Data analytics for weak spot detection in power distribution grids

Hassan Mashad Nemati

Supervisors:
Sławomir Nowaczyk
Anita Sant'Anna

D O C T O R A L T H E S I S | Halmstad University Dissertations no. 55

Data analytics for weak spot detection in power distribution grids

Halmstad University Dissertations no. 55
ISBN 978-91-88749-18-5 (printed)
ISBN 978-91-88749-19-2 (pdf)
Publisher: Halmstad University Press, 2019 | www.hh.se/hup
Printer: Media-Tryck, Lund

Abstract

This research aims to develop data-driven methods that extract information from the available data in distribution grids for detecting weak spots, including the components with degraded reliability and areas with power quality problems. The results enable power distribution companies to change from reactive maintenance to predictive maintenance by deriving benefits from available data. In particular, the data is exploited for three purposes: *(a)* failure pattern discovery, *(b)* reliability evaluation of power cables, and *(c)* analyzing and modeling propagation of power quality disturbances (PQDs) in low-voltage grids.

To analyze failure characteristics it is important to discover which failures share common features, e.g., if there are any types of failures that happen mostly in certain parts of the grid or at certain times. This analysis provides information about correlation between different features and identifying the most vulnerable components. In this case, we applied statistical analysis and association rules to discover failure patterns. Furthermore, we propose a visualization of the correlations between different factors representing failures by using an approximated Bayesian network. We show that the Bayesian Network constructed based on the interesting rules of two items is a good approximation of the real dataset.

The main focus of reliability evaluation is on failure rate estimation and reliability ranking. In case of power cables, the limited amount of recorded events makes it difficult to perform failure rate modeling. Therefore, we propose a method for interpreting the results of goodness-of-fit measures with confidence intervals, estimated using synthetic data.

To perform reliability ranking of power cables, in addition to the age of cables, we consider other factors. Then, we use the proportional hazard model (PHM) to assess the impact of the factors and calculate the failure rate of each individual cable. In reliability evaluation, it is important to consider the fact that power cables are repairable components. We discuss that the conclusions about different factors in PHM and cables ranking will be misleading if one considers the cables as non-repairable components.

In low-voltage distribution grids, analyzing PQDs is important as we are moving towards smart grids with the next generation of producers and consumers. Installing Power Quality and Monitoring Systems (PQMS) at all the nodes in the network, for

monitoring the impacts of the new consumer/producer, is prohibitively expensive. Instead, we demonstrate that power companies can utilize the available smart meters, which are widely deployed in the low-voltage grids, for monitoring power quality events and identifying areas with power quality problems. In particular, several models for propagation of PQDs, within neighbor customers in different levels of the grid topology, are investigated. The results show that meters data can be used to detect and describe propagation in low-voltage grids.

The developed methods of *(a) failure pattern discovery* are applied on data from Halmstad Energi och Miljö (HEM Nät), Öresundskraft, Göteborg Energy, and Växjö Energy, four different distribution system operators in Sweden. The developed methods of *(b) reliability evaluation of power cables* and *(c) analyzing and modeling propagation of PQDs* are applied on data from HEM Nät.

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Associate Prof. Sławomir Nowaczyk and Assistant Prof. Anita Sant'Anna. Your knowledge and insightful comments have been invaluable. I have learned a lot from you and I am grateful for your guidance, encouragement, and patience. I also would like to thank my advisers, at Halmstad Energi och Miljö (HEM Nät), Alexander Örning and Peter Addicksson for their support and inspiring cooperation.

Many thanks to Prof. Antanas Verikas, for providing a pleasant working environment in the Department of Intelligent Systems and Digital Design. Many thanks to Associate Prof. Stefan Byttner for his great support and provision of helpful suggestions particularly in the PhD education. I would like to also acknowledge the support from Prof. Magnus Jonsson, who leads the Embedded and Intelligent Systems Industrial Graduate School (EISIGS).

I would like to thank the support committee members, Prof. Ulf Johansson and Associate Prof. Robert Bass for providing valuable feedback on my research.

I am grateful to Associate Prof. Björn Åstrand. I learned a lot from you during my master thesis project and even after during my PhD education. I would like to thank Assistant Prof. Sepideh Pashami for all your valuable advices and support during my PhD.

I would like to give special thanks to my friends Maytheewat Aramrattana, Pablo del Moral, and Ece Calikus for always being there for me whenever I needed. I also would like to thank Jenifer David, Yuantao Fan, Saeed Gholami Shahbandi, and Kevin Hernandez. Thank you for all the great memories.

I would like to thank every member of the Department of Intelligent Systems and Digital Design specially Roland Thörner, Prof. Thorsteinn Rögnvaldsson, Prof. Josef Bigun, Assistant Prof. Tommy Salomonsson, and Assistant Prof. Martin Cooney.

I would like to express my deepest appreciation to my parents for their love, guidance and support through my life. Special thanks to my brother Hossein for his guidance during my education and research.

This project is funded by Knowledge Foundation and HEM Nät.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

The reliability of electric power grids (generation, transmission, and distribution) is important for both utilities (providers) and consumers. Industries, infrastructure, and citizens rely on electric power, and power interruption can have disastrous effects. Furthermore, many governing bodies are increasing requirements put on power companies to improve the overall reliability of electricity grids.

Electric power distribution grids are the final stages in the delivery of electric power to the end users where the electricity is consumed. Distribution grids are usually facing problems such as aging infrastructure, poor design, high exposure to environmental conditions, and irregular electricity usage by customers, and are therefore addressed as the most vulnerable sector in power grids [1, 2, 3, 4].

Smart distribution grids (SDGs) are designed to be the next generation of power distribution grids. The term “smart” is addressed in different ways by different actors. However, two different types of definitions can be encountered (Swedish government inquiry on smart grids [5]). The first type focuses on the technologies that are included in smart grids (we call it *features*); and, the second focuses on the problems that can be solved with smart grids (we call it *capabilities*). Some of the features include two-way communication and control, distributed computing and Advanced Metering Infrastructure (AMI), equipment installed on the premises of network user. Some of the capabilities include economically efficient and sustainable power systems with low interruption, high level of power quality, and facilitating the large-scale use of renewable energy sources.

Although SDGs contain features that provide improvement mechanisms, they require significant additional operational automation to achieve their full capabilities [6]. For example, SDGs can automatically detect and resolve momentary disturbances by opening and closing interruption devices. However, several long-duration disturbances caused by a failed component need manual work to be precisely localized and fixed [7]. This manual work is usually very costly and time consuming.

Furthermore, in SDGs, many new consumers and producers are continuously being connected to the network, such as electric vehicles and various distributed generators, which will impact the power quality in various ways. Some of these impacts are voltage fluctuation in feeder lines, malfunction of voltage regulation equipment, possibility of overload in power cables, variation of reactive power flow due to malfunction of capacitor bank devices, and malfunction of over-current and over-voltage protection devices [8]. In this case, it is important to find solutions for monitoring the disturbances caused by the new consumer/producer and methods for analyzing their impacts on other components in the grid.

Currently, the majority of improvement mechanisms in SDGs are devoted to analyzing the data from high-frequency sensors. However, large amounts of data from low-frequency measurements, failure records, and grid information are available but rarely used for reliability improvement.

In general, the available data in SDGs can be categorized into three groups: 1) sensor readings, 2) failure records, and 3) grid and components information (see Table 1.1). The sampling rate of the sensor readings can be high, e.g., Supervisory Control And Data Acquisition (SCADA) systems and Power Quality Monitoring Systems (PQMS) or low, e.g., smart meters (SMs). The failure records corresponds to the information about failures and repairs on different components in the grid. Some of this information includes, date and time, cause of failure, affected components, and duration of disturbances. Grid information contains additional data about components, customers, and the grids topology.

Table 1.1: Different type of data in SDGs

Data in smart distribution grids				
Type of data	Sensor readings		Failure records	Grid Information
Resolution	**High-sampling rate** for SCADA usually 1 sample every 2-4 seconds, and for PQMS 10-60 samples per seconds	**Low-sampling rate** for load data usually 1 sample every hour, and for alarm data is event-based	Event-based	Static
Quantity	High high resolution of data but low number of available sensors	High low resolution of data but high number of available SMs	Low	Static
Quality (relevant information to the grid's reliability)	High	Low	High	High
Example	SDACA, PQMS	SMs data (alarm, load profile)	Previous faults, Maintenance history	Components inventory, Grid topology, Customers information

Realizing the full capabilities of smart grids requires using the available features to solve problems in SDGs. This includes extracting information from the available data and applying improvement actions accordingly.

Furthermore, to mitigate the number and severity of power interruptions there is a need to switch from reactive maintenance (repair after failures occur) to predictive maintenance (repair before failures occur). The goal in predictive maintenance is to estimate the condition of a system or a component and perform maintenance accordingly. Predictive maintenance differs from preventive maintenance in that it is determined by the condition of equipments rather than the expected life-time, recommended by the component manufacturers. Applying predictive maintenance on different components in the grid can prevent catastrophic equipment failures [9].

1.1 Objectives

This research aims to develop data-driven methods that extract information from the available data in distribution grids for detecting weak spots, including the components with degraded reliability and areas with power quality problems. The main purpose of this project is to enable power distribution companies to change from a reactive mode to predictive mode by deriving benefit from existing data.

To achieve the objectives, this dissertation focuses on the development of methods for: (a) fault analysis and failure pattern discovery, (b) lifetime modeling and reliability ranking of underground power cables, and (c) smart meters data analysis.

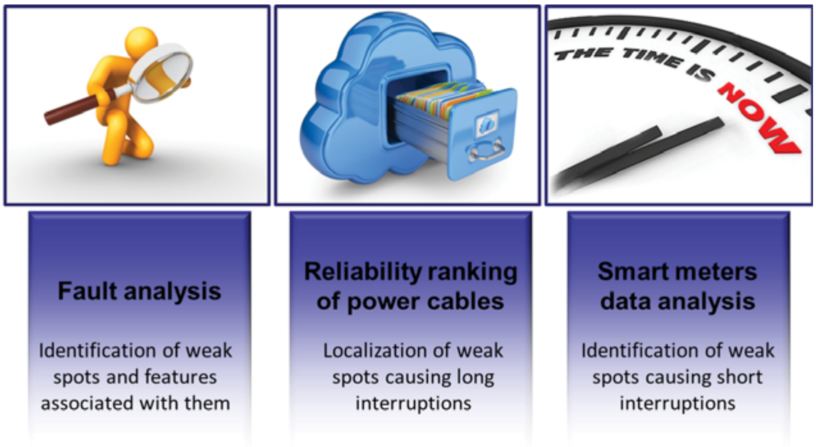


Figure 1.1: Objectives of the thesis.

Fault analysis Finding patterns in the failure records and summarizing them with quantitative models is a step towards turning data into information and turning information into knowledge. The results of this analysis facilitate reasoning about different features associated with faults and can be used by maintenance staff at distribution companies.

Reliability ranking of power cables High-voltage and low-voltage underground power cables are heavily affected by ionization, as well as thermal and mechanical stresses [10]. Cable failures usually create long outages. In case of failures, both pinpointing and repairing faults are very costly and time consuming due to the difficulty in accessing them. Mining and analyzing the available data enables us to design reliability evaluation methods that can estimate the life time of power cables and identify the weak ones. Then, power distribution companies can directly target those vulnerable cables for inspection and preventive repair actions.

Smart meters data analysis In low-voltage distribution grids, analyzing power quality disturbances (PQDs) is important as we are moving towards smart grids. SMs are generally deployed at all customers and continuously measure several features. Currently, SMs are mainly used for billing energy usage; however, they have additional features such as monitoring voltage, current, frequency, and power quality. Analyzing PQDs using SMs data can provide useful insights of the areas in the low-voltage side of the grid with power quality problems. Distribution companies can benefit from the results for strengthening their grid and eliminating the power quality problems for customers.

1.2 Challenges

Real-world data - real-world data, quite often, are faced with the challenge of sparse or no ground truth when it comes to power grids. In these situations it is incredibly valuable to be able to accurately determine whether the recorded data represent the "actual" failures of a system or component. Even in those situations where some of the labels might be available (sparse ground truth), it is often to be expected that there will be new forms of failures that do not fit into any previous patterns. This means that they will not be discovered by methods that attempt to match with already occurring phenomena for which there are labels. In this case, we can investigate answering questions such as which systems, components, or factors are causing a specific failure, or increasing the probability of an event; how different features describing the failures are correlated; and, how these correlations can be represented.

Quantity and quality of data - in SDGs, large amount of data is recorded in different ways, but not all of them can be used for weak spot detection. Among those that have less relevant information about the grid's reliability, we can refer to customers' electricity consumption data. Consumption data are mainly used for billing purposes and their relation with components reliability is very weak. On the other hand, the history of previous faults is highly relevant, but the amount of historical failure information is usually small. This lack of data makes it difficult to perform reliability evaluation with reasonable confidence.

Impact of multiple factors on reliability evaluation - our evidence shows that targeting only the aged components for replacement is not an optimal strategy for reliability improvement. In fact, age is only one factor among many that impact failure rates. Other factors include previous maintenance, geographical position, length, etc. Some of these factors can be captured directly from datasets, although the corresponding information may be imprecise. In this case, the major challenge is related to identifying which factors should be included and estimating how significant is the impact of each factor on components' reliability.

SMs data for analyzing power quality - the main challenge in using SMs for monitoring power quality is the limitations in data storage and communication protocols. In fact, SMs cannot provide the high-resolution waveform-based monitoring for power quality. This makes it difficult to perform the commonly used analysis which mainly consider the shape of the waveforms in the three-phases with precision in milliseconds to detect the root cause of power quality problems.

