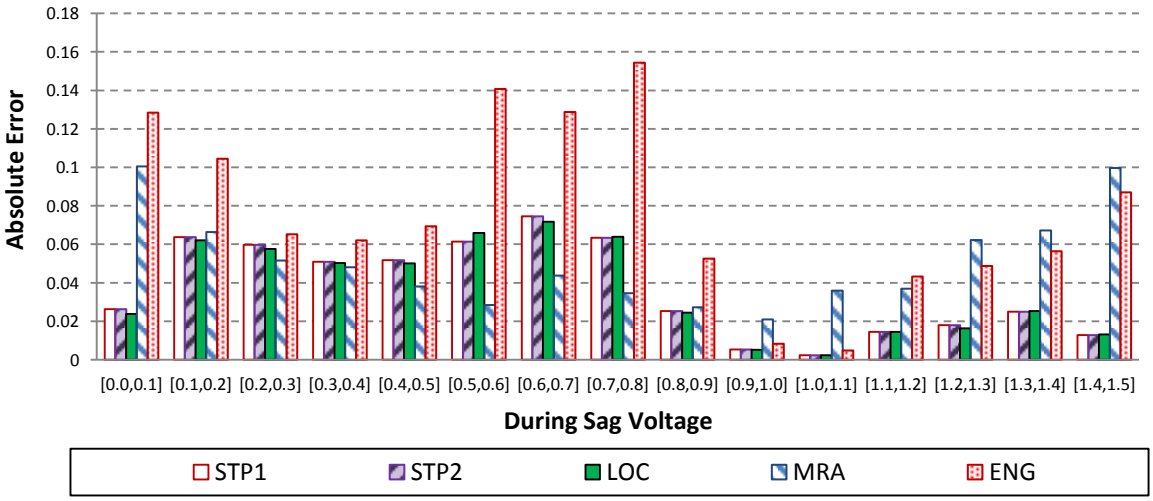


Fig. 7.4. The absolute error in estimating voltages at non-monitored buses for the three monitoring solutions split into 0.1 per unit intervals.

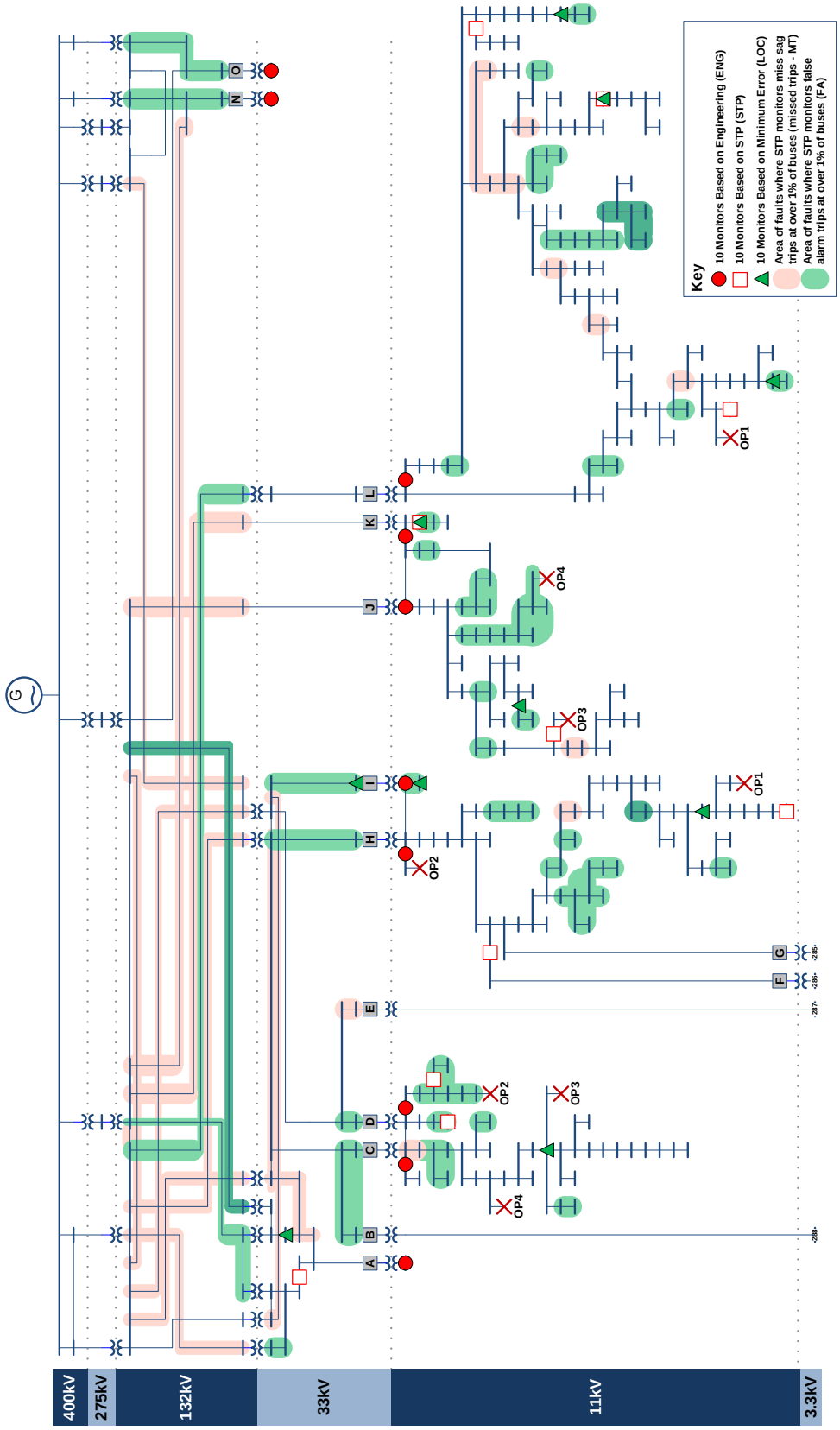


Fig. 7.5. The 3 best monitoring solutions: ENG, LOC and STP in the 295 generic distribution system network. The areas highlighted show the areas where if a fault occurs, the STP monitor will incorrectly label over 1% of the buses with missed trips or false alarms

The *LOC*, STP_1 and STP_2 solutions outperform the *ENG* solution across all ranges of sag magnitudes. The *LOC* and the *STP* solutions are also very closely matched, and there is only a small amount (approx. ±10%) of difference in their performance for voltage magnitude estimation. The *MRA* solution shows higher errors in voltage magnitude estimation at high and low voltages, although it does outperform all methods at around 0.5 to 0.8 per unit.

7.5 Future Network Uncertainties

The second aim of this research is to show that by identifying a wide variety of near optimal solutions, there is a higher potential of finding a higher quality monitoring solution when considering future network uncertainties. To establish whether or not this is the case, the 420 (10×42) solutions found by the BCA were compared against the other monitoring solutions across a range of network loading and topology uncertainties.

Each solution found by the BCA in the *S'* dataset was subjected to testing using the $Test_D$ dataset to form a vector *U* Each element of *U* represents the quality *U* of a solution $s' \in S'$ at a particular loading and topology. *U* vectors were also formed for the STP_2, *LOC*, *MRA* and *ENG* monitoring solutions.

Each set of solutions found by the BCA were then subjected to the Pareto optimality tests. These tests yielded a set of solutions known as *S'**. Nine Pareto optimal solutions were found.

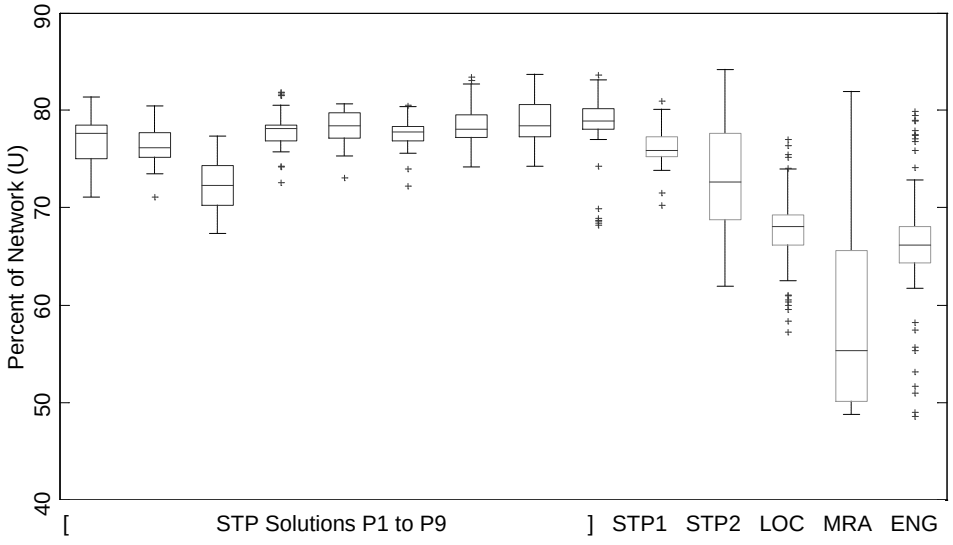


Fig. 7.6. The distribution of U for all expected future network loading conditions.

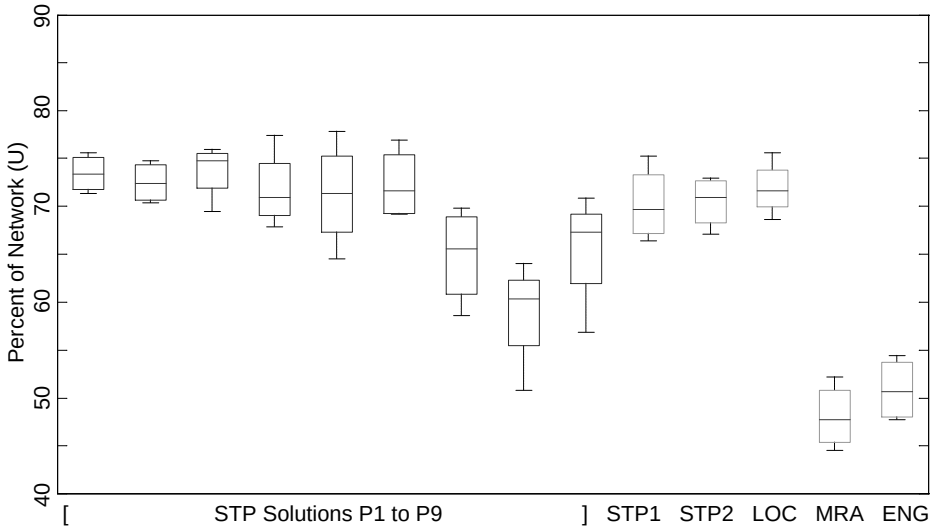


Fig. 7.7. The distribution of U for all expected future network topologies where each network topology is equally likely to occur.

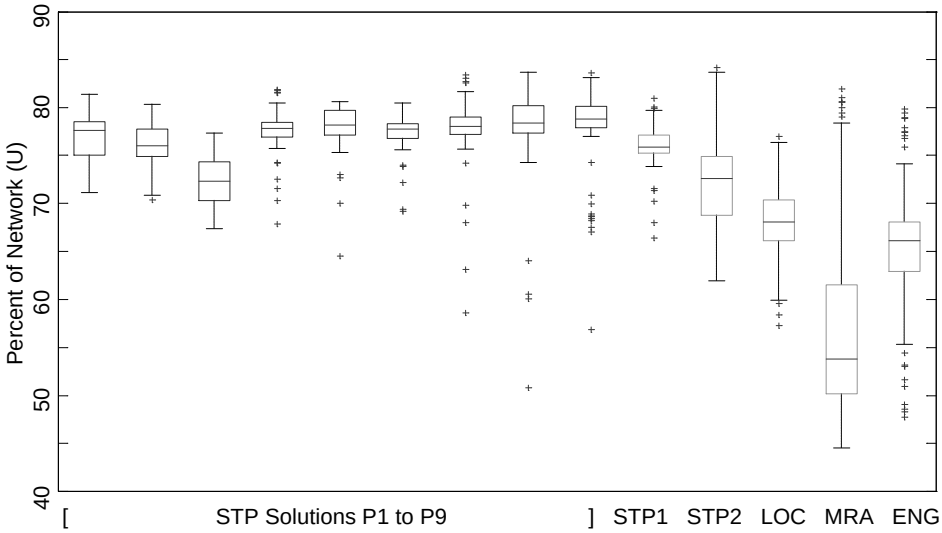


Fig. 7.8. The distribution of U for all expected future network topologies and loading conditions.

Fig. 7.6, Fig. 7.7 and Fig. 7.8 shows the distribution of *U* for variations in network load, topology and both load and topology together. The figures clearly show that the Pareto solutions have a high mean and low range compared with all other monitoring solutions. These Pareto solutions are therefore the most insensitive to network loading changes and achieve the highest overall levels of performance.

Fig. 7.7 shows a comparison of the solutions across all the topologies. In this figure, each topology is equally likely to occur. The results show large variations in all of the monitoring solutions performance, as evidenced by the long bars against each of the monitoring solutions. In this study, the base topology is expected to operate for 96% of the time. The results in Fig. 7.8 are therefore strongly influenced by loading, but only mildly influenced by topology changes. This is easy to see by comparing Fig. 7.8 with

Fig. 7.7 and noting that the bars in Fig. 7.8 are generally shorter than Fig. 7.7, although outliers can be seen in Fig. 7.8 caused by topology variation.

7.6 Analysis of Results

For the task of estimating sag trip probabilities (STPs), Table 7.2 shows that STP focused optimization procedure produces solutions of a higher quality than other monitor placement methods. Even when the optimization procedure is constrained (as in solution STP_2), the STP optimized monitor set is better at predicting at which buses equipment are likely to trip than any of the other compared techniques. The difference in performance between the STP and the LOC methods is the smallest (as shown in Table 7.3), and in a practical sense, the difference between the methods would only result in a small difference in the total number of trips caused by voltage sags. Fig. 7.4 also shows that the STP solutions are almost as good at estimating during fault voltages as monitoring solutions specifically designed for voltage profile estimation, such as MRA and LOC. It should be pointed out that although trips were the focus of this research, the number of trips may not always be the best criterion to use. Converting trips into economic costs might be more appropriate in some circumstances.

The optimization methodology presented in this chapter uses a BCA optimization technique. It was hoped, that the BCA near optimal solutions would perform better than a single optimum solution over a range of network topology and loading uncertainties. Fig. 7.6, Fig. 7.7 and Fig. 7.8 clearly show that set of Pareto optimal solutions (S'^*) derived from the BCA are the highest quality monitoring solutions, when both future load and topology are incorporated. The results support the hypothesis that the best and most robust monitoring solution (in terms of future network uncertainties) is not necessarily the optimum solution at a particular operating scenario. Indeed, the Pareto solutions found by the BCA all have a smaller range (indicating they are more insensitive to future uncertainties) and higher median (indicating they predict STP better on average) than the optimal solution at base load and topology (STP_1). All the Pareto solutions significantly outperform the other techniques, which show they are very sensitive to changes in loading and topology.

Fig. 7.6, Fig. 7.7 and Fig. 7.8 also show that because all of the solutions in S'^* are of high quality, any of these solutions could be selected as the chosen monitoring solution. This gives a DNO flexibility and freedom to choose from a variety of monitoring solutions rather than a single optimal solution. This is important, as the selection made

by an operator can then also be based on other factors such as accessibility, communications infrastructure and physical security.

Fig. 7.7 shows that certain monitoring solutions are strongly influenced by changes in network topology. However, the results also show that Pareto solutions *STP-P1* to *STP-P6* all show robust performance when topology changes are considered. This demonstrates that the methodology can be successfully applied to produce robust monitoring solutions when network topology uncertainties are considered.

7.7 Summary

This chapter presented a new robust methodology for placing monitors within a distribution network. The optimized solutions were analyzed for robustness across a range of uncertain future network loading and topology scenarios. The results show how the methodology can be applied to a 295 bus generic distribution network to generate robust solutions in a network with uncertain load growth and topology over the next 15 years. The results were tested against other techniques and shown to compare strongly.

The task of robustly positioning monitors was achieved using an immune inspired B-cell algorithm (BCA). The BCA was configured to optimize sag trip probabilities (STPs) that penalize reconstruction estimates based on the probability that equipment attached to a busbar will trip. The BCA generated a diverse pool of solutions which were optimized over future loading and topology uncertainties using a Pareto optimization procedure.

The research showed a considerable promise for using an immune inspired BCA for optimal monitor placement. The solutions produced by the BCA were more robust at estimating STPs than those based on a single optimal monitoring solution. The development of a range of near optimal solutions also gives DNOs flexibility to choose from a suite of solutions rather than choose a single optimal solution. This is perhaps the single biggest advantage of using an immune inspired population based technique, and it is expected that this advantage could be exploited for other studies.

In a unique advancement of existing research, the monitors were placed with the objective of estimating the likelihood of equipment trips during voltage sags rather than focusing on estimating the voltage sag profile itself. A second advancement shows how to obtain a robust set of monitoring solutions which are able to estimate the effects of voltage sags over future loading and topology uncertainties.

The methodology presented in this chapter is a practical and flexible way of placing monitors to estimate the affects of voltage sags. It takes into account future network loading and topology uncertainties and can be configured to work with the diverse range of customers and both equipment and process sensitivities. It focuses directly on the effects of voltage sags, rather than simply estimating voltages.

The optimal monitoring results presented in this chapter are used throughout this thesis as the basis of the studies presented for voltage sags and unbalance. It should be pointed out thought that none of the techniques developed in this thesis require an optimal monitoring set. The results presented in this chapter however show that if one is available, it can enhance the performance of a monitoring algorithm. The following chapter explores how monitoring information from these monitoring sets can be used to estimate the source and effects of voltage unbalance in distribution networks.

8 Estimating the Source and Effects of Unbalance

8.1 Introduction

Voltage unbalance is generated through the emission of zero and negative sequence current at sources throughout the network. Voltage unbalance emission sources include asymmetric un-transposed or partially transposed transmission lines [23], single phase and dual-phase loads and unbalanced three phase loads. Locating all of the multiple interacting emission sources is not a straightforward task [75], but it can be achieved using techniques such as [76], which identify the level and contribution made by asymmetrical lines and loads towards voltage unbalance. There has been limited research however, into real-time identification of unbalance sources and their associated effects in distribution networks using a limited set of power quality monitors.

In Chapter 2, the tool for distribution system state estimation (DSSE) was introduced to estimate the state of a power network. The DSSE approach used in Chapter 5 predicted pre-fault voltages at non-monitored busbars and assumed that the network was balanced and the pseudo-measurements were uncorrelated.

In this chapter, a three phase DSSE model will be developed to estimate the location and impact of unbalance within the network, without assuming balanced loading. The DSSE model will deal with the correlative nature of three phase DSSE measurement errors [46] which is critical to understand how a three phase phenomenon like unbalance spreads throughout the network [71-73].

An important aspect of three-phase DSSE is the correlation between measurement errors. In [46] the authors touched upon the correlative nature of three phase DSSE measurement errors. Correlation in multi-phase networks is covered in more detail in [71-73], where the later reference describes how probabilistic three phase load can be used to estimate the probabilistic distribution of voltage unbalance. References in this area are limited, and the current body of research lacks a rigorous method to incorporate three phase correlated measurement errors into a real-time DSSE.

The research presented in this chapter advances the current body of research on DSSE and real-time unbalance detection. The developed methodology is able to locate and estimate the impact of unbalance on a real distribution network using statistical techniques. Correlated measurement errors are incorporated into a three phase DSSE by analysing the correlation between different types of estimates. The DSSE is formulated to make use of the diverse range of measurements available in a typical distribution network by performing case studies using a real network with some real data.

In a similar way to the techniques developed for monitoring voltage sags, the DSSE techniques for unbalance monitoring must be able to synthesise data using real monitoring information. This includes being able to deal with a set of arbitrary devices at different locations, each reporting a variety of measurement inputs to a pre-defined level of accuracy. Specifically, for this case study, this will involve using information from single phase monitors to clarify three phase system behaviour.

The research presented in this chapter has two principal aims:

1. To provide a framework for performing three phase state estimation on a network where measurement information is incomplete, and must be estimated.
2. To highlight how the results of DSSE can be used to statistically estimate the location, level and impacts of unbalance on a typical distribution network, and thus ultimately help fix unbalance related issues.

8.2 Study Background

The inspiration for the study network described in this chapter comes from conversations with engineers at a UK distribution network operator in the UK. A section of their network (shown in Fig. 8.1) was experiencing voltage unbalance. The network comprises 14, 33kV busbars, and 10 11kV busbars. The 11kV busbars are all loaded, whilst all of the 33kV busbars are unloaded. Each 11kV busbar is connected to

the 33kV network via a Δ-Y transformer, with an on load tap changer. The network's topology is shown in Fig. 8.1.

The most problematic busbar in the network in terms of unbalance was considered (by DNO network engineers) to be bus 15. The VUF at this busbar was directly measured at two different instances as 0.7% and 1.9%. The unbalance was causing tripping of industrial processes in customer facilities connected to this busbar. Buses 15 and 23 are the most heavily loaded busbars with 8-10 MVA, whilst the other busbars range in loading from 0.5 MVA to 3 MVA. Each busbar has an average of 2% industrial, 13% commercial and 85% residential customers. An accurate understanding of the source of the unbalance, its estimated impacts on customers and the wider network was required in order to fix the problems caused by unbalance in the network.

The DNO's plan was to install a series of three phase monitors into the network to monitor both the level and impact of the unbalance on customers within the network. It was hoped that these monitoring measurements would allow them to be able to ultimately fix the unbalance problem.

At the time of writing this thesis, there were still no fully operational three phase monitors been installed in the network and therefore three phase monitoring data was not available. The case studies presented in this chapter are therefore manually synthesised simulations which take into account the likely location of both monitors and unbalance sources. Although three phase monitors were not available, the network did contain a series of single phase line to line meters which measured voltage magnitude, current magnitude and apparent power magnitude at 11kV.

8.2.1 Network Description

Fig. 8.1 shows that the network has ten single phase monitors installed which measure line to line voltage magnitude and current. Voltage magnitude, $|V^{(xy)}|$, is measured through a VT between two arbitrary phases *x* and *y* at 11kV. Current flow magnitude is measured through a CT attached to an arbitrary phase z, $|I^{(z)}|$. Using these two measurements, and assuming the network is balanced, an estimate for the total 3-phase apparent power consumption, $S_{3\phi}$, can be calculated. The angle between the voltage and current was not recorded.

The topology and impedances of the network shown in Fig. 8.1 is described in detail in Appendix B. In the studies that follow, it was assumed that the topology of the network was balanced, implying no coupling between positive, negative or zero sequences. No

zero sequence current path exists in the 33kV network (shown as a purple in Fig. 8.1) since all transformers are Δ-Y connected, prohibiting the flow of zero sequence currents from 11kV to 33kV. This assumption is only valid if there are no 33kV loads or in-feeds which could inject zero sequence currents into the meshed 33kV network, and thus assumptions listed in Chapter 2 are not violated.

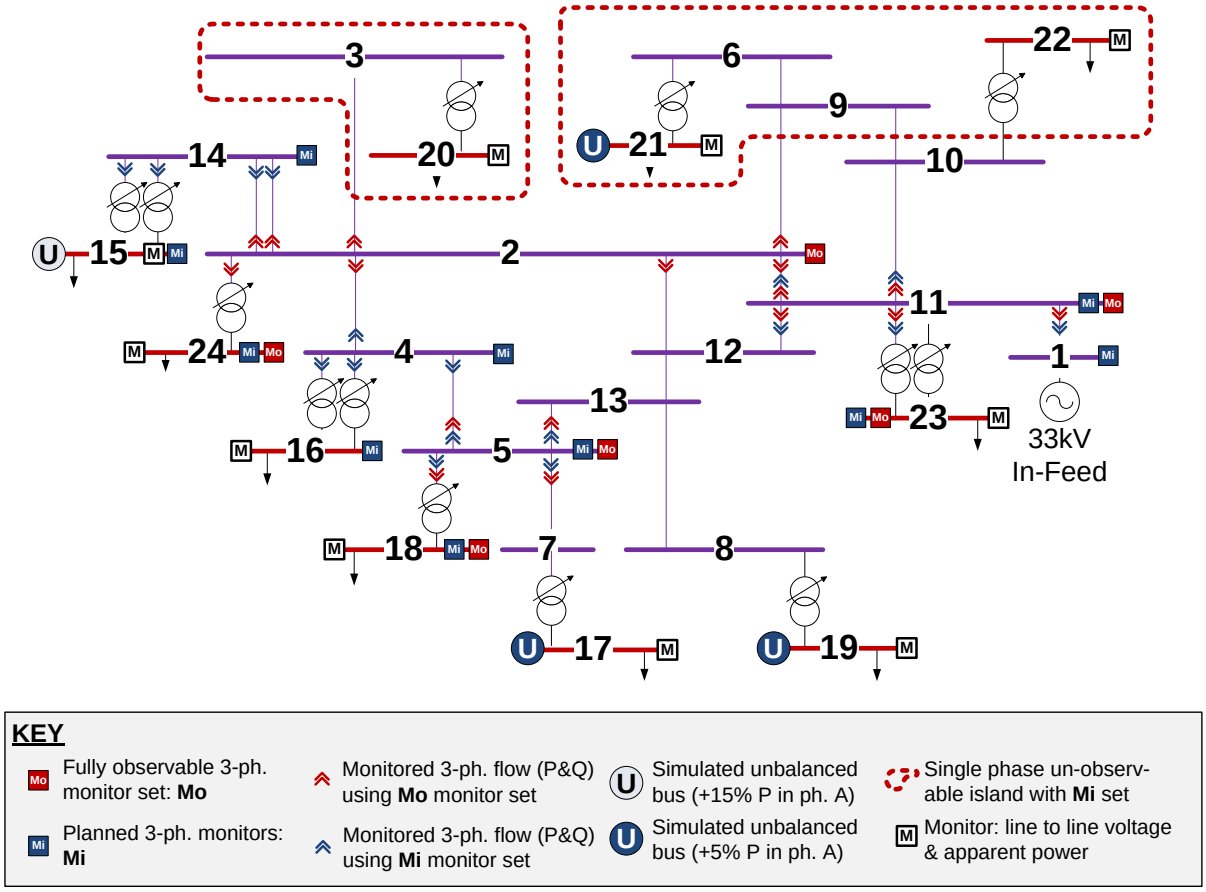


Fig. 8.1. The 24 bus section of the UK distribution network.

8.2.2 Monitoring in the Network

Fig. 8.1 shows the locations of two sets of monitors: M_i and M_o. These sets will be used alongside the single phase monitors to estimate the unbalance level, location and impact of unbalance in the network. M_i is representative of set of monitors which the DNO will install into the network in the near future. M_o represents a set of monitors which is able to observe the full state of the network using single phase state estimation [123]. All of the monitors in the M_i and M_o sets measures phase to neutral voltage magnitude, real power and reactive injections and real and reactive line flows in all three phases.

8.2.3 Voltage Unbalance Issues in the Network

Like all DNOs in the UK, it is important for the operators of this network to have a strong understanding of the level of unbalance within the network as this will affect three phase customers, network losses, heating, operational limits (of three phase machines, cables and lines) and influence customer minutes lost (CML). The DNO also has a regulatory requirement to operate their network within statutory unbalance limits (<2% in the UK [10]). It is important for the DNO to understand whether the unbalance is being generated from within its own network, or from unbalanced customer loads.

From a consumer's perspective, the effects of unbalance include three phase equipment trips, heating of industrial machinery (caused by increased negative and zero sequence currents) and torsional oscillations. The presence of unbalance may also necessitate that three phase machinery is de-rated to operate at less than full load.

The network shown in Fig. 8.1 consists of a mixture of single phase loads connected at 415V and three phase loads connected at both low voltage and 11kV. Single phase loads are not affected by unbalance unless the unbalance causes over or under voltages which exceed statutory limits. The most sensitive loads to unbalance are three phase motors [168]. There is a large industrial customer with a three phase motor connected to bus 15 (shown in Fig. 8.1). Simulated unbalance sources are shown in Fig. 8.1 as originating from unbalanced loading of three 11kV busbars.

The methodology discussed in this chapter will help to tackle these problems by formulating a distribution system state estimator which can provide insight into the flow of negative sequence energy and the level of negative sequence currents and voltages. An accurate understanding of the source of the unbalance and its estimated impacts on customers and the wider network will be required to fix the unbalance problem in the network.

8.3 Distribution System State Estimation

Weighted least squares distribution system state estimation (DSSE) was introduced in Chapter 2. In this chapter, Chapter 2's DSSE formulation will be extended to take into account all three phases, the extra information from single phase monitoring devices and the intrinsic correlation between sets of three phase pseudo-measurement estimates.

8.3.1 Power & Voltage Measurements

All the three phase monitors in M_i and M_o (shown Fig. 8.1) were assumed to measure real and reactive power at all connected busbars in all three phases as well as the voltage magnitude in all three phases. Each real measurement was modelled with a standard deviation of 0.2% of the mean (or scheduled) value of the measurement. Measurement errors were assumed to be independent from one another and other types of measurements. Further details on the assumed measurement error model can be found in Chapter 2.

8.3.2 Virtual Measurements

Virtual measurements are used when the value of the real and reactive power injection at a busbar is defined. The real and reactive power injections at all unloaded busbars in Fig. 8.1 were modelled as virtual measurements. Virtual measurements were modelled with an absolute standard deviation of 2×10^{-7} [69]. Virtual measurement errors were assumed to be independent from one another and other types of measurements.

8.3.3 Pseudo-Measurements

Constructing pseudo-measurements for three phase state estimation is complicated by the intrinsic correlation in three phase systems. Correlated random variables can be used to take account of this problem [71, 73].

8.3.3.1 Scheduled Power (SP)

The most basic form of a pseudo-measurement is to assume a value based on a historical average at that point in the network. Scheduled power (SP) has the advantage of being easy to estimate. However, the variances of the pseudo-measurement errors are high, and there is a high level of correlation between pseudo-measurements at different busbars (see Table 8.1).

SP estimates can be used for three phase state estimation by assuming power consumption in each phase is balanced, and equal to a third of the overall three phase power. The power factor is also assumed constant across all phases.

8.3.3.2 Load Estimation (LE)

Load estimation (LE) concerns estimating the nodal load level [169, 170] at a bus. Many factors will influence the load profile at a bus including ambient weather conditions, time of day, season, and the types of customers (for example residential,

commercial or industrial) connected to a bus. Utilities store several pieces of information which can be used to estimate the load at a busbar including peak load data, customer billing data and load curves for different customer types [170].

In [67, 69] and [170] the load at a particular busbar was estimated by decomposing the load profiles into several customer classes. The load at a bus can then be estimated by proportionately combing the diversified load profiles of each customer type and scaling profile by the recorded maximum power consumption.

In this research, load estimates were created using the method described in [69], with the assumption that the diversity (and hence power factor) of customers across all sites was constant, and distributed evenly across each phase. An estimate for the load profile was obtained by normalising a set of load profiles recorded at busbars with similar customer types to a maximum value of 1 per unit. The profiles were summed together and averaged, before being re-scaled by the maximum amount of loading at that busbar. The loading in each phase was assumed to be balanced; thus the pseudo-measurement estimate for the power consumption in each phase was assumed to be equal.

8.3.4 Mixed Model (MM) Pseudo Measurements

In some situations, a limited amount of monitoring may be available at a site. For example, as in the network shown in Fig. 8.1, monitors may be installed which measure power consumption in phase A only. In this scenario, a mixed model (MM) can be created using a mixture of real and estimated measurements.

Knowledge of the power consumption in one phase can provide significant insight into the estimated power consumption in the other two. A distribution system that supplies only three phase loads will show a high level of correlation between real and reactive power in each phase. A system which supplies single phase loads will show lower levels of correlation between the phases. A real distribution system contains a mixture of single and three phase loads.

In [171], a correlation model for real and reactive power was developed by assuming that power consumptions can be modelled as a set of jointly correlated Gaussian random variables. This assumption is valid as long as the number of electrical appliances supplied by the network is large enough, which is usually the case.

In a similar way to [171], the expected level of correlation between real and reactive power can be estimated by considering the correlation between phases at a typical bus. Let $P^{(a)}$, $P^{(b)}$, $P^{(c)}$, $Q^{(a)}$, $Q^{(b)}$ and $Q^{(c)}$ be the real and reactive power respectively flowing

in each phase, with each measurement normalized by the factor required to make the mean power consumption in phase $P^{(a)}$ and $Q^{(a)}$ equal to 1MW and 1MVAr respectively. The aim is to construct a multi-variate normal distribution that encompasses dependencies of the variation of each of these variables. Using this distribution, the active and reactive power consumption in phase *i* ($P^{(i)}$ and $Q^{(i)}$ respectively) can be determined. The multi-variate normal distribution is defined in (8.1).

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{8.1}$$

where $\mathbf{X}$ is a vector of random variables, $\boldsymbol{\mu}$ is a vector of means, and $\boldsymbol{\Sigma}$ is the covariance matrix.

Using these assumptions a six dimensional covariance matrix $\boldsymbol{\Sigma}$ was estimated using the random variables in the vector $\mathbf{X}$ (8.2).

$$\mathbf{X} = [P^{(b)}, \mathrm{Q}^{(b)}, P^{(c)}, Q^{(c)}, P^{(a)}, Q^{(a)}]^T \tag{8.2}$$

The covariance matrix should ideally be built using measured per phase data from several buses of representative size and customer type, averaged over a defined time interval. Estimating the covariance across several different buses ensures that the variation in fixed offset power consumption (where one phase is permanently loaded more than the other two) is incorporated into the covariance matrix estimate.

The covariance matrix used in this research was estimated using data recorded over a 21 day period at a 15MW bus. The dataset comprises 1007, 30 minute samples of all 6 variables. Information on the data used to construct this model can be found in Appendix C.

The mean vector $\boldsymbol{\mu}$ is represented as:

$$\boldsymbol{\mu} = [\overline{P}^{(b)'}, \overline{Q}^{(b)'}, \overline{P}^{(c)'}, \overline{Q}^{(c)'}, 1, 1]^T \tag{8.3}$$

where $\overline{P}^{(i)'}$ and $\overline{Q}^{(i)'}$ are the mean real and reactive power consumptions in the *i*th phase. If there is no evidence of an individual phase being persistently overloaded, $\boldsymbol{\mu}$ will be very close to a unit vector.

If a measurement is available for power consumption in one phase only (e.g. $\hat{P}^{(a)}, \hat{Q}^{(a)}$), a conditional multivariate distribution [172] can be derived as $(\mathbf{X} | [P^{(a)}, Q^{(a)}] = [\hat{P}^{(a)}, \hat{Q}^{(a)}] \sim N(\overline{\boldsymbol{\mu}}, \overline{\boldsymbol{\Sigma}})$, where $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\Sigma}}$ are shown in (8.4) and (8.5).

$$\overline{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\Sigma_{22}^{-1}\left\{\begin{bmatrix} 3\hat{P}^{(a)}/P_{3\Phi} \\ 3\hat{Q}^{(a)}/Q_{3\Phi} \end{bmatrix} - \boldsymbol{\mu}_2\right\} \tag{8.4}$$

$$\overline{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{8.5}$$

$$\boldsymbol{\mu}_1 = [\overline{P}^{(b)'}, \overline{Q}^{(b)'}, \overline{P}^{(c)'}, \overline{Q}^{(c)'}]^T \tag{8.6}$$

$$\boldsymbol{\mu}_2 = [1,1]^T \tag{8.7}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \tag{8.8}$$

where $\boldsymbol{\Sigma_{11}}$, $\boldsymbol{\Sigma_{12}}$ and $\boldsymbol{\Sigma_{22}}$are regions of $\boldsymbol{\Sigma}$ (as shown in (8.8)).

$\hat{P}^{(a)}$ and $\hat{Q}^{(a)}$ are estimates for the power consumption in one phase, and may be derived from a load estimation or scheduled power pseudo-measurement (as described in Chapter 2) or a measurement from a real monitor. Fig. 8.2 illustrates the correlative model using two example values for the power in phase A.

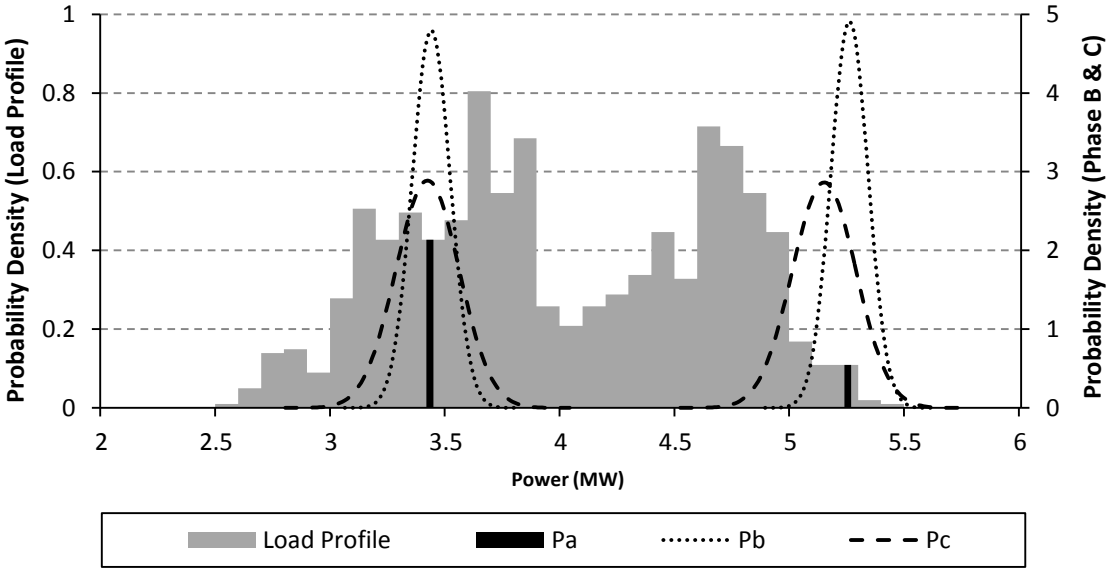


Fig. 8.2. Two Gaussian distributions obtained using the correlative model and data from the UK distribution network. The dashed lines represent the probability density function of Pb and Pc conditional on a value of Pa.

8.3.5 Measurement Error Covariance Matrix: R

The measurement error models can be incorporated into the state estimation equations (described in Chapter 2) by altering the measurement error covariance matrix **R**.

The errors in the pseudo-measurement models described in Chapter 2 and the mixed-pseudo measurement model developed in this chapter are all highly correlated. The correlation can be conceptualised by considering a simple example. Consider a scheduled power (SP) pseudo-measurement made at two busbars in an arbitrary network. At times of peak load, the SP pseudo-measurement is likely to under-estimate both loads, and at times of light load, the SP pseudo-measurement is likely to over-

estimate both loads. Thus, the errors in the pseudo-measurement error will be strongly correlated.

It is therefore important to take into account the covariance of both inter and intra busbar measurement errors [46], as well as understanding their typical variance, and whether or not the errors are normally distributed.

8.3.5.1 Analysis of the Pseudo Measurement Errors

The scheduled power (SP), load estimation (LE) and mixed-model (MM) pseudo-measurements were all analysed to assess 5 properties of the measurement errors:

- Variance of errors in each phase: $\sigma_{X_i^{(a)}}$, $\sigma_{X_i^{(b)}}$ and $\sigma_{X_i^{(c)}}$
- Real power error correlation, all phases, at bus *i*: $corr(e_{P_i^{(p)}}, e_{P_i^{(m)}})$
- Reactive power error correlation, all phases, at bus *i*: $corr(e_{Q_i^{(p)}}, e_{Q_i^{(m)}})$
- Real and reactive power error correlation , all phases, at bus *i*: $corr(e_{P_i^{(p)}}, e_{Q_i^{(m)}})$
- Real-real, reactive-real, reactive-reactive error correlation, at bus *i*, *j*: $corr(e_{X_i^{(p)}}, e_{X_j^{(m)}})$

The measurement models were tested using a combination of single phase and three phase data of real and reactive power demands recorded at four sites in the wider UK distribution network. Single phase data was recorded every 30 minutes for 10 days on phase A at all sites. Three phase data was available at one site over the same period. The load diversity of each site was comparable with each site comprising a mixture of industrial (1% to 2%), commercial (11% to 18%) and residential (77% to 88%) customers.

The standard deviations of the measurement models are shown in columns 2 to 4 of Table 8.1. The inter-phase error correlations (columns 5 to 7) were obtained by averaging correlation coefficients of the measurement errors when applied to the three phase site data. The cross-busbar correlation coefficients (column 8) were obtained by averaging the correlation coefficients obtained when cross correlating the measurements errors of real and reactive power for phase A between two sites. The errors were tested for normality at the 5% significance level using the Anderson-Darling test [173] (column 9).

Table 8.1 shows that the SP, LE, and MM pseudo-measurements errors are highly correlated. There is also considerable correlation between the SP and LE models across different busbars (column 8 of Table 8.1). The SP measurement model failed the normality test, and the LE model passed the test for only 2 out of 4 sites. The properties of the errors shown in Table 8.1 can be used to form the covariance matrix for the state estimator.

TABLE 8.1. MEASUREMENT ERROR VARIANCES AND CORRELATIONS OF THE DIFFERENCE TYPES OF PSEUDO-MEASUREMENT MODELS

Pseudo-Measurement Type	$\sigma_{X_i^a}$	$\sigma_{X_i^b}$	$\sigma_{X_i^c}$	$corr(e_{P_i^p}, e_{P_i^m})$	$corr(e_{Q_i^p}, e_{Q_i^m})$	$corr(e_{P_i^p}, e_{Q_i^m})$	$corr(e_{X_i^p}, e_{X_j^m})$	**Normally Distributed?**
	As a percentage of scheduled power at the busbar							
Real Measurements	0.02%	0.02%	0.02%	0	0	0	0	Yes
Sch. Power (SP)	16%	16%	16%	0.99	0.96	0.78	0.87	No
Load Est. (LE)	7%	7%	7%	0.93	0.61	0	0.52	Yes (2/4 sites)
Mixed Model (MM)	0.02%	10%	10%	0.98	0.96	0.78	0	Yes

σ is the standard deviation, P and Q are real and reactive power measurements, X corresponds to an arbitrary real or reactive power measurement, p and m are arbitrary phases (a, b or c) and i, j are arbitrary busbars.

8.3.6 Building a Correlated Measurement Error Covariance Matrix

The correlation coefficient $\rho_{i,j}$ defines the relationship between the covariance of the two measurements X_i and X_j normalized by the product of their standard deviations. This is described in (8.9) and (8.10).

$$\rho_{i,j} = \text{cov}(X_i, X_j) \big/ \sigma_i \sigma_j \tag{8.9}$$

$$\text{cov}(X_i, X_j) = \sigma_{ij}^2 \tag{8.10}$$

The covariance matrix can be constructed using the correlation coefficients in Table 8.1. The diagonal terms are defined by the measurement type (either pseudo or real) and its associated standard deviation (columns 2 to 4 of Table 8.1). The off-diagonal terms are derived from the results in columns 5 to 8 of Table 8.1 and calculating the covariance using (8.10).

The covariance matrix will also incorporate error estimates for voltage magnitude and line flow measurements. In this research, voltage magnitude and line flow measurements were only present when a monitor was available to monitor these

parameters. The errors in these measurements were therefore assumed to be independent of other pseudo or real measurements errors.

8.3.7 Customer Types

Each of the pseudo-measurement models discussed above will be influenced by the types of customers attached to each modelled section of network. Past analysis [67] of the UK distribution network divided the network into 4 customer types: domestic / unrestricted, domestic / economy, commercial and industrial. Each of these types of customers were found to have average lagging power factors of 0.95, 0.99, 0.98 and 0.90 respectively [67].

In this research, the data used to construct the pseudo-measurements models and target network to be studied have a broadly similar proportion of customer types attached to each busbar. It was therefore assumed that the average power factor would be roughly the linear combination of 1% industrial customers, 16% commercial and 83% residential (split 50% economy and 50% unrestricted). The modelled average power factor at every site was therefore 0.97.

8.4 Methodology

Three case studies were considered to illustrate the performance of the three phase distribution system state estimation approach.

Firstly, the state of the network was estimated with full load on all busbars using an installed set of three phase monitors (M_i, representative of a set installed by a DNO) and single phase monitors (represented by MM pseudo-measurements) at all other 11kV busbars.

Secondly, the state of the network was estimated at full load on all busbars using a set of three phase monitors (M_o) selected by single phase observability analysis along with single phase monitors (represented by MM pseudo-measurements) at all other 11kV busbars.

Lastly, the state of the network was estimated over a 24 hour period using the M_i monitor set with each of the three types of pseudo-measurements at all non-monitored 11kV busbars.

The M_o monitor set was selected by performing single phase observability analysis on the network [58]. M_o was selected by evaluating single phase observability of the network for each of the $N = \sum_{i=1..24} 24Ci \approx 1.7 \times 10^7$ combinations of monitors, and recording

the combination with the smallest number. Single phase observability analysis also reveals that the M_i cannot observe some sections of the 11kV and 33kV network (as shown in Fig. 8.1). Single phase observability analysis is representative of the kind of technique a DNO may use when considering where to place monitors in their network.

Three phase observability analysis highlights that both the M_i and M_o monitor sets can't observe three-phase voltages at unmonitored 11kV busbars. This is caused because the Δ-Y transformers halt the propagation of zero-sequence currents. With no knowledge of the zero sequence currents, their impact upon the level of unbalance on secondary side of the transformer is undefined. Single phase monitors and MM pseudo-measurements are required at all 11kV busbars to make the network three phase observable.

In the third case study, the level of unbalance was estimated using the M_i monitor set. SP, LE and MM pseudo-measurements were added to all unmonitored 11kV busbars in a sequential fashion to create three separate sets of results. All of the busbars were loaded according the recorded historical data recorded over 24 hours on 1st March 2008. Full details of the loading on each busbar can be found in Appendix B.

Unbalance was simulated at 4 busbars in the network across all three case studies. 5% real and reactive power was added to the red phase (A) of busbars 17, 19 and 21, and 15% real and reactive power on phase A of bus 15. This is shown in Fig. 8.1.

The aim of the third study is to estimate the impact of unbalance on both the DNO and consumers over a longer period of time using each different pseudo-measurement model. This was evaluated by estimating the statistical distribution of the level of unbalance, the extra losses caused by unbalance, the expected duration that the level of unbalance exceeds statutory limits, and the expected derating of three phase machines in the network.

A summary of the case studies is shown in the Table 8.2.

TABLE 8.2 PERFORMANCE OF EACH PAIR OF MONITORS WHEN MEASURING THE VUF

Case	Mon.	PM	Description	Results
1	M_i	MM	Estimate level and location of unbalance at peak load.	Unbalance factor at buses 15, 20 & 21. Likely location of unbalance.
2	M_o	MM	As above.	As above.
3.1	M_i	MM	Estimate the distribution of unbalance for a 24 hour time frame.	Unbalance factor bus 21. Total losses. Regulatory limits (time >2%). Machine derating at bus 15.
3.3	M_i	SP	As above.	As above.
3.3	M_i	LE	As above.	As above.

8.4.1 Monte Carlo Distribution State Estimation

All of the results were obtained using a Monte Carlo simulation. 100 Monte Carlo scenarios were generated by running the state estimator with 100 sets of measurements each distributed as defined by their pseudo-measurement or real measurement error properties.

8.5 Results & Analysis

In each of the following case studies, a persistent level of load unbalance was added to buses 15, 17, 19 and 21. Fig. 8.3 shows the true level and location of the unbalance. Even though the emission of negative sequence energy is constrained to these four busbars, voltage unbalance was present at all of the buses in the network.

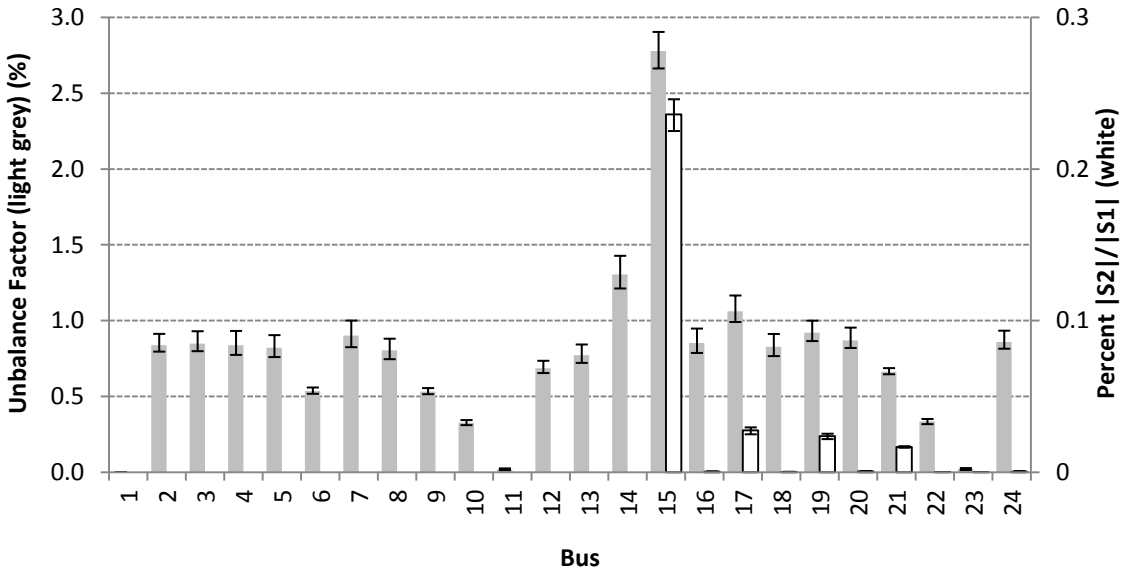


Fig. 8.3. The magnitude of the VUF in % (light grey) and the magnitude of the injected negative sequence apparent power (white) at each bus in the network, with error bars representing the range of values of 90% mutual impedance errors.

In all of the case studies it was assumed that the network was balanced and symmetric based on topological information obtained from the UK DNO. The effect of this assumption on the location and level of unbalance in the network was explored by adjusting the mutual impedance between phases AB, BC and CA in all lines in the

network to 90% of the mutual impedance between the other two phases [174]. The error bars in Fig. 8.3 show the range of variation which was observed whilst applying the asymmetric topologies. The purpose of this analysis is to highlight the variation in VUF and $|S^{(2)}|/|S^{(1)}|$ caused by realistic network topology asymmetries. Fig. 8.3 shows that the asymmetry in topology causes the level of unbalance to vary by ±0.13% and also varies the magnitude of injected negative sequence apparent power by up to ±0.01%, both at bus 15. Thus, topological variations of this magnitude do not heavily influence either the location or level of unbalance for this study.

Estimating the level and location of unbalance as close to the results shown in Fig. 8.3 is the goal of the first two case studies.

8.5.1 Case 1: Estimating Unbalance with M_i

Fig. 8.4 (A), (B) and (C) highlight the results of the first case study by showing the VUF at important busbars (A), the percentage ratio of $|S^{(2)}|/|S^{(1)}|$ (B), and its associated inter-quartile range (IQR) (C) (where $S^{(2)}=V^{(2)}I^{(2)*}$ and $S^{(1)}=V^{(1)}I^{(1)*}$ is the apparent negative and positive sequence power respectively). Fig. 8.4 (A) shows the distribution of the VUF at both buses 15 and 21 is fairly narrow, indicating a high level of accuracy in the VUF estimate. The spikes at 2.8% and 0.7% correspond with the true level of unbalance of 2.8% and 0.66% at buses 15 and 21 respectively.

Fig. 8.4 (A) also shows the level of unbalance at bus 20. The unbalance here has a wider inter-quartile range than the other two busbars. This highlights that the monitor set was less ideally positioned to estimate the unbalance factor at this busbar. This was expected, since bus 21 resides within a single phase un-observable area of the network (see Fig. 8.1).

Fig. 8.4 (B) shows that M_i was able to detect voltage unbalance appearing at buses 15, 17 and 19 with a high level of certainty. However, there was ambiguity as to whether there was unbalance at bus 20, as indicated by the wide distribution at bus 20. This is highlighted in Fig. 8.4 (C), where a spike at bus 20 indicates that there was a large degree of uncertainty in the ratio of $|S^{(2)}|/|S^{(1)}|$.

The ambiguity present at bus 20 means that the location of the unbalance source at bus 21 was only loosely identified within a region of three buses (labelled 20, 21 and 22) using monitor set M_i. This information would be useful to a network operator to begin troubleshooting an unbalance problem, but more information (and hence more monitors) would be required to further define the exact unbalance source in the network.

It should be noted that this is not a deficiency of the proposed technique, but a validation that the method's outputs depend on the quality of its inputs.

8.5.2 Case 2: Estimating Unbalance with Mo

Fig. 8.5 (A), (B) and (C) shows the results obtained with the M_o monitor set and MM pseudo-measurements at 11kV busbars. In contrast to Fig. 8.4 (A), the VUF distributions of Fig. 5 (A) are narrower, especially at bus 20. This is because unlike M_i, M_o has single phase observability over all busbars in the network. Fig. 8.5 (B) and (C) show that the uncertainty in the estimate of $|S^{(2)}|/|S^{(1)}|$ at bus 20 is much reduced when compared with monitor set M_i.

One interesting aspect of Fig. 8.5 (C) is the relatively high level of IQR at bus 21 when compared with other busbars. Bus 21 and its 33kV primary at bus 6 are not directly monitored. Their state estimates are made from multiple monitors (at buses 11 and 2) whose measurement errors compound together to increase the error at bus 21. If M_o is augmented to include an additional three phase monitor at bus 21, the inter-quartile range at bus 21 declines by a factor of 14 from 0.014% to 0.001%. This adds certainty that there is definitely an unbalance source at bus 21. This analysis highlights that in certain situations additional monitors may be required to increase precision.

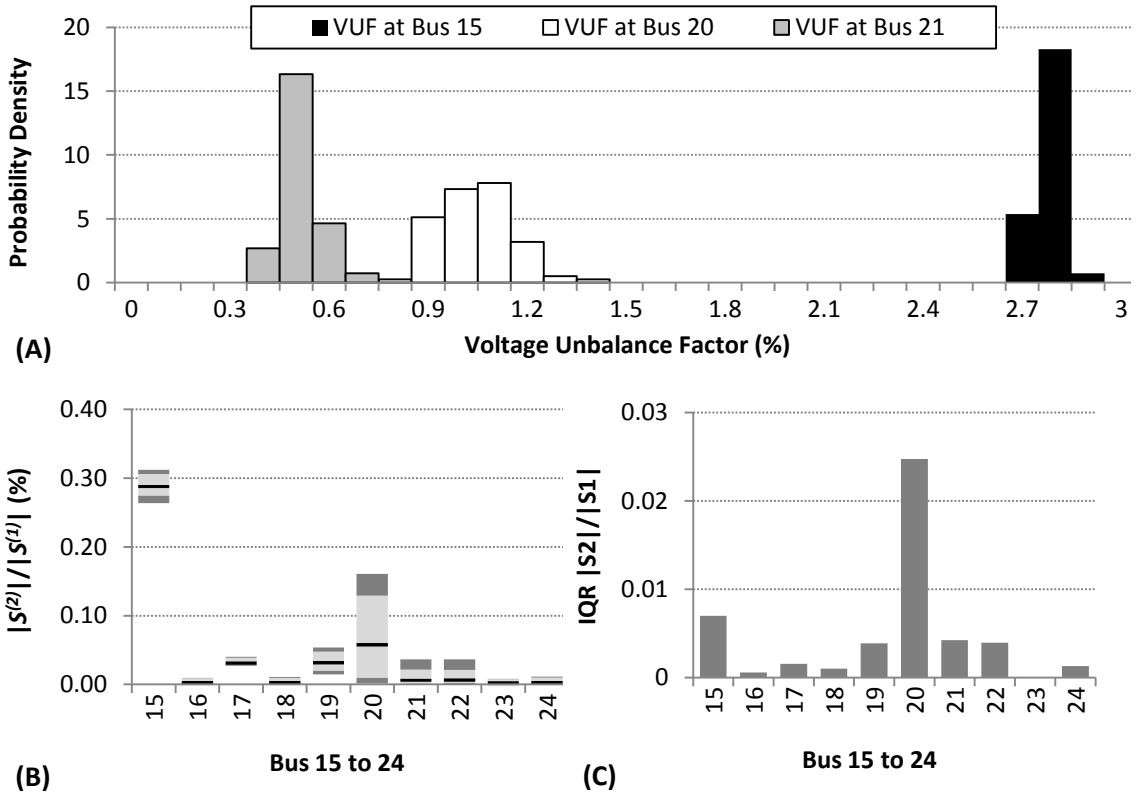


Fig. 8.4. Case study 1: estimating the level and location of unbalance using monitor set M_i (installed monitor set).

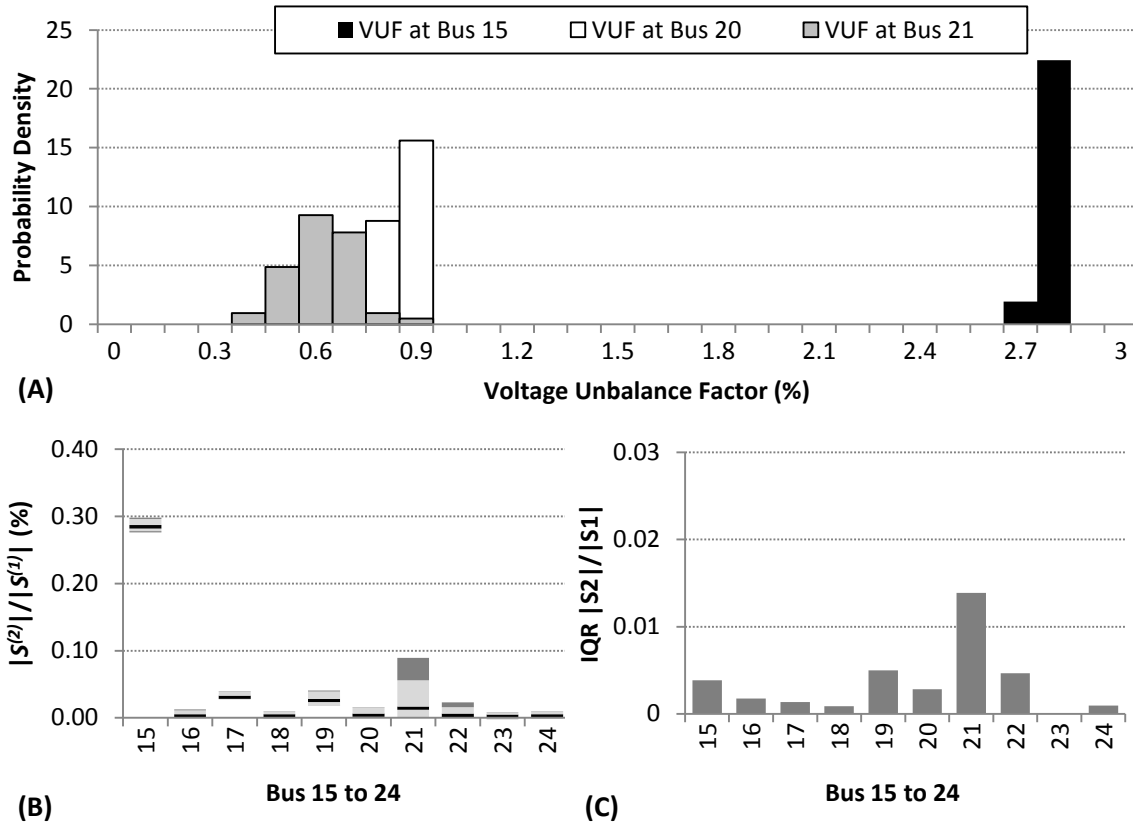
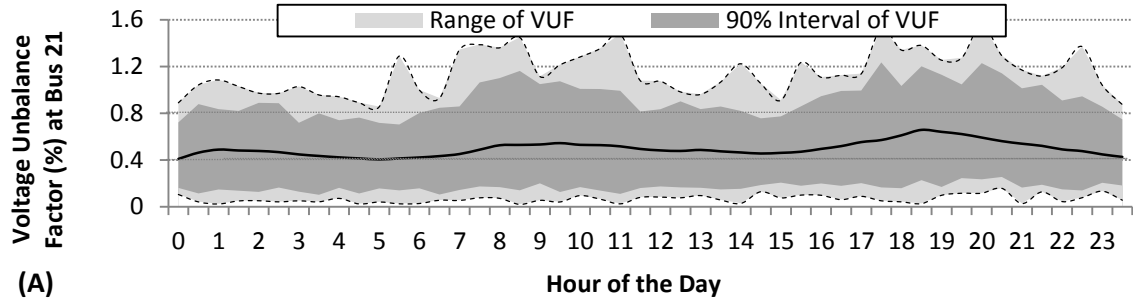


Fig. 8.5. Case study 2: estimating the level and location of unbalance using monitor set M_o (fully observable monitor set).

8.5.3 Case 3: Estimating 24 Hours of Unbalance with M_i

The last case study highlights the proposed method using each of the pseudo-measurement types on an arbitrary winter day. The reported results for this case study focus on the distributions of the VUF at bus 21 and the impact of voltage unbalance on customers at bus 15.

Bus 21 was analysed because it is single phase un-observable using the M_i monitor set and the accuracy of its state estimate will directly depend on the accuracy of the chosen pseudo-measurement model. Bus 15 was analysed because three phase machines connected to this busbar are highly sensitive to unbalance.



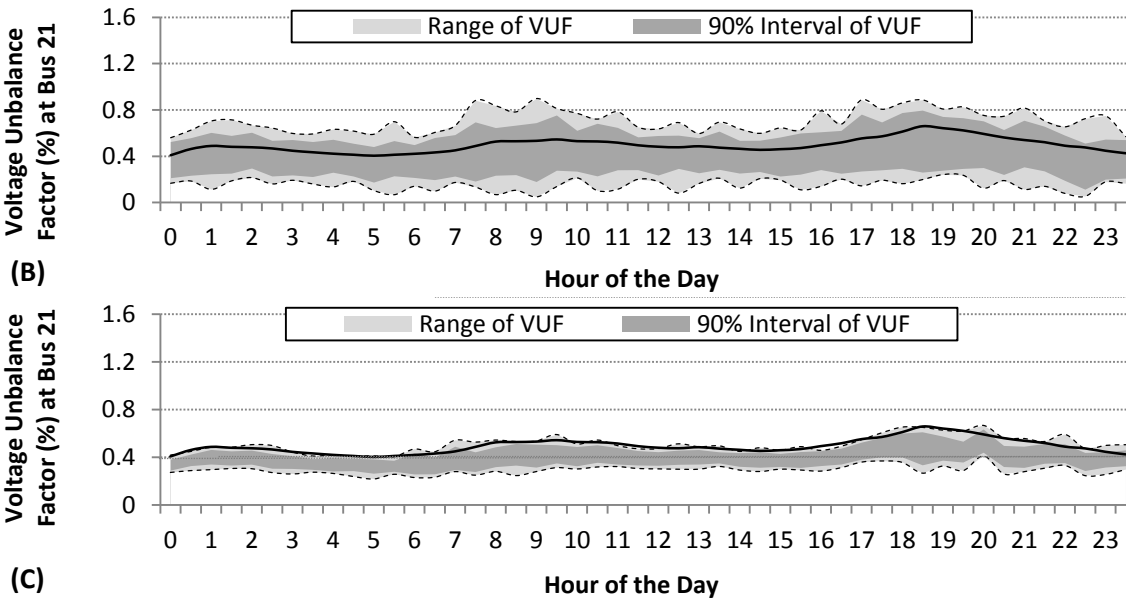


Fig. 8.6. VUF estimated over a 24 hour period using SP (graph A), LE(graph B) and MM (graph C) pseudo-measurements at bus 21. The dark line is the true VUF at bus 21 over 24 hours.

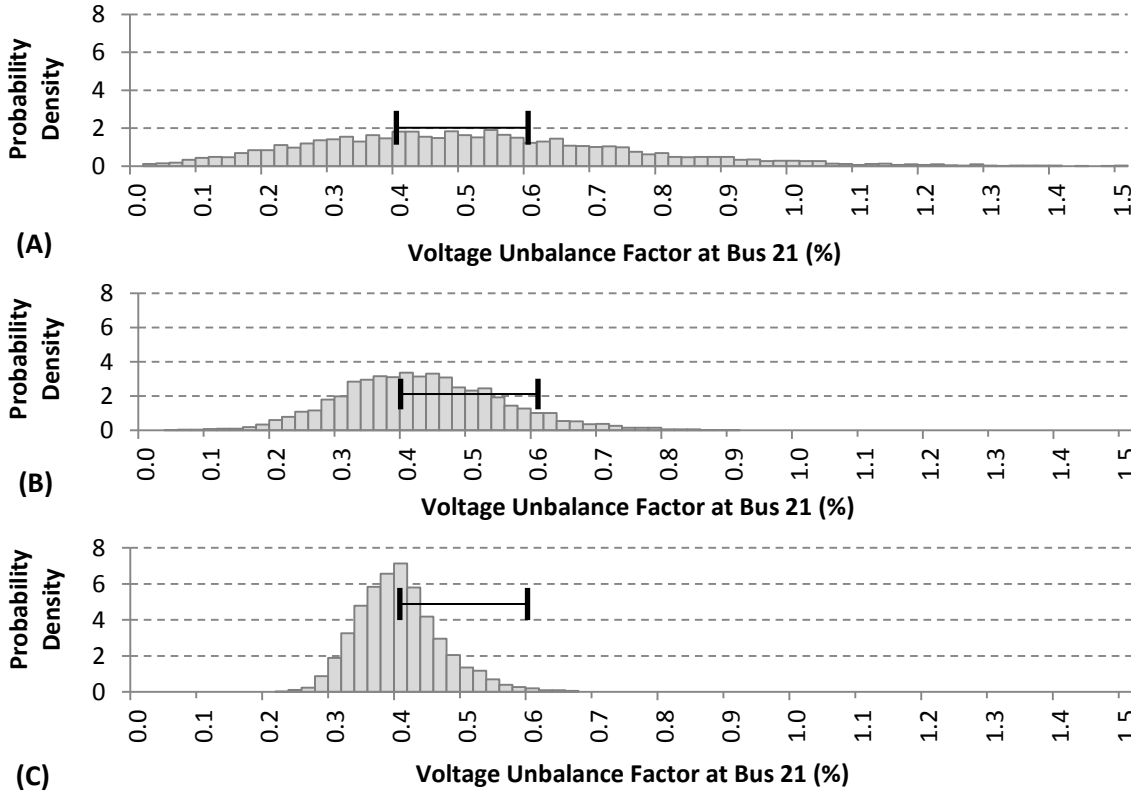


Fig. 8.7. PDF of the VUF at bus 21 over 24 hours using SP (graph A), LE (graph B) and MM pseudo-measurement (graph C) with error bars showing the true range of VUF (0.4% to 0.66%) over the 24 hours.

8.5.3.1 Unbalance Factor at Bus 21

Histograms of the VUF at bus 21 over the 24 hour period using each of the pseudo-measurement models are shown in Fig. 8.6 and Fig. 8.7.

The VUF at bus 21 varies from 0.40% to 0.66% throughout the 24 hour period. The MM pseudo-measurement ((C) of Fig. 8.6 and Fig. 8.7) model shows the tightest estimated distribution of the VUF of the three pseudo-measurement models. Fig. 8.6 and Fig. 8.7 show that the estimated distribution for all of the pseudo-measurement models correctly overlaps the true VUF.

Using any of the pseudo-measurement models, the right hand tails of the histograms of Fig. 8.7 (A), (B) and (C) all show that the VUF at bus 21 is not likely to exceed 1.5% within the 24 hour period. This information would allow a DNO to confidently predict that the level of unbalance at bus 21 did not exceed statutory limits over the past 24 hours, using any of the three pseudo-measurement models.

Fig. 8.7 (C) shows the expected distribution of the VUF is shifted roughly 0.1% lower than the true range of the VUF. This is shown again in Fig. 8.6 (C), where the 5% / 95% range for the mixed pseudo-measurement model consistently falls under the true VUF by approximately 0.1%.

The under-estimation at bus 21 can be explained by analysing the estimated VUF at bus 22. Like bus 21, bus 22 falls within a single phase un-observable island (as shown in Fig. 8.1) when using the M_i monitor set and no pseudo-measurements. The exact composition of the phase unbalance at these two busbars is therefore uncertain, and is estimated using pseudo-measurements. Unlike bus 21, bus 22 does not emit unbalance into the network. With no prior knowledge of the ratio of unbalance emissions between bus 21 and 22, the state estimator evenly apportions unbalance across both busbars. The level of unbalance in bus 21 is therefore underestimated (Fig. 8.7 (C)), and the level of unbalance at bus 22 is overestimated. The median VUF at bus 22 over the 24 hour period is 0.17%, whereas the estimated median is 0.23%; an overestimate of 0.06%.

8.5.3.2 Total Losses Caused by Unbalance

The fraction of losses caused by unbalance is shown in Fig. 8.8. The fractional losses are defined as in (8.11).

$$P_{fi} = \{(P_i - P_b)/P_b\} \times 100 \tag{8.11}$$

where P_{fi} is the percentage fraction of losses caused by unbalance, P_i are the total real losses in the network with unbalanced loads and P_b are the total nominal losses in the network when each of the loads are perfectly balanced.

Fig. 8.8 shows that expected losses from unbalance vary from 0.3% to 0.8% of nominal losses for the 24 hour period. Fig. 8.8 also shows that the estimation of total network losses is minimally affected by the choice of pseudo-measurement model.

Total losses of between 0.3% to 0.8% are slightly less than those reported in [1], which stated that losses caused by unbalance in a typical network could be between 1% to 4% of nominal losses. The loss profile in Fig. 8.8 would change significantly if the number of unbalanced loads increased or if the cables and lines involved in the network were also unbalanced. The transformers in the network also affect the overall unbalance losses. Δ-Y transformers stop the flow of zero sequence currents whilst generating losses through a circulating current in the Δ windings.

It is expected that the loss estimates would be significantly affected by the choice of PM model if the single phase un-observable islands were larger or contained loads which with a higher MW rating.

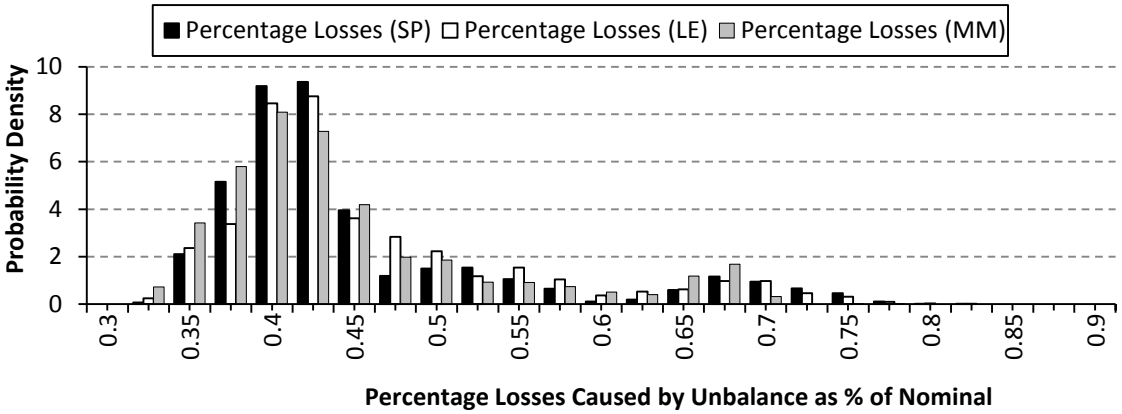


Fig. 8.8. The percentage losses caused by unbalance as a percentage of the total nominal losses for each of the pseudo-measurement models.

8.5.3.3 Regulatory Limits: Estimated Time >2%

2% voltage unbalance is an important regulatory limit for UK distribution networks. The state estimator can be used to estimate the expected time that the unbalance factor exceeds 2% for any busbar in the network using equation (8.12).

$$P(Y > 2) = \int_{2}^{\infty} f(y)dy \tag{8.12}$$

where Y is a random variable that describes the VUF at an arbitrary busbar with probability density function $f(y)$. Fig. 8.7 shows that at bus 21, $P(Y>2)=0$, as the distribution does not exceed 2% for the any of the pseudo-measurement models.

A similar calculation was performed at bus 15 which revealed that $P(Y>2)=28\%$. This indicates that the DNO was in-violation of statutory limits for 28% of the day.

8.5.3.4 Machine De-rating

The state estimator can be used to statistically assess the impact of unbalance on three phase machinery placed anywhere within the network. For example, a three phase machine connected at bus 15 will have to be de-rated if the level of unbalance exceeds machine tolerances.

Fig. 8.9 shows the probability density function for machine derating at bus 15. This was obtained by combining the probability density function of the VUF with the NEMA machine derating curve [175].

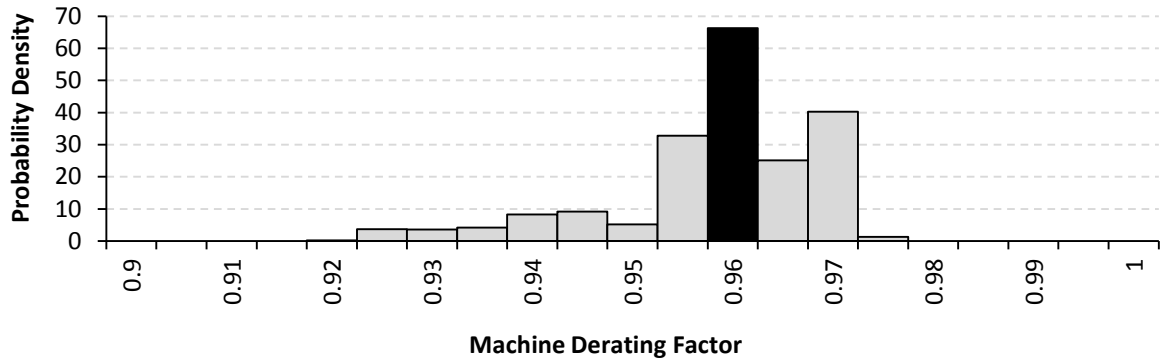


Fig. 8.9. PDF of machine derating at bus 15 during the 24 hour period. The highlighted bar indicates the median derating: 96% of full load.