1.3 Research Questions

Our main objective is to identify weak spots which we have set out to be components with degraded reliability and areas with power quality problems. Based on the available data with the limitations stated in section 1.2, several questions have to be answered. Here, the main question is *what information can be extracted from the available data which is relevant to the reliability of systems and components in the distribution grids?* Historical failure records database is a valuable source for extracting information about the reliability of the grid. However, the amount of this data for components failures is usually limited. In this case, there are other sources of information available in power grids that can be utilized and additional factors that may have correlation with components failures can be extracted. Furthermore, in order to perform reliability analysis, the characteristic of components (if they are repairable or non-repairable) and the limitations of different methods must be considered; otherwise, the conclusions about the reliability analysis would be incorrect. Moreover, SMs which are distributed throughout the distribution grids can provide information directly at the points of customer connection. Methods for extracting this information and then transforming it to knowledge about the weak spots are needed to be investigated.

Therefore, based on the aim of the project and previously highlighted challenges, the following research questions are addressed in this thesis:

1. How to discover and represent failure patterns from historical failure data of a SDG?
2. How to perform reliability evaluation with confidence bounds while dealing with limited historical failure data?
3. How to design a robust method for reliability evaluation considering the impact of the repairability characteristic of power cables and selection of different factors?
4. How to analyze and model the propagation of power quality disturbances in low-voltage grids using smart meters alarm data?

1.4 Research Contributions

The main contributions of this thesis are summarized as follows:

1. Proposing a method for visualization of the correlations between different features representing failures by using an approximated Bayesian network (**Paper B**). Patterns are considered as the correlation between failure and other features such as season, weekday, time, and outage duration. The outcomes of failure pattern discovery can be used for identifying the most vulnerable components (e.g. underground cables) or factors that are important for further analysis. The cause of the identified correlations and ways to limit different failures are some of the topics that can be investigated by utilities. The generality of the method is evaluated by applying the method on historical data from other distribution companies (see Chapter 3.1.2).

2. Proposing a methodology for power cables lifetime modeling with confidence intervals to deal with limited failure data (**Paper A**). We investigated six different models estimating the probability of failures for in-service underground cables. In many practical cases, the amount of data available is very limited, and it is difficult to know how much confidence one should have in the goodness-of-fit results. Therefore, we focused on a methodology for evaluating how well different models fit the historical data and represent the results by confidence bounds. In this analysis only the age of the cables is considered (not the impact of additional factors).

3. Demonstrating the importance of considering the reparability characteristic of power cables on reliability estimation (**Paper C**). In particular we compared three case scenarios depending on how to consider power cables and their failures: as nonrepairable components, as repairable but decommissioned after the last failure, and as repairable components which survive until censoring time. For power cables, the first and second scenarios are incorrect but often used, and we discussed that conclusions about reliability analysis will be misleading if the evaluation is carried out based on these approaches.

4. Developing a method for ranking repairable power cables based on the impact of different factors (**Paper C**). We used the proportional hazard model (PHM) to assess the impact of different factors and calculate the failure rate of each individual cable. Then we ranked cables based on their failure rate. The method is applied considering the restoration characteristic of power cables.

5. Analyzing and modeling propagation of smart meters' alarms in low-voltage grids (**Paper D**). We show that the existence of propagation of sag/swell disturbances in the low-voltage grids can be discovered using data from SMs. Furthermore, several models are designed based on different scenarios and synthetic data generated according to them. These models include the grid with: no propagation, propagation within neighbor customers at the same delivery-point, propagation within neighbor customers at the same branch, and propagation in both delivery points and branches. The similarity of the artificial data generated based on each models are then compared with the real data. The results

demonstrate that the models which include propagation within both delivery points and branches represents the real data better. The collected data for this work, containing power quality alarms from over 1000 customers for a four months period, is made publicly available.

1.5 List of Appended Publications

The appended publications in this thesis are listed in the following:

Paper A - **Hassan M. Nemati**, Anita Sant'Anna, Sławomir Nowaczyk (2015). Reliability Evaluation of Underground Power Cables with Probabilistic Models. *The 11th International Conference on Data Mining (DMIN'15)*, July 2015.

Paper B - **Hassan M. Nemati**, Anita Sant'Anna, Sławomir Nowaczyk (2016). Bayesian Network Representation of Meaningful Patterns in Electricity Distribution Grids. *IEEE International Energy Conference (ENERGYCON)*, April 2016.

Paper C - **Hassan M. Nemati**, Anita Sant'Anna, Sławomir Nowaczyk, Jan Henning Jürgensen, Patrik Hilber (2018). Reliability Evaluation of Power Cables Considering the Restoration Characteristic. *International Journal of Electrical Power and Energy Systems*, 105, pp 622-631, February 2019.

Paper D - (submitted to Data Mining and Knowledge Discovery) **Hassan M. Nemati**, Anita Sant'Anna, Sławomir Nowaczyk (2019). Analyzing and Modeling Propagation of Smart Meters' Alarms in Low-Voltage Grids.

I am the main author of **Papers A-D** where I conducted the research and authored the papers. Sławomir and Anita supported and reviewed the research. In **Paper C**, Jan and Patrik supported and reviewed the paper.

1.6 Other Related Publications

1. **Hassan M. Nemati**, Anita Sant'Anna, Sławomir Nowaczyk (2014). Overview of Smart Grid Challenges in Sweden. *28th annual workshop of the Swedish Artificial Intelligence Society (SAIS)*, May 2014.

2. **Hassan M. Nemati**, A. Laso, M. Manana, Anita Sant'Anna, Sławomir Nowaczyk (2018). Stream Data Cleaning for Dynamic Line Rating Application. *Energies 11(8)*, pp 2007, August 2018.

Chapter 2
Background

Power distribution grid problems usually relate to two main concerns: reliability and power quality (PQ). Typically, the reliability corresponds to power interruptions, and PQ refers to deviations in current and voltage waveforms. Although, the deviations in voltage waveforms are not a complete loss of voltage, the consequences of sensitive equipment can be as severe as those caused by an interruption [11].

2.1 Reliability analysis

In general, reliability is defined as the probability of a system or component performing its purpose adequately for the period of time intended under the operating conditions [12].

There are several ways to describe reliability such as failure probability distribution, cumulative distribution, reliability function, and failure rate. In power grids, reliability is normally expressed by failure rate (FR), which is the number of expected failures per unit in a given time interval [13]. Figure 2.1 shows a commonly

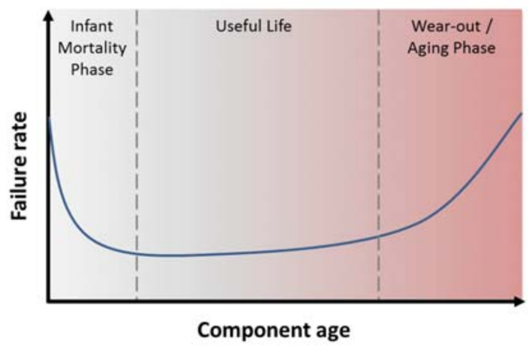


Figure 2.1: Bathtub curve of typical failure rate for components.

used failure rate model known as the bathtub curve [14]. The model begins with a high FR (infant mortality), followed by fairly constant FR (useful life). Finally, the FR increases again as the component reaches the end of its life (wear-out).

In order to perform reliability evaluation, some important aspects should be considered including how to obtain the required data, which statistical analysis should be chosen, and what are the constraints and limitations of reliability evaluation due to the specific characteristic of power components (see Figure 2.2).

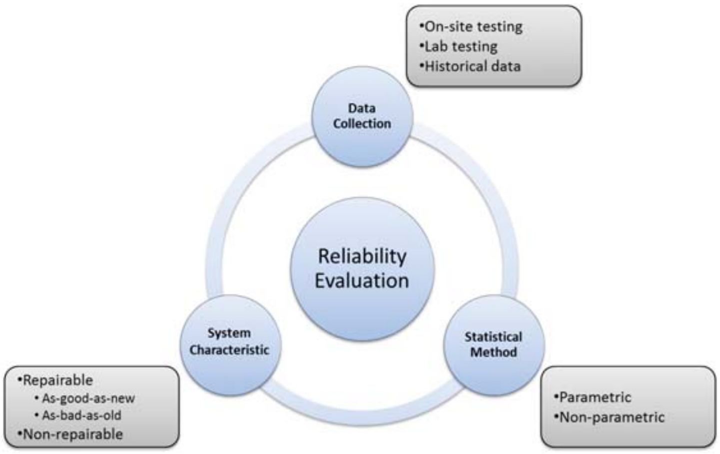


Figure 2.2: Important aspects in reliability evaluation of power components.

2.1.1 Data Collection

One way to collect the required data for reliability analysis of power systems is to perform on-site testing [15, 16]. In this approach, the actual condition of the system can be *measured* while the system is in operation, based on some suitable and measurable indication of component deterioration.

For power systems, on-site testing may damage the insulation of the tested component [17] and therefore laboratory tests are commonly used instead of on-site testing. In laboratory tests [18, 19, 20], a new component first undergoes accelerated aging processes to simulate the condition of aged ones. Then the component deterioration indicators are measured. Hence, instead of measuring the actual status of the in-service component, the condition of the component will be *estimated* based on off-site (laboratory) tests.

In case of power cables, both on-site testing and laboratory testing methods are costly and complex processes. Moreover, the condition of power cables can be *estimated* in a cheaper way by analyzing historical data. In general, utility companies keep records of historical data such as previous events and inventory data (manufacturer information) which can be used for reliability evaluation of power cables [21, 22, 23, 24, 25].

One of the problems with analyzing historical information is the limited amount of failure data, which makes it difficult to estimate the reliability of power cables with reasonable confidence. To mitigate this problem, some previous works have used expert knowledge to manually refine the estimated lifetime of power cables. In [26, 27] expert knowledge was used to obtain uncertainty range for the following variables: impact of previous failures, impact of environmental and operational stressors, probability of cable condition over time, test accuracy of the cables condition, and impact of previous repairs. These approaches, however, are error prone. In addition, such approaches cannot be easily generalized from one grid to another.

If every feeder line is under observation until a failure, the reliability measures may be estimated simply by computing the fraction of lines surviving at each age. However, some feeder lines have not been destroyed or discarded by the end of the experiment (data collection). In this case, the data is “right censored” [28], which affects the reliability analysis and need to be considered.

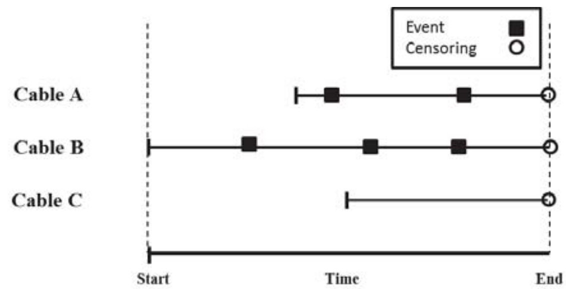


Figure 2.3: Schematic plot for events and censoring in feeder lines.

2.1.2 System Characteristics

In order to use reliability measures such as failure rate, the nature of the system and the limitations of different methods must be considered [29]. In reliability evaluation, there is a crucial difference between the statistical treatment of repairable and non-repairable systems. A repairable system or component can be restored to satisfactory operation after a failure by repair actions; while, a non-repairable system or component is removed permanently (replaced with a new system or component) after a failure.

The replacement of an underground power cable is very costly and it is not economically efficient to change the entire cable after a failure. Therefore, in case of failures, only the faulty point is replaced by a new segment and the rest of the cable stays untouched. This restoration characteristic of power cables allows distribution companies to keep them in service for more than the manufacturers’ recommended lifetime. In fact, as long as the frequency of failures in a specific cable is not high (tradeoff between the cost of multiple repairs and replacing the entire cable), these companies tend to keep the cable in service. Furthermore, power cables after a failure and repair are usually as-bad-as-old, i.e., the repair after each failure does not materially change the condition of the entire cable.

2.1.3 Statistical Method

In a very broad sense, failure data can be evaluated statistically using either parametric methods, or nonparametric methods [14]. Parametric methods make assumptions about the underlying population from which the data are obtained. On the other hand, non-parametric methods, also called "distribution-free" methods, do not assume any particular family for the distribution of the data [30].

In both parametric and nonparametric methods, the failure processes are described as random events. These time of events are then considered as random variables that can have a continuous or discrete characteristic [31].

If components are considered as non-repairable, the FR or hazard rate (HR) function are usually used to estimate the remaining useful life of the components [14, 32, 33]. The failure time is a random variable T described by a single *time to failure*. The order of failure times does not matter, i.e. the random variables T are not chronologically ordered. In this case, FR is the relative rate of failure of components surviving until time T (conditional).

If components are considered as repairable, stochastic point process (SPP), renewal process (RP) model, or reliability growth analysis are usually used to estimate the expected number of events over time [34, 35, 36, 37, 38, 39, 40, 41, 42]. If the failure time T represents the time between successive failures, it is called inter-arrival time. Here it is assumed that the repair action materially changes the condition of the component (the condition of the component after repair is "as-good-as-new"). But, if the repair action does not materially change the condition of the component (the condition of the component after repair is "as-bad-as-old"), the time of failure compared to time 0 represents the failure time T. For repairable systems, the rate of occurrence of failure (ROCOF) and mean time between failure (MTBF) are usually used to represent the expected number of cumulative failures at each time stamp. ROCOF is the absolute rate at which system failures occur (unconditional).

The following measures can be used for reliability evaluation [43] of non-repairable components:
Cumulative distribution function: the probability that any randomly chosen component fails within the interval $(0, t]$

$$F(t) = P(T \leq t) = \int_0^t f(u)du \qquad \text{for } t > 0 \tag{2.1}$$

Here, it is assumed that the time to failure T is continuously distributed with probability density function $f(t)$.

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t} \tag{2.2}$$

Reliability function: also known as survival function, the probability that any randomly chosen component does not fail within the interval $(0, t]$

$$R(t) = P(T > t) = 1 - F(t) = 1 - \int_0^t f(u)du \qquad \text{for } t > 0 \tag{2.3}$$

Failure rate function: the probability that an observed component (the component has not failed yet) fails in the time interval $(t, t + \Delta t]$

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{R(t) - R(t + \Delta t)}{\Delta t \cdot R(t)} = \frac{f(t)}{R(t)} \tag{2.4}$$

Parametric methods

In parametric methods, depending on the type of the random variable (continuous or discrete), a distribution model is fitted to the data. With different goodness-of-fit tests the model parameters can be estimated. The amount of available data influences the confidence bounds of the performed analysis.

For power system components the Weibull model is commonly used to fit the failure rate data points [44]. The failure rate function of the Weibull model with shape parameter $\alpha > 0$ and scale parameter $\eta > 0$ is:

$$\lambda(t) = \frac{\alpha}{\eta} \left(\frac{t}{\eta}\right)^{\alpha - 1} \tag{2.5}$$

Other reliability measures such as probability density function $f(t)$, and cumulative distribution function $F(t)$ can be calculated using the following formulas.

$$f(t) = \frac{\alpha}{\eta} \left(\frac{t}{\eta}\right)^{\alpha - 1} \cdot \exp\left[-\left(\frac{t}{\eta}\right)^{\alpha}\right] \tag{2.6}$$

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\alpha}\right] \tag{2.7}$$

Nonparametric methods

Nonparametric methods make no assumption about the underlying distribution model, meaning that the distribution of a components' life time is unknown. Kaplan-Meier and Nelson-Aalen estimators are two examples of nonparametric methods. These estimators are used for calculating survival function $\hat{S}(t)$ and cumulative FR function $\hat{\Lambda}(t)$ by the following equations [45]. Here the censoring in data is also considered.

$$\hat{S}(t_i) = \prod_{j=1}^{i} \frac{n_j - d_j}{n_j} \tag{2.8}$$

where d_j is the number of events at time t_j and n_j is the number of subjects "at risk". Note that the estimator $\hat{S}(t_i)$ drops only at times when a failure has been observed, not at times when censoring occurs.

$$\hat{\Lambda}(t_i) = \sum_{j=1}^{i} \frac{d_j}{n_j} \tag{2.9}$$

Intuitively, this expression is estimating the failure at each distinct time of event t_j as the ratio of the number of failures d_j to the number of components "at risk" n_j. Therefore $\hat{\Lambda}(t_i)$ estimator is an increasing right-continuous step function with increments $\frac{d_j}{n_j}$ at the observed event time. The components that are censored are not counted as "at risk".

2.2 PQ analysis using SMs data

The data from SMs provides a number of benefits for electricity users such as managing energy consumptions to reduce electric bills [46]. For utilities, on the other hand, employing SMs data to increase efficiency on the delivery side of the power grids (low-voltage grid) has became an important topic worldwide [47]. The main reason lies in the extension of SMs installed in the grid and in the possibility of collecting information directly at the points of customer connection.

Several projects have been established on SMs data analytics by different institutes such as National Science Foundation (NSF) [48], Innovation Center in Denmark (CITIES) [49], Bits to Energy Lab [50], and ESSnet Big Data [51]. The main focuses in these projects are smart grid big data analytics, machine learning techniques for SMs data, SMs data analytics, and methodologies for SMs data analytics, respectively. In these projects, the application areas include load analysis, load forecasting, load management, real-time pricing, customer segmentation, asset management, energy trending, theft detection, demand respond program, and marketing.

Most of the recent SMs data analytics are based on usage data. However, the SMs are able to collect additional features such as monitoring voltage, current, frequency, and more interestingly, PQ. SMs, distributed throughout the low-voltage grid, allow for continuous collection of PQ events.

In general, PQ is described by a set of features that quantify the deviation of current and voltage waveforms from a perfect sinusoid [52]. These features include power system harmonics, flicker, dips and swells in the voltage, transients, frequency variations and voltage unbalance between the power system phases [52]. In power grids, voltage sags and swells are important because the increase and decrease in RMS voltage can cause malfunction in electrical equipment [53].

In the low-voltage distribution grids, many disturbances are caused by consumers. These disturbances can propagate to other customers and have impact on PQ. Lack of high-frequency sensors in the distribution part of the grid makes detection and modeling of the propagation of these power quality disturbances (PQDs) difficult.

Table 2.1: The sub-categories of the articles related to PQ

Name of the Category	Number of articles	References
PQ classification and localization	2	[58, 59]
PQ monitoring	12	[54, 55, 56, 57, 60, 61, 62, 63, 64, 65, 66, 67]
Impact of different components on PQ	8	[68, 69, 70, 71, 72, 73, 74, 75]
Harmonic compensation for microgrids	3	[76, 77, 78]

Here, our goal is to demonstrate that data coming from SMs, despite being sparse, can be used for identifying areas with PQ problems and analyzing the propagation of PQDs in the low-voltage grids.

To provide an overview of the existing research in PQ analysis using SMs data, a bibliometric investigation was conducted on 26 October 2018 using the Scopus database. The following query was used to search for relevant articles: TITLE-ABS-KEY ("smart meter") AND TITLE-ABS-KEY ("data") AND (TITLE-ABS-KEY ("power quality") OR TITLE-ABS-KEY ("alarm") OR TITLE-ABS-KEY ("sag") OR TITLE-ABS-KEY ("swell") OR TITLE-ABS-KEY ("harmonics") OR TITLE-ABS-KEY ("disturbance") OR TITLE-ABS-KEY ("propagation")). In total 153 articles (journal and conference papers) were found which were published after 2009.

By scanning through the title, abstract, and keywords, the articles are grouped into 12 different categories: PQ, Communication, Energy management, Design, Appliance load monitoring, Data management, Faulty data detection, Fault localization, Electricity theft detection, Reliability, Review and Other topics. We conducted review of the articles related to PQ and grouped them into four sub-categories (see Table 2.1).

Among the identified articles, the articles related to PQ classification, localization, and monitoring were reviewed.

Papers [54] and [55] mainly just describe the potential and importance of SMs for aspects related to smart grids observability, control, automation, and outage identification in the grid. The authors of articles [56] and [57] went further than just describing the benefits of SMs. In these papers, even-though the authors emphasize the capabilities of SMs for PQ monitoring, they are not using real data from conventional SMs. In [56] the analysis is performed in a laboratory scale simulation using one SM, and in [57], the authors developed their own "intelligent meter".

In articles [60, 61, 62, 63, 64], the analysis are waveform-based and mainly consider the shape of the waveforms in the three-phases with precision in milliseconds. In [60] the authors present a method for estimation of time-varying harmonics of voltage/current signals. In [61] a web service, based on processing the waveform of voltage sag, recorded by SMs is proposed. In [62] a method for measuring the synchronous condition of voltages and power in a grid with high share of renewable generation sources is presented. Their analysis is performed on the waveform of voltage, active power and reactive power over the intervals of 60 and 900 seconds. The authors in [63] proposed a method for phase-balancing and voltage quality enhancement using load curves. In this study, they used $11,000$ individual load curves from low-voltage grids. In [64] an algorithm based on Hilbert Transform is used to detect voltage sags in synthetic dataset.

In [58] the authors proposed a PQDs classification method by using synthetic waveform signals from SMs. Then, in [59], they proposed a feature extraction method for analyzing the propagation of PQDs. The features are extracted from SMs' voltage signal and are used for localizing the region where the voltage sag source is located. The location of voltage sags are determined by analyzing the features of the RMS waveform. In this work, they used the tree-structure of the grid topology for the propagation analysis. The method was validated by using a simulation model of IEEE 13-Bus Test Feeder [79] (13 meters at 13 nodes), with simulated single-phased short-circuit fault. The results show that their method is capable of localizing the area of the voltage sag source. However, in this work the authors considered special type of SMs that can capture the voltage RMS waveform signal. Moreover, the work is evaluated only in the simulated scenarios.

In [65] and [66], the authors employed conventional SMs for analyzing propagation of PQDs and proposed an integrated power quality monitoring system that collects PQ data from different sources in a distribution grid such as Supervisory Control and Data Acquisition (SCADA), PQMS, and SMs. The data is used for analyzing propagation and the root source detection of PQDs. Furthermore, the benefits and challenges of using SMs for PQ monitoring are discussed. However, in these works, the authors only considered a limited amount of SMs to collect the data from. They analyzed 5 SMs located at a School, Shopping center, Administration building, Residential building, and Petrol station. Moreover, in their analysis the authors only considered one specific propagation scenario: cable failure in the grid where the voltage drops to zero volt. Therefore, PQ deviations near the norms are not analyzed.

Chapter 3
Methodology and Results

This chapter presents our proposed data-driven method for weak spot detection, using available data in SDGs. This methodology contains three steps: fault analysis, reliability ranking of power cables, and SMs data analysis. These steps are illustrated in Figure 3.1 and explained in the following.

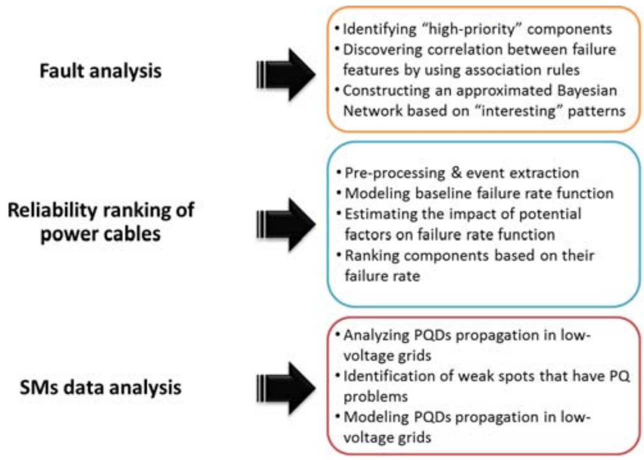


Figure 3.1: The procedure of the proposed data-driven method for weak spot detection in SDGs.

3.1 Fault analysis

For this step, mainly the *failure records* data is utilized. This data source contains information about events, cessation of a system or components' ability to perform its required function.

3.1.1 Method

In the fault analysis step, a preliminary investigation of event history is used to identify “high-priority” components. Components that fail more frequently, cause long outages, are more costly to repair, or affect large number of customers can be considered as high-priority components. Furthermore, potential factors that have impact on failures and the correlations of these factors with each other are specified in this step.

In general, utility companies keep records of previous faults which contain features describing the event. To analyze failures, it is important to discover which failures have common features, e.g., if there are any types of failures that happen mostly in certain parts of the grid or at a certain time. One approach to discover these correlations is employing association rules [80, 81, 82, 83]. Association rules are based on the frequency of the co-occurrence of features and conditional dependency between them.

The objective of mining association rules is to find the most frequently occurring combinations of features. Let $I = \{I_1, I_2, ..., I_m\}$ be a set of features (items). An association rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, and A, B are disjoint itemsets, i.e. $A \cap B = \emptyset$. In this case the itemset $A = \{a_1, a_2, ...\}$ is the prior and the itemset $B = \{b_1, b_2, ...\}$ is the posterior of the rule. Now assume that $X = \{x_1, x_2, ..., x_n\}$ is a set of random variables representing the list of observations (failures) in a dataset. Each observation x_i in the dataset X may or may not contain a specific item, e.g., $x_1 = \{I_1, I_2, I_5\}$ only contains items I_1, I_2, I_5.

The interestingness of an association rule $A \Rightarrow B$ is often expressed in terms of support, confidence, and lift.

- The *support* of a rule is the percentage of observations in the dataset that contain both A and B.
- The *confidence* of a rule is the percentage of examples containing A that also contain B. In other words, a fraction that shows how frequently B occurs among all the observations containing A.
- The *lift* of a rule is a ratio of the confidence of the rule to the frequency of observations containing B. It is a value between 0 and infinity that measures the deviation of a rule from statistical independence.

The confidence value of each association rule corresponds to the strength of the conditional dependence between features. Therefore, these confidences can be used to automatically build a Bayesian Network.

Bayesian Networks [84, 85, 86] are graphical representation of probabilistic relationships over a set of variables, constructed using probability distribution over a set of variables in a dataset. If we consider features of failure events as probabilistic variables, a Bayesian Network captures the conditional relations between those features over a set of events.

An association rule $A \Rightarrow B$ can be seen as a connection from one itemset to another. If $I = \{I_1, I_2, ..., I_t, ..., I_m\}$ is a set of features such that $A = \{I_1, I_2, ..., I_t\}$

and $B = \{I_{t+1}, ..., I_m\}$, the Bayesian Network representation for all the connections between feature set A and B is shown in Figure 3.2, where items in set A are parent nodes and items in set B are child nodes.

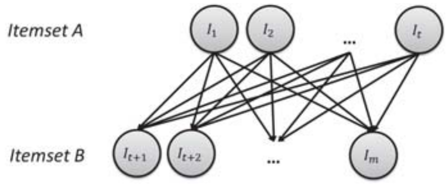


Figure 3.2: The Bayesian Network representing association rule ($A \Rightarrow B$)

If the itemset A is the parent of itemset B, the joint probability distribution represented by the network can be written as:

$$P(I_1, I_2, ..., I_t, ..., I_m) = \prod_{i=1}^{t} P(I_i) \prod_{j=t+1}^{m} P(I_j | I_1, I_2, ..., I_t) \tag{3.1}$$

Each of the terms $P(I_j | I_1, I_2, ..., I_t)$ corresponds to the confidence of the rule $\big((I_1, I_2, ..., I_t) \Rightarrow (I_j)\big)$.

3.1.2 Results

Table 3.1 shows the relative frequency (in percent) of different causes of failures (top) and the faulty components (bottom) from 2009 until the end of 2014 at HEM Nät [1]. According to these tables, the most common failure during these years is caused by “Fabrication fault” with a frequency of 34.6%. The mean-time-between-failures or MTBF refers to the amount of time that elapses between one failure and the next. To calculate MTBF, the total length of time (in here the number of days from 2009 until 2015) is divided by the total number of failures of the same type [87]. According to the MTBF presented in Table 3.1, the “Fabrication fault” had occurred in average every 5.7 days.

According to HEM Nät, any fault that is related to aging of components is recorded under label “Fabrication fault”. The “affected component” statistics show that underground feeder cables are one of the most common faulty components in the grid, with a frequency of 27%. Among these feeder cables, breakdown in PILC cables and joints because of aging is the most common cause of failure.