Fig. 8.9 shows that a three phase machine connected to bus 15 will have to be de-rated, perhaps by as much as 8% during some periods of the day. On average, it was found that the median machine derating at bus 15 during the 24 hour period was 0.96. This implies that a machine connected to this busbar will be de-rated by an average of 4%.

Probability distributions for machine de-rating can also be obtained at busbars which are un-observable using statistical distributions such as those in Fig. 8.7. All of the pseudo-measurement models correctly predict that a machine connected to bus 20 or 21 would not have to be de-rated during the 24 hour study period.

8.6 Summary

In this chapter, an online state estimator was developed which is capable of identifying the level, location and effects of voltage unbalance in a section of UK distribution network.

The first aim of this research was to develop a framework for performing three phase state estimation on a network where measurement information is incomplete. This aim was satisfied by developing a coherent methodology using distribution system state estimation (DSSE).

Unlike the DSSE approach, developed in Chapter 2, the DSSE model in this chapter is three phase and incorporates the expected correlations between pseudo-measurement errors. The proposed methodology was shown to be practical, and gives guidance on how limited information from different monitoring sources (for example single phase monitors) can be pieced together to form a reliable three phase estimate for the state of the network.

The second aim of this research was to estimate the level, location and impacts of unbalance on a typical distribution network. To achieve this goal, the methodology was validated on a section of the UK distribution network using real measurement data. It was shown to provide reliable statistical estimates for the level and location of unbalance and its effects on regulatory limits, losses and machine de-rating.

The technique discussed in this chapter shares a number of similarities with the research developed for voltage sags in Chapters 4, 5 and 6. Like the voltage sag research, the developed methodology can work with any number of arbitrarily accurate monitoring devices. The output of the both the voltage sag profile estimation and the unbalance research are statistical distributions. The results of the unbalance and voltage sag profile estimation process can also be used to estimate the likely impact of these phenomena on customers within the network.

In the next chapter, the techniques developed in this chapter will be merged with the earlier discussions on voltage sags. The aim will be to create a single impression of the network which can estimate the impact of both voltage sags and unbalance on customers within the network.

9 Weakest Areas & Worst Served Customers for Sags and Unbalance

9.1 Introduction

The discussion so far has focussed on how to independently monitor a network for voltage sags and unbalance. This chapter will draw the techniques together by showing how to identify both the weakest areas as well as the worst served customers for both phenomena with a single set of limited monitoring devices.

The aim of this chapter is to demonstrate how the worst served customers and the weakest areas can be identified in feeder L of the 295 bus distribution network (Fig. 4.3). This task will be accomplished using both a full and a limited set of monitors. The results presented in this chapter illustrates how the proposed approaches could be used to help distribution network operators (DNOs) plan infrastructure investment and optimise maintenance decisions, using only a limited set of monitors.

It is expected that voltage sags (being the most prevalent power quality phenomenon) will contribute most towards negative effects on customers within the network. As unbalance detrimentally affects three phase machinery, it is expected to contribute less towards negative effects on customers.

9.2 Weakest Areas and Worst Served Customers

Each of the techniques developed in this thesis have been shown to be capable of helping to identify the sources and impacts of both voltage sags and unbalance. The sources of power quality issues can be considered as *weak areas* and the customers affected can be regarded as the *worst served customers*.

Chapters 4, 5 and 8 present techniques which can be used to identify weak areas whilst Chapters 6 and 8 present techniques which can be used to identify the worst served customers.

9.2.1 Worst Served Customers

Identifying the worst served customers for unbalance and voltage sags should summarise the resultant impact of both sags and unbalance at a particular location in a network. This is synonymous with understanding the power quality performance of a specific bus in a network.

An ultimate common consequence of both a voltage sag and / or unbalance is an equipment trip. Trips due to unbalance are rarer and typically limited to three phase machinery. The configuration of the protection attached to a three phase machine determines whether or not it will trip during high levels of unbalance. Voltage sags however affect single phase and three phase loads, and their impacts can spread throughout the network. The resultant impact of both these power quality issues can be summarised by estimating the expected number of CIs at a busbar. If knowledge is available on the duration of an interruption caused by a sag or unbalance, it is then subsequently possible to determine the total customer minutes lost (CML). CIs and CML can also be segmented by customer type, to identify which types of customers are being affected most severely.

The number of trips due to voltage sags can be determined using an immunity curve, for example the ITIC curve shown in Fig. 6.2. There is no equivalent immunity curve concept which defines when loads sensitive to unbalance sensitive will trip. It is reasonable to assume that trips are likely to start occurring after equipment has been exposed to a level of unbalance greater than regulatory limits (of say 2%), for greater than 10 minutes (as per EN50160 [10]). This concept is defined as a *nominal trip*. The unbalance immunity curve used to define nominal trips is shown in Fig. 9.1.

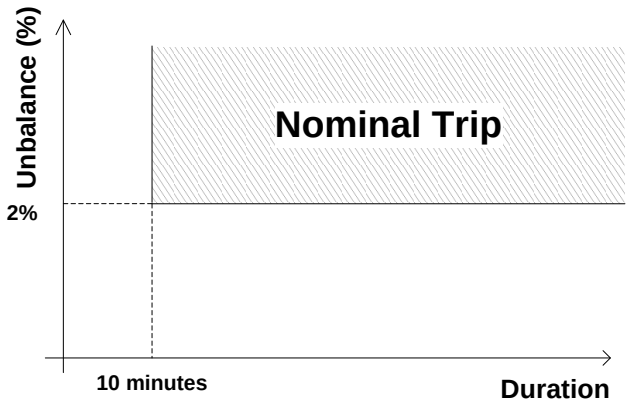


Fig. 9.1. The unbalance immunity curve used to define nominal trips.

There are also other ways of defining the weakest areas of the network for both voltage sags and unbalance independently. For example, the number of sags at a busbar could be represented by the SARFI index [11], or a Generalized Sag Table [59]. The level of unbalance in % provides an indication of the severity of the unbalance problems at a busbar, as does amount of time unbalance exceeds a regulatory limit (such as 2%). The electrical loss caused by unbalance is another measure which DNOs would find useful when assessing the impact of unbalance in their network.

The resultant impact of a trip is that electrical energy is not supplied to an area of network for a period of time. The total amount of electrical energy not supplied (EENS) is the total CML multiplied by the size of the load. Using additional knowledge of the types of customers attached to a busbar and therefore the average cost of an interruption for a given customer type, EENS, CML and trips can be combined to obtain the total economic cost of the interruption [2].

Identifying the worst served customers within the network is important from a regulatory and planning perspective. Regulators are interested in power quality performance metrics to ensure customers are being delivered a high quality service. Network planners are keen to understand potential power quality issues before new loads are connected to the network.

9.2.2 Weakest Areas of the Network

The most basic way of assessing the weakest areas of the network for voltage sags and unbalance is to identify the areas within the network with the highest fault rate and the areas with the most unbalanced loads. This analysis can be improved by considering the resultant impact of the voltage sags or unbalance in terms of customer interruptions (CIs), customer minutes lost (CML) or economic costs to either the DNO or the

customer. By assigning the impact of the power quality issue to its source, the relative strength of the network can be assessed.

Identifying the weak areas in a network is most relevant to DNOs. It is important for DNOs to be able to identify weak areas so they can plan their re-enforcement strategy. Weak area analyses could also be used to help identify the root cause of specific problems impacting on one or a number of customers in a specific area of a network.

9.3 Summary

Table 9.1 summarises some of the metrics which can be used to assess the weakest areas and the worst served customers within the power network for voltage sags and unbalance. The bullet points highlighted in bold in Table 9.1 will be analysed in more detail on a case study network.

TABLE 9.1. METRICS USED TO ASSESS A NETWORK'S WEAKEST AREAS AND WORST SERVED CUSTOMERS

Power Quality Issue	**Worst Served Customers**	**Weakest Areas of the Network**
Voltage Sags	• **Number of Trips (number)** • Customer minutes lost (CML) • Economic impact of CML, no. of trips and energy not supplied (£) • SARFI Index • Generalized Sag Table	• **Source of Customer Interruptions (CIs)** • Economic impact to the DNO (£) • Economic impact to the customer (£) • Customer minutes lost (CML) • Total MW not supplied (MW)
Unbalance	• **Number of Nominal Trips (number)** • Machine trips caused by unbalance (number) • Number of customer minutes lost due to unbalance nominal trips (CML) • Machine de-rating (MW not supplied) • Economic impact of CML, no. of trips and energy not supplied (£) • Level of VUF (%) • **Time >2% VUF (time)** • Asset damage caused by torsional oscillations and heating (£)	• **Source of Customer Interruptions (CIs)** • **Percent negative sequence energy injection $S^{(2)}/S^{(1)}$ (%)** • Negative sequence energy injection (MVA) • Additional losses caused by unbalance (MW) • Asset damage caused by torsional oscillations and heating (£) • Customer minutes lost (CML) • Economic impact to the customer (£) • Total Economic impact to DNO (£)

9.4 Case Study

The aim of the case studies is to demonstrate how the number of trips, the number of CIs the level of unbalance and the location of unbalance (shown in bold in Table 9.1) can be obtained using limited monitoring and the techniques developed in this thesis.

9.4.1 Network

It was proposed that the metrics described in Table 9.1 could be analysed using feeder L of the 295 bus generic distribution network (first described in Chapter 4). Feeder L consists of 95 11kV buses and 1 33kV busbar. The network was assumed to be fed through a balanced slack bus at 33kV modelled at bus 265.

Feeder L was assumed to contain both residential and industrial customers. Three industrial customers were assumed to be connected to buses 137, 225 and 232, with the rest of the loads assumed to be a mixture of commercial and residential customers. Industrial loads were modelled to follow load profile 8 in Appendix D and residential and commercial loads were modelled to follow load profile 1 in Appendix D. The numbers and types of customer attached to each busbar are described in more detail in Appendix E. The power factor at each busbar was assumed to remain constant.

9.4.2 Monitors

Two sets of results were generated for each of the case studies. The first set assumes that monitors were installed at every bus in the network, whilst the second set assumes that only four monitors are installed in the network. The comparison of these two sets of results will highlight the accuracy of the techniques proposed in this thesis.

Three phase power quality meters measuring 3-phase voltage magnitude ($|V|$) and real (P) and reactive power (Q) at each busbar were assumed to be installed at buses 130, 204, 217 and 232. Single phase monitors were assumed to be installed at every busbar in the network which measure power consumption in one phase only.

Three of the four monitors (130, 204 and 217) were derived from the optimal sag monitoring set determined in Chapter 7. An additional monitor installed on the secondary side of the primary substation (bus 232). This location was selected as an additional site for monitoring as DNOs have historically preferentially installed monitors at important substations to monitor both the busbars attached to 33kV to 11kV transformers.

9.4.3 Unbalance

The network was loaded for 24 hours (as described in Appendix B), with the exception of buses 130, 132, 163, 173, 209 and 220 where 5 times the 3 phase power demand was assumed to flow through phase A only. This modification was required to generate unbalance of greater than 2% in some areas of the network.

Unbalance was estimated using three phase and mixed model pseudo-measurements generated from three phase and single phase monitoring devices respectively. Unbalance was estimated using DSSE as outlined in Chapter 8.

9.4.4 Voltage Sags

Voltage sags were assumed to originate from single phase to ground (SLG) faults. Faults were simulated on 11kV overhead lines and cable. Three phase, line to line and double line to ground faults were not modelled because SLG faults are by far the most common type of fault, and to improve the interpretability of the results.

Faults were simulated in proportion to the number expected to occur within a 24 hour period. Each cable was assumed to have a fault rate of 3.8×10^{-2} faults / day and each overhead line was assumed to have a fault rate 4.7×10^{-2} faults / day [129]. Each SLG fault was assumed to occur in the centre of each line, and modelled using the sag duration model described in section 6.4.3.

It should be noted that by simulating faults in this way, the total number of faults recorded over the 24 hour study period will not be an integer number. The 24 hour study period was selected for this demonstration to simplify the loading of the network and eliminate factors which will otherwise complicate the results. Over a 24 hour period, the electrical load profile follows a broadly cyclical pattern. Over longer periods of time (such as a year), network loading is dependent on other factors which otherwise complicate the proposed study such as weather, time of year and type of day (e.g. holiday or working day). In reality, the performance of a network is more likely to be judged over a longer period of time. The objective of this study is to show that the proposed methods work accurately over short time periods so that they can be confidently applied to longer study intervals.

Voltage sags were detected and estimated using the three phase monitoring devices and voltage sag profile estimation (as described in Chapter 4 to 6).

9.5 Assessing the Worst Served Customers

Worst served customer analyses focus on the customer who is most affected by the power quality issues. The worst served customers were assessed using three of the metrics shown in bold in Table 9.1:

- Percentage of 24 Hour Period >2% VUF (*Unbalance*)
- Number of Nominal Trips Caused by *Unbalance*

- Number of Trips Caused by *Voltage Sags*

It should be noted that any of the metrics outlined in Table 9.1 could be used to identify the worst served customers for voltage sags and unbalance using the techniques presented in this thesis. The metrics presented in this chapter were chosen to illustrate the flexibility of the proposed methods.

9.5.1 Percentage of 24 Hour Period >2% Unbalance

The amount of time that each bus rose above 2% unbalance was estimated using DSSE and calculating the percentage of the 24 hour period when the level of unbalance was above 2%. This is important information which enables a DNO to asses which buses are violating regulatory limits and also provide information on which busbars would be likely to generate equipment trips.

9.5.2 Number of Trips Caused by Unbalance

The total number of trips caused by unbalance was estimated using DSSE (as described in Chapter 8). All residential loads were assumed to be immune to the effects of unbalance, since these customers were assumed to be connected to one phase only. It was also assumed that increases or declines in voltage caused by unbalance on any one phase were assumed to be lower than the threshold that would cause single phase equipment to fail or malfunction. The three phase industrial loads attached to buses 137, 225 and 232 were assumed to trip if the level of unbalance at any of these busbars exceeded 2% for any 30 minute period.

9.5.3 Number of Trips Caused by Voltage Sags

The total number of trips caused by voltage sags was estimated using voltage sag impact estimation as described in Chapter 6. All residential and industrial loads were assumed to follow the ITIC immunity model outlined in Chapter 6. A trip was assumed if the voltage and duration violated the ITIC immunity curve in any of the three phases attached to a busbar.

9.6 Assessing the Weakest Areas of the Network

Weak area analyses focus on the source of the power quality issue: faults for voltage sags and unbalanced loads for voltage unbalance. Faults can be located using the techniques described in Chapter 5 and unbalanced loads can be found using the techniques described in Chapter 8.

Unbalance sources are not independent, and it is beyond the scope of this research to identify how the various sources of unbalance interact. The impacts of unbalance were assigned to unbalanced loads in proportion to the amount of negative sequence apparent power injected into the network and these results were visualised graphically. The unbalanced loads with the highest resultant impact on customers were defined as the *weakest areas for unbalance*.

Voltage sags are independent stochastic events, and thus each sag incident can be clearly associated with a resultant impact (in terms of, for example, CIs, CMLs or economic costs). By locating the source of a voltage sag, and assigning its impact to its source, a map of the network was obtained showing the areas where sources of voltage sags were that generate the worst effects on customers. These areas were defined as the *weakest areas for voltage sags*.

There were three metrics that were considered to identify weak areas of the network for voltage sags and unbalance:

- Negative Sequence Energy Injection (*Unbalance*)
- Sources of Customer Interruptions (CIs) Caused by *Unbalance*
- Sources of Customer Interruptions (CIs) Caused by *Voltage Sags*

Assigning an incident's impact to its source is much more informative than simply describing the areas with the incident rate. By assigning the incident's impact to its source, the weakest areas of the network are identified as regions where faults occur and subsequently affect the most customers. This idea can be extended to find the areas of the network with the faults which cause the highest economic impact on customers in the network.

9.6.1 Sources of Negative Sequence Energy Injection

The location of sources of negative sequence energy injection can be used to identify sources of unbalance in the network. The methodology from Chapter 8 and DSSE was used to estimate the median percentage ratio of negative to positive sequence apparent power ($|S^{(2)}|/|S^{(1)}|$) over the 24 hour period.

9.6.2 Sources of Customer Interruptions (CIs) Caused by *Unbalance*

CIs caused by unbalance were measured using the same method as defined for the worst served customer analysis. The CIs were then allocated to unbalanced loads in proportion to the estimated absolute level of negative sequence energy injection at each bus.

The analysis on the sources of negative sequence energy injection focussed on the ratio of $|S^{(2)}|/|S^{(1)}|$ to locate potential sources of unbalance. $|S^{(2)}|/|S^{(1)}|$ is a relative measure which is dependent on the size of the load. Therefore, it was decided that impacts (CIs) should be assigned in proportion to $|S^{(2)}|$, rather than $|S^{(2)}|/|S^{(1)}|$. The results of this style of analysis should provide a DNO with a prioritised list of the areas which generate the most significant unbalance impacts on their network.

9.6.3 Sources of Customer Interruptions (CIs) Caused by *Voltage Sags*

CIs caused by voltage sag were assessed using the same procedure as defined to identify the worst served customers. Probabilistic fault location was applied (Chapter 5) to estimate each voltage sag's location. The CIs for each voltage sag were then assigned to the estimated fault locations in proportion to the fault location's probability for each line in the network.

9.7 Results: Worst Served Customers

9.7.1 Percentage of 24 Hour Period >2% Unbalance

Fig. 9.2 and Fig. 9.3 show the percentage of the 24 hour period where the voltage unbalance factor (VUF) was estimated to be over 2%. Fig. 9.2 shows the results of the estimated time that unbalance was greater than 2%, whilst Fig. 9.3 shows the same metric estimated using limited monitoring. The areas with the most significant unbalance are located around the unbalanced loads. The unbalance located around bus 209 propagates over the largest area, whilst the unbalance at buses 130 and 132 affects fewer customers for smaller amounts of time.

The areas estimated using only four monitors (Fig. 9.3) are larger than those estimated using monitoring at every busbar (Fig. 9.2), highlighting only four monitors may not be an ideal solution. Nevertheless, the two maps of the network broadly match, in particular the critical areas with the highest unbalance, and both of the heat maps show that the level of unbalance is above 2% in a large area of the network.

Both Fig. 9.2 and Fig. 9.3 show that the level of unbalance at the busbars with sensitive three phase industrial loads is only above 2% at bus 137 (shown in the bottom left corner of the network). Because single phase customers are not affected by voltage unbalance, the only customers likely to be impacted by unbalance in this network will be limited to the industrial customers attached to bus 137.

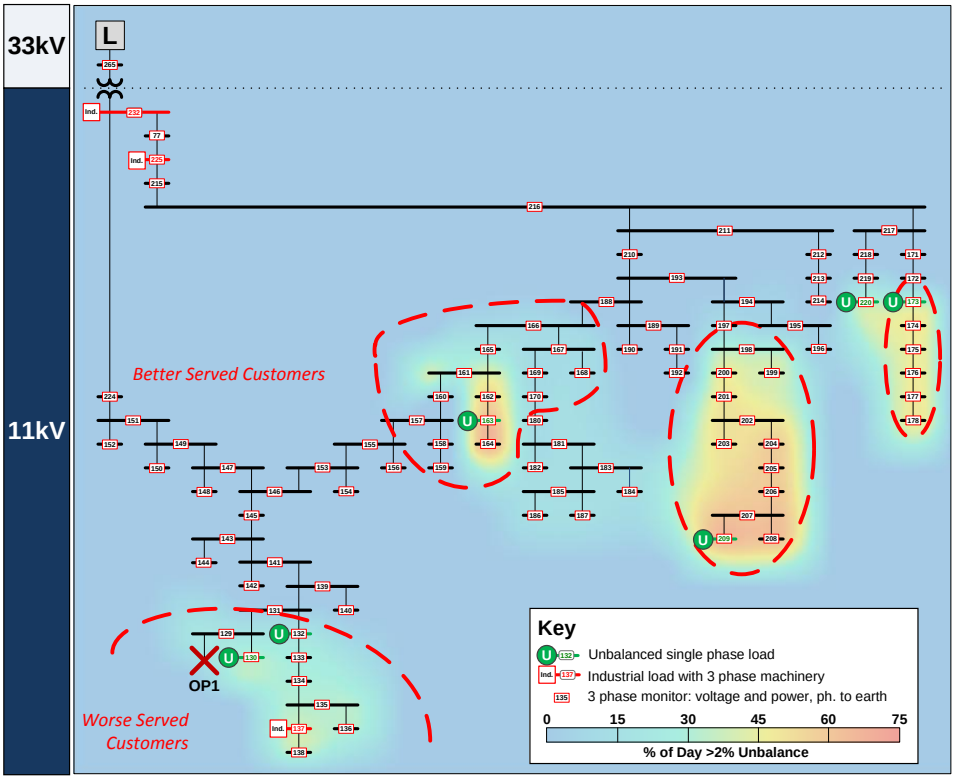


Fig. 9.2. A heat map of the 96 bus network showing the worst served customers in the network as described by the expected amount of time a bus is >2% unbalanced using monitors at all busbars in the network.

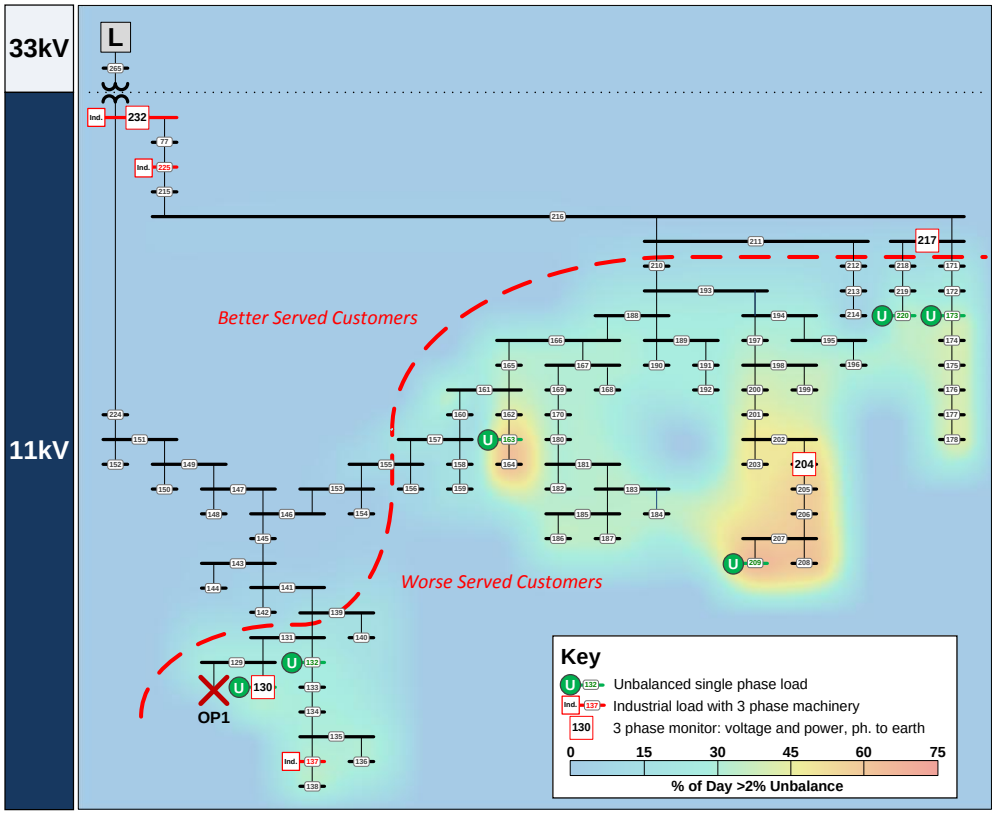


Fig. 9.3. A heat map of the 96 bus network showing the worst served customers in the network as described by the expected amount of time a bus is >2% unbalanced using monitors at 4 busbars in the network.

Insight into the uncertainty of the limited monitoring estimate can be obtained by reviewing the distribution of the voltage unbalance factor (VUF) as estimated at each of the busbars in the network at peak load. Fig. 9.4 shows the distribution of the voltage unbalance factor estimated using four monitors at peak load. The 90% interval for buses 144 to 154, 225 and 232 are all contained below 2%, thus it is unlikely that a trip will ever occur at any of these busbars based on a 2% immunity model. Fig. 9.4 shows that the inter-quartile range for the VUF estimated at peak load ranges from 0.05% to 0.86%. Fig. 9.4 confirms that the accuracy of the estimation results obtained using limited monitoring varies depending on which bus is being estimated.

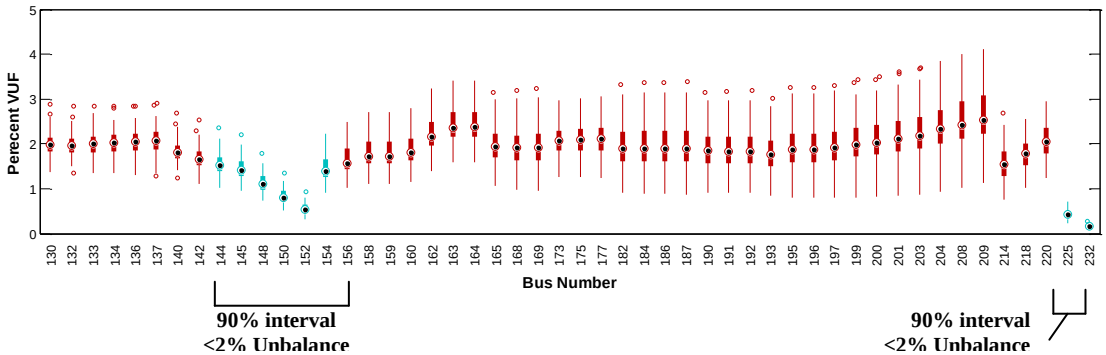


Fig. 9.4. The distribution for the voltage unbalance factor at all loaded busbars in the network estimated for a 30 minute interval at peak load. The boxes represent 25% to 75% percentiles, and the lines cover all data points that are not deemed to be outliers.

9.7.2 Number of Trips Caused by *Unbalance*

Unbalance trips are localised to sensitive three phase loads. The worst served customers for unbalance will always be three phase customers since single phase customers are generally not affected by unbalance (as discussed before).

Fig. 9.5 and Fig. 9.6 show the expected number of trips caused by unbalance over the 24 hour period. As expected, the trips are localised to bus 137 which is the only three phase load located within the region where the level of unbalance exceeds 2%.

The estimated number of trips shown in Fig. 9.6 (obtained using 4 monitors) is very similar to the true number of nominal trips estimated with monitoring at every busbar. With 4 monitors, the estimated number of trips at bus 137 is 14.6, whereas with all monitors, this number increases to 17 trips in 24 hours. This supports the conclusion that four monitors can sufficiently accurately monitor this network for voltage unbalance.

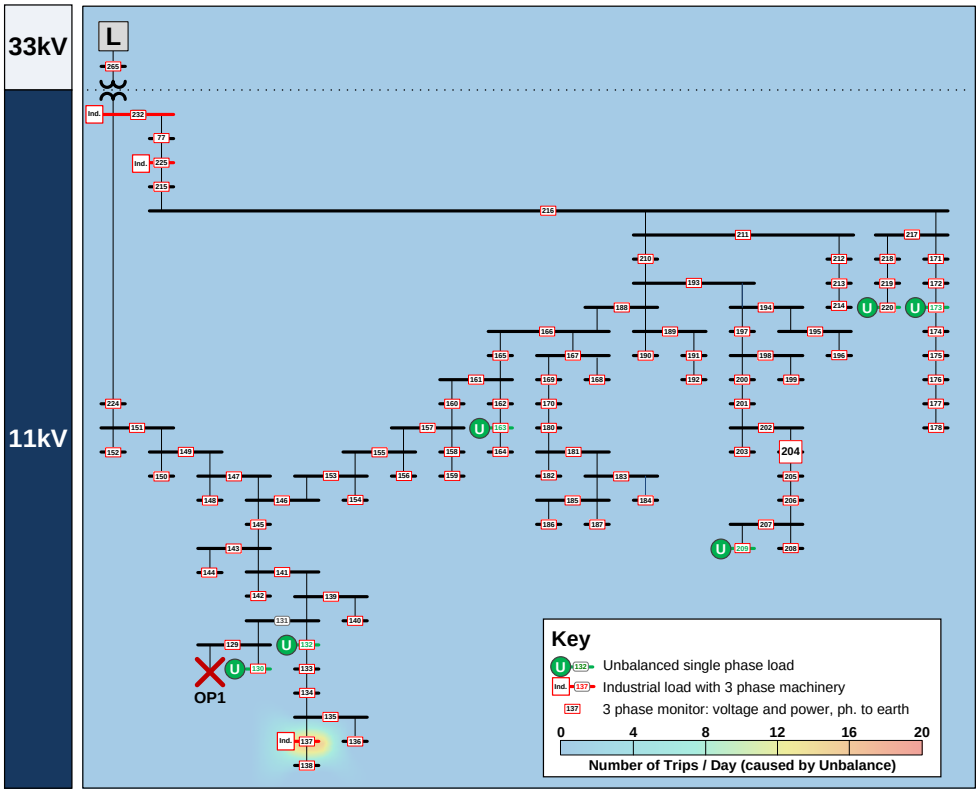


Fig. 9.5. A heat map of the 96 bus network showing the worst served customers described by the number of customer interruptions caused by voltage unbalance over a 24 hour period using monitors at all busbars.

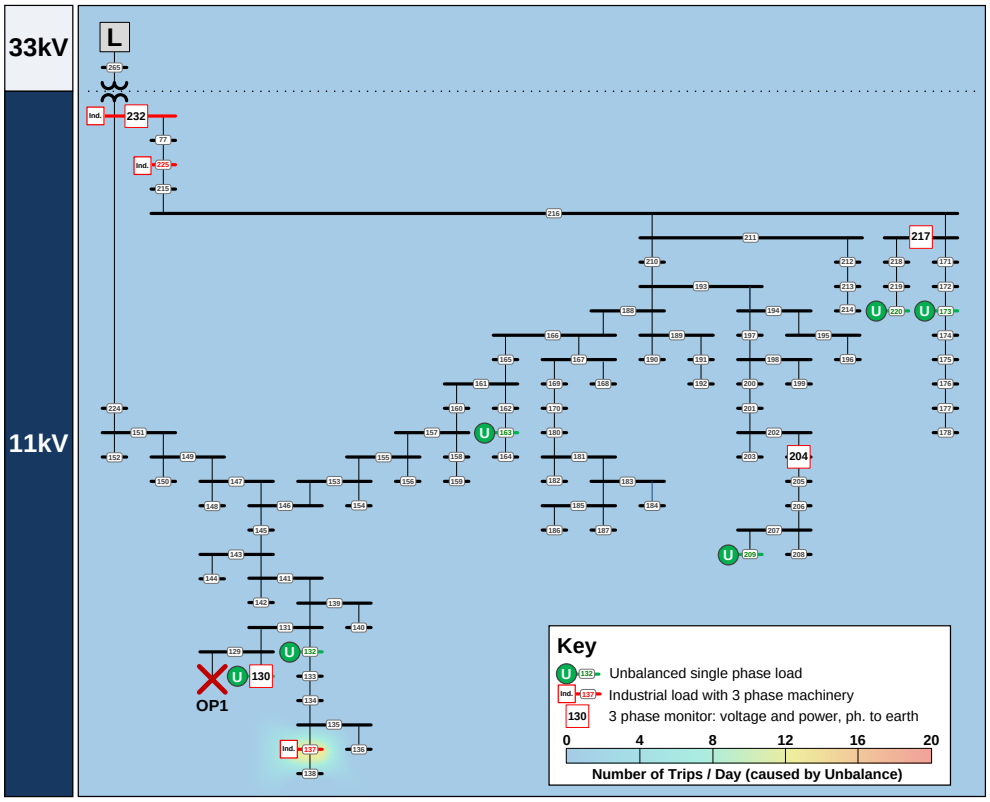


Fig. 9.6. A heat map of the 96 bus network showing the worst served customers described by the number of customer interruptions caused by voltage unbalance over a 24 hour period using monitors at 4 busbars.

Information on the accuracy of the trip estimates obtained at bus 137 using limited monitoring can be obtained by reviewing the distribution of the estimated VUF at bus 137 during the past 24 hours. This is the same technique which was applied in Chapter 8 to estimate the level of unbalance in a 24 bus network.

Fig. 9.7 shows the distribution of the voltage unbalanced factor estimated over the full 24 hour period using 4 monitors. The average inter-quartile range across the period is 0.35%. It can be concluded from Fig. 9.7 that any unbalance trips will be likely to occur between 6am and midnight, and unlikely to occur between midnight and 6am.

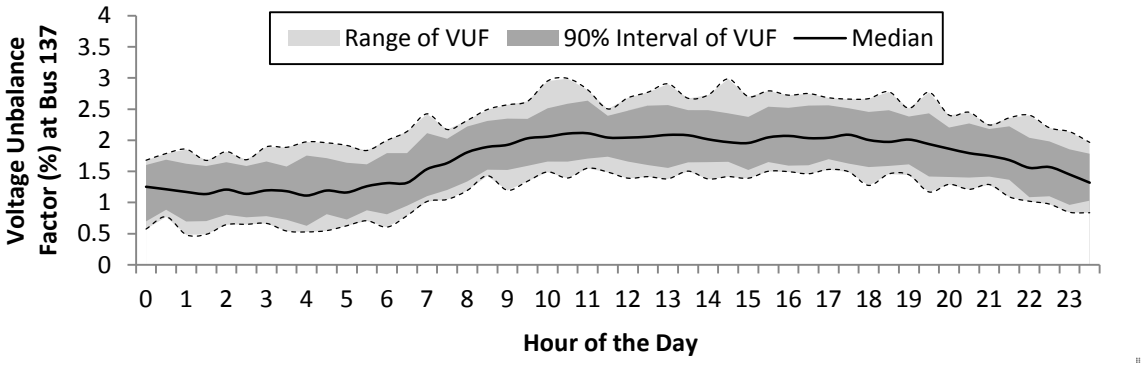


Fig. 9.7. The distribution of the VUF at bus 137 over the 24 hour study period estimated using 4 monitors.

9.7.3 Number of Trips Caused by *Voltage Sags*

Fig. 9.8 and Fig. 9.9 show the expected number of trips caused over a 24 hour period at each of the busbars in the network using monitors at all busbars and monitors only at a four busbars in the network. The number of trips can be broadly split into two sets of customers as shown by dotted red lines. These regions broadly separate customers into two groups: customers with more than 0.006 trips per day and customers with less than 0.006 trips per day. The size of the group estimated with limited monitoring (Fig. 9.9 where greater than 0.006 trips per day occur) is larger than the same group obtained using a complete set of monitors. The results in Fig. 9.9 highlight that limited monitoring can provide an accurate estimate for the voltage sag performance in a region of network where monitoring is incomplete.

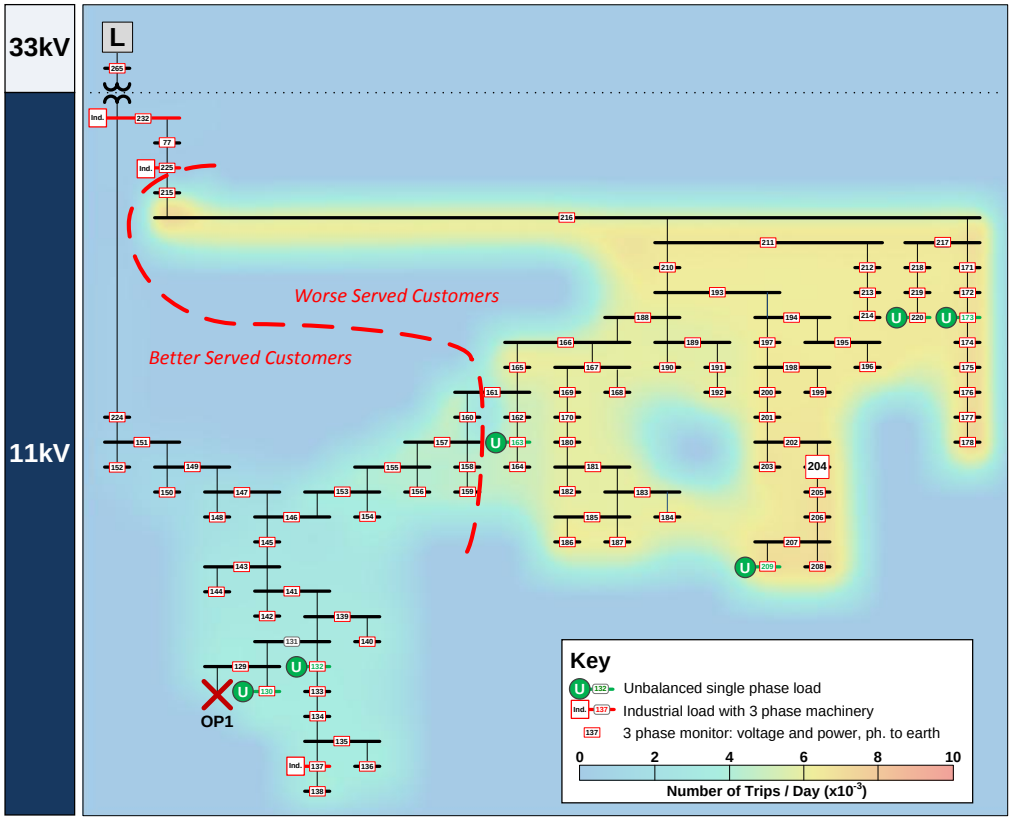


Fig. 9.8. A heat map of the 96 bus network showing the worst served customers in the network as described by the expected number of trips caused by single line to ground faults over a twenty four hour period using monitors at all busbars.