The association rules with “high” support and confidence, which also have lift greater than 1, are considered as interesting rules. One example of the identified “interesting rules” is $Fusebreak \Rightarrow UngPillar$. According to its confidence ($C = 79\%$),

[1] A distribution system operator (DSO) in the south of Sweden

Table 3.1: Statistics of the failure records dataset for the time interval between 2009 and 2015

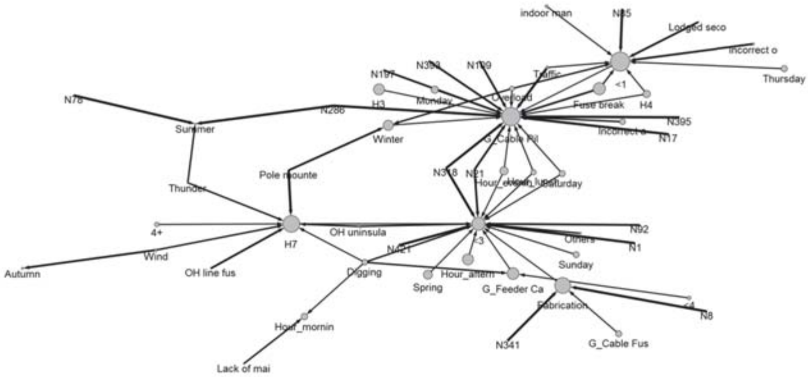
Type of Failure	Cause of Failure	Frequency(%)	MTBF(days)
Operational Failure (846 failures)	Fabrication fault	34.59%	5.7
	Fuse break	24.95%	7.91
	Incorrect installation	7.12%	27.72
	Overload	5.59%	35.32
	Incorrect operation	1.35%	136.88
	Lack of maintenance	1.44%	146
	Others	1.17%	168.46
Non-Operational Failure (227 failures)	Digging	14.41%	13.69
	Traffic	1.71%	115.26
	Weather	3.42%	57.63
	Animal	0.72%	273.75
	Others	0.18%	1095
Type of Failure	**Affected Component**	**Frequency(%)**	**MTBF(days)**
All Type of Failure (1110 failures)	Underground cable pillar	48.11%	4.1
	Underground feeder cable	26.94%	7.32
	Underground cable fuse	10.09%	19.55
	Concr.sec.substation indoor man	4.32%	45.63
	OH uninsulated free line	2.70%	73
	Others	7.84%	25.17

we can interpret this rule as: *the probability that an underground cable pillar is the affected component knowing that the cause of failure is fuse break is* 79%.

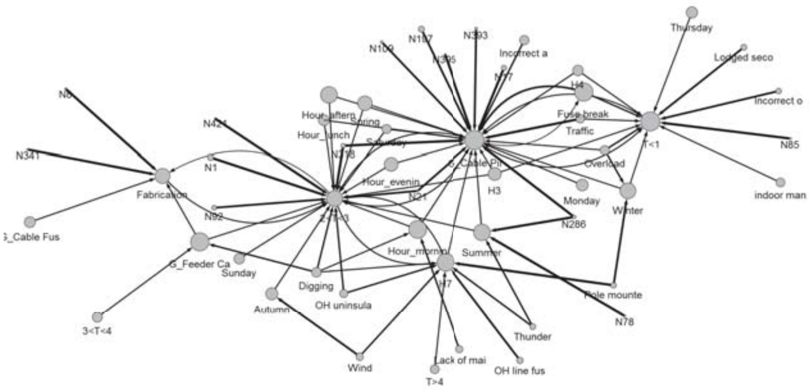
Finally, we use association rules of length two for constructing Bayesian Network (approximated Bayesian Network). For this purpose the lists of priors and posteriors of each rule correspond to the network nodes, and the confidence of the rule (conditional dependency) corresponds to the connections between nodes. Figure 3.3 illustrates three networks constructed from the association rules but with different thresholds for confidence and support. In **Paper B** we show that the Bayesian Network constructed based on the interesting rules of two items is a good approximation of the real dataset.

The methodology for discovering failure patterns and failure statistics, presented in **Paper B**, was also applied on historical data from Öresundskraft, Göteborg Energy, and Växjö Energy. These companies are DSOs which are located in Sweden. The results show that the most vulnerable component in these distribution grids is either underground cable or overhead line. Furthermore, interesting failure patterns are discovered using association rules and represented by Bayesian Networks.

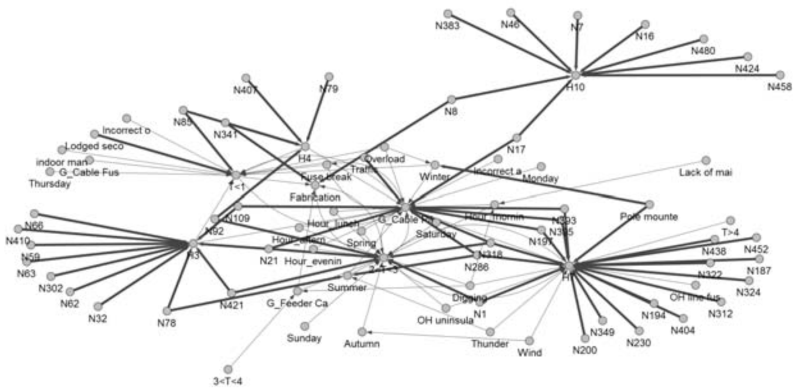
One of the constrains in automatically constructing approximated Bayesian Network from association rules is the manual threshold setting for confidence and support. For HEM Nät data, the fully connected failure network contains 859 connections between all the nodes from different categories. However, we assume that some of the



(a)



(b)



(c)

Figure 3.3: Approximated Bayesian Network representation using different threshold values. These figures show how the complexity of the network varies depending on the thresholds for confidence and support.

items are independent or the dependencies can be neglected, since they are very weak (confidence and support smaller than a certain threshold). In fact, the selected thresholds for confidence and support specify whether to consider a rule as an "interesting rule" or not. This manual setting can be tuned by considering the complexity and accuracy of the network. Figure 3.4 shows the relation between the complexity of the Bayesian Network (as the number of connections between nodes) and the thresholds for confidence and support. The darker the area, the higher the thresholds and the smaller number of connections. If the thresholds are too small, the number of connection in the network is high and consequently interpreting the result will be difficult. In the other case if the thresholds are too high, the number of connections is too small to capture the "interesting" correlations.

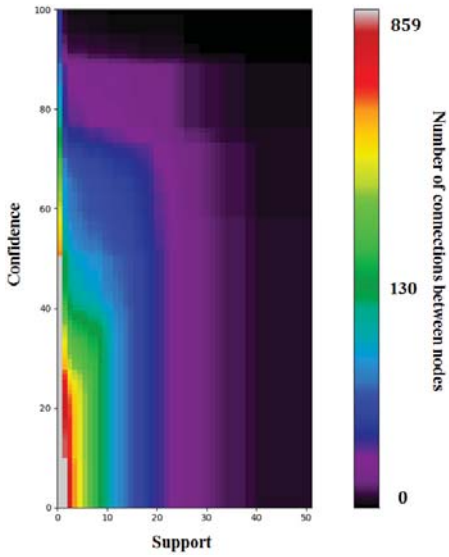


Figure 3.4: From the historical failures in the network information system dataset, 859 pairwise association rules are generated corresponding to 58 nodes. This figure shows the relation between the complexity of the approximated Bayesian Network and the thresholds for confidence and support. The darker the area, the higher the thresholds and the smaller number of connections.

The procedure of constructing Bayesian Networks can also be used for other type of databases. Figure 3.5 shows the result of Bayesian Network, constructed based on event history in *PQMS* database. This dataset contains information about recorded events occurred in one main station with two transformers. In this figure, the thicker connections represent high confidence (above 80%). Several observation can be made from this network. For example, most of the events of type "sag" happened during Saturday, Sunday, Hour night (between $23:00$ and $07:00$), or Summer. The transformer $T31$ has more recorded events during Autumn and the transformer $T32$, during

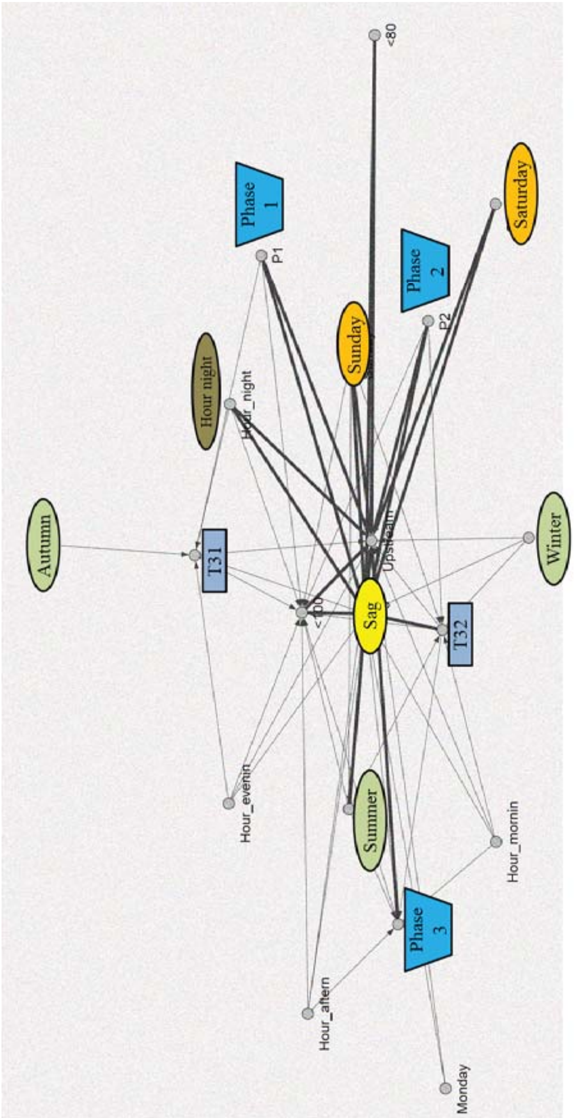


Figure 3.5: Bayesian Network representation constructed based on the event history of the *power quality measurements system* database.

Summer. This representation of the events provides a visualization of the failure patterns.

The results enable distribution companies to discover failure patterns and accordingly mitigate the conditions that increase the probability of failures.

3.2 Reliability ranking of power cables

The main focus of this step is modeling the failure rate of underground cables (the identified high-priority component from the previous step) and ranking the cables based on their failure rate.

3.2.1 Method

Our methodology for reliability analysis contains four major steps: pre-processing and event extraction, modeling baseline failure rate function, estimating the impact of relevant factors on failure rate function, and ranking components based on their failure rate.

Here, in addition to the data from *failure records* (including previous faults and sub-station maintenance history), the data from *grid information* (including cable inventory) are used. The *cable inventory* dataset contains historical information about the in-service cables that have been installed since 1929. Each cable is described with an ID and the unique feeder line name to which it belongs, as well as additional information such as insulation type, conductor size, length, etc. The *sub-station maintenance history* contains information about the previous maintenance carried out on the sub-stations and connected feeder lines.

In order to estimate failure rate for underground feeder lines, taking into account their age, we require information about failure time of each cable. The previous failure data, however, does not contain the identifiers such as id, type, or age of the cables. In addition, the amount of failure data is limited to about 15 years. Therefore, other database such as *cable inventory* is considered and methods that can be used for feeder lines failure extraction are explored.

Usually in case of failures, only the faulty part will be replaced by a new short segment and the rest of the cable stays untouched. Therefore, we make the assumption that *short* cable sections (length smaller than 20 meters) in any given line are artifacts of previous repairs. The failure is assumed to have taken place in the year of the installation of the short cable section, and to take place in the oldest cable within this line. Therefore, the year difference between the installation of the oldest cable and the new *short* section is considered as the *age* of the cable, when the failure occurred.

Based on this assumption alone, some feeder lines would be incorrectly linked to a few extra failure events. For example, several short cable sections may be installed because of upgrading a sub-station, and not because of a failure. In order to eliminate these extra failures, we exploit the *sub-station maintenance history* database. According to this database, some of the lines have short sections because of a maintenance on the sub-station and not because of a failure. Those assumptions are not fully accurate, but we have confirmed through discussions with domain experts that they are realistic. By combining these two databases, it is possible to identify id, type, and age of the feeder lines when the events have occurred.

Baseline failure rate is a function that estimates the variation of failure rate over age ($\lambda_0(t)$). In estimating the baseline failure rate, only the age of cables is considered as a factor that impacts failure rate. In **Paper A**, a methodology for estimating the baseline failure rate while dealing with limited amount of failure data is explained. It is important to note that, while goodness-of-fit (GOF) results can be compared directly, it is often difficult to properly interpret the results, especially when the data is of limited quantity. Therefore we propose a method for interpreting the results of GOF measures with confidence intervals, estimated using synthetic data. Five different models are used to fit with the empirical data. For each model, a number (e.g. 100) of synthetic data sets are generated by drawing random points from a normal distribution with mean equal to the failure function at each age and variance computed from the empirical failure rate data points. Then the GOF of the synthetic data are computed in compare with all other models to determine how well a data generated from one model can be fitted by another model. These comparisons help us to draw conclusions about how well each model fits the empirical data points.

One way to estimate the impact of additional factors on failure rate function is to use a Proportional Hazard Model (PHM). PHM is a statistical regression model which was first introduced by Cox in 1972 [88]. PHM is based on the assumption that the failure rate of a system or component consists of two multiplicative coefficients: the baseline failure rate λ_0, and an exponential function, capturing the effect of explanatory factors:

$$\lambda(t, \boldsymbol{X_t}) = \lambda_0(t) \cdot exp\,(\boldsymbol{\beta} \cdot \boldsymbol{X_t}), \tag{3.2}$$

where $\lambda_0(t)$ is the baseline failure rate that is dependent on time t, $\boldsymbol{X_t}$ is a row vector representing the factors at time t, and $\boldsymbol{\beta}$ is a column vector representing the regression parameters. The vector of factors $\boldsymbol{X_t}$ can be time-dependent or time-independent [89]. The vector $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_n]$ corresponds to the weight or impact of different factors. β_i could be positive (positive correlation with the failure), or negative (negative correlation with the failure). If $\beta_i = 0$ then the expression $exp\,(\beta_i \cdot \boldsymbol{X_t})$ for the ith factor is equal to 1, meaning that the factor does not affect the failure rate.

In this work, for simplicity, the factors are assumed to be time-independent. Therefore, equation 3.2 can be simplified as follows:

$$\lambda(t, \boldsymbol{X}) = \lambda_0(t) \cdot exp\,(\boldsymbol{\beta} \cdot \boldsymbol{X}), \tag{3.3}$$

where $\lambda_0(t)$ depends on time but not the factors, and $exp\,(\boldsymbol{\beta} \cdot \boldsymbol{X})$ depends on the factors but not time. $X = [X_1, X_2, ..., X_n]$ is a row vector corresponding to the factors that can have impact on cable failure.

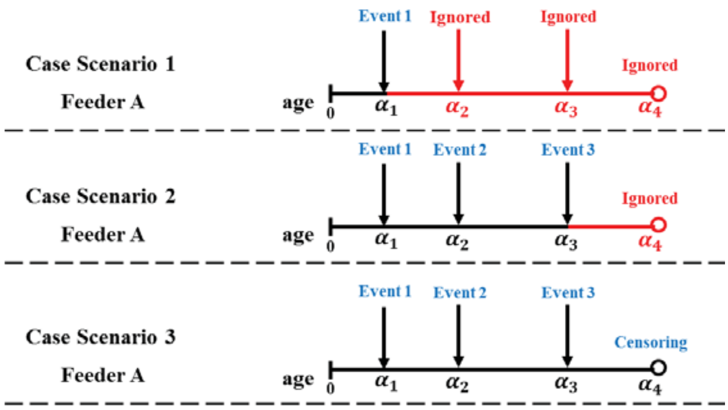


Figure 3.6: An illustration of the three case scenarios.

An estimate of the β_i values, without making any assumption about the baseline failure rate, can be calculated using maximum likelihood [90].

After modeling the baseline failure rate and measuring the influence of different factors, we calculate the failure rate for each individual feeder line. The feeder lines are then ranked from the highest failure rate to lowest. The feeder lines with higher rank indicate greater vulnerability and the need for remedial actions.

As mentioned in the Chapter 2, it is important to determine if a component is repairable or non-repairable. To show how this would impact the results of cable ranking and conclusions about different factors, we consider three case scenarios (see Figure 3.6).

Case scenario 1 in this scenario the feeder lines are considered as non-repairable components, meaning that after one failure they are decommissioned. For the cables with more than one failure only the time-to-the-first failure is used, and the feeder line is then removed from the list of "in-service" feeders. For example, if a feeder line A has three failure at ages α_1, α_2, and α_3, this line will be removed from the list of in-service feeders after the first failure α_1. This failure type is what is often simulated by laboratory stress tests.

Case scenario 2 in this scenario the feeder lines are considered as repairable components and can have more than one failure. Only after their last recorded failure the feeder lines are removed from the list of in-service feeders. For the feeder line A in the previous example, the line will be removed from the list of in-service feeders after the third failure at age α_3, and the censoring time α_4 will be ignored.

Case scenario 3 in this scenario the feeder lines are considered as repairable components and even after the last recorded failure the feeder lines are still in-service until their censoring time. This type of analysis is also known as recurrent event analysis [91]. For the feeder line A in the previous example, all the failure times α_1, α_2, and α_3 and the censoring time α_4 will be considered. This scenario is more realistic for underground power cables compared to the other two scenarios because failures can occur more than once in a particular feeder line, and after a repair, the cables usually go back into service.

Here, we are introducing three different cases which might be used by researchers for reliability analysis of underground cables. When analyzing the long time history of power cables' failures, the case scenario 1 is incorrect because it does not take into account the repair of cables. Case scenario 2 is also incorrect because it does not consider the survival time after the last failure, but we know that many cables are continuously repaired and stay in service. However, the case scenario 3 is more appropriate given the way that utility companies use the power cables and keep them in service.

3.2.2 Results

In **Paper A** we investigate five different models (linear, piecewise linear, exponential, constant, and piecewise constant) to model baseline failure rate and evaluate how well, each model fits empirical failure rate data points. We interpret the results by comparing the obtained GOF measures with *expected* GOF and *confidence intervals*, estimated using synthetic data. Observe that, we do not specifically consider the "infant mortality" period in this analysis.

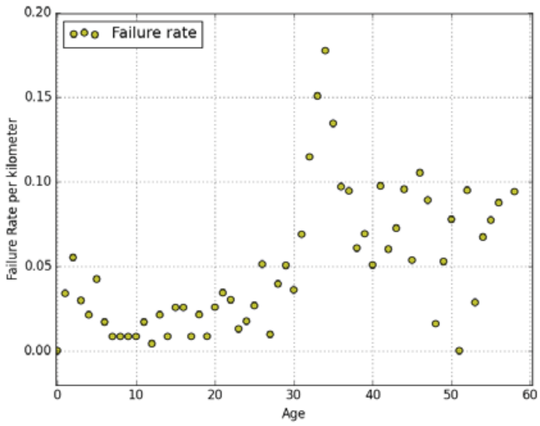


Figure 3.7: Empirical failure rate per kilometer as function of age, for high voltage PILC cables.

The result of calculating empirical failure rates at each age, for high voltage PILC cables, is shown in Figure 3.7. In this figure the failure rate at ages between 32 and 35 have very high values (unexpected values compare to other data points). Our investigations show that some feeder lines are incorrectly linked to a few extra failure events because of upgrading a sub-station, and not because of a failure. Therefore, we eliminate these extra failures for further analysis.

After removing the superfluous failures, we fit six different models with the empirical failure data points. The result of fitting these models are shown in Figure 3.8 and the GOF measures presented in Table 3.2.

Note that, in Figure 3.7 the empirical failure rate is calculated based on the number of failures per year per *kilometer of in-service cables*; however, in Figure 3.8 the failure rate is estimated by the number of failures per year per *number of in-service cables*.

According to the Table 3.2, the constant and piecewise constant models (models A and B) are statistically different from the rest of the models. Furthermore, there is no statistically significant difference between GOF of the data points, neither the empirical nor synthetic, between linear, piecewise linear, exponential, and Weibull models. This indicates that these four models are virtually identical.

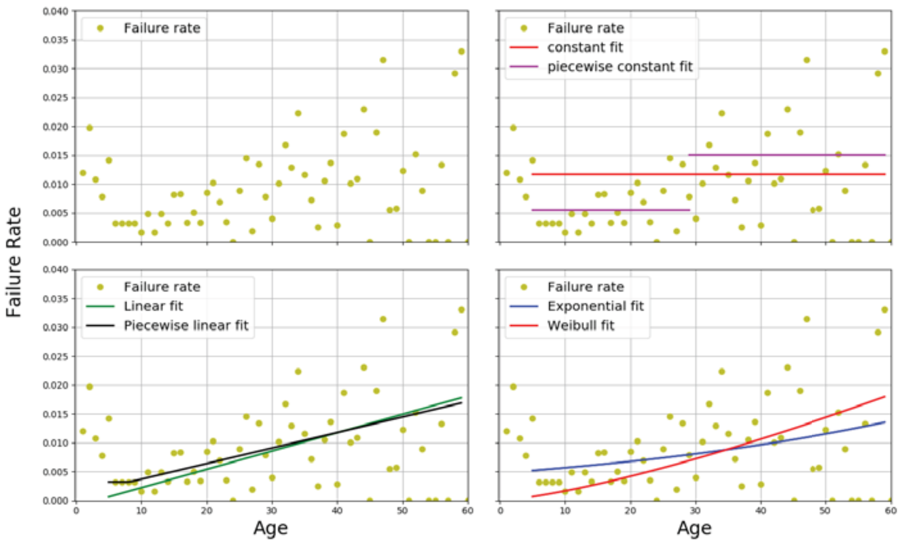


Figure 3.8: Six different failure rate models, fitted to the empirical data after removing the incorrect failures.

In **Paper C** we discuss that, when analyzing the long time history of failures for power cables, the first and second scenarios are incorrect, and conclusions about different factors in PHM and cables ranking will be misleading if they are used.

Table 3.3 shows the ranking results for the 10 highest ranked feeder lines based on case 1. Then the corresponding ranking values and ranking positions for these lines according to case 2 and 3 are added for comparison. Accordingly, for some feeder cables such as $H200$ the ranking position is the same in all three cases; some other lines such as $H105$ have slightly different position; and finally, there are some feeder lines such as $H996$ which are ranked completely different (in positions 10, 18, and 111 for cases 1, 2, and 3, respectively). It is not very important if the line is ranked in

Table 3.2: Goodness-of-fit measurement by using MSE test

(1+e3 mse)	Data A	Data B	Data C	Data D	Data E	Data F	Real Data
Constant (Model: A)	0.0162 ±0.0027	0.0387 ±0.0058	0.0451 ± 0.0063	0.0377 ± 0.0063	0.0302 ± 0.0048	0.0522 ±0.0077	*0.0704*
P.W. Constant (Model: B)	0.0411 ±0.0062	0.0239 ±0.0042	0.0234 ± 0.0047	0.0216 ± 0.0040	0.0261 ±0.0042	0.0240 ± 0.0042	*0.0648*
Linear (Model: C)	0.0572 ±0.0077	0.0318 ±0.0052	0.0190 ± 0.0038	0.0201 ± 0.0036	0.0233 ± 0.0045	0.0167 ± 0.0034	*0.0570*
P.W. Linear (Model: D)	0.0426 ±0.0063	0.0267 ±0.0051	0.0179 ± 0.0032	0.0175 ± 0.0033	0.0189 ± 0.0037	0.0176 ± 0.0038	*0.0537*
Exponential (Model: E)	0.0349 ±0.0060	0.0335 ±0.0053	0.0253 ± 0.0039	0.0230 ± 0.0038	0.0164 ± 0.0035	0.0247 ± 0.0040	*0.0548*
Weibull (Model: F)	0.0613 ± 0.0079	0.0392 ± 0.0055	0.0216 ± 0.0039	0.0246 ± 0.0045	0.0244 ± 0.0047	0.0187 ± 0.0038	*0.0598*

Table 3.3: Part of the ranking results for HV cables

Id	Case 1		Case 2		Case 3	
	Rank	FR	Rank	FR	Rank	FR
H200	1	0.063	1	0.127	1	0.128
H306	2	0.035	2	0.052	3	0.039
H784	3	0.03	3	0.044	5	0.032
H105	4	0.025	5	0.035	8	0.025
H843	5	0.023	7	0.031	6	0.031
H205	6	0.02	6	0.033	7	0.026
H971	7	0.019	4	0.036	4	0.038
H206	8	0.018	8	0.029	12	0.023
H996	9	0.017	13	0.024	56	0.017
H914	10	0.015	18	0.021	111	0.014

position e.g. 4, 5, and 8 in the three lists, but the very high difference between ranks e.g. 10 and 111 (changing from a very high rank to a very low rank) is significant and can not be neglected.

When considering different factors, we believe that the amount of energy consumption has also impact on failure rate. To investigate this, we need to find the load information of the energy consumers who are connected to each feeder line. From the data available in the *customer information* and *grid topology* it is possible to extract the number of customers and the *annual energy consumption* of the customers who are connected to HV lines. Then we used this information in the PHM. Contrary to the expectations, the result does not show any significance impact of these factors on failure rate. This shows that the total annual energy consumption is not a good representation of the important patterns in the energy usage. To capture the impact of such

Table 3.4: Three Faulty HV feeder lines and their corresponding rank

	Rank case 1	Rank case 2	Rank case 3
Cable 1	80 (% 17.5)	88 (% 19.2)	52 (% 11.4)
Cable 2	103 (% 22.5)	102 (% 22.3)	67 (% 14.6)
Cable 3	78 (% 17.0)	63 (% 13.8)	45 (% 9.8)

patterns we need more detailed information such as hourly consumption or PQ data that can be collected from smart meters.

To evaluate the results of the ranking lists, created based on the three case scenarios, we consider two options. The first option is to search for recent failures in the grid and look into the position of the faulty feeder lines in each list. The other option is to perform some tests such as Time Domain Reflectometry (TDR) on the high-ranked lines.

TDR is a method to localize the faulty part of a feeder line by sending a low-energy signal through the line. The printout of TDR, also known as "trace", is a graphical representation of the return signal which gives an approximate location of impedance variations.

During the four month period after creating the list, three failures (caused by aging of the cables) have occurred in HV feeder lines. The ranks of these three lines based on the case scenarios are shown in Table 3.4. Among 458 HV feeder lines these failures have happened on the lines which are ranked in the first 15% of the case scenario 3 list. These three examples are not enough to qualify the performance of the ranking approach but it gives us some real observation of faulty lines and their position in the ranking lists based on the three cases.

From the ranking list of case scenario 3, the top 20% of the high ranked feeder lines are selected. In addition, to specify which line should be selected for measurement and test we consider two other factors: number of customers, and average annual consumption of customers who are connected to each feeder line. Then we used Pareto frontier which is defined as the graphical representation of the tradeoffs within a set of parameters. The result is shown in Figure 3.9. In this figure, the data point correspond to the feeder lines. The bigger the data point, the higher the failure rate. Based on the Pareto frontier (tradeoff between the number of customers and the average annual consumption) and discussion with experts from the company, five feeder lines are selected for testing.

3.3 SMs data analysis

SMs are distributed throughout the low-voltage side of the grid, but currently they are only used for billing electricity consumption. Our ultimate goal is to find correlation between SMs data (load profile and power quality data) and cables' reliability. This

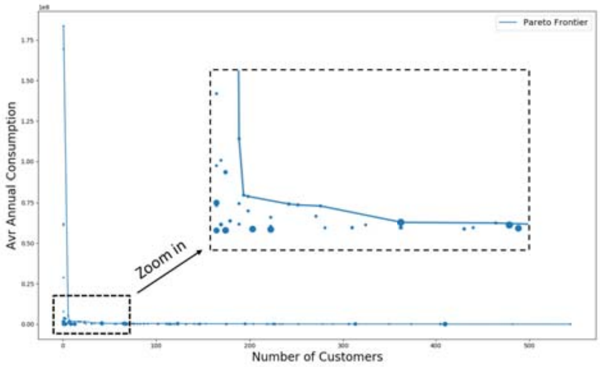


Figure 3.9: Tradeoff between the number of customers and the average annual consumption for the first 20% cables in the case scenario 3 ranking list.

requires reconfiguration of SMs to collect more data, characterization of disturbances in low-voltage grid, and investigation of methods for discovering and modeling the impacts of disturbances on other components in the grid.