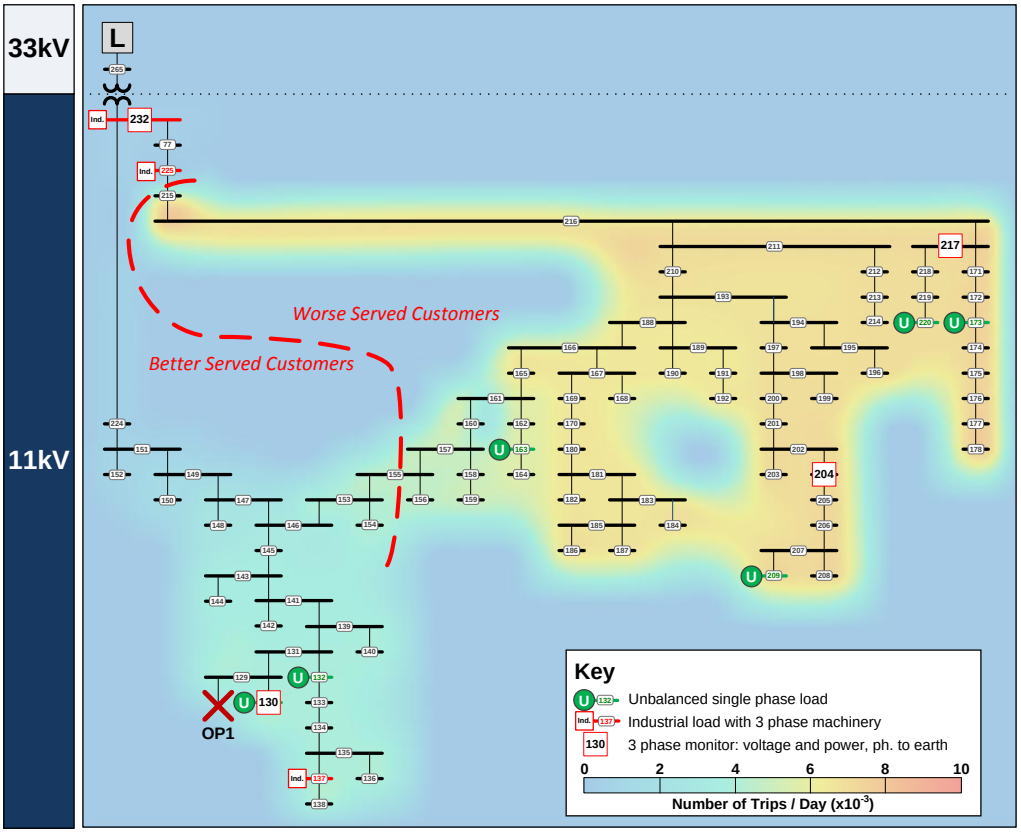


Fig. 9.9. A heat map of the 96 bus network showing the worst served customers in the network as described by the expected number of trips caused by single line to ground faults over a twenty four hour period using monitors at 4 busbars.

Comparing Fig. 9.8 and Fig. 9.9 highlights the uncertainty inherent in estimating the number of trips using a limited monitoring set. Fig. 9.9 shows that the approximated estimate for the number of trips in the network matches closely to the true number of trips (Fig. 9.8). The median error across all busbars in number of trips was 4%. The largest inaccuracy in sag trip estimation was 17% at bus 148, where the number of interruptions was estimated as 2.3×10^{-3} trips per year, and with four monitors it was estimated as 2.7×10^{-3} trips per year.

Further uncertainty analysis could be conducted by analysing the spread of the estimated during fault voltage profile using the techniques described in Chapter 6 and 7.

9.8 Results: Weakest Areas of the Network

9.8.1 Negative Sequence Energy Injection

The level of negative sequence energy injected into the network can be visualised by estimating the ratio of negative to positive sequence energy injected at each of the nodes in the network. Fig. 9.10 shows the areas of the network which are the most likely sources of unbalance using 4 monitors. The more overlap there is between regions where the green unbalanced loads and red hot spots are, the better the estimation results.

Fig. 9.10 correctly highlights the areas where there are likely sources of unbalance. All the methods developed in this research are statistical, so it is also important to assess the distribution of $|S^{(2)}|/|S^{(1)}|$ unbalance at each busbar to narrow down where operators should start diagnosing problems.

Fig. 9.11 shows the distribution of errors of the estimated $|S^{(2)}|/|S^{(1)}|$ injections across the network for a single half hour interval. A single operating point (represented by a half hour interval) was chosen to ensure that only the variation in error of $|S^{(2)}|/|S^{(1)}|$ was captured in the results. Selecting a longer period would introduce the natural variation of $|S^{(2)}|/|S^{(1)}|$ which would make it difficult to discriminate between variation in error and natural variation. The highlighted busbars are those with unbalance.

Upon reviewing Fig. 9.11 it is possible to discern that buses 163, 173, 220, 130 and 132 all have relatively low inter-quartile ranges compared with other busbars but they still inject a reasonable level of unbalance into the network. This means the state estimator is more certain that these buses are a definite source of unbalance than other locations. These busbars would therefore be a good candidate to begin an investigation to ultimately rebalance the network.

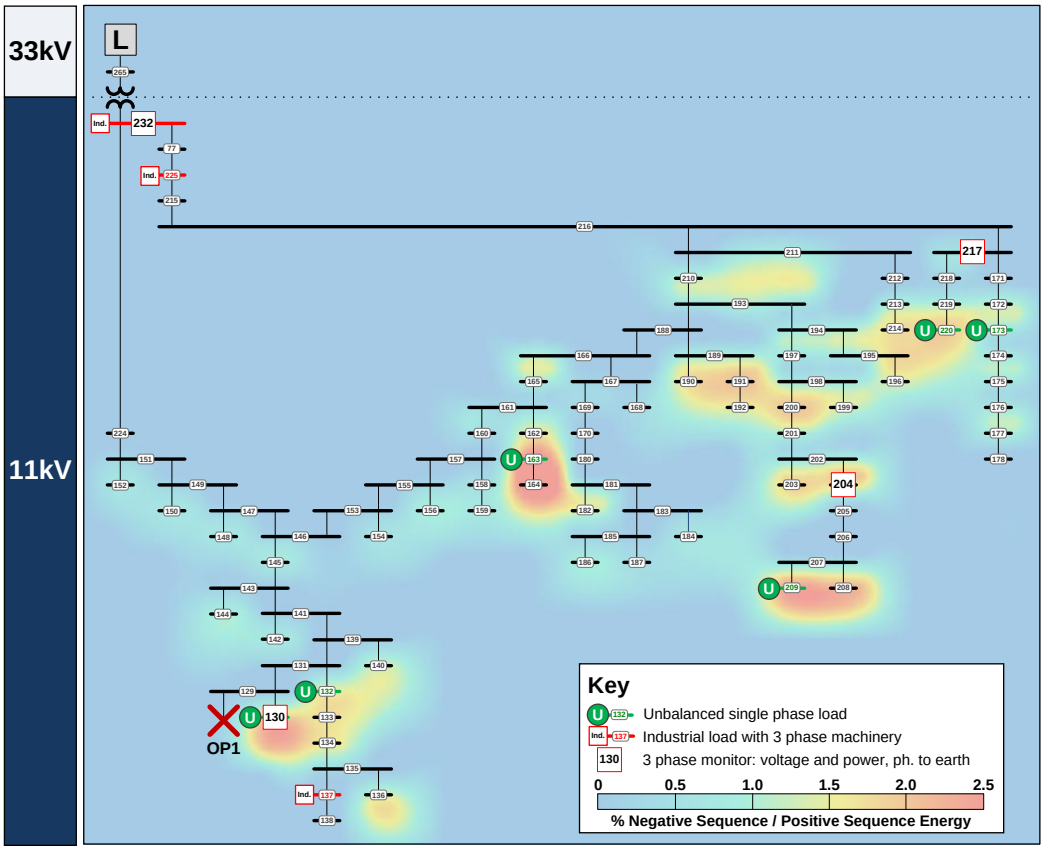


Fig. 9.10. The median percentage $|S^{(2)}|/|S^{(1)}|$ injected into the network as estimated by 4 monitors.

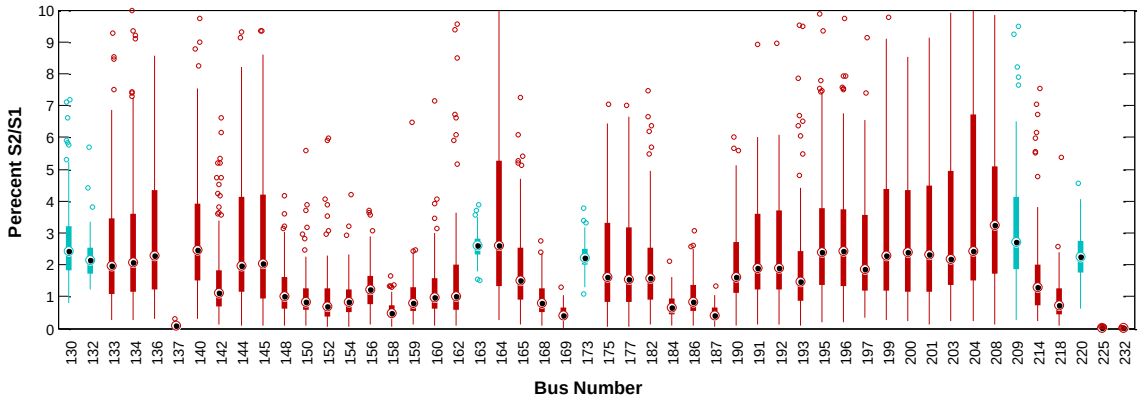


Fig. 9.11. The distribution of percentage $S^{(2)}/S^{(1)}$ across all unloaded busbars. The boxes represent 25% to 75% percentiles, and the lines cover all data points that are not deemed to be outliers. The unbalanced busbars are 130, 132, 163, 173, 209 & 220.

9.8.2 Sources of Customer Interruptions Caused by *Unbalance*

The sources of unbalance in the network interact together to produce trips at customer busbars. Fig. 9.12 and Fig. 9.13 show the areas in the network which are most likely to generate the most number of CIs. Fig. 9.12 and Fig. 9.13 both show that the load at bus

163 generates the most numbers of CIs. The weakest area of the network for unbalance can therefore be defined as bus 163, shortly followed by 173.

Unlike Fig. 9.10, the estimated regions shown in Fig. 9.13 are limited to unbalanced loads. Although this suggests the limited monitoring techniques are more accurate at estimating $|S^{(2)}|$ rather than $|S^{(2)}|/|S^{(1)}|$, this result is caused because the unbalanced loads at busbars 132, 163, 209, 130, 220 and 173 are up to 4 times as larger than any of the other loads shown in Fig. 9.10. Thus, the level of $|S^{(2)}|$ dominates at these busbars.

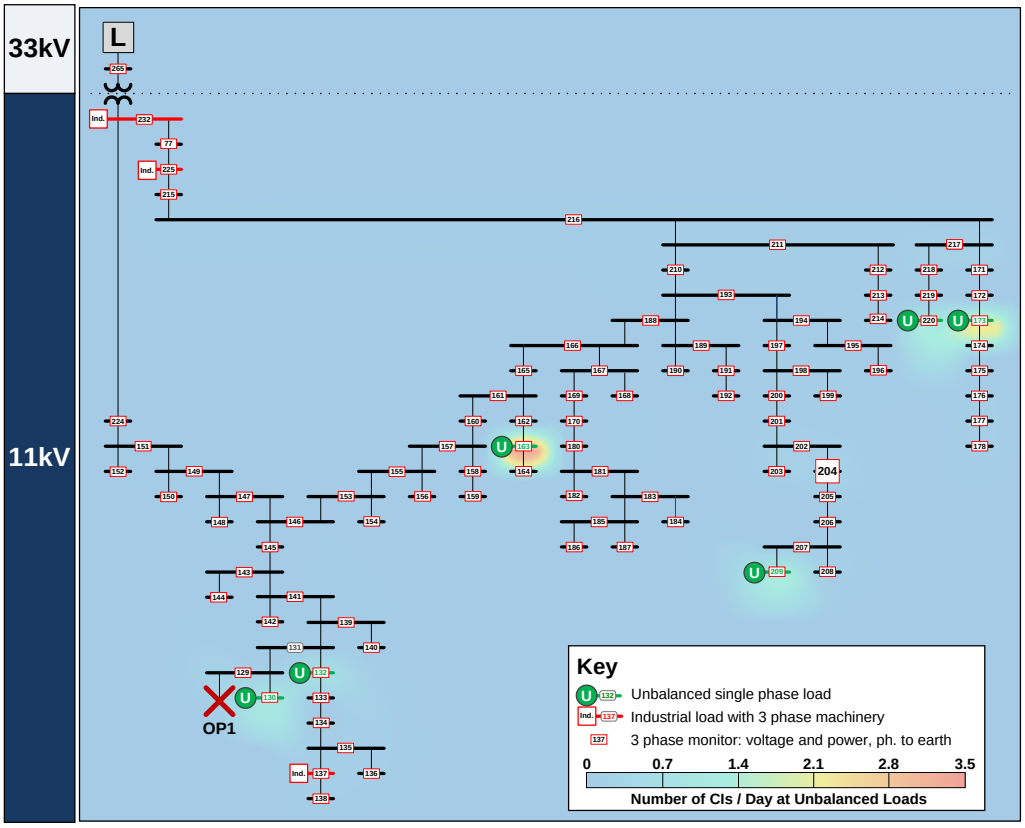


Fig. 9.12. The estimated number of customer interruptions caused by unbalanced loads allocated in proportion to the estimated amount of $|S^{(2)}|$ at each load estimated using monitors at all busbars.

9.8.3 Sources of Customer Interruptions (CIs) Caused by *Voltage Sags*

The number of customer interruptions caused by voltage sags can be visualised by assigning the interruptions to their source. Fig. 9.14 and Fig. 9.15 show the number of customer interruptions estimated over the 24 hour period using monitors at all busbars and monitors at only four busbars. The total number of CIs caused by faults in the weakest areas of the network reaches 0.01 customer interruptions (CIs) per day. In the strongest areas of the network, this figure is averages 0.002 CIs per day.

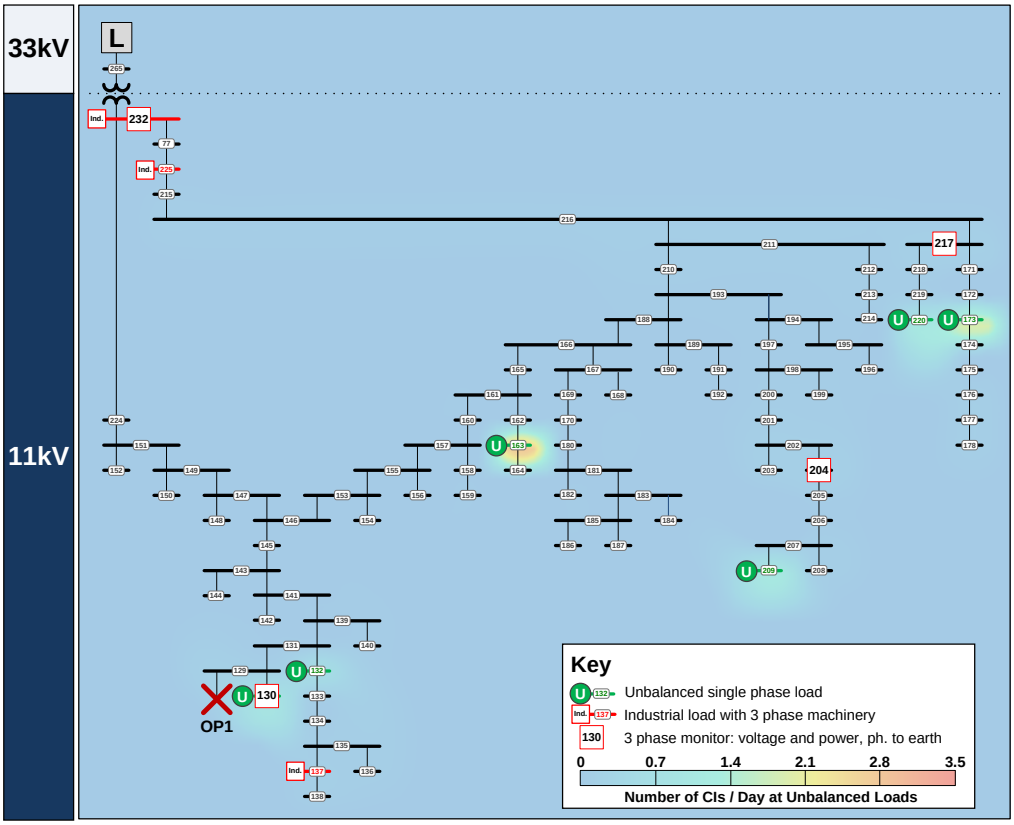


Fig. 9.13. The estimated number of customer interruptions caused by unbalanced loads allocated in proportion to the estimated amount of $|S^{(2)}|$ at each load estimated using four monitors.

In Fig. 9.14 and Fig. 9.15, the network has been divided into two sets of regions: strong areas and weak areas. Weak areas are broadly defined as regions where faults occur and subsequently cause more than 0.008 CIs. Strong areas are defined as regions where faults can occur and cause less than 0.002 CIs per day.

Through visually analysing Fig. 9.15 and Fig. 9.14 it can be seen that they correlate strongly, and the same areas of the network are identified as weak. In both diagrams, the weakest region of the network is found in the centre of the network. It can be inferred that the stronger areas of the network are typically located at the end of long feeders. This conclusion can be explained by considering that faults occurring at the ends of feeders are likely to have only a limited localised effect, and thus only affect customers close to the fault.

Uncertainty in the voltage sag weak area assessment will be caused by inaccuracies in trip estimation technique and inaccuracies fault location algorithm.

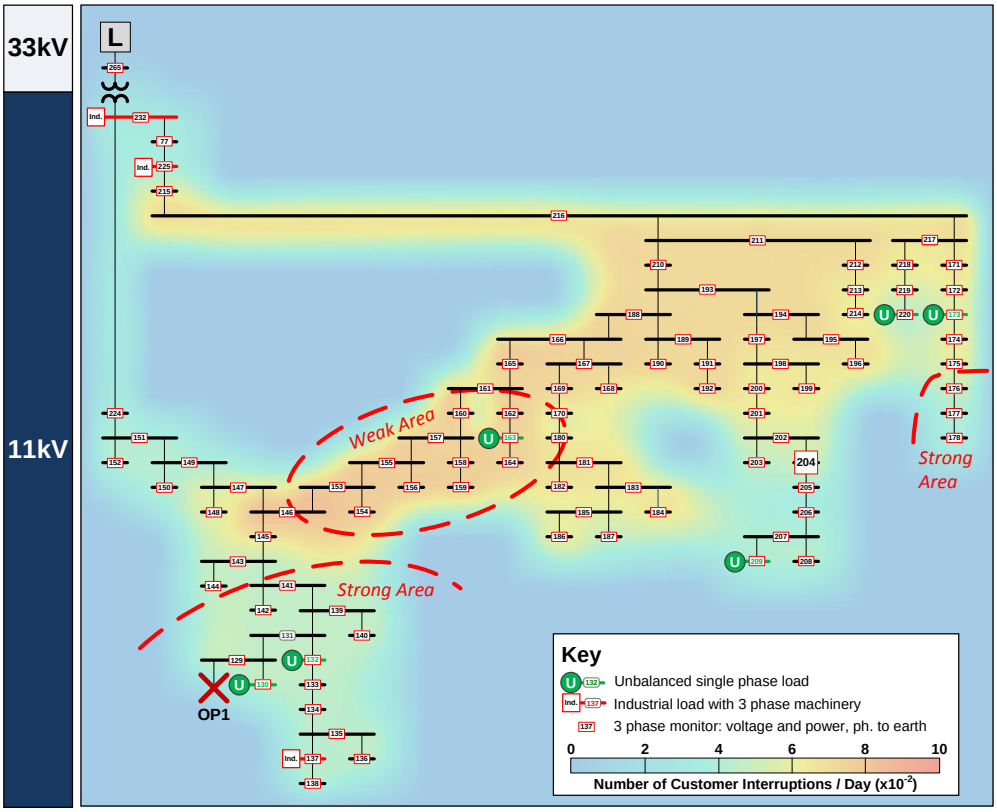


Fig. 9.14. A heat map of the 96 bus network showing the weakest areas of the network expected number of customer interruptions caused by single line to ground faults over a twenty four hour period using monitors at all busbars.

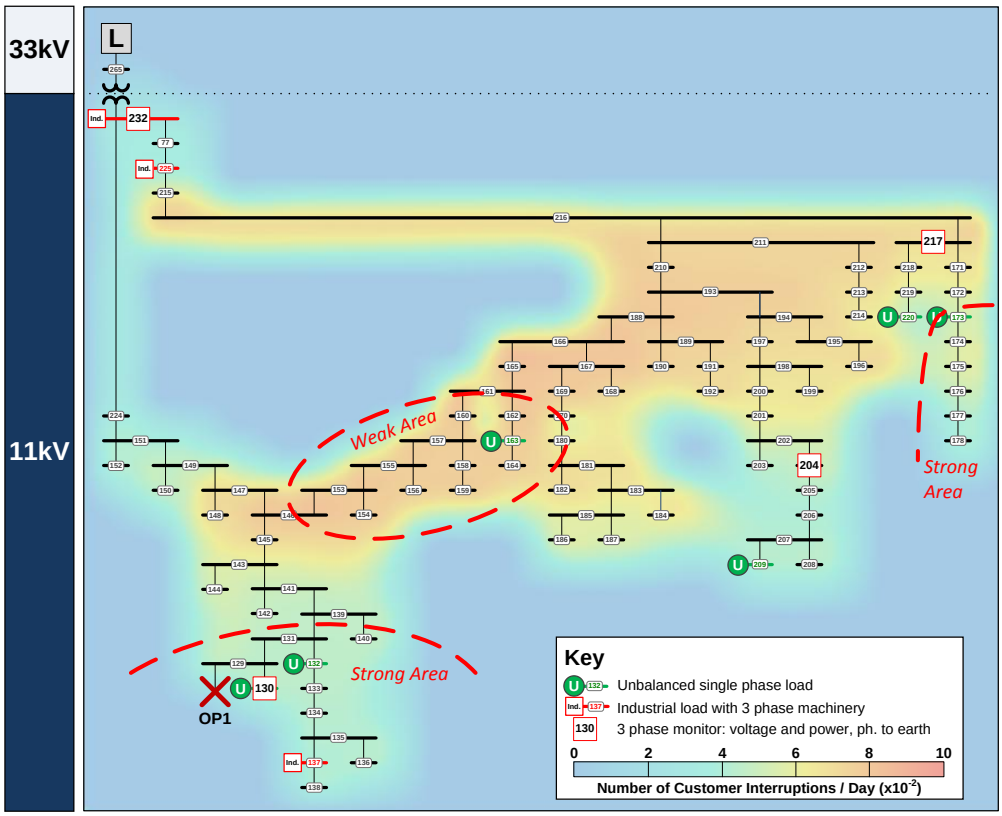


Fig. 9.15. A heat map of the 96 bus network showing the weakest areas of the network expected number of customer interruptions caused by single line to ground faults over a twenty four hour period using four monitors.

The accuracy of the fault location algorithm will vary depending on where the fault occurred in the network. For each fault, it is possible to determine the number of lines where the fault location estimated that the fault could have occurred with a probability of greater than 0.01. This is shown in the heat map of Fig. 9.16.

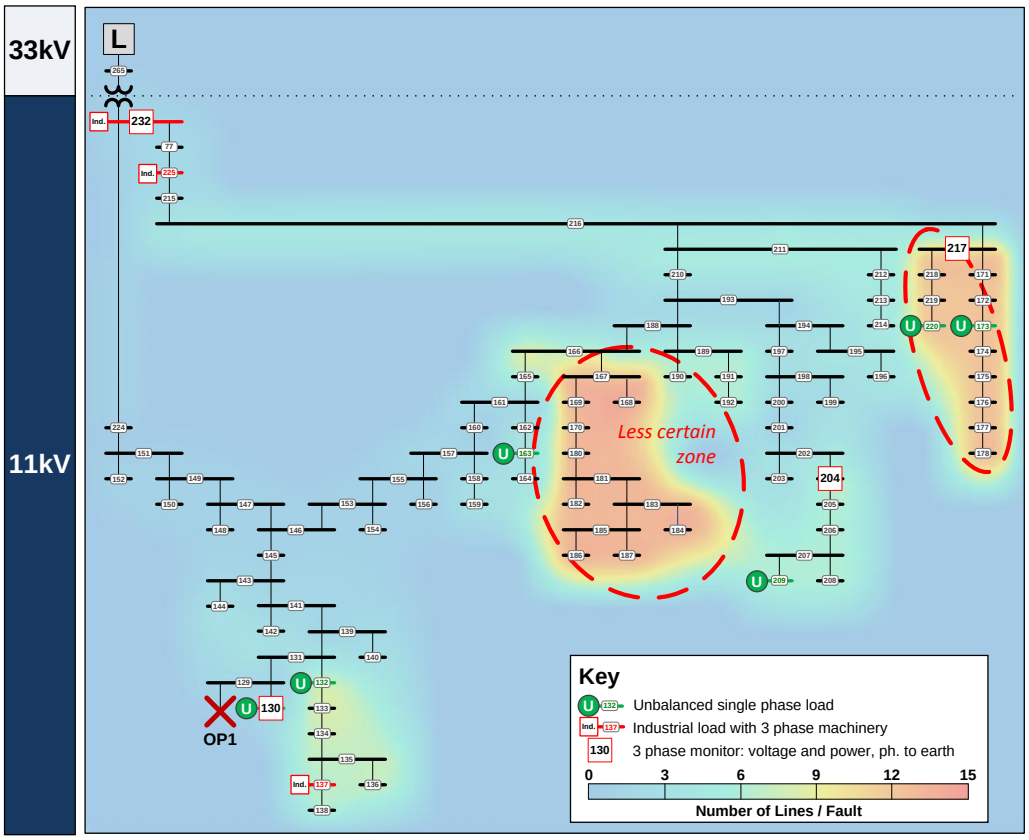


Fig. 9.16. The number of lines with a probability of > 0.01 where the fault location algorithm and 4 monitoring devices estimated that faults could occur.

Fig. 9.16 highlights two regions where if a fault occurs, they will be localised to around 12 lines in the network. Fig. 9.16 provides a guide for where the results of weak area analyses with four monitors may be most unreliable.

9.9 Globally Worst Served Customers & Weakest Areas

The results presented in this chapter provide sufficient evidence to be able to identify the worst served customers and the weakest areas of the network for both unbalance and voltage sags together.

When combining results, both measures must be on the same scale. One way of defining the globally worst served customers is to consider the busbars with the highest number of customer interruptions. Global weakest areas could be considered as source locations which generate the largest numbers of CIs.

9.9.1 Global Worst Served Customers

For this case study, the globally worst served customers will be heavily dominated by the trips caused by unbalance (Fig. 9.12 and Fig. 9.13). This is because the immunity model assumed for the three phase load at bus 130 assumes that a trip will occur if the VUF exceeds 2% for a 30 minute period and could therefore reach a maximum of 48 trips in a 24 hour period. Because unbalance trips were localised to bus 130, the number of trips caused in the rest of the network will follow the same pattern as observed for voltage sags (Fig. 9.8 and Fig. 9.9).

Table 9.2 shows the top ten globally ranked buses as defined in terms of the number of trips for voltage sags and unbalance over a 24 hour period.

TABLE 9.2. THE GLOBALLY RANKED TOP TEN WORST SERVED CUSTOMERS

	Unb. Trips	***Voltage Sag Trips***								
Bus Ranking (All Monitors)	137	205	206	207	208	209	204	175	176	177
Number of Trips / 24 Hours (All Monitors)	17*	0.0070	0.0070	0.0070	0.0070	0.0070	0.0070	0.0069	0.0069	0.0069
Bus Ranking (4 Monitors)	137	172	171	219	205	203	206	217	208	174
Number of Trips / 24 Hours (4 Monitors)	14.6*	0.0070	0.0070	0.0070	0.0070	0.0070	0.0070	0.0070	0.0070	0.0070

* Unbalance trips only affect three phase machinery.

The locations of the buses ranked using 4 monitors are in similar locations to the buses ranked using monitors at every busbar. For example, buses 173 and 174 are very close to buses 175, 176 and 177. This result corroborates with the heat map for voltage sag trips shown in Fig. 9.9 which shows that a network wide estimation of the number of trips using 4 monitors discriminates the network wide variation in trip rates which can be picked up with full monitoring (Fig. 9.8). Table 9.2 also shows that the trip rates for voltage sag trips are very similar, which implies that it will be very difficult to obtain a perfect match between the bus ranking estimated using 4 monitors (when incorporating measurement errors) and the bus ranking obtained using all monitors.

Using all monitors, the top 10 worst served customers are located in two distinct regions of the network at the end of two feeders: one section containing buses 205, 206, 207, 208, 209 and 204 and the other section contains buses 175, 176 and 177. Using four monitors, the 10 worst served customers are located in broadly similar regions; the 5^{th} to 7^{th} worst served customers in a region containing busbars 203 and 205, 206 and the 2^{nd}, 3^{rd}, 4^{th}, 7^{th} and 10^{th} buses are contained within a region containing buses 172, 171, 219,

217 and 174. Both of these two areas identified by the two monitor sets are towards the end of the two feeders shown on the far right of Fig. 9.8.

Care must be taken when using these results to guide the deployment of new loads in the network. Although bus 130 is the busbar where trips are most likely to occur, trips will only occur if three phase machinery is attached to the network at this busbar. To overcome this issue, the worst served customers could be segmented by load type.

9.9.2 Global Weakest Areas of the Network

The global weakest areas of the network can be considered as the areas of the network which cause the greatest number of customer interruptions (CIs). Unlike worst served customers, the weakest area analysis takes into account the number of customers attached to each busbar and focus on the source of the power quality incident.

In this case study, voltage sags were assumed to occur on lines whilst unbalance was simulated as originating from loads attached to busbars. To allow a comparison between these two sets of locations each fault location was moved from the centre of the line and evenly assigned to the bus at the start and end of the line. Thus, the locations for the weakest areas of the network were consistently defined in terms of areas defined by busbars.

Table 9.2 shows the top ten globally ranked buses as defined in terms of the areas which are most likely to cause the largest number of customer interruptions. The number of CIs caused by voltage sags is much larger than the number of trips in Table 9.2, since the number of CIs also incorporates the numbers of customers attached to each busbar.

TABLE 9.3. THE GLOBALLY RANKED TOP TEN WEAKEST AREAS IN THE NETWORK

	Dominated by Unbalance CIs						***Dominated by Voltage Sag CIs***			
Rank	1	2	3	4	5	6	7	8	9	10
Bus (All Monitors)	209	163	173	220	130	132	146	153	155	156
Number of CIs Caused by this Area / 24 Hours (All Monitors)	3.02	2.85	2.44	2.39	2.36	2.33	0.081	0.077	0.076	0.076
	Dominated by Unbalance CIs									
Bus (4 Monitors)	163	173	220	209	130	132	129	204	217	219
Number of CIs Caused by this Area / 24 Hours (4 Monitors)	3.52	2.15	1.33	1.25	0.90	0.74	0.24	0.22	0.20	0.18

The top ten weakest areas were heavily dominated by unbalance, in the same way as the worst served customers were located at busbars sensitive to unbalance. This is because

the resultant number of trips caused by the negative sequence energy injections was much higher than any trips caused by voltage sags.

The top four weakest areas for voltage sags are located in the central section of the network. These busbars can all be found in the ellipsoid area shown in Fig. 9.14 and Fig. 9.15. Although the busbars ranked using 4 monitors are different to those obtained using monitoring at all busbars, the buses are located in the same area , so both monitoring approaches can successfully identify the same important areas of the network whilst prioritising within that areas may be more difficult.

9.10 Summary

The aim of this chapter was to show how the approaches proposed in this thesis could be applied to help DNOs plan infrastructure investment and optimised maintenance decisions using only a limited set of monitors. It was hoped that a realistic case study would go some way to validating that the proposed approaches could be used in a practical manner.

The worst served and weakest areas of the network were obtained by considering a number of power quality metrics. The metrics selected for this research were focussed on trips, and customer interruptions. The study focussed on these trips in order to limit the scope of the demonstration and also maintain its relevance to the previous discussion in this thesis. It is straightforward to see how a study focussing on trips could be extended to economic appraisal using economic interruption costs available in [2].

The results presented in this chapter show that even with a limited number of monitors, the worst served customers and weakest areas of the network for voltage sags and unbalance can be successfully identified. The results using a limited monitoring are not, nor can they be, as accurate as those that can be obtained with full monitoring. There is always more uncertainty in limited monitoring estimates and the estimates therefore include additional errors.

The inaccuracy in limited monitoring metric estimates were found to either mask subtle parameter variations, or broaden the estimated regions where problems were likely to be found. When estimating the number of trips at each busbar caused by voltage sags, the limited monitoring results failed to pick out the variation across the network where trips were likely to occur. Applied to estimate the source of unbalanced loads, the limited monitoring estimates yielded results that spilled beyond the problem busbars.

The techniques proposed in this thesis are all probabilistic, and therefore a level of certainty can be assigned to each estimate of a parameter. This was used in the chapter to show how the worst served customers for unbalance could be identified. The existing body of power systems research mainly focuses on estimating metrics in a deterministic manner, thus, the ability to attach uncertainty to calculated metrics is an important advancement. By defining the level of uncertainty associated with a monitoring result, it should also be possible for DNOs to more accurately quantify the risk that intervention or investment is required.

The study scenario considered in this chapter was fabricated based on the likely loading of a typical UK distribution network. It is therefore difficult to derive any reliable conclusions between the relative importance of unbalance and voltage sag power quality issues, other than noting that unbalance only typically affects three phase machinery and voltage sags affect all types of electrical load. Applying the techniques developed in this chapter (and thesis) to a real section UK distribution would provide further guidance on the extent to which these two power quality issues affect customers.

10 Conclusions & Future Work

10.1 Conclusions

The core objective of the research presented in this thesis was to formulate a series of new and novel techniques which are capable of identifying the weakest areas and the worst served customers for voltage sags and unbalance using a limited set of power quality monitors. Monitoring a power network with a limited set of monitors is a pertinent problem for modern distribution network operators, because of the current international trends towards tighter regulation, greater use of power quality and reliability contracts and the drive towards increased distribution network operational efficiency as part of the smart grid agenda. Monitoring for power quality issues such as voltage sags and unbalance is important because these issues cause significant economic losses to customers as well affecting a distribution network operator's (DNO's) operational and regulatory performance. By effectively managing these issues, a DNO has the potential to benefit from regulatory incentives, decreased losses, increased operational performance and greater return on investment. The key challenge from a research perspective is how to interpret the vast quantities of data emanating from distribution networks to yield knowledge about network performance that can guide network operation and investment.

Though it is never possible to fully explore every aspect and detail of a research area within a limited time and scope project, the research presented in this thesis develops several new and novel techniques which advance the topic of power quality monitoring. The research goes some way to satisfying all of the aims and hypotheses which were identified at the start of the discussion.

The research presented in Chapters 4, 5 and 6 described a new way of processing monitoring information to interpret voltage sag performance in terms of customer trips. The aim of Chapter 4 was to demonstrate how voltage sags can be first detected and classified using artificial intelligence classification algorithms and a limited set of potentially erroneous monitors. The method was shown to work with very few monitors (less than 5 in a 295 bus network), achieve classification accuracies in excess of 99%, robustly deal with measurement error as well as working with varying fault impedances and pre-fault loading conditions. Chapter 4 also compared a variety of classification techniques and provided some evidence that the Artificial Immune Recognition System (AIRS) classifier is a poor classification algorithm when compared with other more conventional techniques. Understanding where new classification algorithms (like artificial immune systems) could be advantageously applied to power systems is an important current research topic, and the results shown in this thesis show areas where their application is less advantageous.