In addition, propagation analysis and modeling of PQDs is important especially in the context of the future smart distribution grids. In smart grids, many new consumers and producers are continuously being connected to the grid, which will impact PQ in various ways.

In this section, the methodology for detecting the propagation of PQDs in the low-voltage grids using SMs alarm data is presented. Then, several scenarios for modeling the propagation of disturbances are introduced.

3.3.1 Method

In SMs, whenever the PQ parameters are outside the configured thresholds, an "alarm" will be recorded. The precision of these recorded alarms in SMs is in seconds. Therefore, many alarms are recorded at exactly the same time, or within one second of each other. This makes it difficult to distinguish the sequence of SMs recording the alarms. In addition, it is not possible to rely on exact synchronization of clocks in different SMs. Therefore, *burst* of alarms is used instead of sequence of alarms. We defined a burst as series of SMs for which the time difference between their alarms' time-stamp is less than or equal to one second.

A burst represents the *co-occurrence* of alarms for a set of customers. This set of customers may have triggered the alarm: (a) randomly at the same time (independent to each other), (b) because of the same cause (propagation from a faulty component to customers), or (c) one after the other in the cascading order (propagation from one customer to others) but because of the low time resolution in SMs, they are recorded together. All of these conditions, except the random co-occurrence, are important for propagation analysis.

In order to test whether the SM alarms within the bursts correspond to propagation of PQDs (alarms are dependent) or they are co-occurred randomly (alarms are independent), the relative positions of the SMs are considered. If those SMs are equally likely to be located far away as they are to be close, then one can conclude that the co-occurrences are random. If, however, close customers are co-occurring more often than distant ones, one can conclude that PQDs propagate.

We define the null-hypothesis as: every customer generates alarms independently of other customers. In this case, the alternative hypothesis is: they are not independent, i.e., alarms can propagate from one customer to another or they are triggered by the same cause.

To test the null-hypothesis two sets of data are considered and their corresponding distributions are compared using the two-sample Kolmogorov-Smirnov (KS) test. These two probability distributions are the co-occurrence probabilities of neighbor customers and non-neighbor customers, among all the bursts. We consider the neighbor customers to be the ones which are connected to the same branch in the tree-structure of the grid topology; and, the non-neighbor customers to be the ones which are connected to different branches.

If every customer generates alarms independently, i.e., alarms do not propagate, we expect that the statistical test does not show a significant difference between the two distributions. On the other hand, if alarms propagate to neighbor customers, we expect the difference between the two distributions to be significant.

Let $B = \{B_1, B_2, ..., B_k\}$ be the list of branches in a tree-structure of a sub-station, and $C = \{c_1, c_2, ...c_m\}$ be the list of customers in the sub-station. Every burst o_i in the list of bursts $O = \{o_1, o_2, ..., o_n\}$ represents the co-occurrences of alarms for a set of customers. For an example, the burst $o_5 = \{c_3, c_7, c_9\}$ implies that customers c_3, c_7, and c_9 generated alarms relatively at the same time. The probability $p(c_x)$ is defined as the percentage of bursts containing customer c_x.

$$p(c_x) = \frac{|\{o \in O : c_x \in o\}|}{n}, \tag{3.4}$$

where $|\{o \in O : c_x \in o\}|$ is the number of bursts containing customer c_x.

The joint probability $p(c_x, c_y)$ represents how frequently customers c_x and c_y co-occurred over n:

$$p(c_x, c_y) = \frac{|\left\{o \in O : c_x \in o \wedge c_y \in o\right\}|}{n}, \tag{3.5}$$

where $|\left\{o \in O : c_x \in o \wedge c_y \in o\right\}|$ is the number of bursts containing both customers c_x and c_y.

For each customer c_i from branch B_j, we calculate the pairwise probability $p(c_i, c_x)$ for all the customers c_x, where c_x is from the same branch B_j. The distribution of these probabilities is then compared with the pairwise probability $p(c_i, c_y)$, where c_i and c_y are from different branches.

If the distribution of these two sets of probabilities are not significantly different from each other, then one can conclude that the customers generate alarm independently.

$$H_0 : p(c_i, c_x) = p(c_i, c_y),$$

where $c_i, c_x \in B_j$ and $c_y \notin B_j$.

To compare the two distributions, the KS test is used. This test is nonparametric, i.e., it does not assume any underlying distribution. The KS test compares the maximum distance between the cumulative distribution (CDF) of the two data sets.

The intuition behind modeling propagation with artificial data is to produce data based on a specific model and then compare it with the real data. By repeating the experiment with different models and comparing the results, one can find a model that best represents the real data. Then this model can be used for reasoning about the characteristics of the alarms propagation in the low-voltage grids.

When generating synthetic data, we assume that the probability of observing an alarm for each customer within one second interval (at each time-stamp) has a Bernoulli distribution $x \in Bern(p)$. The amount of artificial data to be generated is the same as the real data available.

The following cases based on different conditions are considered. These cases differ in assigning the baseline probability for the Bernoulli distribution.

1. In the first case, the probability of alarm for each customer in the artificial data is considered to be the same as their probability of alarm in the real data.

2. The second case assumes all the customers have the same baseline probability in generating alarm.

$$p_{c_i} = p_{base}, \text{for all } c_i \in C$$

3. Third, most of the customers have the same baseline probability, except for those in a single designated branch. This condition is included because of observing similar behavior in the real data, namely, we noticed that in one of the sub-stations the customers in one branch have relatively higher probability of alarm. This scenario is considered to model this particular behavior.

$$p_{c_i} = \begin{cases} p_{base1} & \text{if } c_i \in B_j \\ p_{base2} & \text{if } c_i \notin B_j \end{cases}, \text{for all } c_i \in C$$

4. Finally, every customer generates alarm with different baseline probability.

A greedy optimization algorithm is used to find the best values of the parameters. In these cases, the search algorithm starts with a small initial value as the p_{base} and calculate the distance (maximum distance between the CDFs) between the generated data and the real data. This distance is captured from KS test and is used as an indication of how far the two distributions are from each other. The search continues by increasing the value p_{base} in an iterative manner to find the probability parameter that minimize the distance between the two distributions.

In addition to these cases for modeling the baseline probability, three settings for PQDs propagation are considered. To artificially include propagation we assume the neighbor customers generate alarm simultaneously with a certain probability. Therefore, for neighbor customers, we generate additional alarm at the same time-stamp with that probability. In each of the scenarios, the neighborhood relation is defined differently.

In total, 16 different models based on the combination of the following settings are considered:

A) no PQ propagation - These models assume that there is no propagation of PQDs, i.e., each customer generates alarms independently. Within this setup, we consider the four different cases (1-4) presented before.

B) propagation within delivery-points - In this case the customers who are connected to the same delivery-points are assumed to be neighbor which generate alarm simultaneously with probability P_D.

C) propagation within branches - Here the customers who are connected to the same branch are assumed to be neighbor which generate alarm simultaneously with probability P_B.

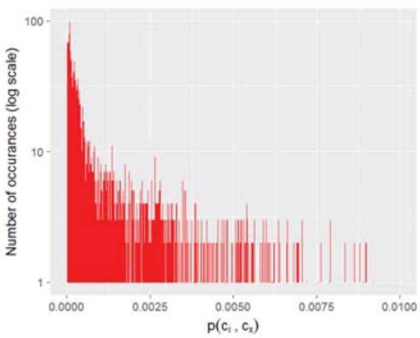
D) propagation within both delivery-point and branch - This includes generating alarms based on both B and C.

For each model, the pairwise probability $p(c_i, c_x)$ for all the customers within a branch is calculated. Then the maximum distance (from the KS test) between the CDF of the pairwise probabilities in the real data and artificial data are compared. Furthermore, the optimization algorithm is used to find the best parameters.

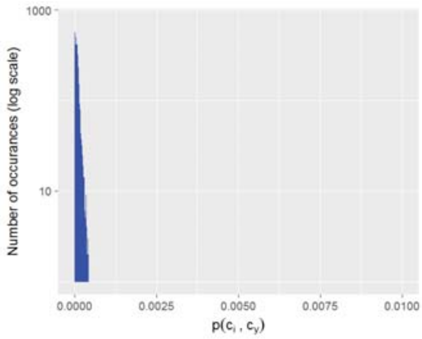
3.3.2 Results

The result of comparing the co-occurrence probabilities of neighbor customers $p(c_i, c_x)$ and non-neighbor customers $p(c_i, c_y)$ among all the bursts are shown in Fig. 3.10. As can be seen in sub-plots (a) and (c), customers from the same branch are more probable to have an alarm in the same burst, compared to the customers from different branches. However, there are also clear differences across different parts of the network (the two sub-stations S and T).

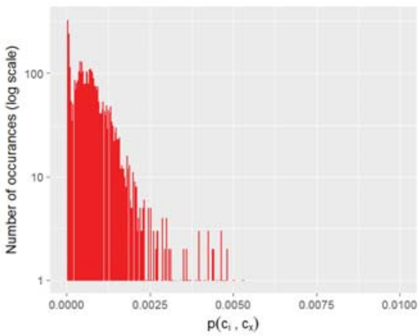
The goodness-of-fit results of the KS test shows that the lowest level of significance at which the null-hypothesis could have been rejected is smaller than 0.0001 ($p - value < 0.0001$). Therefore, we can reject the null-hypothesis with 99% confidence, i.e., the two data sets come from different distributions. In other words, the alarms in customers are not independent.



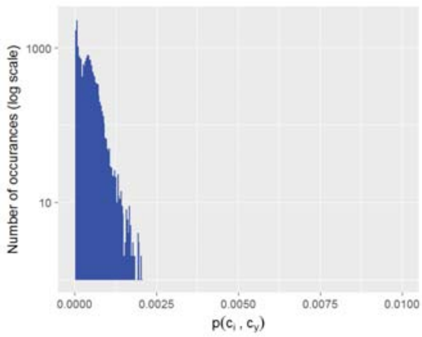
(a) sub-station S - neighbor customers



(b) sub-station S - non-neighbor customers



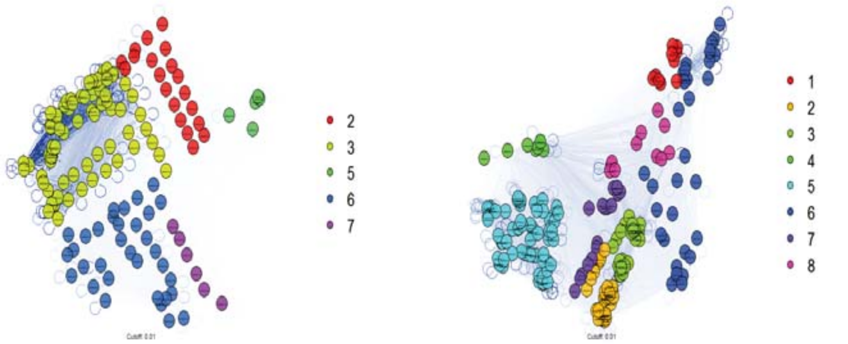
(c) sub-station T - neighbor customers



(d) sub-station T - non-neighbor customers

Figure 3.10: Distribution of the real data for two different sub-stations. In left: distributions of co-occurrence probabilities of neighbor customers $p(c_i, c_x)$, and in right, co-occurrence probabilities of non-neighbor customers $p(c_i, c_y)$, where $c_i, c_x \in B_j$ and $c_y \notin B_j$.

Fig. 3.11 illustrates the pairwise co-occurrences probability as a *co-occurrence network* in which the nodes of the network represent the customers, and the widths of the links between them correspond to the probability $p(c_i, c_j)$. The relative position of the nodes is obtained from the GPS coordinate information of the SMs location. The customers connected to the same branch in the tree-structure are shown with the same color. In this figure, the darker links correspond to probability of co-occurrence more than 0.01 (Cutoff = 0.01). The links from a customer to itself corresponds to the probability of alarm for that customer only. According to the Fig. 3.11, the customers from branch 3 of the sub-station S generate more alarms compare to other customers in this station (this is the reason behind introducing case 3). Furthermore, the customers from the same branch have higher probability of co-occurring in the same burst. It means that when a customer triggers and alarm, there is a higher probability that the neighbor customers will also trigger an alarm, compared to the non-neighbor customers. In this figure, this situation is less significant for sub-station T but still can be observed in some of the customers from branch 6. This figure is an illustration of the areas in the grid which have more PQ problems (SMs with darker links) compare to the other part of the grid.



(a) sub-station S with 141 customers **(b)** sub-station T with 213 customers

Figure 3.11: The probability of the pairwise co-occurrences of alarms for two different sub-stations. The darker links correspond to probability of co-occurrence more than 0.01 and the links from a customer to itself corresponds to the probability of alarm for that customer only. The customers connected to the same branch in the tree-structure are shown with the same color.

After demonstrating the existence of propagation, we need to identify which of the proposed models has the most similar distribution for co-occurrence probabilities of neighbor customers $p(c_i, c_x)$ compared to the real data.

The results of comparing all the combinations of different conditions (models with and without propagation) with the real data from sub-station S are presented in Table 3.5. For example, the result of comparing the distributions in Fig. 3.10 (a) and (b) is presented in this table in the row 1 column 2, with value 0.36.

Several observation can be made from this table. The artificial data generated based on the model A (without propagation and cases 1-4) has the highest distance with the real data. By including propagation within different levels of the grid topology, i.e., propagation within both delivery-point and branch, the difference is the lowest. Among all the models, the closest one to the real data is when: every customer generates alarms with different baseline probability, and the alarms propagate within both the delivery-point and the branch (model D case 4).

The *optimized* distributions of the co-occurrence probabilities of customers (within a same branch) for different conditions are illustrated in Fig. 3.12. According to the figure, without including propagation, the data generated based on model A cases 1-4 are visually very different from the real data. Hence, the models which are based on generating alarms independently for every customer are not a good representation of

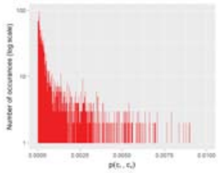
Table 3.5: The minimum value of the maximum distance between the cumulative distribution of pairwise probabilities $p(c_i, c_x)$ in the real data and artificial data, where $c_i, c_x \in Br_i$.

	Case 1	Case 2	Case 3	Case 4
Model A	0.78	0.36	0.29	0.22
Model B	0.437	0.349	0.273	0.179
Model C	0.143	0.217	0.215	0.135
Model D	0.131	0.195	0.182	0.107

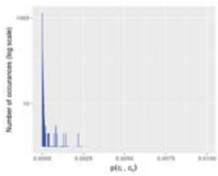
Case 1: equal probability to the real data; Case 2: same probability for all; Case 3: same probability for all except one particular branch; Case 4: different probability for all.

the real data. Among these, the model A case 4 has the smallest maximum distance between the CDF of the two data sets.

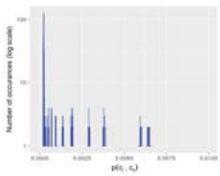
One way to improve the modeling is to include propagation in additional levels of the grid topology, such as propagation within customers connected to the same delivery-point. There could be other scenarios that can be considered. One can include propagation in correlation with the voltage magnitude of the deviation. Moreover, propagation of voltage sag and swell can be separated and investigated individually.



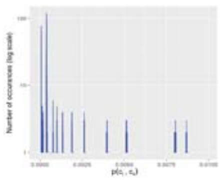
(a) Real data



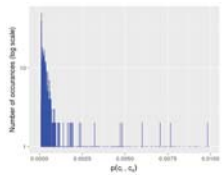
(b) Model A case 1



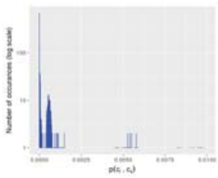
(c) Model A case 2



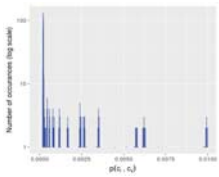
(d) Model A case 3



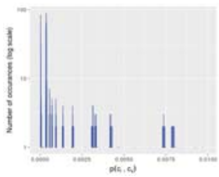
(e) Model A case 4



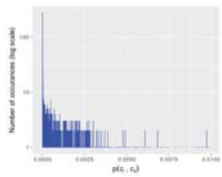
(f) Model B case 1



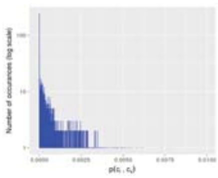
(g) Model B case 2



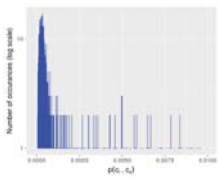
(h) Model B case 3



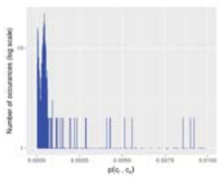
(i) Model B case 4



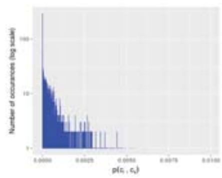
(j) Model C case 1



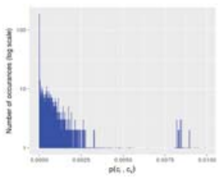
(k) Model C case 2



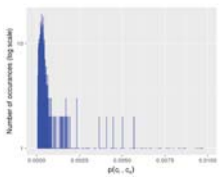
(l) Model C case 3



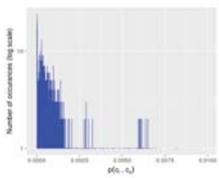
(m) Model C case 4



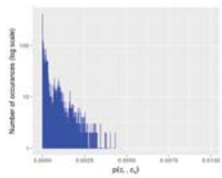
(n) Model D case 1



(o) Model D case 2



(p) Model D case 3



(q) Model D case 4

Figure 3.12: The distribution of the co-occurrence probabilities of neighbor customers for different conditions.

Chapter 4
Summary of papers

4.1 Paper A - Reliability Evaluation of Underground Power Cables with Probabilistic Models

In this paper, we focus on the methodology for evaluating how well different models fit empirical failure rate while dealing with limited amount of failure data. We analyze five different models to estimate the relationship between the age and failure rate in underground high voltage cables. As is common in this domain, the amount of failure data is limited, and it is difficult to know how much confidence should one have in the GOF results. The proposed methodology is based on interpreting the results by comparing the obtained GOF measures with expected GOF and confidence intervals, estimated using synthetic data.

In addition to commonly used models, we also consider constant and piecewise constant models. For each model, a number of synthetic data sets are generated by drawing random points from a normal distributions with mean equal to the failure function at each age and variance computed from the empirical data points.

According to the result of GOF tests, the linear, piecewise linear, and exponential models do not show significant difference. This indicates that those three models are virtually identical when they are used to fit empirical failure rate data. On the other hand, the piecewise constant model fits the failure rates better, in a statistically significant way, than other models.

As it is described in Chapter 3.2.2, some feeder lines are incorrectly linked to a few extra failure events because of upgrading a sub-station. These “incorrect” failures affect the results of GOF and evaluating the models. After identifying these failures, removing them, and fitting different models with the empirical failure data points, we observe that the constant and piecewise constant models are statistically different from the rest of the models. Furthermore, for our data points there is no statistically significant difference between choosing either of linear, piecewise linear, exponential, or Weibull models.

4.2 Paper B - Bayesian Network representation of meaningful patterns in electricity distribution grids

In **Paper B**, three different ways for detecting fault patterns in an electricity distribution grid using history of event data are presented: failure statistics, association rules, and Bayesian networks. We proposed a simplified representation of the association rules by using approximated Bayesian Networks. We show that a small subset of the most interesting rules is enough to obtain a good and sufficiently accurate approximation of the original dataset.

In general, utility companies keep records for previous faults that contain features describing the event such as time, date, cause, faulty component, etc. To analyze failures characteristic it is important to discover which failures have common features, e.g., if there are any types of failures that happen mostly in certain parts of the grid or at certain times. Primary evaluation of the failure records is used for analyzing the frequency of occurrence for each failure in an electricity distribution grid. Statistical analysis and association rules are applied to discover correlation between the features. Association rules are based on the frequency of the co-occurrence of features and conditional dependency between them. Their interestingness is often expressed in terms of probability. If we consider features in recorded events as probabilistic variables, a Bayesian Network captures the conditional relations between those features over a set of events.

The results provide a practical representation of features associated with events that can be used by maintenance staff at electricity distribution companies. The outcomes of the proposed method for discovering failure pattern facilitate the choice of considering the most vulnerable components, e.g., underground cables or important factors for further analysis.

4.3 Paper C - Reliability Evaluation of Power Cables Considering Restoration Characteristics

In this paper we show that it is important to consider the repairability characteristics of power cables and choose the reliability analysis which is designed for repairable systems. We demonstrate that the methods which estimate the time-to-the-first failure (for non-repairable components) may lead to incorrect conclusions about reliability of power cables.

We use proportional hazard model (PHM) to assess the impact of different factors and calculate the failure rate of each individual cable. After modeling the PHM baseline and the influence of different factors, we calculate failure rate for each individual feeder line, and rank them from the highest failure rate to the lowest. In particular we compare three case scenarios depending on how to consider power cables and their failures: as non-repairable components, as repairable but decommissioned after the last failure, and as repairable components which survive until censoring time. In principle, for power cables, the first and second scenarios are incorrect, and we show that conclusions about different factors in PHM and cables ranking will be misleading if they are used.

The results show that the significance level of the factors in PHM is different considering each case scenarios. Furthermore, the variation between the ranking lists shows that the case scenarios produce different outcomes for reliability ranking. This variations between the lists are not negligible.

By ranking the components importance to system reliability, the awareness of the grid status can be improved and actions can be taken to reduce the risk.

4.4 Paper D - Analyzing and Modeling Propagation of Smart Meters' Alarms in Low-Voltage Grids

In this paper we demonstrate that the existence of propagation in the low-voltage grids can be detected using smart meters alarm data. In particular, several models for propagation of disturbances, within neighbor customers in different levels of the grid topology, are investigated. These models include the grid with: no propagation, propagation within neighbor customers at the same delivery-point, propagation within neighbor customers at the same branch, and propagation in both delivery points and branches. A method for measuring how the reality corresponds to each of the models, by measuring the similarity between real data and synthetic data, created according to the models, is proposed. Furthermore, the paper presents smart meters alarm dataset (SMAData), an open-source dataset containing power quality disturbances of over 1000 meters.

The results show that the PQDs propagate throughout the low-voltage grid; and, they impact customers within the same branch frequently. Furthermore, the models which include PQDs propagation in both delivery-points and branches are better representation of the real data.

The analysis, can be used for identifying the areas in the grid with higher probability of generating PQDs compared to other parts of the grid. These "problem areas" can be analyzed to determine if the PQDs are generated by a customer or they are symptoms of fault initiation in one of the grids components in the area. In both cases, the distribution companies can benefit from the results for strengthening their grid and eliminating the PQ problems for customers.

Chapter 5
Conclusion and Perspectives

5.1 Summary and conclusion

This thesis and appended papers present data-driven methods that exploit available data in distribution grids for detecting weak spots. In particular, the data is used for three purposes: failure pattern discovery, reliability evaluation of power cables, and analyzing and modeling the propagation of power quality disturbances (PQDs).

In distribution grids, the majority of reliability improvement techniques are devoted to deviation detection based on real-time streaming data. However, large amounts of data related to measurement readings, previous faults, repairs, and manufacturer information are also recorded but rarely used for predictive maintenance. Approaching the full capabilities of smart grids including low interruption and high level of power quality require using all the available data. Mining and analyzing this data enables us to evaluate reliability and detect problem areas. With this knowledge, power companies can directly target the most vulnerable components for inspection and preventive repair actions.

The proposed reliability evaluation method provides estimation of failure rate for every individual feeder line, in spite of the limited failure data. The outcome of this analysis is a list of feeder lines which are ranked from the highest to the lowest failure rate and can be used for maintenance prioritization. In particular, we consider three case scenarios, based on how failures are defined, to estimate failure rate and impact of different factors. The results show that the significance level of the factors and the lists are different for each case scenario. Furthermore, comparing the position of the recently failed high-voltage feeder lines in the ranking lists of the three cases indicates higher priority of the lines in the case scenario three: the feeder lines are considered as repairable components which stay in-service until their censoring time.

Smart meters, distributed in low-voltage grids, have the potential to provide valuable information. In this thesis we demonstrate an application of using these meters for identifying power quality events, detecting propagation of PQDs, and modeling them. Considering the use of renewable energy sources, utilizing smart meters is more economically efficient than installing additional equipments such as PQMS. We show that these meters can be used as a sensor for detecting power quality deviations and identifying how PQDs can propagate through low-voltage grids and impact other components.

The main contributions of this thesis are summarized in the following:

- Proposing a visualization for the correlations between different features representing failures by using a Bayesian network.
- Proposing a methodology for power cables lifetime modeling with confidence intervals to deal with limited failure data.
- Demonstrating the importance of considering the reparability characteristic of power cables on reliability estimation.
- Developing a method for ranking repairable power cables based on the impact of different factors.
- Developing a methodology for modeling the propagation of sag/swell disturbances in the low-voltage grids using SMs alarm data.

5.2 Future work

The future work is grouped into two categories. The first is related to testing cables and validating the reliability ranking method. Improving the analysis by considering additional factors such as type of the soil where the cables are buried and the process of how the cables are installed is required. Furthermore, considering the proportional hazard model in a time-dependent setting by including some features of the customers who are connected to each feeder lines (captured from smart meters) is needed. In addition, planning for maintenance is an essential part in improving reliability. Based on the results of the reliability analysis, maintenance can be conducted on cables with higher failure rank and higher importance, e.g., number or type of customers. Investigating methods and techniques for maintenance planning based on the relationship between failure rank, importance, and cost should be considered in future applications.

The second category is related to finding correlations between power failures which are recorded in smart meters and other data. In the alarm data the event "primary power down" corresponds to power interruption in the meters. We can investigate the correlation of this event with other power quality events and *load profile* data. The *load profile* in meters corresponds to a collection of different features including, RMS voltage (for three phases), RMS current (for three phases), power factor (for three phases), frequency, active power and reactive power. In total there are 12 different features with resolution of one sample every hour. In this case, developing methods for detecting any existing correlation, identifying an appropriate time interval before the event for improving the correlation detection, and specifying important features which have significant impact on detecting the event are some of the planned activities for future.

Furthermore, we believe that it would be possible to find “symptoms” of the actual cable failures in the low-voltage grids by using smart meters data. However, the amount of recorded cable failures for the four months period (the duration which we have collected the data from smart meters) is limited to 2 faults. This lack of data makes identifying the symptoms of failures very difficult. Therefore we are considering two options. The first option is collecting data from smart meters for a longer duration, e.g., one year, to observe more cable failures. The second option is including expert knowledge to identify symptoms of cable failures.

References

[1] B. Xu, T. Li, and Y. Xue, "Smart distribution grid and distribution automation," *Automation of Electric Power Systems*, vol. 33, no. 17, pp. 38–41, 2009. (Cited on page 1.)

[2] R. N. Allan and R. Billington, *Reliability evaluation of power systems*. Springer Science & Business Media, 2013. (Cited on page 1.)

[3] R. E. Brown, A. P. Hanson, H. L. Willis, F. A. Luedtke, and M. F. Born, "Assessing the reliability of distribution systems," *IEEE Computer Applications in Power*, vol. 14, no. 1, pp. 44–49, 2001. (Cited on page 1.)

[4] R. E. Brown, *Electric power distribution reliability*. CRC press, 2008. (Cited on page 1.)