The voltage sag research continues in Chapter 5 where a new technique was presented which can robustly localise and estimate the voltage sag profile using a limited set of monitors whilst dealing with measurement error. The method was shown to be able to formulate a statistical estimate for a voltage sag's location and the voltage sag's magnitude profile at non-monitored busbars. This is in contrast to existing voltage sag localisation and profile estimation techniques which tend to focus on producing a single deterministic estimate for both of these properties. The method further differentiates itself from existing techniques by being able to synthesize information from an arbitrary set of monitoring devices such as relays, power quality meters, disturbance recorders and phasor measurement units, all of which have potentially varying levels of accuracy. The developed technique also helps to eliminate the problem of multiple impedance path estimates by selecting the statistically most probable fault locations. A further useful aspect of the developed technique is its ability to identify potentially faulty monitors by using statistical tests.

The final aspect of the voltage sag performance monitoring research was a method capable of estimating the impact of voltage sags on customers in the network (which was presented in Chapter 6). The method builds upon the voltage sag profile estimation algorithm (developed in Chapter 5) to yield a probabilistic assessment of the likelihood of a trip. A key contribution of this aspect of the research is a generalised measure known as the sag trip probability (*STP*) which defines the probability of an equipment

trip. The *STP* can be additionally used to derive the number of trips caused by voltage sags over a series of events or a defined time period. A key difference between the *STP* and other existing sag performance monitoring indices is that the measure takes into account both measurement and equipment uncertainties. Most existing measures for voltage sag performance deterministically measure the technical performance of a busbar without determining the impact of a sag on end users.

One of the aims of this thesis was also to present a new method which would allow power system operators to be able to add new monitors to existing monitor deployments whilst taking into account future uncertain load and topology scenarios. Chapter 7 presented a practical method which is capable of optimally locating monitors to observe the effects of voltage sags using the *STP* (developed in Chapter 6). The methodology used an immune inspired optimisation process (known as the B-Cell algorithm) which was shown to yield promising results. It was thought that the immune inspired technique was particularly suited to the task of optimal monitor placement because it is a population based algorithm which is capable of maintaining a range of near optimal solutions. This adds robustness into the optimisation solutions when they are exposed to topology and loading uncertainties and also gives network operators the flexibility to choose from a suite of locations for future monitor placements. Another novel aspect of the developed monitor placement technique was its ability to focus on accurately estimating impacts of voltage sags rather than simply estimating voltages. This is important, since it allows operators to place monitors in locations best able to estimate the economic impacts of voltage sags on end users. This topic is of particular relevance in a regulatory environment which is increasingly looking to holistically monitor voltage sag performance from the customer's perspective.

The voltage unbalance monitoring technique developed in this thesis used a newly developed distribution system state estimation (DSSE) method to provide a statistical estimate for the level, location and effects of unbalance in a network. This method was developed and tested on a generic distribution network and the results were presented in Chapter 8. In a similar way to the methods developed for voltage sags (in Chapters 4 to 6), the framework was shown to work effectively even when measurement information was incomplete. Measurement information was estimated by creating pseudo-measurements which took into account the correlated nature of three phase unbalanced loads and incorporated other measurement information such as single phase data. These

types of technique should play an important role in monitoring unbalanced distribution networks where monitoring information is limited.

The last major component of the research presented in this thesis was a series of case studies demonstrating how all of the developed techniques could be applied to a generic section of a UK distribution network. The aim was to identify the globally worst served customers and the weakest areas of the network for both voltage sags and unbalance using a limited set of power quality monitors and then subsequently visualise the results using a set of heat maps. The results highlighted that weak areas and worst served customers could be identified effectively in a 100 bus network using only 4 monitors. Comparing the results obtained using limited monitoring with those obtained using full monitoring highlighted that limited monitoring can sometimes mask subtle variations in the results of unbalance and voltage sag analysis. The results demonstrated that although there is always a loss of accuracy with a limited number of monitors, the developed techniques can always identify the broad area where problems are found or customers are being affected. The case studies also demonstrate the power of the statistical nature of the developed techniques by showing that a level of certainty can be applied to each of the estimation results. The new way of visualising results using a heat map is a powerful vindication that the developed methods could be implemented to observe the power quality performance of modern distribution networks.

All of the results presented in this thesis have validated using a 295 bus generic distribution network or a real 24 bus model of the UK's distribution. The methods presented in this thesis are general in nature, and therefore can be applied to any type of network. However, the results of monitor placement, voltage sag profile estimation or unbalance estimation may vary depending on network topology, varying X/R ratios, line lengths or conductor types.

Perhaps the best way to appraise the success of the research presented in this thesis would be to return to the original hypotheses, stated in the introduction. The following hypotheses were asserted:

1. Simple measurements from a limited number of variable accuracy metering devices measuring properties such as voltage magnitude, reactive power and real power can be synthesised to yield an intelligent customer centric picture of the power network in terms of its performance for voltage sags, and unbalance.

2. Performance estimation for voltage sags and unbalance can be quantified to optimise the monitoring locations of any arbitrary number of monitors, perhaps much less than the number of nodes in a network.
3. Artificial immune system heuristic based techniques are required to monitor the state of the power network for voltage sags and unbalance and can help with monitor placement in the networks.

It is clear from the preceding discussion that the results for both voltage sags and unbalance strongly support the first hypothesis. Simple measurements such as voltage, current and power were used for all of the developed techniques and the methods were shown to be capable of producing a intelligent customer centric picture of the network.

The second hypothesis is supported by the research on optimal monitor placement for voltage sags. This technique showed how the monitoring performance of a network could be enhanced and optimised through additional monitor placement. However, it should be noted that there is no comparable method presented for voltage unbalance monitor placement. This will be highlighted in the discussion on future work that follows.

The last hypothesis asserted that artificial immune system heuristic based techniques are required for monitoring of power network for voltage sags and unbalance. Perhaps the most challenging aspect of this hypothesis is proving that a technique is definitely required to perform a certain task; artificial intelligence techniques are readily deployed to tackle all kinds of problems but it is often difficult to prove why they are required over more established conventional techniques. The methodology developed for classifying and detecting voltage sags highlighted a requirement for a classification algorithm, and therefore perhaps an artificial intelligence based classifier. However, the results showed that the selected immune inspired technique performed poorly when attempting to classify voltage sag events. Promising results were presented in Chapter 8 where voltage sag monitor placement was optimised using an immune inspired algorithm. However, for the specific task of monitoring a power network, the results do not conclusively show that immune inspired techniques perform any better than more established algorithms.

10.2 Future Work

The path for future work which naturally follows on from this thesis can be split into two areas. The first of these is future academic work which adds to the ideas and

methods developed in this thesis. The second area is more practical, and concerns future research work which could be conducted in cooperation with a distribution network operator.

There are plenty of academic research extensions which can be added to the ideas presented in this thesis. A valuable addition to the unbalance research would be the construction of an immunity curve which incorporated overheating, protection (for trips), losses, de-rating and the reduction in an asset's life time. The unbalance state estimator could also be enhanced to cover the location of unbalanced lines and cables. An investigation into the time-scales associated with unbalance would also be useful. The development of algorithms which can place a limited number of monitors to optimally monitor a distribution network using three phase distribution system state estimation would also be of significant further value; such a technique would add to the current body of unbalance and state estimation research.

The research on voltage sags could be extended by incorporating other causes of voltage sags into the methodology, such as transformer energizing and starting of induction motors. Dynamic load modelling would also provide additional insight into the effect of generators and motors on voltage sag profile estimation. Further research could also be conducted into voltage sag immunity curves to build a standardised probabilistic family of curves for each type of load and / or customer type. Probabilistic immunity curves could then be readily fed into the sag trip probability concept developed in this thesis.

All of the algorithms developed in this thesis assume that the network's impedance can be derived accurately. A significant further addition would be extensions to the developed techniques to take into account network topology and impedance uncertainties.

It would be interesting to investigate the economic consequences of both unbalance and voltage sags on both customers and DNOs. Power quality (PQ) economic frameworks such as [1] and [2] allow the impacts of PQ events to be quantified in monetary terms, and it is a logical progression to combine these with the probabilistic approach proposed in this thesis.

AIS provide a powerful paradigm, but it is unclear whether they have any specific advantages over simpler and more comprehensible techniques. AIS optimization algorithms proved that they may be of some merit for optimal monitor placement, although other population based techniques exist which exploit AIS' advantage of

maintaining a range of sub-optimal solutions. Further investigation in AIS techniques is required to establish their viability for power systems research.

Lastly, the project could be extended by broadening the statistical methodologies developed for sags and unbalance to other power quality phenomena, thus creating a universal *power quality state estimator*. Harmonics, flicker, transients, under-voltages and over-voltages are all serious power quality issues. It would be advantageous if these issues could be estimated in networks where measurement information is incomplete to understand the impacts of these problems on both customers and the DNO.

Future research conducted in cooperation with a distribution network operator could consist of a pilot project to implement the algorithms developed in this project on a section of a distribution network. The pilot would probably need to cover a reasonable area of network, and be located in a network where a significant amount of monitoring was being planned. A pilot could have several objectives, including:

- Estimating the maintenance efficiency gains (in £) which could be made from PQ monitoring for sags and unbalance.
- Providing DNOs with visibility of regulatory breaches from unbalance, trips and sags.
- Providing DNOs with insight into the impacts of sags and unbalance on their customers.
- Creating a user friendly software application (with visualization software) to help identify PQ problems.
- Validating the results of probabilistic estimation methods with the actual number of customer trips and customer complaints.
- Building a UK power quality monitoring database for future research and development studies.
- Practically validating the modelling assumptions used throughout this project.
- Establishing the usefulness of multi-purpose methodologies (e.g. distribution system state estimation) for alternative purposes such as voltage / VAR control, load estimation, offline studies (e.g. power flow) and network planning.
- Assessing the feasibility of integrating the developed methodologies within a DNO's third party DMS.

11 References

[1] "JWG CIGRE-CIRED C4.107," Economic Framework for Power Quality, 2010.

[2] J. Y. Chan, "Framework for Assessment of Economic Feasibility of Voltage Sag Mitigation Solutions," *Faculty of Engineering and Physical Sciences*, 2010.

[3] R. Targosz and J. Manson, "Pan European LPQI Power Quality Survey," in *19th International Conference on Electricity Distribution (CIRED)* Vienna, 2007.

[4] J. Y. Chan and J. V. Milanovic, "Risk Based Assessment of Financial Losses Due to Voltage Sag," *IEEE Transactions on Power Delivery*, vol. 26, No. 2, April 2011 2011.

[5] K. K. Kariuki and R. N. Allan, "Evaluation of Reliability Worth and Value of Lost Load," *IEE Proceedings Generation Transmission Distribution*, vol. 143, No. 2, 1996.

[6] "Electricity Distribution System Losses - Non-Technical Overview," Sohn Associates, 2009, http://bit.ly/w2N82d.

[7] "Electricity Distribution Loss Percentages by Distribution Network Operator (DNO) Area," 2010.

[8] Y. Zhang and J. V. Milanovic, "Voltage Sag Cost Reduction with Optimally Placed FACTS Devices," in *9th International Conference on Electrical Power Quality and Utilisation*, 2007, pp. 1-6.

[9] "Capacity to Customers," Manchester: Electricity North West, 2011, http://www.enwl.co.uk.

[10] "EN50160," Voltage Characteristics of Electricity Supplied by Public Distribution Systems, July 2010.

[11] M. H. J. Bollen, *Understanding Power Quality Problems Voltage Sags and Interruptions*: Piscataway, 1999.

[12] M. F. McGranaghan, D. R. Mueller, and M. J. Samotyj, "Voltage Sags in Industrial Systems," *IEEE Trans. Ind. Appl.*, vol. 29, No. 2, pp. 397-403, 1993.

[13] Juan A. Martínez-Velasco, Nick Abi-Samra, Myo T. Aung, Math Bollen, Sasa Z. Djokiç, Nikos Hatziargyriou, Roberto C. Leborgne, J. V. Milanović, Gabriel Olguin, and Melanie Schilde, "Voltage Dip Evaluation and Prediction Tools," in *CIGRE TF C4.1.02 Working Group*, 2009.

[14] D. Chapman, *The Cost of Poor Power Quality*: Copper Development Association, November 2001.

[15] J. Arrillaga, M. H. J. Bollen, and N. R. Watson, "Power Quality Following Deregulation," *Proceedings of the IEEE*, vol. 88, No. 2, pp. 246-261, 2000.

[16] M. H. J. Bollen, M. R. Qader, and R. N. Allan, "Stochastical and Statistical Assessment of Voltage Dips," in *IEE Colloquium on Tools and Techniques for Dealing with Uncertainty (Digest No. 1998/2000)*.

[17] G. Olguin, F. Vuinovich, and M. H. J. Bollen, "An Optimal Monitoring Program for Obtaining Voltage Sag System Indexes," *IEEE Transactions on Power Systems*, vol. 21, No. 1, p. 378 to 384, 2006.

[18] J. V. Milanović and C. P. Gupta, "Probabilistic Assessment of Financial Losses due to Interruptions and Voltage Sags - Part I: The Methodology," *IEEE Transactions on Power Delivery*, vol. 21, No. 2, pp. 918-924, 2006.

[19] K. Keller and B. F. C. Franken, "Quality of Supply and Market Regulation; Survey within Europe," KEMA Consulting, 2006, 30630505-TDC 06-59095A.

[20] "JWG CIGRE-CIRED C4.110," Voltage Dip Immunity of Equipment and Installations, April 2010.

[21] A. Jouanne and B. Banerjee, "Assessment of Voltage Unbalance," *IEEE Transactions on Power Delivery*, vol. 16, No. 4, pp. 782-790, October 2001.

[22] "IEC / TR 61000-3-13: Electromagnetic Compatibility (EMC) - limits - assessment of emission limits for the connection of unbalanced installations to MV, HV and EHV power systems," 2008.

[23] Z. Emin and D. S. Crisford, "Negative Phase-Sequence Voltages on E&W Transmission System," *IEEE Transactions on Power Delivery*, vol. 21, No. 3, 2006.

[24] "SEMI F47-0606," Specification for Semiconductor Processing Equipment Voltage Sag Immunity.

[25] G. Olguin, "Voltage Dip (Sag) Estimation in Power Systems based on Stochastic Assessment and Optimal Monitoring," *Department of Energy and Environment*, Doctor of Philosophy, 2005.

[26] P. Pohjanheimo and M. Legtonen, "Equipment Sensitivity to Voltage Sags - Test Results for Contactors, PCs and Gas Discharge Lamps," in *10th International Conference on Harmonics and Quality of Power* Finland, 2002.

[27] "NEMA MG 1," Motors and Generators, NEMA, 2009.

[28] Council of European Energy Regulators (CEER), "4th Benchmarking Report on Quality of Electricity Supply," 2008.

[29] "Electromagnetic compatibility (EMC) - Part 2-2: Environment - Compatibility levels for low-frequency conducted disturbances and signalling in public low-voltage power supply systems," IEC 61000-2-2 ed2.0, 2002.

[30] "Electromagnetic compatibility (EMC) - Part 2-12: Environment - Compatibility levels for low-frequency conducted disturbances and signalling in public medium-voltage power supply systems," IEC 61000-2-12 ed1.0 B:2003, 2003.

[31] P. Paranavithana, "Contributions towards the development of the Technical Report IEC/TR 61000-3-13 on voltage unbalance emission allocation," *University of Wollongong*, PhD Thesis, 2009.

[32] G. V. Moodley, D. Dama, and R. Vajeth, "Consideration of Electromagnetic Induction During Transposition Studies," in *7th AFRICON Conference*, 2004.

[33] J. Hughes, "The Integrated Energy and Communication Systems Architecture Volume II," in *Permanent Power Quality Measurement*: Electric Power Research Institute (EPRI), 2004, http://smartgrid.epri.com/UseCases/PermanentPowerQualityMeasurement.pdf.

[34] M. H. J. Bollen and I. Y. H. Gu, *Signal Processing of Power Quality Disturbances*: IEEE Press, 2006.

[35] V. Giovanni, D. O. Luigi, S. Sergio, and D. B. Ettore, "Power Quality Survey In Progress On The MV ENEL Network: Analysis Of Gathered Data," in *CIRED*, 2009.

[36] C. Noce and S. Sartore, "The new Enel Distribuzione power quality data warehouse and its applications for smart grids," in *Harmonics and Quality of Power (ICHQP)* Bergamo, 2010.

[37] J. Fan and S. Borlase, "The Evolution of Distribution," in *IEEE Power and Energy Magazine*. vol. March / April: IEEE Power and Energy Systems (PES), 2009.

[38] GE, "GENe," 2011, http://www.gedigitalenergy.com/uos/catalog/GENeDMS.htm.

[39] ABB, "Network Manager SCADA / DMS," 2011, http://bit.ly/vsAEXu.

[40] "Distribution Management System Open++ Opera v.3.3," 1999.

[41] C.-S. Chen, C.-H. Lin, and H.-Y. Tsai, "A Rule-Based Expert System With Colored Petri Net Models for Distribution System Service Restoration," *IEEE Transactions on Power Systems*, vol. 17, No. 4, pp. 1073-1080, 2002.

[42] C.-H. Lin, H.-J. Chuang, C.-S. Chen, C.-S. Li, and C.-Y. Ho, "Fault Detection, Isolation and Restoration Using a Multiagent-based Distribution Automation System," in *Industrial Electronics and Applications*, 2009.

[43] H. M. Gill, "Smart Grid Distribution Automation for Public Power," in *Transmission and Distribution Conference and Exposition* New Orleans, LA, USA: IEEE PES, 2010.

[44] D. G. Hart, "Using AMI to Realize the Smart Grid," in *Power and Energy Society General Meeting*, Pittsburgh, 2008.

[45] C. N. Lu, J. H. Teng, and W. H. E. Liu, "Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 10, No. 1, pp. 229-240, 1995.

[46] K. Li, "State Estimation for Power Distribution Systems and Measurement Impacts," *IEEE Transactions on Power Systems*, vol. 11, No. 2, pp. 911-916, 1996.

[47] K. Nareshkumar, M. A. Choudhry, J. Lai, and A. Feliachi, "Application of Mult-Agents for Fault Detection and Reconfiguration of Power Distribution Systems," in *Power & Energy Society General Meeting*, 2009.

[48] Y. Liao, "Fault Location Observability Analysis and Optimal Meter Placement Based on Voltage Measurements," *Electric Power Systems Research*, vol. 79, No. 7, pp. 1062-1068, 2009.

[49] W. W. Dabbs, D. D. Sabin, T. E. Grebe, and H. Mehta, "PQView - A Power Quality Data Management and Analysis System," in *IEEE Computer Applications in Power Apparatus*, 1994.

[50] eMeter, "Outage Event Management," 2010, http://www.emeter.com/products/our-applications/outage-event-management/.

[51] Oracle, "Oracle Utilities Meter Data Management," 2010, http://www.oracle.com/us/industries/utilities/046897.html.

[52] Siemens, "Meter Data Management (MDMS)," 2010, http://www.energy.siemens.com/us/en/services/power-transmission-distribution/mdms.htm.

[53] H. Farhangi, "The Path of the Smart Grid," in *IEEE Power and Energy Magazine*. vol. January / February: IEEE Power and Energy Systems (PES), 2010.

[54] D. S. Dorr, T. M. Gruzs, M. B. Hughes, R. E. Jurewicz, G. Dang, and J. L. McClaine, "Interpreting Recent Power Quality Surveys to Define the Electrical Environment," Industry Applications Conference, Ed., 1996, pp. 2251-2258.

[55] B. Paszkier, C. Santander, and J. Gauthier, "Improving Power Quality By Databases Crossing," in *CIRED* Vienna, 2007.

[56] "A Complete Power Quality Solution," Alpes Technology, 2009, http://www.leonardo-energy.org/webfm_send/3811.

[57] EMS, "Embedded Monitoring Systems - Sub.net," 2011, http://www.emsni.com/.

[58] B. Gou and A. Abur, "A Direct Numerical Method for Observability Analysis," *IEEE Transactions on Power Systems*, vol. 15, No. 2, pp. 625-630, 2000.

[59] S. Z. Djokic, J. V. Milanovic, D. J. Chapman, M. F. McGranaghan, and D. S. Kirschen, "A New Method for Classification and Presentation of Voltage Reduction Events," *IEEE Transactions on Power Delivery*, vol. 20, No. 4, pp. 2576 - 2584, 2005.

[60] "UK Grid Code," National Grid, 2009.

[61] B. Wang, W. Xu, and Z. Pan, "Voltage Sag State Estimation for Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 20, No. 2, 2005.

[62] A. Abur and A. G. Exposito, *Power System State Estimation*: Marcel Dekker, 2004.

[63] J. J. Grainger and J. William D. Stevenson, *Power Systems Analysis*: Mcgraw Hill, 1994.

[64] B. Xu and A. Abur, "Observability Analysis and Measurement Placement for Systems with PMUs," in *Power Systems Conference and Exposition*, 2004.

[65] C. W. Hansen, "Power System State Estimation Using Three Phase Models," *IEEE Transactions on Power Systems*, vol. 10, No. 2, pp. 818-824, 1995.

[66] M. E. Baran and A. W. Kelley, "State Estimation for Real-Time Monitoring of Distribution Systems," *IEEE Transactions on Power Systems*, vol. 9, No. 3, 1994.

[67] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution System State Estimation Through Gaussian Mixture Model of the Load as Pseudo-Measurement," *IET Generation, Transmission and Distribution*, vol. 4, No. 1, pp. 50-59, 2009.

[68] A. K. Ghosh, D. L. Lubkeman, M. J. Downey, and R. H. Jones, "Distribution Circuit State Estimation Using a Probabilistic Approach," *IEEE Transactions on Power Systems*, vol. 12, No. 1, pp. 45-51, 1997.

[69] R. Singh, B. C. Pal, and R. B. Vinter, "Measurement Placement in Distribution System State Estimation," *IEEE Transactions on Power Systems*, vol. 24, No. 2, May 2009 2009.

[70] V. Thornley, N. Jenkins, and S. White, "State Estimation Applied to Active Distribution Networks with Minimal Measurements," in *15th Power Systems Computational Conference*, 2005.

[71] P. Caramia, G. Carpinelli, P. Varilone, and P. Verde, " Probabilistic Three Phase Load Flow," *Electrical Power and Energy Systems*, vol. 21, No. 2, pp. 55-69, 1999.

[72] P. Caramia, G. Carpinelli, M. Pagano, and P. Varilone, "Probabilistic Three-Phase Load Flow for Unbalanced Electrical Distribution Systems with Wind Farms," *IET Renewable Power Generation*, vol. 1, No. 2, 2007.

[73] Y. J. Wang, "Modelling of Random Variation of Three-Phase Voltage Unbalance in Electric Distribution Systems Using Trivariate Gaussian Distribution," *IEE Proc. Generation Transmission Distribution*, vol. 148, No. 4, 2001.

[74] C. M. Kuan, "Generalized Least Squares Theory," in *Introduction to Econometric Theory*: Institute of Economics, Academia Sinica, Taipei 115, Taiwan, 2004.

[75] P. Paranavithana and S. Perera, "Location of Sources of Voltage Unbalance in an Interconnected Network," in *Power & Energy Society General Meeting*: IEEE, 2009.

[76] P. Paranavithana and S. Perera, "A Systematic Approach Towards Evaluating Voltage Unbalance Problem in Interconnected Sub-Transmission Networks: Separation of Contribution by Lines, Loads and Mitigation," in *International Conference on Harmonics and Quality of Power (ICHQP)*, 2008.

[77] J. V. Milanović and C. P. Gupta, "Probabilistic Assessment of Financial Losses due to Interruptions and Voltage Sags - Part II: Practical Implementation," *IEEE Transactions on Power Delivery*, vol. 21, No. 2, 2006.

[78] J. Wang, S. Chen, and T. T. Lie, "System Voltage Sag Performance Estimation," *IEEE Transactions on Power Delivery* vol. 20, No. 2, pp. 1738-1747, 2005.

[79] P. Heine, P. Pohjanheimo, M. Lehtonen, and E. Lakervi, "A Method for Estimating the Frequency and Cost of Voltage Sags," *IEEE Transactions on Power Systems*, vol. 17, No. 2, pp. 290-296, 2002.

[80] Y. Liao, "Generalized Fault-Location Methods for Overhead Electric Distribution Systems," *IEEE Transactions on Power Delivery*, vol. 26, No. 1, 2011.

[81] Y. Liao, "Fault location for single-circuit line based on bus impedance matrix utilizing voltage measurements," *IEEE Transactions on Power Delivery*, vol. 23, No. 2, pp. 609-617, April 2008.

[82] R. Krishnathevar and E. E. Ngu, "Generalized Impedanced-Based Fault Location for Distribution Systems," *Power Engineering Letters*, vol. 27, No. 1, 2011.

[83] L. Xu and M.-Y. Chow, "A Classification Approach for Power Distribution Systems Fault Cause Identification," *IEEE Transactions on Power Systems*, vol. 21, No. 1, pp. 53-60, 2006.

[84] L. Xu and M.-Y. Chow, "Distribution fault diagnosis using a hybrid algorithm of fuzzy classification and artificial immune systems," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008.

[85] M. Gui, A. Pahwa, and S. Das, "Anomaly Detection in Animal-Related Failures in Overhead Distribution Systems," in *Power Symposium, 2007. NAPS '07. 39th North American*, 2007, pp. 498-504.

[86] M. Gui, A. Pahwa, and S. Das, "Analysis of Animal-Related Outages in Overhead Distribution Systems With Wavelet Decomposition and Immune Systems-Based Neural Networks," *IEEE Transactions on Power Systems*, vol. 24, No. 4, 2009.

[87] A.Watkins, "AIRS: A Resource Limited Artificial Immune Classifier," Master of Science, Mississippi State University, 2001.

[88] A. Jamehbozorg and S. M. Shahrtash, "A Decision-Tree-Based Method for Fault Classification in Single-Circuit Transmission Lines," *IEEE Transactions on Power Delivery*, vol. 24, No. 4, 2010.

[89] L. Breiman, "Random Forests," *Machine Learning*, vol. 1, No. 1, pp. 5-32, 2001.

[90] J. Timmis, M. Neal, and J. Hunt, "An artificial immune system for data analysis," *Biosystems*, vol. 55, No. 1-3, pp. 143-150, February 200 2000.

[91] Y. Zhong, L. Zhang, B. Huang, and P. Li, "An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery," *IEE Transactions on Geoscience and Remote Sensing*, vol. 44, No. 2, pp. 420-431, February 2006 2006.

[92] E. G. Carrano, F. G. Guimaraes, R. H. C. Takahashi, O. M. Neto, and F. Campelo, "Electric Distribution Network Expansion Under Load-Evolution Uncertainty Using an Immune System Inspired Algorithm," *IEEE Transactions on Power Systems*, vol. 22, No. 2, pp. 851-861, May 2007.

[93] A. Ahuja, S. Das, and A. Pahwa, "An AIS-ACO Hybrid Approach for Multi-Objective Distribution System Reconfiguration," *IEEE Transactions on Power Systems*, vol. 22, No. 3, pp. 1101-1111, Auguest 2007.

[94] D. Dasgupta and S. Forrest, "An Anomaly Detection Algorithm Inspired By The Immune System," *Artificial Immune Systems and Their Applications*, No. 1, pp. 262-277, 1998.

[95] D. Dasgupta and B. A. G. Osorio, "Neuro-Immune and Self-Organizing Map Approaches to Anomaly Detection: A Comparison," in *1st International Conference on Artificial Immune Systems*, 2002.

[96] A. Watkins, J. Timmis, and L. Boggess, "Artificial immune recognition system (AIRS): An immune inspired supervised machine learning algorithm," *Genetic Programming and Evolvable Machines*, vol. 5, No. 3, pp. 291-317, 2004.

[97] A. A. Freitas and J. Timmis, "Revisiting the Foundations of Artificial Immune Systems for Data Mining," *IEEE Transactions on Evolutionary Computation*, vol. 11, No. 4, August 2007 2006.

[98] R. T. Alves, M. R. Delgado, H. S. Lopes, and A. A. Freitas, "An Artificial Immune System for Fuzzy-Rule Induction in Data Mining," *Lecture Notes in Computer Science*, vol. 3242/2004, No. 1, 2004.

[99] L. N. d. Castro and F. J. V. Zuben, "An Evolutionary Immune Network for Data Clustering," in *IEEE Brazillian Symposium on Artificial Neural Networks*, Rio de Janeiro, 2000, pp. 84-89.

[100] M. Kezunovic and B. Peunicic, "Fault Location," in *Wiley Ency*. vol. 7: Wiley Encyclopedia of Electrical and Electronics Terminology, 1999, pp. 276-285.

[101] T. Takagi, Y. Yamakoshi, Y. Yamaura, R. Kondow, and T. Matsushima, "Development of a New Type of Fault Locator Using the One-Terminal Voltage and Current Data," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, No. 8, pp. 2892-2898, 1981.

[102] J. Mora-Florez, J. Melendez, and G. Carillo-Caicedo, "Comparison of Impedance Based Fault Location methods for Power Distribution Systems," *Electric Power Systems Research*, vol. 78, No. 20047.

[103] M. McGranaghan, T. Short, and D. Sabin, "Using PQ Monitoring Infrastructure for Automatic Fault Location," in *19th International Conference of Electricity Distribution*, 2007.

[104] M. Kezunovic, "Smart Fault Location for Smart Grids," *IEEE Transactions on Smart Grid*, vol. 2, No. 1, pp. 11-22, 2011.

[105] N. Kang and Y. Liao, "Double-Circuit Transmission-Line Fault Location With the Availability of Limited Voltage Measurements," *IEEE Transactions on Power Delivery,* vol. 27, No. 1, pp. 325-336, 2012.

[106] C. P. Gupta and J. V. Milanović, " Probabilistic Assessment of Equipment Trips due to Voltage Sags," *IEEE Transactions on Power Delivery,* vol. 21, No. 2, 2006.

[107] Y. Moon, Y. H. Moon, J. B. Choo, and T. W. Kwon, "Design of Reliable Measurement System for State Estimation," *IEEE Transactions on Power Systems,* vol. 3, No. 3, pp. 830-836, 1988.

[108] B. Gou, "Optimal Placement of PMUs by Integer Linear Programming," *IEEE transactions on Power Systems,* vol. 23, No. 3, pp. 1525-1526, 2008.

[109] E. W. Palmer and G. Ledwich, "Optimal Placement of Angle Transducers in Power Systems," *IEEE Transactions on Power Systems,* vol. 11, No. 2, pp. 788-793, 1996.

[110] A. M. Almutairi and J. V. Milanovic, "Comparison of Different Methods for Optimal Placement of PMUs," in *Power Tech* Bucharest, Romania, 2009.

[111] C. Madtharad, S. Premrudeepreechacharn, N. R. Watson, and R. S. Udom, "An Optimal Measurement Placement Method for Power System Harmonic State Estimation," *IEEE Transactions on Power Systems,* vol. 20, No. 2, 2005.

[112] S. Chakrabarti, E. Kyriakides, and D. G. Eliades, "Placement of Synchronized Measurements for Power System Observability," *IEEE Transactions on Power Delivery,* vol. 24, No. 1, pp. 12-19, 2009.

[113] R. Emami and A. Abur, "Robust Measurement Design by Placing Synchronized Phasor Measurements on Network Branches," *IEEE Transactions on Power Systems,* vol. 25, No. 1, pp. 38-43, 2010.

[114] B. Milosevic and M. Begovic, "Nondominated Sorting Genetic Algorithm for Optimal Phasor Measurement Placement," *IEEE Transactions on Power Systems,* vol. 18, No. 1, pp. 69-75, 2003.

[115] C.-Y. Ho, T.-E. Lee, and C.-H. Lin, "Optimal Placement of Fault Indicators Using the Immune Algorithm," *IEEE Transactions on Power Systems,* vol. 26, No. 1, pp. 38-45, 2011.

[116] X. Bian and J. Qiu, "Adaptive Clonal Algorithm and Its Application for Optimal PMU Placement," in *International Conference on Communications, Circuits and Systems Proceedings*, Guilin, 2006.

[117] F. Aminifar, C. Lucas, A. Khodaei, and M. Fotuhi-Firuzabad, "Optimal Placement of Phasor Measurement Units Using Immunity Genetic Algorithm," *IEEE Transactions on Power Delivery,* vol. 24, No. 3, 2009.

[118] A. A. Ibrahim, A. Mohamed, H. Shareef, and G. P. Ghosal, "Optimal power quality monitor placement in power systems based on particle swarm optimization and artificial immune system," in *3rd Conference on Data Mining and Optimization (DMO)*, 2011.

[119] F. Aminifar, M. Fotuhi-Firuzabad, M. Shahidehpour, and A. Khodaei, "Probabilistic Multistage PMU Placement in Electric Power Systems," *IEEE Transactions on Power Delivery*, vol. 26, No. 2, pp. 841-849, 2011.

[120] G. Olguin, "An Optimal Trade-off Between Monitoring and Simulation for Voltage Dip Characterization of Transmission Systems," in *IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific* China, 2005.

[121] B. Xu and A. Abur, "Observability Analysis and Measurement Placement for Systems with PMUs," in *IEEE Power Engineering Society Power Systems Conference Exposition*, New York, 2004.

[122] C. Muscas, F. Pilo, G. Pisano, and S. Sulis, "Optimal Allocation of Multichannel Measurement Devices for Distribution State Estimation," *IEEE Transactions on Instrumentation and Measurement,* vol. 58, No. 6, pp. 1929-1937, 2009.

[123] J. Arrillaga and N. R. Watson, *Computer Modelling of Power Systems*: John Wiley & Sons, 2001.

[124] "PowerLogic PM700 Series," Schneider Electric, Ed., 2011, http://www.powerlogic.com/literature/3020HO0701.pdf.

[125] "Instrument transformers - Part 1: Current transformers," IEC 60044-1 ED. 1.2, 2003.

[126] Siemens, "9610 Power Quality Meter," 2011, http://bit.ly/xnC1nG.

[127] GE, "EPM 9650 Power Quality Meter," 2011, http://www.gedigitalenergy.com/multilin/catalog/epm9650.htm.

[128] C. L. Fortescue, "Method of Symetrical Co-Ordinates to the Solution of Polyphase Networks," in *34th Annual Convention of the AIEE (American Institute of Electrical Engineers)*, Atlantic City, New Jersey, 1918.

[129] Y. Zhang, "Techno-economic Assessment of Voltage Sag Performance and Mitigation," PhD, 2008.

[130] F. Corcoles, L. Sainz, J. Pedra, J. Sanchez-Navarro, and M. Salichs, "Three-phase Transformer Modelling for Unbalanced Conditions," *IET Electric Power Applications,* vol. 2, No. 2, p. 99 to 112, 2007.

[131] S. Civanlar and J. J. Grainger, "Volt / VAR Control on Distribution Systems with Lateral Branches Using Shunt Capacitors and Voltage Regulators Part II: The Solution Method," *IEEE Transactions on Power Apparatus and Systems,* No. 1985.

[132] I. Moghram and S. Rahman, "Analysis and Evaluation of Five Short-Term Load Forecasting Techniques," *IEEE Transactions on Power Systems,* vol. 4, No. 4, 1989.

[133] D. E. Goodman, L. Bogess, and A. Watkins, "Artificial Immune System classification of multiple-class problems," in *Intelligent Engineering Systems*, 2002, pp. 179-184.

[134] F. M. Burnet, "The Clonal Selection Theory of Acquired Immunity," *Cambridge University Press,* No. 1959.

[135] C. Berek and M. Ziegner, "The Maturation of the Immune Response," *Immunology Today,* vol. 14, No. 8, pp. 400-404, 1993.

[136] J. Timmis, "Artificial immune systems—today and tomorrow," *Natural Computing,* vol. 6, No. 1, pp. 1-18, 2007.

[137] L. N. d. Castro and J. Timmis, *Artificial Immune Systems: A New Computational Intelligence Approach*: Springer-Verlag, 2002.

[138] A. B. Watkins, "Exploiting Immunological Metaphors in the Development of Serial Parallel and Distributed Learning Algorithms," *Department of Computer Science,* 2005.

[139] A. Watkins and J. Timmis, "Artificial immune recognition system (AIRS): Revisions and refinements," in *Proc. 1st Int. Conf. Artificial Immune Systems (ICARIS)*, J. Timmis and P. J. Bentley, Eds., 2002, pp. 173-181.

[140] C. Cortes and V. Vapnik, "Support Vector Networks," in *Machine Learning*, 1995, pp. 273-297.

[141] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Computation,* vol. 15, No. 7, pp. 1667-1689, 2003.

[142] E. A. Leonidaki, D. P. Georgiadis, and N. D. Hatziargyriou, "Decision Trees for Determination of Optimal Location and Rate of Series Compensation to Increase Power System Loading Margin," *IEEE Transactions on Power Systems,* vol. 21, No. 3, 2006.

[143] B. V. Dasarathy, *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*: IEEE Computer Society, 1991.

[144] J. Timmis, P. Andrews, and N. Owens, "An Interdisciplinary Perspective on Artificial Immune Systems," *Evolutionary Intelligence,* vol. 1, No. 1, pp. 5-26, 2008.

[145] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*: Wiley-Interscience, 1990.

[146] J. Kelsey and J. Timmis, "Immune Inspired Somatic Contiguous Hypermutation for Function Optimisation," *Lecture Notes in Computer Science,* vol. 2723/2003, No. 2003.

[147] K. Trojanowski and S. T. Wierzchon, "Immune-based Algorithms for Dynamic Optimization," *Information Sciences,* vol. 179, No. pp. 1495-1515, 2009.

[148] J. Kelsey, J. Timmis, and A. N. W. Hone, "Chasing Chaos," in *Congress on Evolutionary Computation* Canberra, Australia, 2003, pp. 413-419.

[149] R. Rosin-Arbesfeld, F. Townsley, and M. Bienz, "The APC Tumour Suppressor has a Nuclear Export Function," *Letters to Nature,* No. 406, pp. 1009-1012, 2000.

[150] H. Lamlum and e. al., "The Type of Somatic Mutation at APC in Familial Adenomatous Polyposis is Determined by the Site of the Germline Mutation: A New Facet to Knudson's 'two-hit' Hypothesis," *Nature Medicine,* No. 1071-1075, 1999.

[151] T. U. o. Waikato, "WEKA - http://www.cs.waikato.ac.nz/ml/weka/," 2009, http://www.cs.waikato.ac.nz/ml/weka/.

[152] C.-C. Chang and C.-J. Lin, "LIBSVM - A Library for Support Vector Machines," 2008, http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[154] S. Geisser, *Predictive Inference*: Chapman and Hall, 1993.

[155] "A Practical Guide to Support Vector Classification," 21st May 2008, http://www.csie.ntu.edu.tw/~cjlin.

[156] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin,* vol. 1, No. 6, pp. 80-83, 1945.

[157] A. Vargha and H. Delaney, "A Critique and Improvement of the CL Common Language Effect Size Statistics of McGraw and Wong," *Journal of Educational and Behavioural Statistics,* vol. 25, No. 2, pp. 101-132, 2000.

[158] J. Zhu, D. L. Lubkeman, and A. A. Girgis, "Automated Fault Location and Diagnosis on Electric Power Distribution Feeders," *IEEE Transactions on Power Delivery,* vol. 12, No. 2, April 1997 1997.

[159] M. E. Payton, M. H. Greenstone, and N. Schenker, "Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance?," *Journal of Insect Science,* vol. 3, No. 34, 2003.

[160] I. T. I. Council, "ITI (CBEMA) Curve Application Note," http://www.itic.org/archives/iticurv.pdf.

[161] J. Kyei, R. Ayyanar, G. Heydt, R. Thallam, and J. Blevins, "The Design of Power Acceptability Curves," *IEEE Transactions on Power Delivery,* vol. 17, No. 3, 2002.

[162] C. N. Lu and C. C. Shen, "Voltage Sag Immunity Factor Considering Severity and Duration," in *IEEE Power Engineering Society General Meeting*, 2004.

[163] S. Bahadoorsingh, J. V. Milanovic, Y. Zhang, C. P. Gupta, and J. Dragovic, "Minimization of Voltage Sag Costs by Optimal Reconfiguration of Distribution Network Using Genetic Algorithms," *IEEE Transactions on Power Delivery,* vol. 22, No. 4, pp. 2271-2278, 2007.

[164] J. A. Martinez and J. Martin-Arnedo, "Voltage sag stochastic prediction using an electromagnetic transients program," *IEEE Transactions on Power Delivery,* vol. 19, No. 4, pp. 1975-1982, October 2004.

[165] N. C. Woolley, M. Avendano-Mora, and J. V. Milanovic, "A Comparison of Voltage Sag Estimation Algorithms Using Optimal Monitoring Locations," in *International Conference On Harmonics And Quality Of Power (ICHQP)* Milan, Italy, 2010.

[166] A. Abur and A. G. Exposito, *Power System State Estimation Theory and Implementation*: Marcel Dekker Inc., 2004.

[167] M. A. Eldery, E. F. El-Saadany, M. M. A. Salama, and A. Vannelli, "A Novel Power Quality Monitoring Allocation Algorithm," *IEEE Transactions on Power Delivery,* vol. 21, No. 2, pp. 768-777, 2006.

[168] C. A. Reineri, J. C. Gomez, E. B. B., and M. Felici, "Revision of Concepts and Approaches for Unbalance Problems in Distribution," in *PES Transmission & Distribution*, 2006.

[169] J. Wan and K. N. Miu, "Weighted Least Squares Methods for Load Estimation in Distribution Networks," *IEEE Transactions on Power Systems,* vol. 18, No. 4, pp. 1338-1345, 2003.

[170] R. P. Broadwater, A. H. Khan, H. E. Shaalan, and R. E. Lee, "Time Varying Load Analysis to Reduce Distribution Losses Through Reconfiguration," *IEEE Transactions on Power Delivery,* vol. 8, No. 1, pp. 294-300, 1993.

[171] Y. J. Wang, "Simulation of random variation of three-phase voltage unbalance resulting from load fluctuation using correlated Gaussian random variables," in *Proc. Natl. Sci. Counc.*, 1999.

[172] A. C. Rencher, *Methods of Multivariate Analysis*: Wiley, 2002.

[173] T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain 'Goodness-of-fit' Criteria Based on Stochastic Processes," *Ann. Math. Stat.*, vol. 23, No. pp. 193-212, 1952.

[174] S. Zhong and A. Abur, "Effects of Non-transposed Lines and Unbalanced Loads on State Estimation," in *IEEE Power Engineering Society Winter Meeting*, 2002, pp. 975-979.

[175] "MG 1-2009, Motors and Generators," 2009.

[176] S. Chauhan and M. P. Dave, "Kohonen Neural Network Classifier for Voltage Collapse Margin Estimation," *Electric Power Components and Systems*, vol. 33, No. 10, pp. 607-619, 1997.

[177] H. B. Wan, Y. H. Song, and A. T. Johns, "Kohonen Neural Network based Approach to Voltage Buses / Areas Identification," *IEEE Proceedings on Generation, Transmission and Distribution*, vol. 144, No. 3, pp. 340-344, May 1997.

[178] T. An, S. Zhou, and J. Yu, "Application of Fuzzy Maximal Tree Clustering for Weak Voltage Buses Grouping," in *The Sixth World Congress on Intelligent Control and Automation, 2006. WCICA 2006.* , 2006, pp. 7462-7465.

[179] O. Alsac and B. Stott, "Optimal Load Flow with Steady State Security," *IEEE Transactions on Power Apparatus and Systems*, vol. 93, No. 3, 1974.

[180] I. Dabbagchi and R. Christie, "IEEE 14 Bus Test Case," 1993, http://www.ee.washington.edu/research/pstca/.

[181] J. H. Chow, "Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence," Springer 2005.

[182] R. D. Zimmerman, C. E. Murillo-Sanchez, and D. D. Gan, "MATPOWER," 2007, http://www.pserc.cornell.edu/matpower/.

[183] C. W. Taylor, *Power System Voltage Stability*: McGraw-Hill Education, 1998.

[184] P. Rousseaux and T. V. Cutsem, "Quasi steady-state simulation diagnosis using Newton method with optimal multiplier," in *IEEE Power Engineering Society General Meeting*, 2006.

[185] K. Iba, H. Suzuki, M. Egawa, and T. Watanabe, "Calculation of Critical Loading Condition with Nose Curve Using Homotopy Continuation Method," *IEEE Transactions on Power Systems*, vol. 6, No. 2, pp. 584-593, 1991.

[186] V. Ajjarapu, *Computational Techniques for Voltage Stability Assessment and Control*. Iowa: Springer, 2006.

[187] D. R. Wilson and T. R. Martinez, "Improved Heterogeneous Distance Functions," *Journal of Artificial Intelligence Research*, vol. 6, No. 1, pp. 1-34, 1997.

[188] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, No. 1, pp. 37-46, 1960.

[189] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*: John Wiley & Sons Inc, 2001.

[190] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*: Prentice-Hall, London, 1982.

[191] A. Seeker and A. A. Freitas, "WAIRS: improving classification accuracy by weighting attributes in the AIRS classifier," in *IEEE Congress on Evolutionary Computation*, 2007, pp. 3759-3765.

[192] J. Donald E. Goodman, L. Boggess, and A. Watkins, "An Investigation into the Source of Power for AIRS, an Artificial Immune Classification System," in *Proceedings of the International Joint Conference on Neural Networks, 2003*, 2003, pp. 1678-1683.

12 Appendices

Appendix A: Derivation of Distribution System State Estimation Equations

The equations in this appendix describe the components the Jacobian required to solve three phase load flow or state estimation.

In this research, the three phase equations were formulated using the real and imaginary components of the voltage, rather than voltage magnitude and angle. The real component of the voltage is defined in equation (12.1) and the imaginary component of the voltage is defined in equation (12.2):

$$\phi_{\mathrm{ic}}^{\mathrm{m}} = V_i^m \cos\theta_i^m \tag{12.1}$$

$$\phi_{\mathrm{is}}^{\mathrm{m}} = V_i^m \sin\theta_i^m \tag{12.2}$$

The measurement vector **z** is made up of three types of quantities:

- Real and reactive power injections
- Real and reactive power line flows
- Voltage magnitudes

Real and reactive power injections are required for three phase load flow. In addition to power injections, distribution system state estimation can also utilise line flows and voltage magnitude.

The Jacobian matrix ($\mathbf{H}_x$) can be calculated by considering the partial differentiation of each of the state non-linear equations defining power injections, power flows and voltages with respect to the state variables in equations (12.1) and (12.2).

The *i*th, *j*th element of the Jacobian matrix is defined as shown in equation (12.3):

$$H_{\mathbf{x}(i,j)} = \frac{\partial H_{(i)}(\mathbf{x_k})}{\partial x_{(j)}} \tag{12.3}$$

Where $H_{(i)}(\mathbf{x_k})$ is the *i*th equation defining the injections, line flows or voltages in the network and the $x_{(j)}$ is the *j*th state variable; either one of equation (12.1) or (12.2).

The derivations that follow all of the elements needed to formulate a full three phase Jacobian matrix.

A.1 Power Injections

For both the real and reactive power injections, **G** and **B** are the real and imaginary components of the three phase admittance matrix **Y**.

A.1.1 Real Power Injections

Equation (12.4) describes real power flow in a transmission line:

$$P_i^p = \sum_{k=1\ldots N} \sum_{m=\{a,b,c\}} V_k^m [G_{ik}^{pm} \cos(\theta_{ij}^{pm}) + B_{ik}^{pm} \sin(\theta_{ik}^{pm})] \tag{12.4}$$

If $\phi_{\mathrm{ic}}^{\mathrm{m}} = V_i^m \cos\theta_i^m$ and $\phi_{\mathrm{is}}^{\mathrm{m}} = V_i^m \sin\theta_i^m$, then the real power injection equations can be re-written as:

$$P_i^p = \sum_{k=1}^{N} \sum_{m=\{a,b,c\}} G_{ik}^{pm}\left[\emptyset_{ic}^p \emptyset_{kc}^m + \emptyset_{is}^p \emptyset_{ks}^m\right] + B_{ik}^{pm}\left[\emptyset_{is}^p \emptyset_{kc}^m - \emptyset_{ic}^p \emptyset_{ks}^m\right] \tag{12.5}$$

$$\begin{aligned} P_i^p = {} & G_{ii}^{pp}\left[{\emptyset_{ic}^p}^2 + {\emptyset_{is}^p}^2\right] + B_{ii}^{pp}\left[\emptyset_{is}^p \emptyset_{ic}^p - \emptyset_{ic}^p \emptyset_{is}^p\right] \\ & + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ii}^{pm}\left[\emptyset_{ic}^p \emptyset_{ic}^m + \emptyset_{is}^p \emptyset_{is}^m\right] + B_{ii}^{pm}\left[\emptyset_{is}^p \emptyset_{ic}^m - \emptyset_{ic}^p \emptyset_{is}^m\right] \\ & + \sum_{\substack{k=1 \\ k\neq i}}^{N} \Bigg\{ G_{ik}^{pp}\left[\emptyset_{ic}^p \emptyset_{kc}^p + \emptyset_{is}^p \emptyset_{ks}^p\right] + B_{ik}^{pp}\left[\emptyset_{is}^p \emptyset_{kc}^p - \emptyset_{ic}^p \emptyset_{ks}^p\right] \\ & + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ik}^{pm}\left[\emptyset_{ic}^p \emptyset_{kc}^m + \emptyset_{is}^p \emptyset_{ks}^m\right] + B_{ik}^{pm}\left[\emptyset_{is}^p \emptyset_{kc}^m - \emptyset_{ic}^p \emptyset_{ks}^m\right] \Bigg\} \end{aligned} \tag{12.6}$$

A.1.1.1 Real Power Injection Derivatives

The derivatives with respect to real $(\phi_{\mathrm{ic}}^{\mathrm{m}})$ and imaginary $(\phi_{\mathrm{is}}^{\mathrm{m}})$ component of the voltage for real line flow are shown below:

$$\frac{\partial P_i^p}{\partial \emptyset_{is}^p} = 2G_{ii}^{pp}\emptyset_{is}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ii}^{pm}\emptyset_{is}^m + B_{ii}^{pm}\emptyset_{ic}^m + \sum_{\substack{k=1, \\ k\neq i}}^{N} \left\{ G_{ik}^{pp}\emptyset_{ks}^p + B_{ik}^{pp}\emptyset_{kc}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ik}^{pm}\emptyset_{ks}^m + B_{ik}^{pm}\emptyset_{kc}^m \right\} \tag{12.7}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{ic}^p} = 2G_{ii}^{pp}\emptyset_{ic}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ii}^{pm}\emptyset_{ic}^m - B_{ii}^{pm}\emptyset_{is}^m + \sum_{\substack{k=1, \\ k\neq i}}^{N} \left\{ G_{ik}^{pp}\emptyset_{kc}^p - B_{ik}^{pp}\emptyset_{ks}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ik}^{pm}\emptyset_{kc}^m - B_{ik}^{pm}\emptyset_{ks}^m \right\} \tag{12.8}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{js}^p} = G_{ij}^{pp}\emptyset_{is}^p - B_{ij}^{pp}\emptyset_{ic}^p \tag{12.9}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{jc}^p} = G_{ij}^{pp}\emptyset_{ic}^p + B_{ij}^{pp}\emptyset_{is}^p \tag{12.10}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{is}^q} = G_{ii}^{pq}\emptyset_{is}^p - B_{ii}^{pq}\emptyset_{ic}^p \tag{12.11}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{ic}^q} = G_{ii}^{pq}\emptyset_{ic}^p + B_{ii}^{pq}\emptyset_{is}^p \tag{12.12}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{js}^q} = G_{ij}^{pq}\emptyset_{is}^p - B_{ij}^{pq}\emptyset_{ic}^p \tag{12.13}$$

$$\frac{\partial P_i^p}{\partial \emptyset_{jc}^q} = G_{ij}^{pq}\emptyset_{ic}^p + B_{ij}^{pq}\emptyset_{is}^p \tag{12.14}$$

A.1.2 Reactive Power Injections

Equation (12.15) describes real power flow in a transmission line:

$$Q_i^p = \sum_{k=1\ldots N} \sum_{m=\{a,b,c\}} V_k^m [G_{ik}^{pm} \sin(\theta_{ij}^{pm}) - B_{ik}^{pm} \cos(\theta_{ik}^{pm})] \tag{12.15}$$

If $\phi_{\mathrm{ic}}^{\mathrm{m}} = V_i^m \cos\theta_i^m$ and $\phi_{\mathrm{is}}^{\mathrm{m}} = V_i^m \sin\theta_i^m$, then the real power injection equations can be re-written as:

$$Q_i^p = \sum_{k=1}^{N} \sum_{m=\{a,b,c\}} G_{ik}^{pm}\left[\emptyset_{is}^p\emptyset_{kc}^m - \emptyset_{ic}^p\emptyset_{ks}^m\right] - B_{ik}^{pm}\left[\emptyset_{ic}^p\emptyset_{kc}^m + \emptyset_{is}^p\emptyset_{ks}^m\right] \tag{12.16}$$

$$
\begin{aligned}
Q_i^p = {} & G_{ii}^{pp}\left[\emptyset_{is}^p\emptyset_{ic}^p - \emptyset_{ic}^p\emptyset_{is}^p\right] - B_{ii}^{pp}\left[{\emptyset_{ic}^p}^2 + {\emptyset_{is}^p}^2\right] \\
& + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ii}^{pm}\left[\emptyset_{is}^p\emptyset_{ic}^m - \emptyset_{ic}^p\emptyset_{is}^m\right] - B_{ii}^{pm}\left[\emptyset_{ic}^p\emptyset_{ic}^m + \emptyset_{is}^p\emptyset_{is}^m\right] \\
& + \sum_{\substack{k=1, \\ k\neq i}}^{N}\left\{ G_{ik}^{pp}\left[\emptyset_{is}^p\emptyset_{kc}^p - \emptyset_{ic}^p\emptyset_{ks}^p\right] - B_{ik}^{pp}\left[\emptyset_{ic}^p\emptyset_{kc}^p + \emptyset_{is}^p\emptyset_{ks}^p\right] \right. \\
& \left. + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ik}^{pm}\left[\emptyset_{is}^p\emptyset_{kc}^m - \emptyset_{ic}^p\emptyset_{ks}^m\right] - B_{ik}^{pm}\left[\emptyset_{ic}^p\emptyset_{kc}^m + \emptyset_{is}^p\emptyset_{ks}^m\right] \right\}
\end{aligned} \tag{12.17}
$$

A.1.2.1 Reactive Power Injection Derivatives

The derivatives with respect to real ($\phi_{\text{ic}}^{\text{m}}$) and imaginary ($\phi_{\text{is}}^{\text{m}}$) component of the voltage for real line flow are shown below:

$$
\begin{aligned}
\frac{\partial Q_i^p}{\partial \emptyset_{is}^p} = {} & -2B_{ii}^{pp}\emptyset_{is}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ii}^{pm}\emptyset_{ic}^m - B_{ii}^{pm}\emptyset_{is}^m \\
& + \sum_{\substack{k=1 \\ k\neq i}}^{N}\left\{ G_{ik}^{pp}\emptyset_{kc}^p - B_{ik}^{pp}\emptyset_{ks}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} G_{ik}^{pm}\emptyset_{kc}^m - B_{ik}^{pm}\emptyset_{ks}^m \right\}
\end{aligned} \tag{12.18}
$$

$$
\begin{aligned}
\frac{\partial Q_i^p}{\partial \emptyset_{ic}^p} = {} & -2B_{ii}^{pp}\emptyset_{ic}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} -G_{ii}^{pm}\emptyset_{is}^m - B_{ii}^{pm}\emptyset_{ic}^m \\
& + \sum_{\substack{k=1, \\ k\neq i}}^{N}\left\{ -G_{ik}^{pp}\emptyset_{ks}^p - B_{ik}^{pp}\emptyset_{kc}^p + \sum_{\substack{m\epsilon\{a,b,c\} \\ m\neq p}} -G_{ik}^{pm}\emptyset_{ks}^m - B_{ik}^{pm}\emptyset_{kc}^m \right\}
\end{aligned} \tag{12.19}
$$

$$\frac{\partial Q_i^p}{\partial \emptyset_{js}^p} = -G_{ij}^{pp}\emptyset_{ic}^p - B_{ij}^{pp}\emptyset_{is}^p \tag{12.20}$$

$$\frac{\partial Q_i^p}{\partial \emptyset_{jc}^p} = G_{ij}^{pp}\emptyset_{is}^p - B_{ij}^{pp}\emptyset_{ic}^p \tag{12.21}$$

$$\frac{\partial Q_i^p}{\partial \emptyset_{is}^q} = -G_{ii}^{pq}\emptyset_{ic}^p - B_{ii}^{pq}\emptyset_{is}^p \tag{12.22}$$

$$\frac{\partial Q_i^p}{\partial \emptyset_{ic}^q} = G_{ii}^{pq}\emptyset_{is}^p - B_{ii}^{pq}\emptyset_{ic}^p \tag{12.23}$$

$$\frac{\partial Q_i^p}{\partial \emptyset_{js}^q} = -G_{ij}^{pq}\emptyset_{ic}^p - B_{ij}^{pq}\emptyset_{is}^p \tag{12.24}$$

$$\frac{\partial Q_i^p}{\partial \emptyset_{jc}^q} = G_{ij}^{pq}\emptyset_{is}^p - B_{ij}^{pq}\emptyset_{ic}^p \tag{12.25}$$

A.2 Line Flow Equations

For the line flow equations, $G_{p,m}$ and $B_{p,m}$ are elements of 3×6 line admittance matrices $\mathbf{G_L}$ and $\mathbf{B_L}$ relating voltage $(V=[V_i^a,V_i^b,V_i^c,V_j^a,V_j^b,V_j^c]^T)$ and current $(I=[I_{i,j}^a,I_{i,j}^b,I_{i,j}^c]^T)$ between buses *i* and *j* in the form $\mathbf{I}_{ij}=(\mathbf{G_L}+j\mathbf{B_L})\mathbf{V}_{ij}$, where $\mathbf{I}_{ij}$ is a 3×1 vector representing the current flows in phases *a*, *b* and *c* from bus *i* to *j* and $\mathbf{V}_{ij}$ is a 6×1 vector representing the voltages at bus *i* in all three phases and the voltages in bus *j*.

A.2.1 Real Power Flow

Equation (12.26) describes real power flow in a transmission line:

$$P_{ij}^p=\sum_{m=\{a,b,c\}}\begin{matrix}G_{p,m}V_i^mV_i^p[\cos(\theta_i^m-\theta_i^p)]-B_{p,m}V_i^mV_i^p[\sin(\theta_i^m-\theta_i^p)]\\+G_{p,(m+3)}[\cos(\theta_i^m-\theta_i^p)]-B_{p,(m+3)}[\sin(\theta_i^m-\theta_i^p)]\end{matrix} \tag{12.26}$$

If $\phi_{\mathrm{ic}}^{\mathrm{m}}=V_i^m\cos\theta_i^m$ and $\phi_{\mathrm{is}}^{\mathrm{m}}=V_i^m\sin\theta_i^m$, then the power flowing in a line can be re-written as:

$$\begin{aligned}P_{ij}^p=\sum_{m=\{a,b,c\}}&G_{p,m}\left[\emptyset_{ic}^m\emptyset_{ic}^p+\emptyset_{is}^m\emptyset_{is}^p\right]-B_{p,m}\left[\emptyset_{is}^m\emptyset_{ic}^p-\emptyset_{ic}^m\emptyset_{is}^p\right]\\&+G_{p,(m+3)}\left[\emptyset_{jc}^m\emptyset_{ic}^p+\emptyset_{js}^m\emptyset_{is}^p\right]-B_{p,(m+3)}\left[\emptyset_{js}^m\emptyset_{ic}^p-\emptyset_{jc}^m\emptyset_{is}^p\right]\end{aligned} \tag{12.27}$$

A.2.2 Real Power Flow Derivatives

The derivatives with respect to real $(\phi_{\mathrm{ic}}^{\mathrm{m}})$ and imaginary $(\phi_{\mathrm{is}}^{\mathrm{m}})$ component of the voltage for real line flow are shown below:

$$\begin{aligned}\frac{\partial P_{ij}^p}{\partial\emptyset_{is}^p}=\sum_{\substack{m=\{a,b,c\}\\m\neq p}}&\left[G_{p,m}\emptyset_{is}^m+B_{p,m}\emptyset_{ic}^m+G_{p,(m+3)}\emptyset_{js}^m+B_{p,(m+3)}\emptyset_{jc}^m\right]+2G_{p,p}\emptyset_{is}^p\\&+G_{p,(p+3)}\emptyset_{js}^p+B_{p,(p+3)}\emptyset_{jc}^p\end{aligned} \tag{12.28}$$

$$\begin{aligned}\frac{\partial P_{ij}^p}{\partial\emptyset_{ic}^p}=\sum_{\substack{m=\{a,b,c\}\\m\neq p}}&\left[G_{p,m}\emptyset_{ic}^p-B_{p,m}\emptyset_{is}^m+G_{p,(m+3)}\emptyset_{jc}^m-B_{p,(m+3)}\emptyset_{js}^m\right]+2G_{p,p}\emptyset_{ic}^p\\&+G_{p,(p+3)}\emptyset_{jc}^p-B_{p,(p+3)}\emptyset_{js}^p\end{aligned} \tag{12.29}$$

$$\frac{\partial P_{ij}^p}{\partial\emptyset_{js}^p}=G_{p,(p+3)}\emptyset_{is}^p-B_{p,(p+3)}\emptyset_{ic}^p \tag{12.30}$$

$$\frac{\partial P_{ij}^p}{\partial\emptyset_{jc}^p}=G_{p,(p+3)}\emptyset_{ic}^p+B_{p,(p+3)}\emptyset_{is}^p \tag{12.31}$$

$$\frac{\partial P_{ij}^p}{\partial\emptyset_{is}^q}=G_{p,q}\emptyset_{is}^p-B_{p,q}\emptyset_{ic}^p \tag{12.32}$$

$$\frac{\partial P_{ij}^{p}}{\partial \emptyset_{ic}^{q}} = G_{p,q}\emptyset_{ic}^{p} + B_{p,q}\emptyset_{is}^{p} \tag{12.33}$$

$$\frac{\partial P_{ij}^{p}}{\partial \emptyset_{js}^{q}} = G_{p,(q+3)}\emptyset_{is}^{p} - B_{p,(q+3)}\emptyset_{ic}^{p} \tag{12.34}$$

$$\frac{\partial P_{ij}^{p}}{\partial \emptyset_{jc}^{q}} = G_{p,(q+3)}\emptyset_{ic}^{p} + B_{p,(q+3)}\emptyset_{is}^{p} \tag{12.35}$$

A.2.3 Reactive Power Flow

Equation (12.26) describes real power flow in a transmission line:

$$Q_{ij}^{p} = \sum_{m=\{a,b,c\}} \begin{matrix} -B_{p,m}V_i^{m}V_i^{p}[\cos(\theta_i^{m}-\theta_i^{p})] - G_{p,m}V_i^{m}V_i^{p}[\sin(\theta_i^{m}-\theta_i^{p})] \\ -B_{p,(m+3)}[\cos(\theta_i^{m}-\theta_i^{p})] - B_{p,(m+3)}[\sin(\theta_i^{m}-\theta_i^{p})] \end{matrix} \tag{12.36}$$

If $\phi_{\mathrm{ic}}^{\mathrm{m}} = V_i^{m}\cos\theta_i^{m}$ and $\phi_{\mathrm{is}}^{\mathrm{m}} = V_i^{m}\sin\theta_i^{m}$, then the reactive power flowing in a line can be re-written as:

$$\begin{aligned} Q_{ij}^{p} = \sum_{m=\{a,b,c\}} & -B_{p,m}\left[\emptyset_{ic}^{m}\emptyset_{ic}^{p} + \emptyset_{is}^{m}\emptyset_{is}^{p}\right] - G_{p,m}\left[\emptyset_{is}^{m}\emptyset_{ic}^{p} - \emptyset_{ic}^{m}\emptyset_{is}^{p}\right] \\ & - B_{p,(m+3)}\left[\emptyset_{jc}^{m}\emptyset_{ic}^{p} + \emptyset_{js}^{m}\emptyset_{is}^{p}\right] - G_{p,(m+3)}\left[\emptyset_{js}^{m}\emptyset_{ic}^{p} - \emptyset_{jc}^{m}\emptyset_{is}^{p}\right] \end{aligned} \tag{12.37}$$

A.2.3.1 Reactive Power Flow Derivatives

The derivatives with respect to real $(\phi_{\mathrm{ic}}^{\mathrm{m}})$ and imaginary $(\phi_{\mathrm{is}}^{\mathrm{m}})$ component of the voltage for reactive line flow are shown below:

$$\begin{aligned} \frac{\partial Q_{ij}^{p}}{\partial \emptyset_{is}^{p}} = \sum_{\substack{m=\{a,b,c\} \\ m\neq p}} & \left[-B_{p,m}\emptyset_{is}^{m} + G_{p,m}\emptyset_{ic}^{m} - B_{p,(m+3)}\emptyset_{js}^{m} + G_{p,(m+3)}\emptyset_{jc}^{m}\right] \\ & - 2B_{p,p}\emptyset_{is}^{p} - B_{p,(p+3)}\emptyset_{js}^{p} + G_{p,(p+3)}\emptyset_{jc}^{p} \end{aligned} \tag{12.38}$$

$$\begin{aligned} \frac{\partial Q_{ij}^{p}}{\partial \emptyset_{is}^{p}} = \sum_{\substack{m=\{a,b,c\} \\ m\neq p}} & \left[-B_{p,m}\emptyset_{ic}^{p} - G_{p,m}\emptyset_{is}^{m} - B_{p,(m+3)}\emptyset_{jc}^{m} - G_{p,(m+3)}\emptyset_{js}^{m}\right] \\ & - 2B_{p,p}\emptyset_{ic}^{p} - B_{p,(p+3)}\emptyset_{jc}^{p} - G_{p,(p+3)}\emptyset_{js}^{p} \end{aligned} \tag{12.39}$$

$$\frac{\partial Q_{ij}^{p}}{\partial \emptyset_{js}^{p}} = -B_{p,(p+3)}\emptyset_{is}^{p} - G_{p,(p+3)}\emptyset_{ic}^{p} \tag{12.40}$$

$$\frac{\partial Q_{ij}^{p}}{\partial \emptyset_{jc}^{p}} = -B_{p,(p+3)}\emptyset_{ic}^{p} + G_{p,(p+3)}\emptyset_{is}^{p} \tag{12.41}$$

$$\frac{\partial Q_{ij}^{p}}{\partial \emptyset_{is}^{q}} = -B_{p,q}\emptyset_{is}^{p} - G_{p,q}\emptyset_{ic}^{p} \tag{12.42}$$

$$\frac{\partial Q_{ij}^{p}}{\partial \emptyset_{ic}^{q}} = -B_{p,q}\emptyset_{ic}^{p} + G_{p,q}\emptyset_{is}^{p} \tag{12.43}$$

$$\frac{\partial Q_{ij}^{p}}{\partial \emptyset_{js}^{q}} = -B_{p,(q+3)}\emptyset_{is}^{p} - G_{p,(q+3)}\emptyset_{ic}^{p} \tag{12.44}$$

$$\frac{\partial Q_{ij}^{p}}{\partial \emptyset_{jc}^{q}} = -B_{p,(q+3)}\emptyset_{ic}^{p} + G_{p,(q+3)}\emptyset_{is}^{p} \tag{12.45}$$

A.3 Voltage Equations

Equation (12.46) describes the voltage magnitude in terms of its real and imaginary components:

$$V_{i}^{p} = \left\{ \emptyset_{is}^{p\,2} + \emptyset_{ic}^{p\,2} \right\}^{1/2} \tag{12.46}$$

A.3.1 Voltage Magnitude Derivatives

The derivatives with respect to real $(\phi_{\mathrm{ic}}^{\mathrm{m}})$ and imaginary $(\phi_{\mathrm{is}}^{\mathrm{m}})$ component of the voltage for voltage magnitude measurements are shown below:

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{is}^{p}} = \frac{\emptyset_{is}^{p}}{\left\{ \emptyset_{is}^{p\,2} + \emptyset_{ic}^{p\,2} \right\}^{1/2}} \tag{12.47}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{ic}^{p}} = \frac{\emptyset_{ic}^{p}}{\left\{ \emptyset_{is}^{p\,2} + \emptyset_{ic}^{p\,2} \right\}^{1/2}} \tag{12.48}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{js}^{p}} = 0 \tag{12.49}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{jc}^{p}} = 0 \tag{12.50}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{is}^{q}} = 0 \tag{12.51}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{ic}^{q}} = 0 \tag{12.52}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{js}^{q}} = 0 \tag{12.53}$$

$$\frac{\partial V_{i}^{p}}{\partial \emptyset_{jc}^{q}} = 0 \tag{12.54}$$

Appendix B: Description of 24 Bus Network

B.1 Network Topology

Fig 12.1 shows the topology of the 24 bus UK distribution network.

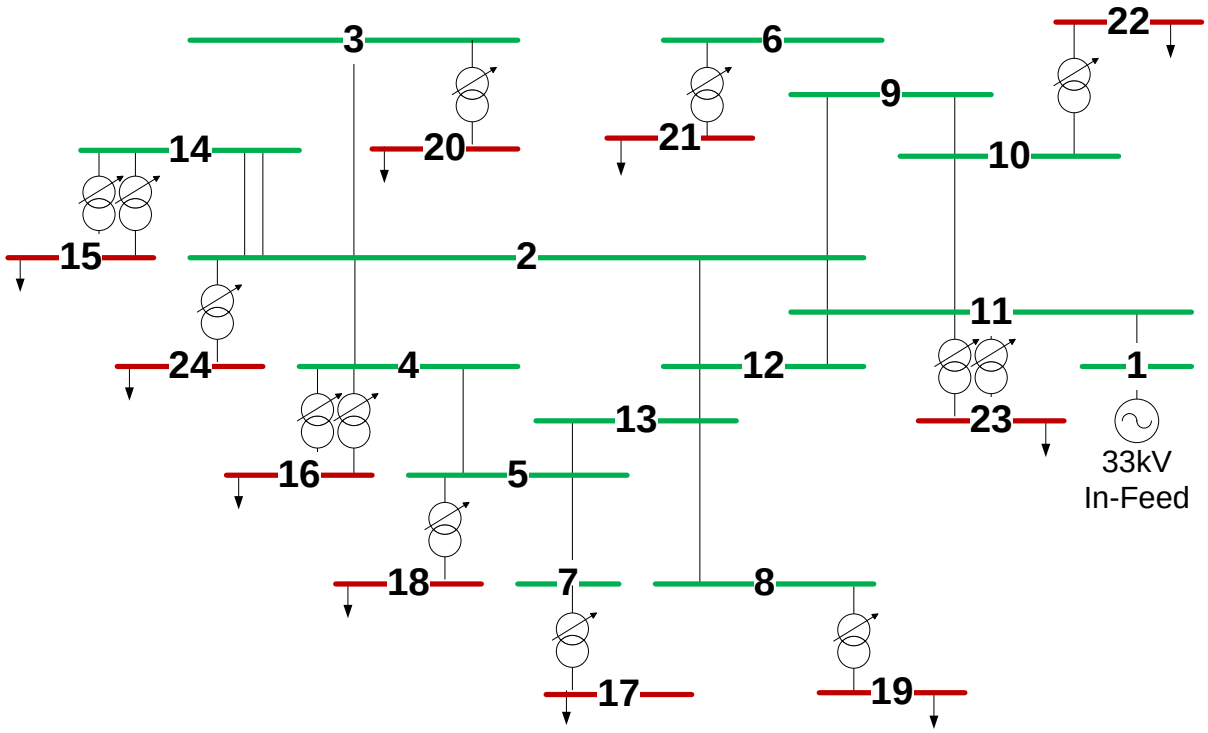


Fig. 12.1. The 24 bus UK Distribution network.

All transformers are Δ-Y connected (primary to secondary). The secondary side of all transformers is 11kV and the primary side is 33kV (line to line).

B.2 Loading

Load data was recorded on 1st March 2008. The load on the network on the 1st March 2008 was estimated by meters installed at each of the 10 11kV busbars which measured the line to line voltage and the current in one phase. From these two measurements, an estimate for three phase power was produced by assuming that the network loading was balanced, and by assuming a power factor based as advised by the team at the UK distribution network. The meters did not measure the angular difference between voltage and current, and therefore no measurement for the power factor was available. All power factors were assumed to be 0.97.

TABLE 12.1. THE ACTIVE AND REACTIVE POWER DEMAND IN THE 24 BUS UK DISTRIBUTION NETWORK ON MARCH 1ST 2008

Time	3 Phase Active Power (MW)										3 Phase Reactive Power (MVar)									
	Bus Number																			
	15	16	17	18	19	20	21	22	23	24	15	16	17	18	19	20	21	22	23	24
00:00	7.76	1.68	1.16	0.39	1.17	0.84	0.49	1.25	7.62	1.01	1.94	0.42	0.29	0.10	0.29	0.21	0.12	0.31	1.91	0.25
00:30	8.50	1.79	1.36	0.58	1.45	1.00	0.52	1.56	7.99	1.24	2.13	0.45	0.34	0.15	0.36	0.25	0.13	0.39	2.00	0.31
01:00	8.89	1.99	1.52	0.65	1.60	1.15	0.52	1.83	8.37	1.35	2.23	0.50	0.38	0.16	0.40	0.29	0.13	0.46	2.10	0.34
01:30	8.73	1.97	1.53	0.65	1.64	1.23	0.52	1.99	8.37	1.35	2.19	0.49	0.38	0.16	0.41	0.31	0.13	0.50	2.10	0.34
02:00	8.65	1.91	1.49	0.65	1.60	1.16	0.52	1.99	7.99	1.33	2.17	0.48	0.37	0.16	0.40	0.29	0.13	0.50	2.00	0.33
02:30	8.50	1.93	1.44	0.65	1.57	1.16	0.52	1.95	7.99	1.35	2.13	0.48	0.36	0.16	0.39	0.29	0.13	0.49	2.00	0.34
03:00	8.22	1.93	1.36	0.62	1.53	1.17	0.49	1.91	7.80	1.30	2.06	0.48	0.34	0.15	0.38	0.29	0.12	0.48	1.96	0.33
03:30	8.02	1.81	1.34	0.62	1.45	1.17	0.50	1.81	7.43	1.21	2.01	0.45	0.34	0.16	0.36	0.29	0.12	0.45	1.86	0.30
04:00	7.87	1.75	1.29	0.55	1.34	1.09	0.50	1.72	7.14	1.13	1.97	0.44	0.32	0.14	0.34	0.27	0.12	0.43	1.79	0.28
04:30	7.83	1.75	1.25	0.55	1.30	1.09	0.46	1.68	7.05	1.08	1.96	0.44	0.31	0.14	0.33	0.27	0.11	0.42	1.77	0.27
05:00	7.72	1.75	1.17	0.52	1.26	1.00	0.46	1.64	7.05	1.02	1.93	0.44	0.29	0.13	0.32	0.25	0.11	0.41	1.77	0.26
05:30	7.77	1.74	1.18	0.53	1.23	1.03	0.49	1.63	7.29	1.01	1.95	0.44	0.30	0.13	0.31	0.26	0.12	0.41	1.83	0.25
06:00	7.81	1.74	1.19	0.54	1.21	1.07	0.52	1.62	7.52	0.99	1.96	0.43	0.30	0.14	0.30	0.27	0.13	0.41	1.88	0.25
06:30	7.89	1.82	1.26	0.48	1.29	1.15	0.55	1.65	7.99	1.05	1.98	0.46	0.32	0.12	0.32	0.29	0.14	0.41	2.00	0.26
07:00	8.04	1.85	1.38	0.51	1.36	1.15	0.58	1.69	8.18	1.11	2.01	0.46	0.35	0.13	0.34	0.29	0.15	0.42	2.05	0.28
07:30	8.19	1.91	1.39	0.52	1.40	1.16	0.75	1.66	8.37	1.08	2.05	0.48	0.35	0.13	0.35	0.29	0.19	0.42	2.10	0.27
08:00	8.73	1.94	1.40	0.52	1.40	1.15	0.79	1.72	8.75	1.14	2.19	0.49	0.35	0.13	0.35	0.29	0.20	0.43	2.19	0.29
08:30	8.81	2.04	1.42	0.54	1.48	1.15	0.77	1.88	9.12	1.24	2.21	0.51	0.35	0.14	0.37	0.29	0.19	0.47	2.28	0.31
09:00	8.96	2.04	1.42	0.61	1.48	1.14	0.74	1.95	9.59	1.25	2.25	0.51	0.35	0.15	0.37	0.29	0.19	0.49	2.40	0.31
09:30	9.20	1.97	1.42	0.57	1.48	1.06	0.73	1.95	9.68	1.25	2.30	0.49	0.35	0.14	0.37	0.27	0.18	0.49	2.43	0.31
10:00	9.19	1.94	1.38	0.54	1.48	0.99	0.67	1.92	9.78	1.25	2.30	0.49	0.35	0.14	0.37	0.25	0.17	0.48	2.45	0.31
10:30	9.08	1.74	1.40	0.49	1.48	0.99	0.70	1.69	9.87	1.25	2.27	0.44	0.35	0.12	0.37	0.25	0.18	0.42	2.47	0.31
11:00	8.96	1.74	1.40	0.49	1.44	0.99	0.68	1.65	9.68	1.23	2.25	0.44	0.35	0.12	0.36	0.25	0.17	0.41	2.43	0.31
11:30	9.00	1.75	1.36	0.49	1.45	0.92	0.55	1.88	9.31	1.11	2.26	0.44	0.34	0.12	0.36	0.23	0.14	0.47	2.33	0.28
12:00	8.89	1.72	1.36	0.45	1.45	1.00	0.52	1.89	8.84	1.14	2.23	0.43	0.34	0.11	0.36	0.25	0.13	0.47	2.21	0.29
12:30	8.81	1.72	1.34	0.49	1.40	1.01	0.52	1.87	8.56	1.14	2.21	0.43	0.34	0.12	0.35	0.25	0.13	0.47	2.14	0.29
13:00	8.97	1.69	1.36	0.45	1.40	1.01	0.52	1.75	8.56	1.12	2.25	0.42	0.34	0.11	0.35	0.25	0.13	0.44	2.14	0.28
13:30	8.89	1.59	1.31	0.42	1.36	0.92	0.49	1.70	8.27	1.07	2.23	0.40	0.33	0.10	0.34	0.23	0.12	0.43	2.07	0.27
14:00	8.73	1.55	1.27	0.42	1.36	0.93	0.49	1.70	8.09	1.09	2.19	0.39	0.32	0.10	0.34	0.23	0.12	0.43	2.03	0.27
14:30	8.65	1.55	1.29	0.42	1.29	0.93	0.46	1.68	8.09	1.07	2.17	0.39	0.32	0.11	0.32	0.23	0.11	0.42	2.03	0.27
15:00	8.65	1.57	1.25	0.45	1.29	0.93	0.49	1.68	7.99	1.01	2.17	0.39	0.31	0.11	0.32	0.23	0.12	0.42	2.00	0.25
15:30	8.81	1.63	1.29	0.49	1.33	1.01	0.49	1.72	7.71	1.07	2.21	0.41	0.32	0.12	0.33	0.25	0.12	0.43	1.93	0.27
16:00	9.04	1.66	1.39	0.55	1.29	1.07	0.55	1.60	7.90	1.12	2.27	0.42	0.35	0.14	0.32	0.27	0.14	0.40	1.98	0.28
16:30	9.20	1.79	1.53	0.57	1.29	1.15	0.62	1.48	7.99	1.26	2.30	0.45	0.38	0.14	0.32	0.29	0.15	0.37	2.00	0.32
17:00	9.57	1.84	1.72	0.61	1.31	1.23	0.64	1.53	8.46	1.38	2.40	0.46	0.43	0.15	0.33	0.31	0.16	0.38	2.12	0.35
17:30	9.58	1.94	1.89	0.68	1.44	1.30	0.71	1.35	9.03	1.49	2.40	0.49	0.47	0.17	0.36	0.33	0.18	0.34	2.26	0.37
18:00	9.88	2.05	1.99	0.70	1.60	1.38	0.78	1.26	9.50	1.63	2.48	0.51	0.50	0.18	0.40	0.35	0.20	0.32	2.38	0.41
18:30	10.31	2.07	2.07	0.70	1.69	1.40	0.85	1.30	10.06	1.74	2.58	0.52	0.52	0.18	0.42	0.35	0.21	0.33	2.52	0.43
19:00	10.12	2.07	2.05	0.71	1.65	1.40	0.82	1.31	9.87	1.78	2.54	0.52	0.51	0.18	0.41	0.35	0.21	0.33	2.47	0.45
19:30	9.90	2.01	1.99	0.71	1.61	1.33	0.83	1.29	9.49	1.69	2.48	0.50	0.50	0.18	0.40	0.33	0.21	0.32	2.38	0.42
20:00	9.75	1.87	1.83	0.65	1.57	1.33	0.76	1.21	8.93	1.64	2.44	0.47	0.46	0.16	0.39	0.33	0.19	0.30	2.24	0.41
20:30	9.35	1.87	1.72	0.65	1.68	1.24	0.75	1.16	8.74	1.59	2.34	0.47	0.43	0.16	0.42	0.31	0.19	0.29	2.19	0.40
21:00	9.12	1.81	1.60	0.62	1.57	1.17	0.72	1.05	8.27	1.43	2.28	0.45	0.40	0.15	0.39	0.29	0.18	0.26	2.07	0.36
21:30	8.97	1.69	1.57	0.62	1.49	1.09	0.69	1.05	8.09	1.42	2.25	0.42	0.39	0.16	0.37	0.27	0.17	0.26	2.03	0.35
22:00	8.69	1.74	1.48	0.58	1.40	1.07	0.62	1.01	7.71	1.33	2.18	0.43	0.37	0.15	0.35	0.27	0.16	0.25	1.93	0.33
22:30	8.42	1.68	1.43	0.48	1.36	1.01	0.62	0.96	7.52	1.25	2.11	0.42	0.36	0.12	0.34	0.25	0.15	0.24	1.88	0.31
23:00	8.19	1.69	1.33	0.45	1.26	1.01	0.56	0.93	7.14	1.20	2.05	0.42	0.33	0.11	0.32	0.25	0.14	0.23	1.79	0.30
23:30	7.95	1.63	1.21	0.42	1.14	0.94	0.52	0.86	6.77	1.10	1.99	0.41	0.30	0.11	0.29	0.23	0.13	0.21	1.70	0.28

B.3 System Parameters

All system parameters were based on data supplied from UK distribution network operators in the form of an IPSA+ model.

B.3.1 Line Impedances

Table 12.2 shows the impedances of the 24 bus network. All measurements are in per unit.

TABLE 12.2. THE IMPEDANCES OF THE 24 BUS UK DISTRIBUTION NETWORK

From Bus	To Bus	$r^{(1)}$	$x^{(1)}$	$b^{(1)}$	$r^{(0)}$	$x^{(0)}$
1	11	0.001	0.005	0	0.001	0.005
11	10	0.102526	0.202248	0.000506	0.242922	0.569747
10	9	0.060713	0.125964	0.001399	0.239232	0.511495
9	6	0.01471	0.033768	0.000191	0.040016	0.151362
9	2	0.081607	0.192149	0.000653	0.179262	0.894319
2	3	0.141867	0.181301	0.001058	0.257722	0.687259
2	14	0.079633	0.21864	0.00382	0.319168	0.699283
2	14	0.108887	0.222799	0.00382	0.212509	1.06867
2	4	0.290454	0.392972	0.000654	0.45775	1.70807
4	5	0.249381	0.311469	0.000804	0.404292	1.41495
5	7	0.206661	0.243562	0.000294	0.308075	1.07704
5	13	0.09273	0.126109	0.000279	0.154289	0.540876
13	8	0.123486	0.133079	0.00011	0.170987	0.503904
13	12	0.129383	0.139371	0.000116	0.179152	0.527967
12	2	0.151812	0.163516	0.00029	0.216645	0.616247
12	11	0.232062	0.59124	0.000492	0.52087	1.63683
11	2	0.23707	0.481094	0.001697	0.492347	2.31225

B.3.2 Transformer Information

All transformers are Δ-Y connected (primary to secondary).

From Bus	To Bus	Type	Voltage	$r^{(1)}$	$x^{(1)}$	$r^{(0)}$	$x^{(0)}$
14	15	Δ-Y	33kV:11kV	0.0442	0.995	10	20.6
14	15	Δ-Y	33kV:11kV	0.0449	0.9917	10	20.6
4	16	Δ-Y	33kV:11kV	0.1558	1.64	0	0
4	16	Δ-Y	33kV:11kV	0.1588	1.612	Inf	0
7	17	Δ-Y	33kV:11kV	0.09976	1.11338	0	11.4049
5	18	Δ-Y	33kV:11kV	0.09351	1.06667	Inf	0
8	19	Δ-Y	33kV:11kV	0.07192	1.08205	Inf	0
3	20	Δ-Y	33kV:11kV	0.08	1.324	16.46	1.1916
6	21	Δ-Y	33kV:11kV	0.27778	2.19667	0	1.97667
10	22	Δ-Y	33kV:11kV	0.08996	1.06667	21	0.90667
11	23	Δ-Y	33kV:11kV	0.04433	1.09115	Inf	0
11	23	Δ-Y	33kV:11kV	0.04066	1.1039	Inf	0
2	24	Δ-Y	33kV:11kV	0.10009	1.21102	Inf	0

B.3.3 Types of Customers

Table 12.3 shows the numbers and types of customers in the 24 bus network as classified by personnel at the UK distribution network.

TABLE 12.3. THE NUMBERS AND TYPES OF CUSTOMERS IN THE 24 BUS NETWORK.

Bus Number	Number of Customers				Percentage of Customers		
	Commercial	Domestic	Industrial	Total	Commercial	Domestic	Industrial
15	1326	7200	124	8650	15%	83%	1%
16	379	1365	26	1770	21%	77%	1%
17	381	2776	11	3168	12%	88%	1%
18	255	1156	12	1423	18%	81%	1%
19	418	1938	59	2415	17%	80%	2%
20	290	1172	18	1480	20%	79%	1%
21	Not available	Not available	Not available	Not available	Not available	Not available	Not available
22	292	1489	36	1817	16%	82%	2%
23	Not available	Not available	Not available	Not available	Not available	Not available	Not available
24	277	2210	21	2508	11%	88%	1%
Mean	**452**	**2413**	**38**	**2904**	**16%**	**83%**	**1%**

Data for buses 21 and 23 was not supplied by personnel at the UK distribution network.

Appendix C: Data for Covariance Model

The three phase loading information can be found in a file attached to this thesis called "**Data for Covariance Model.xlsx**".

Appendix D: Load Profiles Used in this Thesis

Load profile loading information from the UK Energy Research Centre.

- Profile class 1: Domestic Unrestricted (single rate)
- Profile class 8: Non-Domestic Maximum Demand Customers with Load Factor >40%

All profiles are taken from a weekday during the winter.

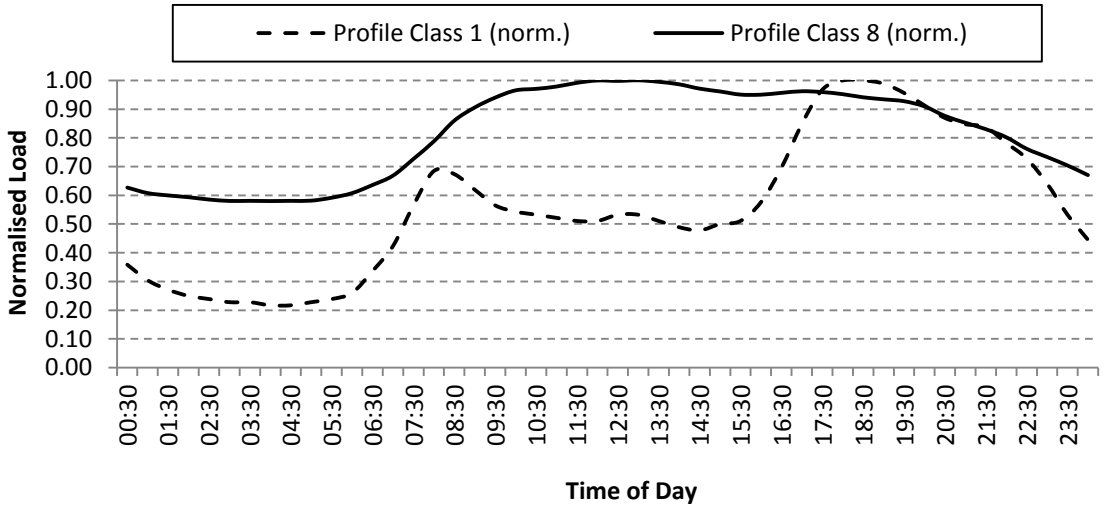


Fig. 12.2. A load profile over a 24 hour period on a typical winter day.

TABLE 12.4. PROFILE CLASS 1 AND PROFILE CLASS 8 EXTRACTED FROM THE UK ENERGY RESEARCH CENTRE.

Time	Profile Class 1	Profile Class 1 (norm.)	Profile Class 8	Profile Class 8 (norm.)
00:30	0.33	0.36	17.95	0.63
01:00	0.28	0.30	17.4	0.61
01:30	0.25	0.27	17.17	0.60
02:00	0.23	0.25	16.99	0.59
02:30	0.22	0.24	16.76	0.59
03:00	0.21	0.23	16.62	0.58
03:30	0.21	0.23	16.63	0.58
04:00	0.2	0.22	16.61	0.58
04:30	0.2	0.22	16.63	0.58
05:00	0.21	0.23	16.65	0.58
05:30	0.22	0.24	16.94	0.59
06:00	0.24	0.26	17.4	0.61
06:30	0.31	0.34	18.21	0.64
07:00	0.39	0.42	19.14	0.67
07:30	0.52	0.57	20.8	0.73
08:00	0.63	0.68	22.55	0.79
08:30	0.62	0.67	24.61	0.86
09:00	0.57	0.62	25.92	0.91
09:30	0.52	0.57	26.9	0.94
10:00	0.5	0.54	27.62	0.97
10:30	0.49	0.53	27.77	0.97
11:00	0.48	0.52	28	0.98
11:30	0.47	0.51	28.37	0.99
12:00	0.47	0.51	28.6	1.00
12:30	0.49	0.53	28.56	1.00
13:00	0.49	0.53	28.62	1.00
13:30	0.47	0.51	28.47	0.99
14:00	0.45	0.49	28.22	0.99
14:30	0.44	0.48	27.79	0.97
15:00	0.46	0.50	27.51	0.96
15:30	0.47	0.51	27.2	0.95
16:00	0.53	0.58	27.19	0.95
16:30	0.64	0.70	27.37	0.96
17:00	0.78	0.85	27.53	0.96
17:30	0.89	0.97	27.45	0.96
18:00	0.92	1.00	27.24	0.95
18:30	0.92	1.00	26.93	0.94
19:00	0.91	0.99	26.73	0.93
19:30	0.88	0.96	26.54	0.93
20:00	0.84	0.91	26.03	0.91
20:30	0.8	0.87	25.08	0.88
21:00	0.78	0.85	24.4	0.85
21:30	0.77	0.84	23.76	0.83
22:00	0.72	0.78	22.96	0.80
22:30	0.67	0.73	21.81	0.76
23:00	0.59	0.64	20.99	0.73
23:30	0.49	0.53	20.15	0.70
00:00	0.41	0.45	19.19	0.67

Appendix E: Customers Attached to the 295 Bus Network

Table 12.5 shows the numbers of customers assumed to be attached to each busbar in the 295 bus network. These customer numbers were obtained by assuming that 1 industrial customer was attached to buses 136, 225 and 232. All of the other loads were assumed to comprise 80% residential customers and 20% commercial customers, divided in proportion to the total load. Each industrial and commercial customer was assumed to consume and average of 1kW. These numbers were based on the analysis of the 24 bus distribution network shown in Appendix B. Only loaded busbars are shown in Table 12.5.

Full loading information on the 295 generic distribution network can be found in [129].

TABLE 12.5. THE NUMBER OF CUSTOMERS OF EACH TYPE CONNECTED TO FEEDER L IN THE 295 BUS NETWORK

Busbar	Residential	Commercial	Industrial
152	728	182	0
150	1608	402	0
148	1512	378	0
145	232	58	0
144	232	58	0
142	1704	426	0
140	767	191	0
132	2824	706	0
133	1440	360	0
134	1360	340	0
137	0	0	1
136	752	188	0
225	0	0	1
218	6312	1578	0
220	4952	1238	0
173	8928	2232	0
175	2776	694	0
177	2848	712	0
214	2560	640	0
193	1440	360	0
197	1208	302	0
195	680	170	0
196	680	170	0
199	1280	320	0
200	1360	340	0
201	1807	451	0
203	1584	396	0
204	600	150	0
208	600	150	0
209	3536	884	0
190	376	94	0
191	376	94	0
192	376	94	0
168	1600	400	0
169	3120	780	0
182	680	170	0
184	912	228	0
186	1520	380	0
187	3184	796	0
165	608	152	0
162	2760	690	0
163	10496	2624	0

164	383	95	0
160	1464	366	0
158	2784	696	0
159	1624	406	0
156	1016	254	0
154	1648	412	0
130	3448	862	0
232	0	0	1

Appendix F: An Immune System Inspired Clustering and Classification Method to Detect Critical Areas in Electrical Power Networks

F.1 Introduction

Regulatory pressures and market conditions are causing power system networks to operate closer to their stability limits. In such a situation, unforeseen scenarios may arise which endanger a network's integrity and lead to voltage instability and / or violation of load carrying limits of transmission lines. Increased monitoring of networks and the ability to predict potentially fatal failures allows for better utilisation of assets within the network.

Power system operators must, through a set of available control actions, maintain all network parameters within acceptable limits by continuously monitoring critical parameters. The proximity of the network to voltage collapse and the loading of overloaded transmission lines are two crucial phenomena that must be monitored in order to avoid blackouts in the system. However, monitoring these phenomena is not a simple procedure as it requires substantial financial and manpower commitments in addition to the technical complexity of the task. The routinely collected parameters which describe the state of the network must be collected and processed to ensure that the network is not dangerously close to its stability or load-ability limit.

Artificial Immune Systems (AIS) and other artificial intelligence (AI) techniques can successfully identify critical areas of the power network by virtue of developing an approximate model that takes account seemingly uncorrelated data and events and as such, enables the avoidance of catastrophic failures that might lead to blackouts.

The idea of using classification and clustering algorithms to identify weak buses in the power system network has been explored before. Research in both [176] and [177] uses a Kohonen self organizing map (SOM) to identify weak voltage buses within the power system network. The results in [177] were compared against those obtained from singular value decomposition of the load flow Jacobian. In [178] weak voltage buses were grouped together using a fuzzy clustering algorithm. In a similar way to these techniques, voltage weak areas and overloaded lines can be grouped together using an immune inspired AIS clustering algorithm. AIS algorithms have been used to solve data clustering problems in the past [91, 99], and [91] demonstrated that for some data sets an AIS algorithm can outperform other clustering algorithms.

AIS algorithms can also be used to deal with labelled data as classifiers [96, 136]. Classification algorithms work by building a model capable of labelling a set of previously unseen testing data. They are ideally suited to identifying critical areas of the power system network. A popular AIS algorithm is the Artificial Immune Recognition System (AIRS) [96]. AIRS uses the classification method from the k nearest neighbours algorithm [143] (kNN) and immunological inspiration to arrive at a classification result. Another popular machine learning classification algorithm is a Support Vector Machine (SVM), a supervised learning algorithm [140].

The research in this appendix describes a novel AIS algorithm based methodology, to identify voltage collapse prone buses and overloaded lines in the power system network. The output of the proposed methodology produces a reliable prediction of voltage weak areas and overloaded lines within the power system network.

The aims of the research in this appendix are therefore:

1. *Demonstrate and experimentally validate a methodology to assist with the quick identification of voltage stability issues and overloaded lines in a generic power system network.*
2. *Show experimentally how AIS and other classification algorithms compare for voltage stability and overloaded line classification.*

The first aim is of specific interest to both applied AIS researchers and power system researchers. Addressing the second aim attempts to enhance the understanding of the merits of applied AI / AIS research in the power systems domain.

The research in this appendix offers a comprehensive insight into the applicability of these methods to two challenging power systems problems and clarifies the extent to which the classification methods are actually producing statistically or scientifically significant results. The experimental comparison presented in this research, in particular the comparison of AIRS (and kNN) and kNN alone, also provide further insight into the types of problems where AIRS and kNN can produce different levels of classification performance.

F.2 Classification & Clustering Algorithms

The methodology developed in this appendix uses the AIRS classification algorithm, an SVM classification algorithm, a kNN classification algorithm and an unsupervised artificial immune classifier (UAIC) clustering algorithm. Full details on the inner mechanisms involved in these algorithms can be found in Chapter 3, and they are not brevity. Chapter 3 also includes a review of some of the similarities between each of the algorithms.

F.3 Problem Definition

F.3.1 Monitoring Voltage Stability & Overloaded Lines with Simulated Data

The first aim of this research is to use both clustering and classification to assist with quick identification of voltage stability issues and overloaded lines in a power system network. In this study, three realistic power system networks were considered, a 30 bus network[179], 14 bus network[180] and a 9 bus network[181]. The 30 bus network is shown in Figure 1.

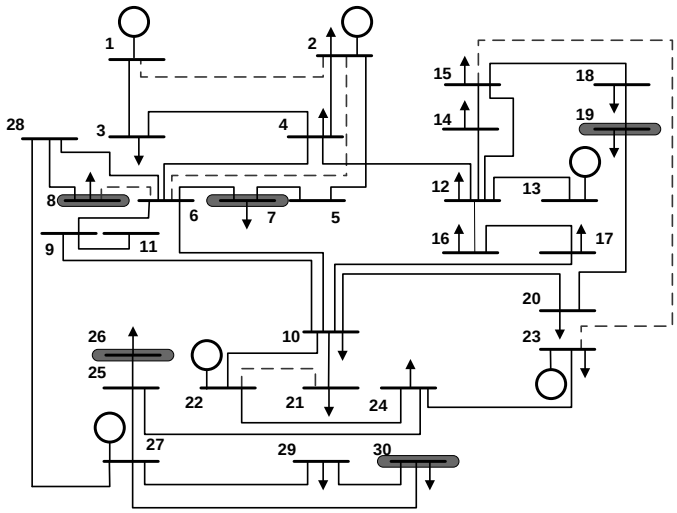


Figure 1 - 30 bus test network(adapted from [21]). The identified 5 weakest voltage buses are enclosed in shaded rounded rectangles and the 5 most commonly overloaded lines are highlighted using thick grey dashed lines.

The power system is loaded according to realistic patterns of electrical demand. This can be achieved by simulating a wide variety of loading scenarios on a power network[18][182].

For this study, sample data was generated by randomly raising the real and reactive power level across the whole network by up to 100% above the base system load. The base system load on the network does not cause any of the lines to be overloaded, nor any bus voltage to be below the statutory limit of $\mp 5\%$. Each load in the network was also randomly varied by up to ±20% of the original base load at that particular bus. This is mathematically shown by equation (1).

$$\boldsymbol{l} = \boldsymbol{l_b} + p\boldsymbol{l_b} + \boldsymbol{l_b}.\boldsymbol{y} \qquad (1)$$

where:

$\boldsymbol{l}$	System load at given operating conditions
$\boldsymbol{l_b}$	Base system load
p	Uniformly distributed random variable between 0 and 1
$\boldsymbol{y}$	Row vector of uniformly distributed random numbers between [-0.2, 0.2]
.	Element by element multiplication of two vectors

F.3.2 Homotopy Continuation Load Flow

Voltage stability is an increasingly common issue in electrical power systems. In simple terms, it is caused when the power demand of the network cannot be met by the generation capability connected to the network. Each of the buses (or nodes) in a power system network has a different loading limit after which the network voltages collapse.

Understanding which of the buses in the network will fail first is of important consideration to power system operators. This information gives system operators insight into which areas of the network are more vulnerable than others.

A standard approach to establish the proximity to voltage collapse in a power system is to monitor the system load flow Jacobian matrix [183]. However, computing the voltage stability limit of a power system using the standard Newton-Raphson method (typically used to solve the load flow problem) causes divergence of the algorithm. Several methods have been discussed to establish the proximity of an individual bus voltage to voltage collapse, including regulating the step length on the Newton-Raphson iterations [184], performing sensitivity analysis on the converged Jacobian matrix [183] or using the homotopy continuation [185] method to trace the power voltage relationship at each bus.

The voltage stability of a power system is assessed in this study by using homotopy continuation load flow analysis (HCLF)[185, 186]. HCLF analysis enables each of the buses (nodes) in the power system network to be assessed for their proximity to voltage collapse.

The results of the (HCLF) analysis will be different depending on the loading conditions of the network. Therefore, simulated data (as described in section 3.1) is required to determine each bus' proximity to voltage collapse.

The (HCLF) method can be described with reference to the standard Newton-Raphson load flow algorithm[63]. A homotopy parameter λ is added to the standard Newton-Raphson load flow to represent load growth.

$$L_s(\lambda) = L_{s0} + \lambda L_d \qquad (2)$$

where $L_s(\lambda)$ is the load / generation corresponding to an increase in lambda, L_{s0} is the load / generation at current operating point ($\lambda = 0$), L_d is the loading and generation

[18] The simulated results were generated using MATPOWER.

that will lead to voltage collapse and λ is the homotopy parameter or the load increase factor (load-ability).

These equations can be combined with the standard load flow equations to form a homotopy function.

$$F(\boldsymbol{x}, \lambda) = L(\boldsymbol{x}) - L_s(\lambda) \tag{3}$$

where $\boldsymbol{x}$ is a vector of voltage magnitudes and angles, $L(\boldsymbol{x})$ is the standard load flow equation for voltage magnitudes and angles $\boldsymbol{x}$ and $F(\boldsymbol{x}, \lambda)$ is the homotopy function.

The homotopy equation can be shown to be equivalent to solving a differential equation [185, 186] in the following form.

$$\boldsymbol{F_x}(\boldsymbol{x}, \lambda)\frac{d\boldsymbol{x}}{d\lambda} = -\boldsymbol{F_\lambda}(\boldsymbol{x}, \lambda) \tag{4}$$

where $\boldsymbol{F_x}$ and $\boldsymbol{F_\lambda}$ are the Jacobians of $F(\boldsymbol{x}, \lambda)$ with respect to $\boldsymbol{x}$ and λ.

The singularity of $\boldsymbol{F_x}$ causes problems at the maximum power point as the standard load flow Jacobian ($\boldsymbol{F_x}$) becomes singular. To alleviate this problem, the load flow Jacobian can be augmented with an extra row and column.

$$\boldsymbol{J_{aug}} = \begin{bmatrix} \boldsymbol{F_x} & \boldsymbol{F_\lambda} \\ \boldsymbol{e_k} & \end{bmatrix} \tag{5}$$

where $\boldsymbol{e_k}$ is a vector of modulus 1 with its *k*th element equal to unity. By carefully selecting k, $\boldsymbol{J_{aug}}$ becomes non singular. To determine the maximum power point on the power voltage curve the load flow equations are modified. The *i*th bus in the system is tested by amending the real and reactive power demand at this bus using:

$$P_{Li}(\lambda) = P_{Li0}[1 + \lambda] \tag{6}$$

$$Q_{Li}(\lambda) = P_{Li0}\tan(\emptyset_i)\,[1 + \lambda] \tag{7}$$

where $\boldsymbol{P_{Li0}}, \boldsymbol{Q_{Li0}}$ are the original active and reactive loads at bus *i, and* $\emptyset_{\boldsymbol{i}}$ is the power factor angle at bus *i*.

The elements described above can be combined together to form a HCLF algorithm that will estimate the load-ability (λ) of each bus in the power network. The iterations of the continuation algorithm use a predictor corrector process.

Firstly, a tangent vector ($\boldsymbol{T}$) is calculated by solving:

$$\begin{bmatrix} \boldsymbol{F_x} & \boldsymbol{F_\lambda} \\ \boldsymbol{e_k} & \end{bmatrix}\boldsymbol{T} = \boldsymbol{e_k}^{\boldsymbol{T}} \tag{8}$$

A predicted solution is estimated using the computed tangent vector and a stepping size σ.

$$\begin{bmatrix} \boldsymbol{x'} \\ \lambda' \end{bmatrix} = \begin{bmatrix} \boldsymbol{x} \\ \lambda \end{bmatrix} + \sigma\boldsymbol{T} \tag{9}$$

Where $\boldsymbol{x'}$ and λ' represent predicted solutions. These solutions are corrected using a modified Newton-Raphson load flow method where the equations have one additional state variable, namely λ.

During the algorithm, the step length σ is rigorously controlled. It is allowed to grow at a rate governed by equation (10), where i is the number of iterations taken by the last

correction of the Newton-Raphson method. If divergence of the Newton-Raphson method is detected, then the step length is reduced using equation (11).

$$\sigma_{new} = 6\sigma/i \tag{10}$$

$$\sigma_{new} = \sigma/3 \tag{11}$$

The continuation parameter $\boldsymbol{e_k}$ is also closely monitored. It is changed to the parameter that yields the largest tangent vector ($\boldsymbol{T}$) component. If ill-conditioning of the augmented Jacobian matrix is detected, the continuation parameter is also changed.

The figure below shows predicted and corrected solutions when the continuation load flow method is applied to the 30 bus test network to measure the load-ability of bus 13. The continuation method is shown tracing the entire power-voltage (PV) characteristic which is not possible with a conventional load flow algorithm.

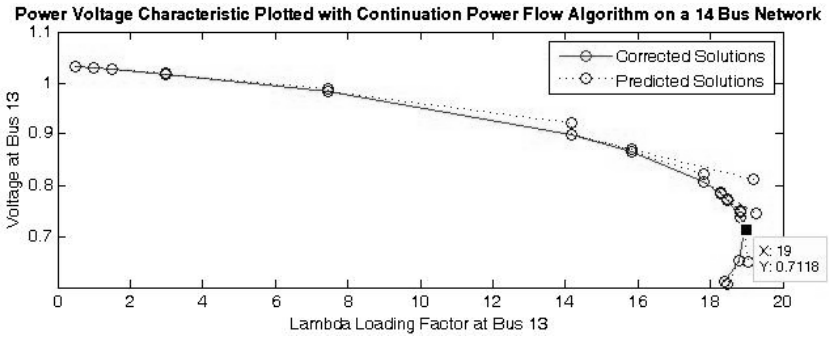


Figure 2 - A trace of a power voltage characteristic curve with the maximum value of lambda highlighted on the graph. This indicates the load-ability of that bus.

Once each bus' load-ability limit (λ) has been calculated for a given operating point, each bus in the network can be ranked. The bus with the lowest value of λ is ranked as the weakest bus. The weakest bus can also be described as the bus which can take the smallest amount of additional power before the voltage at that bus and ultimately the whole system collapses.

F.3.3 Overloaded Lines

In power systems, overloaded lines are defined as those lines where the apparent power rating (usually recorded in MVA) of the line has been exceeded. The apparent power rating of a power line is defined as the maximum amount of apparent power that can flow through the line. The current flowing in the line is proportional to the apparent power flowing in the line.

To test a network for overloaded lines, each operating point must be inspected to find any lines in the network that are overloaded. High levels of current flowing in power lines is the main cause of overheating due to the I^2R losses that the current generates. For this reason, the apparent power must be closely monitored to ensure that no lines in the network are approaching their thermal limits.

F.3.4 Summarised Problem Definition

Analysing the voltage stability using the HCLF method and overloaded lines using apparent power analysis allows a network's vulnerabilities to these phenomena to be identified. To keep the study realistic, a large number of possible operating conditions should be utilised that summarise various typical loadings levels in the network. The key aim of this assessment, is to identify which areas in the power network are most likely to fail, and in what order.

F.4 Methodology

The methodology developed in this research allows the vulnerability of the power system network to be assessed. The aim is not only to present the methodology but also to show experimentally how a variety of AIS and machine learning classification algorithms compare. The fulfilment of both of these objectives should create insight into the applicability of AIS classification algorithms in the power systems domain.

The methodology developed here can be summarised into three key procedures:

1) ***Data Generation***
   Simulated data must be generated for each of the 3 power networks. The simulated databases must represent a diverse and realistic set of operating conditions for the network.
2) ***Clustering***
   Clustering is performed on the simulated data by extracting voltage stability and overloaded line information from the dataset. The dataset is clustered for overloaded lines and voltage weak buses independently, as the two phenomena are quite different. The output of the clustering procedure is two sets of labels for the data set: one for voltage weak buses and one for overloaded lines.
3) ***Classification***
   Classification consists of a two step procedure. Firstly, the chosen classification algorithm is optimised by selecting the configuration parameters for the algorithm which yield the best performance on the data set. Secondly, the optimised classification algorithm is applied to the dataset. The trained classification model should then be capable of placing previously unseen operating conditions into a group which represents its failure mode for both voltage weak buses and overloaded lines. The classification method is applied to both overloaded lines and voltage weak buses independently. The classification algorithms were compared using statistical analysis.
   An overview of the experimental method employed for this study is shown in Figure 3.

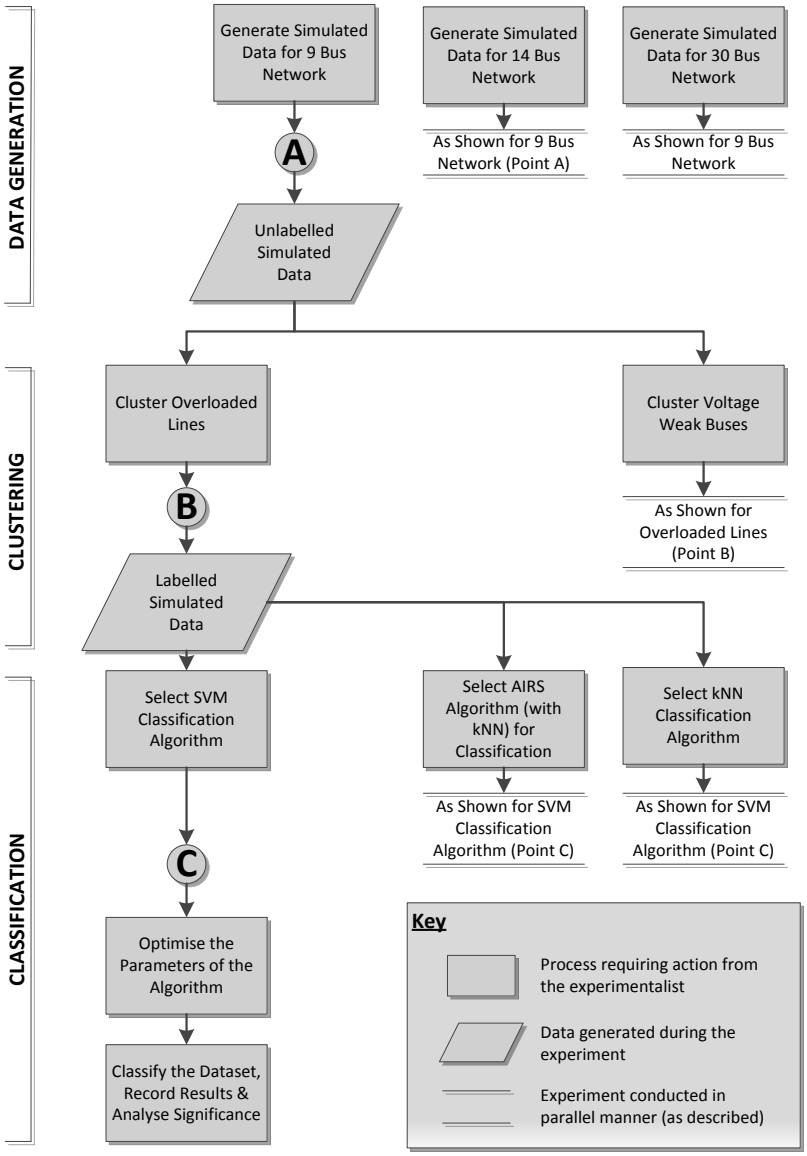


Figure 3 – The salient points of the methodology employed in this study.

F.4.1 Data Generation

2500 operating points were generated for each of test networks in this study. A further set of 100 sets of 100 operating points were generated on the 30 bus, 14 bus and 9 bus networks and a further 500 sets of 100 operating points generated on the 9 bus network.

The large (2500 operating point) data sets are required to test the overall classification accuracy of the developed methodology. The smaller sets (100 sets of 100 operating points and 500 sets of 100 operating points) are used to compare statistical classification performance of the algorithms and to optimise model parameters.

The database generation was carried out as described in section 3.1.

F.4.2 Clustering

The aim of the clustering procedure is to add two class labels to each of the simulated operating points within a dataset. The clustering procedure achieves this by using an unsupervised strategy to group together similar operating conditions in terms of both voltage stability and overloaded lines. In this way, the clustering algorithm is run over a dataset twice: firstly to group together similar operating points in terms of voltage stability and secondly to group together similar operating points in terms of overloaded

lines. The voltage stability and overloaded line groupings that are formed will become class labels for the classification procedure used in the subsequent part of this methodology.

Operating points with similar voltage stability conditions and overloaded line configurations were grouped together using an unsupervised artificial immune classifier (UAIC)[91].

Before the data can be presented to the UAIC, the simulated data must be pre-processed to extract information into feature vectors. For each clustering run (i.e. for voltage weak buses and overloaded lines), a feature vector of information is extracted for every simulated operating point. Each feature vector contains only information necessary for each of the two clustering runs: voltage weak buses or overloaded lines.

In the UAIC, each antigen that is presented to the algorithm, represents a feature vector of extracted information from a simulated operating point. Antibodies and memory cells represent the prospective and evolved cluster centres respectively. Table 1 and Table 2 describe the mapping between the components in the AIS model and their power system equivalents.

TABLE 1 - AIS AND POWER SYSTEM MAPPING FOR VOLTAGE STABILITY CLUSTERING

AIS	**Power System**
Antigen	Feature vector of bus ranks (extracted from the simulated operating point information)
Antibody	Prospective cluster centre
Memory	Evolved cluster centres

TABLE 2 - AIS AND POWER SYSTEM MAPPING FOR OVERLOADED LINE CLUSTERING

AIS	**Power System**
Antigen	Feature vector of overloaded lines (extracted from the simulated operating point information)
Antibody	Prospective cluster centre
Memory	Evolved cluster centres

The UAIC clustering algorithm was configured to find 4 clusters ($N_c = 4$) on both the run of the algorithm for voltage weak buses and the run for overloaded lines. The UAIC algorithm was also configured with a $affinity\ threshold\ scalar$ of 0.1 and a termination switching cluster threshold of 5%, clonal rate (C_r) of 20 and a displace rate (D) of 0.1 (10%). These configuration parameters were determined through a series of tests on the algorithm whilst clustering the overloaded line dataset. They were selected because to ensure each run of the algorithm converged in a reasonable number of iterations and that the clusters found were consistent from one run to the next.

F.4.2.1 Clustering Weak Voltage Buses

A feature vector that contains the rank of each bus (of a simulated operating point) is used as the antigen input for the voltage weak bus run of the UAIC algorithm. The rank of a bus corresponds to an order representing the voltage weakness of each of the buses in the network. The antigen representation is shown in Figure 4.

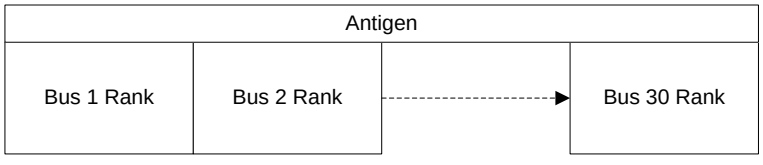


Figure 4 - Antigen structure for clustering weak voltage buses

The similarity between antigens is determined by applying a bespoke distance measure developed to compare similarly ranked operating points. The distance measure defines the number of upwards or downwards movements that each bus would have to make to map one antigen's rank to another's. For example, the antigen (vector) $\boldsymbol{a} = [1\ 2\ 3\ 4]$ of a hypothetical 4 bus system is a ranking distance of 6 away from $\boldsymbol{b} = [4\ 1\ 2\ 3]$. Bus 4 moves up 3, bus 1, 2 & 3 all move down 1, giving a total distance of 6.

F.4.2.2 Clustering Overloaded Lines

A feature vector that contains the network's overloaded lines is used as the antigen input for the second run of the UAIC algorithm. The antigen representation is shown in Figure 5.

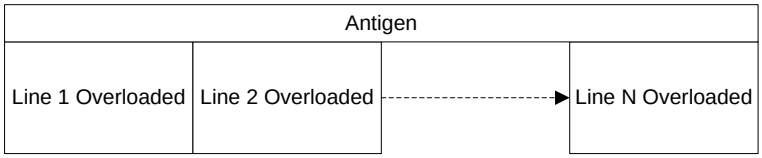


Figure 5 - Antigen for overloaded line clustering

Similarity between antigens was determined using the Manhattan distance measure, defined as[187]:

$$distance = Manhattan(x, y) = \frac{1}{N}\sum_{i=1}^{N}(|x_i - y_i|) \qquad (12)$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are points in $\mathbb{R}^N$.

F.4.2.3 Output of the Clustering Algorithm

Running the UAIC algorithm on the dataset for voltage weak buses and overloaded lines groups the dataset into 2 separate sets of 4 clusters. These 2 sets of 4 clusters are used to label the dataset twice: once for voltage weak buses and once for overloaded lines. These two labels are used by the classification algorithms to perform supervised classification.

F.4.2.4 Validity of the Clustering Algorithm

The UAIC algorithm was selected for this study because of its promising results presented in [91]. Reference [91] also showed that the UAIC algorithm performed better than the simpler k-means algorithm for the chosen clustering task. The clustering problem described in this appendix is similarly complex and will benefit from the added power of the UAIC algorithm.

It should be noted that any clustering algorithm could be used at this stage of the presented methodology. However, it should not be assumed that the UAIC algorithm will provide better levels of clustering performance when compared with other (perhaps simpler) methods such as k-means or a self organizing map (SOM).

A rigorous comparison of the quality of the UAIC algorithm against other clustering algorithms is outside the scope of this appendix. However, the validity of the UAIC algorithm can be explored for this methodology by comparing the UAIC against the k-means algorithm on the overloaded line dataset. The comparison was achieved by configuring k-means to find 4 clusters on the 9, 14 and 30 bus 2500 operating point

overloaded line datasets. The k-means clusters were then compared against the clusters found by using the UAIC algorithm.

The algorithms were compared by using Cohen's kappa statistic[188], described as follows:

$$\kappa = \frac{P_r(a) - P_r(e)}{1 - P_r(e)} \quad (13)$$

where $P_r(a)$ the agreement among algorithms is, $P_r(e)$ is the hypothetical probability of a chance agreement using the data. κ will equal 1 when there is perfect agreement between the two algorithms and <0 when there is no agreement.

It is anticipated that if the UAIC algorithm conducts clustering operation in a similar manner to k-means, the kappa statistic should be high across all three datasets. Clustering techniques can only be compared in this way when it is predicted that the dataset will be split into the same set of clusters using both algorithms[189]. For this reason, the overloaded line dataset is more appropriate than the voltage weak bus dataset, as this dataset is simpler and more likely to split into obvious groups. The comparison of Cohen's kappa statistic across the 9, 14 and 30 bus overloaded line datasets is shown in Table 3.

TABLE 3 – THE COMPARISON OF KAPPA STATISTIC FOR THE UAIC AND K-MEANS CLUSTERING ALGORITHM FOR THE 2500 POINT OVERLOADED LINE DATASETS

Dataset	**Kappa Statistic (κ) of UAIC vs. k-means**
9 Bus 2500 Operating Point	1
14 Bus 2500 Operating Point	1
30 Bus 2500 Operating Point	0.644

The kappa statistics of 1 for both the 14 and 9 bus datasets indicate that k-means and UAIC are partitioning the dataset in exactly the same manner. A kappa statistic of 0.644 for the 30 bus network shows that UAIC and k-means have good agreement in clusters that are formed.

The purpose of this validity check is to show that the UAIC algorithm is capable of clustering simple datasets in a manner similar to k-means. The validity check shows UAIC to be a viable clustering method, and a suitable choice for clustering overloaded areas of a power system network.

F.4.3 Classifying Critical Areas of the Network

The performance of the three selected classification algorithms (AIRS, kNN and SVM) were compared in this study. One of the aims of this research is to discover which algorithm yields the highest classification accuracy for the chosen problem. This was evaluated by comparing the differences in classification performance between the three algorithms and checking for statistical [156] and scientific significance [157].

F.4.3.1 Inputs for Classification Algorithm

The inputs for the classification of voltage weak buses and overloaded lines were a summary of network parameters representing the current load on the network. The parameters used were as follows:

- Real power demand at each bus in the network
- Reactive power demand at each bus in the network
- Real power generation at each bus (node) in the network
- Reactive power generation at each bus (node) in the network
- Total real power losses in the network
- Total reactive power losses in the network

On the 30 bus system a feature vector used for input into the classifier has 40 non-zero attributes for real and reactive power demand, 12 non-zero attributes for real and reactive power generation and 2 attributes for real and reactive total power losses. Each feature vector therefore has 54 attributes, excluding the class label.

The class label attached to each of the feature vectors is either one of the 4 clusters created by clustering voltage weak buses, or one of the 4 clusters created by clustering overloaded lines. Two runs of the selected classification algorithm are performed: one for voltage weak buses and one for overloaded lines.

F.4.3.2 Optimising the Parameters of the Classification Algorithms

F.4.3.2.1 Optimising the Parameters of the SVM Algorithm

The parameters C and σ featuring in the SVM algorithm were optimised using cross validation and a grid search algorithm. The 10 fold cross-validated classification accuracy of the SVM algorithm was tested using the smaller 100 operating point databases across all sizes of network with labelling for overloaded line (OL) detection and voltage collapse (VC) limit detection. The smaller 100 operating point databases are used rather than the large 2500 operating point database because performing a grid search on such a large database would take a prohibitively long length of time.

Different pairs of values of C and σ are chosen until the entire parameter range has been searched. C was chosen to range from 2^{-5} to 2^{15} whilst σ was chosen to range from 2^{-15} to 2^{3}. The step interval for the exponent of the parameters was set to 0.1. Selecting the parameters in the ranges describes above describes a grid of parameters where the SVM algorithm can be tuned[155] for the specific classification problem.

All the tests were run using LibSVM [152] and all features contained within the data were normalised in the region of [-1, 1]. The results of the parameter optimisation procedure are shown in the Table 4.

TABLE 4 – OPTIMISED PARAMETER VALUES FOR THE SVM CLASSIFICATION ALGORITHM.

Size of Network and Labelling Scheme	C	σ
9 Bus Load-ability	32768	0.0004883
9 Bus Overloaded Lines	128	0.5
14 Bus Load-ability	32768	0.0019531
14 Bus Overloaded Lines	32768	0.0004883
30 Bus Load-ability	32768	0.0004883
30 Bus Overloaded Lines	2048	0.0078125

F.4.3.2.2 Optimising the Parameters of the AIRS Algorithm

There are several user configurable parameters available for AIRS. These include:

- k value for the k nearest neighbour algorithm (k)
- Clonal rate (CR)
- Hyper-mutation rate (HR)
- Total resources (TR)
- Stimulation threshold (ST)

The AIRS algorithm was tested using 10 fold cross validation on the 100 operating point databases across all sizes of networks and labelled for both overload line detection and voltage weak bus detection. The optimum value of k for the kNN classifier was determined by using the default algorithm parameters and varying the value of k until an optimum was reached. Each subsequent parameter was then taken in turn and tuned (within the specified range) until the best overall selected classification accuracy was

obtained. This method does not guarantee finding a globally optimal solution for the algorithm's parameters, but ensured that locally optimised parameters were discovered.

The ranges for the optimised parameters are shown in Table 5:

TABLE 5 - PARAMETER OPTIMISATION RANGES FOR THE AIRS ALGORITHM.

Parameter Name	**Min**	**Max**	**Step**
k Nearest Neighbour	1	30	1
Clonal Rate	5	30	1
Hyper-mutation Rate	1	3	0.5
Total Resources	100	500	10
Stimulation Threshold	0.7	0.98	0.1

TABLE 6 – PARAMETER OPTIMISATION RANGES FOR THE AIRS AND KNN ALGORITHMS.

Network & Labelling Scheme	**k (AIRS)**	**K (kNN)**	**CR**	**HR**	**TR**	**ST**
9 Bus Load-ability	5	5	5	1	250	0.98
9 Bus Overloaded Lines	6	6	10	2	400	0.9
14 Bus Load-ability	25	14	10	2	100	0.92
14 Bus Overloaded Lines	6	12	20	3	100	0.9
30 Bus Load-ability	15	18	10	2	350	0.98
30 Bus Overloaded Lines	20	12	10	1.5	300	0.96

F.4.3.2.3 Optimising the Parameter of the k Nearest Neighbours Algorithm

The only variable parameter in the k nearest neighbour classifier is the value of k. Different values of k were tested in the range from 1 to 30 using 10 fold cross validation on the 100 operating point datasets across all sizes of networks and labelled for both overloaded line detection and load-ability limit detection. 30 was the maximum tested value of k, since this represents a large percentage (30%) of the dataset being used for the majority voting classification procedure. The optimum value of the k parameter was recorded as the value of k which gave the highest classification accuracy. This value was selected and used for all further algorithm tests. The optimal values for k in the kNN algorithm are shown in Table 7.

F.4.3.3 Comparing the Classification Algorithms

The comparative performance of the 3 selected classification algorithms was analysed by attempting to reject a series of null hypotheses. The null hypotheses were formulated by creating 9 experiments that compare each of the three chosen algorithms against one another. Each one of the 9 null hypotheses can be stated in a generalised form as follows:

Algorithm A and Algorithm B yield the same distribution of classification performance when classifying voltage weak buses using the methodology described in this appendix on the X bus power system network.

Algorithm A and B represent two of the three algorithms under comparison; for example, AIRS & SVM, or AIRS and kNN. The X bus power network refers to one of the 9, 14 or 30 bus power system network described in section 3.1. 9 null hypotheses are generated each referring to a comparison of one of the 3 algorithms on each of the 3 power networks.

The statistical significance of the differences in classification performance was analysed by applying the Mann-Whitney-Wilcoxon non-parametric rank-sum test[156]. The aim of the test is to ascertain rejection of the null hypothesis; each of the algorithms has identical distributions and medians of classification performance. A significance level of 5% was used which ensures that we are 95% confident that the null hypothesis can be rejected. A p value of <5% shows that the null hypothesis may be rejected.

Given a large enough sample size it is often possible to show a certain level of statistical significance between any two sets of experimental results. However, whether any two sets of experimental results are different enough to be of scientific value depends on the magnitude of difference between the associated result distributions. The magnitude of the difference in two sets of experimental results is known as effect size and this can be measured to assess the scientific significance of a pair of experimental results.

The Vargha-Delany A statistic [157] is a non-parametric measure which assesses the corresponding effect size between two non-parametric distributions. The A value ranges between 0 and 1. 0.5 indicates a non-existent effect size whilst values closer to 0 or 1 indicate an increasingly large effect size. An effect size of <0.36 or >0.64 is taken as a guide from [157] to indicate a scientifically significant effect size, and hence a result with some degree of scientific value.

The statistical significance tests were run using the 100 dataset 100 operating point databases generated in the database generation phase. Only data labelled with clusters relating to voltage weak buses were tested during the statistical and scientific significance comparison.

F.5 Results

F.5.1 Data Generation

Simulations were carried out on the 30 bus network (whose single line diagram is shown in Figure 1), a 14 bus test network [180] and a 9 bus network [181]. A 2500 operating point dataset was generated for each of these three networks as described in section 3.1.

F.5.2 Clustering Using the UAIC Algorithm

The UAIC algorithm clustered each dataset once for voltage weak buses and once for overloaded lines. The UAIC algorithm was set to find 4 clusters in each dataset.

Each cluster does not necessarily contain identical rankings or overloaded lines, there may be some variation. However, each cluster represents the best division of operating points achievable using the UAIC algorithm and distance similarity measures described in the previous section.

Table 7 and Table 8 show the top 10 weakest buses and most prominent overloaded lines respectively for the 30 bus network with the UAIC finding 4 clusters in the 2500 point dataset. The different ranks and lines produced in each of the 4 clusters are the most commonly occurring ranks or lines within that cluster.

TABLE 7 - VARIOUS OPERATING POINTS OF THE 30 BUS POWER NETWORK GROUPED INTO CLUSTERS DEPENDING ON THE RANKING OF (VOLTAGE WEAK) BUSES WITHIN THE NETWORK. THE TABLE ALSO SHOWS THE TOP 10 WEAKEST BUSES BY CLUSTER FOR THE 2500 OPERATING POINT DATASET.

Label / Cluster	No. in Cluster	Top 10 Weakest Buses									
		1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Cluster 1	761	30	8	26	7	19	21	14	17	12	15
Cluster 2	526	8	30	26	7	19	21	17	14	12	24
Cluster 3	610	30	8	26	7	19	17	21	24	12	14
Cluster 4	603	30	8	26	7	19	21	14	17	24	12

TABLE 8 - THE CLUSTERS FOR OVERLOADED LINES IN THE 30 BUS POWER SYSTEM NETWORK ON THE 2500 DATASET.

Label / Cluster	No. in Cluster	No. Of Overloaded Lines	Overloaded Lines
Cluster 1	1624	2 overloaded lines	Bus 6 to 8 & 21 to 22
Cluster 2	188	0 or 1 overloaded lines	Bus 6 to 8
Cluster 3	412	5 or more overloaded lines	Bus 6 to 8, 21 to 22, 1 to 2, 2 to 6, 15 to 23
Cluster 4	276	3 or 4 overloaded lines	Bus 6 to 8, 21 to 22, 15 to 23 & 1 to 2

The single line diagram in Figure 1 shows both the 5 weakest buses and the 5 most overloaded lines that were found from this study.

The 4 clusters found for voltage weak buses and overloaded lines (as shown in the label column of Table 7 and Table 8) were used to label the 2500 operating point dataset with 4 class labels (as described in section 4.2.3).

F.5.3 Classification Performance of the SVM, AIRS and kNN Algorithms

The performance of the parameter optimised SVM, AIRS and kNN classification algorithms were tested using the 2500 dataset using the classification analysis technique of 10 fold cross validation (10 FCV) [190]. In each experiment, the classification accuracy was measured. Classification accuracy is defined as the total number of correctly classified instances over the total number of instances that were attempted to be classified. In the 10 FCV experiment, 250 instances will be classified on each of 10 runs of classification algorithm. The classification accuracy of the SVM, AIRS and kNN algorithms is shown in Table 9 for the 2500 operating point dataset for both voltage weak buses and overloaded lines.

TABLE 9 - CLASSIFICATION ACCURACY COMPARISON OF THE 9 BUS, 14 BUS AND 30 BUS SUPPORT VECTOR MACHINE USING 10 FOLD CROSS VALIDATION AND THE 2500 OPERATING POINT DATASET.

Test Network	SVM	AIRS	kNN
Overloaded Lines (9 Bus Network)	99.24%	97.4%	97.64%
Ranking Voltage Weak Buses (9 Bus Network)	97.28%	75.32%	79.36%
Overloaded Lines (14 Bus Network)	99.40%	96.16%	96.44%
Ranking Voltage Weak Buses (14 Bus Network)	95.88%	79.12%	79.84%
Overloaded Lines (30 bus network)	96.60%	90.06%	92.56%
Ranking Voltage Weak Buses (30 bus network)	91.56%	48.88%	48.04%

The results in Table 9 show the mean percentage classification accuracy of a 10 FCV experiment. For each of the 10 splits of the dataset in 10 FCV, a different percentage classification result will be obtained. The distribution of the classification accuracy for each of the 10 runs, across all algorithms on both datasets is shown in Figure 6 and Figure 7.

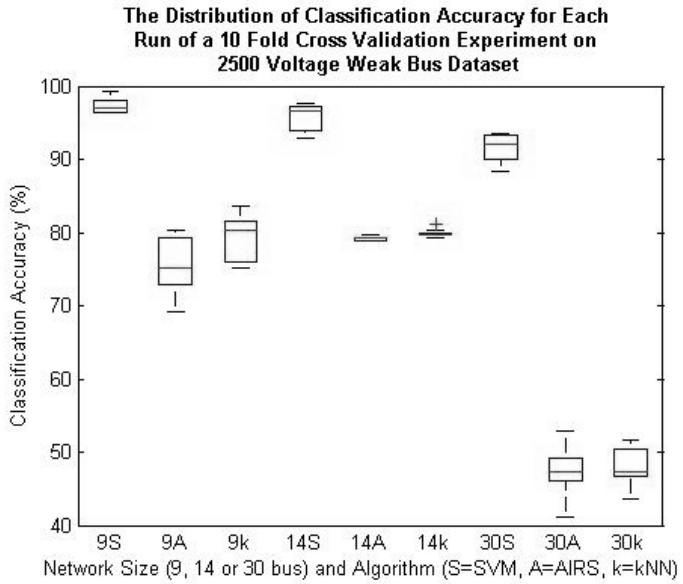


Figure 6 – The distribution of classification accuracy for each run of a 10 fold cross validation experiment with each classification algorithm on the 2500 operating point voltage weak bus dataset.

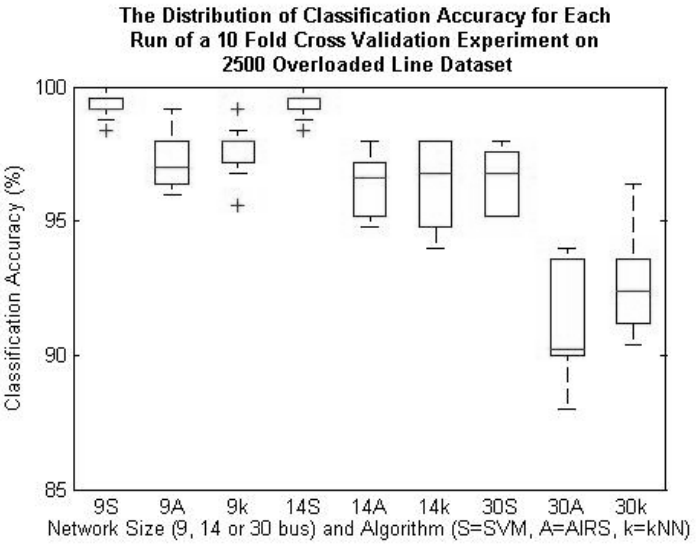


Figure 7 – The distribution of classification accuracy for each run of a 10 fold cross validation experiment with every classification algorithm on the 2500 operating point overloaded line dataset.

F.5.4 Statistical & Scientific Significance of Results

The classification algorithms were compared as described in section 4.3.3 of the methodology. The aim is to compare each of the algorithms to reject 9 null hypotheses which are described in general as:

Algorithm A and Algorithm B yield the same distribution of classification performance when classifying voltage weak buses using the methodology described in this appendix on the X bus power system network.

Table 11 shows the results of the attempted rejection of the 9 null hypotheses. It also shows an overview of whether or not the size of any statistical significance is scientifically significant.

TABLE 11 – THE STATISTICAL (STAT.) AND SCIENTIFIC (SCI.) SIGNIFICANCE OF CLASSIFICATION PERFORMANCE WHEN COMPARING AIRS, SVM AND KNN TO CLASSIFY OVERLOADED AREAS ON THE POWER SYSTEM NETWORK.

Hypothesis Number	**Network & Dataset**	**Comparison**	**Rank-sum *p* Value**	**Vargha-Delany [157] A Statistic**	**Statistical & Scientific Significance**
Hypothesis 1	9 Bus 100 Database	AIRS to kNN	0.0975	0.4322	None
	9 Bus 500 Database	AIRS to kNN	1.28×10^{-99}	0.0466	Sci. & Stat.
Hypothesis 2	9 Bus 100 Database	AIRS to SVM	2.60×10^{-29}	0.0403	Sci. & Stat.
Hypothesis 3	9 Bus 100 Database	SVM to kNN	5.56×10^{-29}	0.0502	Sci. & Stat.
Hypothesis 4	14 Bus 100 Database	AIRS to kNN	3.25×10^{-15}	0.1762	Sci. & Stat.
Hypothesis 5	14 Bus 100 Database	AIRS to SVM	7.25×10^{-34}	0.0120	Sci. & Stat.
Hypothesis 6	14 Bus 100 Database	SVM to kNN	2.37×10^{-28}	0.0460	Sci. & Stat.
Hypothesis 7	30 Bus 100 Database	AIRS to kNN	1.89×10^{-4}	0.3240	Stat.
Hypothesis 8	30 Bus 100 Database	AIRS to SVM	1.05×10^{-34}	0.0875	Sci. & Stat.
Hypothesis 9	30 Bus 100 Database	SVM to kNN	4.01×10^{-17}	0.1544	Sci. & Stat.

The results show that the difference in classification performance for each of the 9 hypotheses (and algorithm comparisons) is mainly both scientifically and statistically significant. However, the differences in effect sizes between the AIRS algorithm and the kNN algorithm are much less than those experienced between SVM and AIRS / kNN. This is also apparent in the classification results obtained on the larger 2500 operating point database.

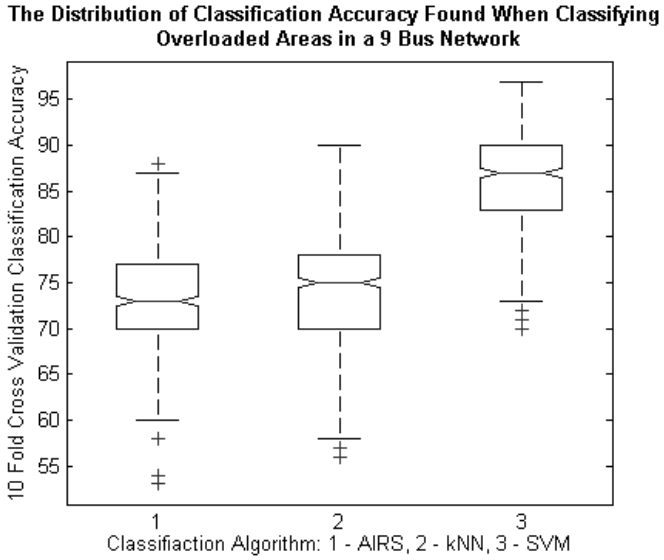


Figure 8 – The classifiaction accuracy of AIRS, kNN and SVM algorithms achieved using 10 fold cross validation and 100 operating points on the 9 bus test network.

Figure 8 describes the distribution of classification accuracy found when running simulations on all three algorithms. The chart shows that the inter-quartile ranges of classification accuracy found when running the AIRS and kNN algorithms are very similar. However, the distribution of classification accuracies found for the SVM algorithm are somewhat different.

F.5.5 Discussion of Results

For the task of grouping together voltage weak buses and overloaded lines, the UAIC clustering algorithm was able to identify several uniquely different areas of the network that were overloaded. This is shown in tables Table 7 and Table 8.

The classification accuracy of the SVM algorithm outperformed both AIRS and the kNN algorithm for each of the tests across all networks (Table 9). Support vector machines are often good at generalising high dimensional data sets like those presented in this study. kNN and AIRS both require a well populated feature space as they are instance based methods of classification. The relatively high dimensional feature space (over 60 dimensions for the 30 bus network) coupled with a low number of instances (2500) may have caused difficulties for both the kNN and AIRS algorithms.

As the number of dimensions (and the size of the network) in the feature space increased, all of the algorithms declined in classification performance (shown in Figure 6 and Figure 7). The most notable declines were for AIRS and kNN where classification performance fell to around 45% (Table 9) when attempting classification of the 30 bus voltage weak bus dataset. The SVM algorithm remained able to classify reliably even with the high dimensions involved in the 30 bus network.

The conclusion that could be drawn from these results is that the SVM algorithm is superior to either the kNN or AIRS when classifying overloaded areas of the power network in this manner. There would also seem to be a strong divergence between the instance based methods and the SVM algorithm as the number of dimensions grew larger (as shown in Figure 6 and Figure 7). This could mean that both AIRS and kNN don't perform well on higher dimensional datasets. For AIRS, this may be because the internal mechanisms of the algorithm do not operate well when the number of dimensions is high.

The results also showed that the classification performance obtained with AIRS was very close to those obtained by the kNN algorithm. In 5 out of the 6 tests on the 2500 operating point database AIRS fell just short of the performance obtained by the kNN algorithm. This corroborates with previous research by Secker et al. [191] that argued that AIRS was simply a pre-processor for a k nearest neighbour classification routine. Indeed the results obtained here show that AIRS performs worse than the kNN classifier on all tests.

Overall, the SVM classification algorithm performed well across all of the conducted tests. In each of 10 fold cross validation experiments (across different networks and for both overloaded lines and voltage weak buses) the overall classification accuracy of the algorithm was always in excess of 91%.

F.5.6 Scientific & Statistical Significance of AIRS, kNN and SVM Classifiers

In nearly all cases scientific and statistical significance were present when comparing the classification accuracy of all the algorithms. Statistical and scientific significance only fell short when comparing AIRS and the kNN algorithm (see Table 11).

These results may indicate the possibility that the best classification performance obtainable between AIRS and kNN for this type of problem will always be very similar. Table 11 shows that AIRS consistently performed in a very similar manner to the kNN algorithm, failing to register scientific significance in two tests and statistical significance in one test.

The results presented in this research seem to show a selection of problems where AIRS has no real performance advantage over the simpler kNN algorithm. However, the fact that there is (in some cases) no significant difference in performance may itself

be interesting. AIRS is an instance creation algorithm that significantly compresses the original dataset into a smaller number of instances. If the performance of AIRS is comparable with kNN, this may show that the internal mechanisms in AIRS are powerful at compressing the number of instances in a dataset and maintaining the accuracy of a kNN classifier.

These results certainly add to the debate amongst the AIS community about where the power of the AIRS algorithm truly lies [192]. They are, however, also a counter example to the research presented in [133] which stated that the performance of AIRS (with the kNN classifier) is often higher than for the kNN classification algorithm alone.

F.6 Conclusion

In this appendix a Support Vector Machine (SVM), Artificial Immune Recognition System (AIRS) and k Nearest Neighbours (kNN) algorithm were used as part of a methodology developed to identify critical areas in power system networks, namely voltage collapse prone areas and overloaded transmission lines.

The methodology described used an unsupervised artificial immune classification (UAIC) algorithm to identify groups (clusters) of voltage weak buses (voltage collapse prone areas in the network) and overloaded transmission lines. The UAIC algorithm is immune inspired and this research highlighted some of the common elements it shares with AIRS. The UAIC algorithm was shown to perform clustering reliably and to perform in a similar manner to the k-means clustering algorithm when applied to a very simple clustering problem. The clusters were used as labels for a supervised classification stage which utilised AIRS, kNN and SVM classification methods to build a model capable of identifying weak buses and overloaded lines in the network. The classification results were tested by comparing the SVM, AIRS and kNN classification performance using rigorous statistical methods.

The research in this appendix had two aims: firstly to present a methodology capable of detecting overloaded areas of a generic power system network and secondly to show experimentally how the three selected classification algorithms compared when applied to this type of problem.

The results show that the methodology is capable of identifying distinct regions within the power network which contain overloaded lines and voltage weak buses, thus satisfying the first aim. Indeed, if the SVM algorithm was selected as the preferred classification method, overloaded areas in a 9 bus, 14 bus and 30 bus network can all be successfully identified with a level of classification accuracy exceeding 91%.

The research presented in this appendix also aimed to present a rigorous comparison of SVM, AIRS and kNN classification methods when applied to detecting overloaded areas of the power system network. The results showed that the SVM significantly outperformed both AIRS and the kNN algorithm. They showed that AIRS and the kNN algorithm were in some cases statistically indistinguishable in terms of classification performance. In the cases where AIRS and kNN showed a slight difference in classification performance, AIRS typically performed slightly worse than the kNN algorithm. The results also showed that altering the number of dimensions in the problem did not significantly alter the performance difference between the kNN and AIRS algorithms. However, increasing the number of dimensions did increase the gap in classification performance between the SVM and both AIRS / kNN.

The research presented in this appendix demonstrated that the presented methodology is a viable and fast way of identifying voltage weak areas and overloaded lines of a power system network. It could easily be applied to studies where the computationally intensive nature (due to the size of the power network) of performing a static voltage stability assessment and load flow is prohibitively expensive.

Appendix G: List of Author's Thesis Based Publications

G.1 Journal Papers

[G1] N. C. Woolley, J. V. Milanović, "An Immune Inspired Clustering and Classification Method to Detect Overloaded Areas in Power System Networks", Journal of Natural Computing, vol. 10, No. 1, March 2011, **Invited Journal Paper**

[G2] N. C. Woolley, M. Avendaño-Mora, J. V. Milanović, "Immune System Inspired Methodology for Robust Monitoring of Voltage Sags Based on Equipment Trip Probabilities" Accepted for Publication in Elsevier Electric Power Systems Research, 2012

[G3] N. C. Woolley, J. V. Milanović, "Estimating the Level, Location and Impact of Voltage Unbalance in Distribution Networks Using Distribution System State Estimation", Accepted for Publication in the IEEE Transactions on Power Delivery, 2012

[G4] R. Preece, N. C. Woolley, J. V. Milanović, "The Probabilistic Collocation Method for Power System Damping and Voltage Collapse Studies in the Presence of Uncertainties", Submitted to IEEE Transactions on Power Systems in 2012

G.2 International Conference Papers

[G5] N. C. Woolley, J. V. Milanović, "Application of Artificial Immune System Algorithms for the Detection of Weak Voltage Buses in the Power System Network", International Conference on Artificial Immune Systems (ICARIS), 2009

[G6] N. C. Woolley, M. Avendaño-Mora, J. V. Milanović, "A Comparison of Voltage Sag Estimation Algorithms Using Optimal Monitoring Locations", International Conference on Harmonics and Quality of Power (ICHQP), 2010

[G7] M. Avendaño-Mora, N. C. Woolley, J. V. Milanović, "On Improvement of Accuracy of Optimal Voltage Sag Monitoring Programs", International Conference on Harmonics and Quality of Power (ICHQP), 2010

[G8] N. C. Woolley, J. V. Milanović, "Estimating the Voltage Unbalance Factor Using Distribution System State Estimation", Innovative Smart Grid Technologies Europe, 2010

[G9] N. C. Woolley, J. V. Milanović, **"**Integration of cost effective bus profiling in distribution networks", International Federation of Automatic Control (IFAC), 2011, **Invited Conference Paper**

[G10] S. Mat Zali, N. C. Woolley, J. V. Milanović, "Development of Equivalent Dynamic Model of Distribution Network Using Clustering Procedure", Power Systems Computational Conference (PSCC), 2011

[G11] N. C. Woolley, M. Avendaño-Mora, J. V. Milanović, "Probabilistic Fault Location Using Erroneous Measurement Devices", Smart Metering for Future Grids (SMFG), 2011