[5] M. Bollen, *Adapting Electricity Networks to a Sustainable Energy System: Smart Metering and Smart Grids*. Energy Markets Inspectorate, 2011. (Cited on page 1.)

[6] "Supporting systems for a smart grid world - the role of workforce, asset and network management systems in supporting an intelligent electric network," *ABB Inc.*, 2015. (Cited on page 1.)

[7] D. Staszesky, D. Craig, and C. Befus, "Advanced feeder automation is here," *IEEE Power and Energy Magazine*, vol. 3, no. 5, pp. 56–63, 2005. (Cited on page 1.)

[8] M. Karimi, H. Mokhlis, K. Naidu, S. Uddin, and A. Bakar, "Photovoltaic penetration issues and impacts in distribution network–a review," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 594–605, 2016. (Cited on page 2.)

[9] U. S. D. of Energy Federal Energy Management Program, "Om best practices guide, release 3.0, chapter 5: Types of maintenance programs," 2010. [Online; Available: www.eere.energy.gov/femp/pdfs/omguide-complete.pdf]. (Cited on page 2.)

[10] R. Bucci, R. Rebbapragada, A. McElroy, E. Chebli, and S. Driller, “Failure prediction of underground distribution feeder cables,” *IEEE Transactions on Power Delivery*, vol. 9, no. 4, pp. 1943–1955, 1994. (Cited on page 3.)

[11] X. Xiangning, T. Shun, B. Tianshu, and X. Yonghai, “Study on distribution reliability considering voltage sags and acceptable indices,” *IEEE Transactions on Power Delivery*, vol. 22, no. 2, pp. 1003–1008, 2007. (Cited on page 9.)

[12] I. Bazovsky, *Reliability theory and practice*. Courier Corporation, 2004. (Cited on page 9.)

[13] A. Chowdhury and D. Koval, *Power distribution system reliability: practical methods and applications*, vol. 48. John Wiley & Sons, 2011. (Cited on page 9.)

[14] R. Jongen, P. Morshuis, E. Gulski, J. Smit, J. Maksymiuk, and A. Janssen, “Application of statistical methods for making maintenance decisions within power utilities,” *IEEE electrical insulation magazine, 22,(6)*, 2006. (Cited on pages 10 and 12.)

[15] P. Cichecki, E. Gulski, J. Smit, and R. Jongen, “Statistical analysis of transmission power cables condition data,” *Proceedings of the 16th International Symposium on High Voltage Engineering*, pp. 851–856, 2009. (Cited on page 10.)

[16] E. Gulski, J. J. Smit, F. J. Wester, and J. W. van Doeland, “Condition assessment of high voltage power cables,” in *International Conference on Power System Technology*, vol. 2, pp. 1661–1666, IEEE, 2004. (Cited on page 10.)

[17] G. Toman, G. Morris, and P. Holzman, “Effects of moisture on the life of power plant cables: Part 1, medium-voltage cables; part 2, low-voltage cables. final report,” tech. rep., Electric Power Research Inst., Palo Alto, CA (United States); Ogden Environmental and Energy Services Co., Inc., Blue Bell, PA (United States); Strategic Technology and Resources, Winchester, MA (United States), 1994. (Cited on page 10.)

[18] C. Stancu, P. V. Notingher, and M. G. Plopeanu, “Electrical lifetime estimation of underground power cables,” *Journal of International Scientific Publications: Materials, Methods & Technologies*, vol. 6, no. part 1, pp. 165–178, 2012. (Cited on page 10.)

[19] M. Duval and J. Dukarm, “Improving the reliability of transformer gas-in-oil diagnosis,” *IEEE Electrical Insulation Magazine*, vol. 21, no. 4, pp. 21–27, 2005. (Cited on page 10.)

[20] J. Fuhr, “Procedure for identification and localization of dangerous pd sources in power transformers,” *IEEE transactions on dielectrics and electrical insulation*, vol. 12, no. 5, pp. 1005–1014, 2005. (Cited on page 10.)

[21] X. Zhang and E. Gockenbach, "Component reliability modeling of distribution systems based on the evaluation of failure statistics," *Dielectrics and Electrical Insulation, IEEE Transactions on*, vol. 14, no. 5, pp. 1183–1191, 2007. (Cited on page 11.)

[22] O. Quiroga, J. Meléndez, and S. Herraiz, "Fault causes analysis in feeders of power distribution networks," in *International Conference in Renewables Energies and Quality Power, ICREP*, vol. 11, 2011. (Cited on page 11.)

[23] D. Saxena, K. Verma, and S. Singh, "Power quality event classification: an overview and key issues," *International Journal of Engineering, Science and Technology*, vol. 2, no. 3, pp. 186–199, 2010. (Cited on page 11.)

[24] L. Qun, "Analysis on fault statistics in wenzhou electric power distribution network," in *Power and Energy Engineering Conference, 2009. APPEEC 2009. Asia-Pacific*, pp. 1–4, IEEE, 2009. (Cited on page 11.)

[25] T. A. Short, *Electric power distribution handbook*. CRC press, 2014. (Cited on page 11.)

[26] J. A. Bloom, "Guidelines for intelligent asset replacement, vol.3, underground distribution cables," tech. rep., EPRI, Palo Alto, CA, 2005. (Cited on page 11.)

[27] J. A. Bloom, C. Feinstein, and P. Morris, "Optimal replacement of underground distribution cables," in *Power Systems Conference and Exposition*, pp. 389–393, IEEE, 2006. (Cited on page 11.)

[28] S. Lagakos, "General right censoring and its impact on the analysis of survival data," *Biometrics*, pp. 139–156, 1979. (Cited on page 11.)

[29] C. Zapata, A. Torres, D. Kirschen, and M. Rios, "Some common misconceptions about the modeling of repairable components," in *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, pp. 1–8, IEEE, 2009. (Cited on page 11.)

[30] M. Nikulin, B. Vilijandas, and K. Julius, "Non-parametric tests for complete data," 2011. (Cited on page 12.)

[31] W. Hauschild and W. Mosch, *Statistical techniques for high-voltage engineering*. No. 13, IET, 1992. (Cited on page 12.)

[32] R. Jongen, P. Morshuis, J. Smit, A. Janssen, and E. Gulski, "Failure analysis of in service failed resin cable joints by means of a statistical approach," in *Electrical Insulation and Dielectric Phenomena, 2006 IEEE Conference on*, pp. 517–520, IEEE, 2006. (Cited on page 12.)

[33] Y. Zhou and R. E. Brown, "A practical method for cable failure rate modeling," in *Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES*, pp. 794–798, IEEE, 2006. (Cited on page 12.)

[34] P. Cichecki, R. Jongen, E. Gulski, J. J. Smit, B. Quak, F. Petzold, and F. Vries, “Statistical approach in power cables diagnostic data analysis,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 6, pp. 1559–1569, 2008. (Cited on page 12.)

[35] C. Zapata, S. Silva, H. Gonzalez, O. Burbano, and J. Hernandez, “Modeling the repair process of a power distribution system,” in *Transmission and Distribution Conference and Exposition: Latin America, 2008 IEEE/PES*, pp. 1–7, IEEE, 2008. (Cited on page 12.)

[36] H. Guo, A. Mettas, G. Sarakakis, and P. Niu, “Piecewise nhpp models with maximum likelihood estimation for repairable systems,” in *Reliability and Maintainability Symposium (RAMS), 2010 Proceedings-Annual*, pp. 1–7, IEEE, 2010. (Cited on page 12.)

[37] V. Kogan and T. Jones, “An explanation for the decline in urd cable failures and associated nonhomogeneous poisson process,” *Power Delivery, IEEE Transactions on*, vol. 9, no. 1, pp. 534–543, 1994. (Cited on page 12.)

[38] M. Kaminskiy and V. Krivtsov, “G1-renewal process as repairable system model,” *arXiv preprint arXiv:1006.3718*, 2010. (Cited on page 12.)

[39] H. P. Barringer, “Use Crow-AMSAA reliability growth plots to forecast future system failures,” *Second Annual International Maintenance Excellence*, 2006. (Cited on page 12.)

[40] P. Barringer, “Predict failures: Crow-AMSAA 101 and Weibull 101,” *Proceedings of IMEC*, 2004. (Cited on page 12.)

[41] J. Y. Gill, *Forecasting Underground Electric Cable Faults Using the Crow-AMSAA Model*. Engineering Information Transfer, 2011. (Cited on page 12.)

[42] E. L. Mariut and E. Helerea, “Enhancing reliability for medium voltage underground power lines,” *7th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, pp. 1–6, 2011. (Cited on page 12.)

[43] M. Rausand and A. Høyland, *System reliability theory: models, statistical methods, and applications*, vol. 396. John Wiley & Sons, 2004. (Cited on page 12.)

[44] Q. Zhang, C. Hua, and G. Xu, “A mixture weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data,” *Mechanical Systems and Signal Processing*, vol. 43, no. 1, pp. 103–112, 2014. (Cited on page 13.)

[45] G. Rodrıguez, “Non-parametric estimation in survival models,” 2005. (Cited on page 13.)

[46] J. Zheng, D. W. Gao, and L. Lin, "Smart meters in smart grid: An overview," in *Green Technologies Conference, 2013 IEEE*, pp. 57–64, IEEE, 2013. (Cited on page 14.)

[47] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, 2018. (Cited on page 14.)

[48] N. S. Foundation, "Smart grids big data." `https://www.nsf.gov/awardsearch/showAward?AWDID=1636772&HistoricalAwards=false`, 2016. (Cited on page 14.)

[49] X. Liu, A. Heller, and P. S. Nielsen, "Citiesdata: a smart city data management framework," *Knowledge and Information Systems*, vol. 53, no. 3, pp. 699–722, 2017. (Cited on page 14.)

[50] "Bits to energy lab projects." `http://www.bitstoenergy.ch/home/projects/`, accessed July 31, 2017. (Cited on page 14.)

[51] "Wp3 overview." `https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WP3overview`, accessed July 31, 2017. (Cited on page 14.)

[52] P. Wright, A. Christensen, P. Davis, B. Larson, T. Lippert, and P. Patel, "Power quality propagation measurements in smart grids," in *Precision Electromagnetic Measurements (CPEM 2016), 2016 Conference on*, pp. 1–2, IEEE, 2016. (Cited on page 14.)

[53] M. H. Bollen and M. H. Bollen, *Understanding power quality problems: voltage sags and interruptions*, vol. 445. IEEE press New York, 2000. (Cited on page 14.)

[54] M. Sănduleac, C. Stănescu, and N. Golovanov, "Power networks observability, control and automation using unbundled smart meters," in *Development and Application Systems (DAS), 2016 International Conference on*, pp. 16–20, IEEE, 2016. (Cited on page 15.)

[55] L. Garpetun, "Smart meters in operation center," 2013. (Cited on page 15.)

[56] E. J. Palacios-Garcia, E. Rodriguez-Diaz, A. Anvari-Moghaddam, M. Savaghebi, J. C. Vasquez, J. M. Guerrero, and A. Moreno-Munoz, "Using smart meters data for energy management operations and power quality monitoring in a microgrid," in *Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on*, pp. 1725–1731, IEEE, 2017. (Cited on page 15.)

[57] K. Suslov, N. Solonina, and V. Stepanov, "A principle of power quality control in the intelligent distribution networks," in *Smart Electric Distribution Systems and Technologies (EDST), 2015 International Symposium on*, pp. 260–264, IEEE, 2015. (Cited on page 15.)

[58] F. A. Borges, R. A. Fernandes, I. N. Silva, and C. B. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 824–833, 2016. (Cited on pages 15 and 16.)

[59] F. A. Borges, I. Silva, and R. A. Fernandes, "Feature extraction-based method for voltage sag source location in the context of smart grids," in *International Symposium on Distributed Computing and Artificial Intelligence*, pp. 71–78, Springer, 2017. (Cited on pages 15 and 16.)

[60] S. K. Jain and S. Singh, "Fast harmonic estimation of stationary and time-varying signals using ea-awnn," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 2, pp. 335–343, 2013. (Cited on page 15.)

[61] F. J. Gomez, J. Melendez, S. Herraiz, E. Gonzalez, M. Lopez-Perea, and C. M. Oconnor, "Web services framework for power quality monitoring," in *Smart Objects, Systems and Technologies (SmartSysTech), Proceedings of 2013 European Conference on*, pp. 1–7, VDE, 2013. (Cited on page 15.)

[62] M. Stifter and P. Palensky, "Smart meter data as a basis for smart control in low voltage distribution networks," in *Industrial Electronics (ISIE), 2013 IEEE International Symposium on*, pp. 1–5, IEEE, 2013. (Cited on pages 15 and 16.)

[63] G. Roupioz, X. Robe, and F. Gorgette, "First use of smart grid data in distribution network planning," 2013. (Cited on pages 15 and 16.)

[64] F. Uçar, Ö. F. Alçin, B. Dandil, and F. Ata, "Hilbert transform based simple detection and indice analyze of voltage sags using synthetic data," in *Smart Grid and Cities Congress and Fair (ICSG), 2017 5th International Istanbul*, pp. 37–40, IEEE, 2017. (Cited on pages 15 and 16.)

[65] M. Music, N. Hasanspahic, A. Bosovic, D. Aganovic, and S. Avdakovic, "Upgrading smart meters as key components of integated power quality monitoring system," in *Environment and Electrical Engineering (EEEIC), 2016 IEEE 16th International Conference on*, pp. 1–6, IEEE, 2016. (Cited on pages 15 and 16.)

[66] M. Music, A. Bosovic, N. Hasanspahic, S. Avdakovic, and E. Becirovic, "Integrated power quality monitoring system and the benefits of integrating smart meters," in *Compatibility and Power Electronics (CPE), 2013 8th International Conference on*, pp. 86–91, IEEE, 2013. (Cited on pages 15 and 16.)

[67] M. Music, A. Bosovic, N. Hasanspahic, S. Avdakovic, and E. Becirovic, "Integrated power quality monitoring systems in smart distribution grids," in *Energy Conference and Exhibition (ENERGYCON), 2012 IEEE International*, pp. 501–506, IEEE, 2012. (Cited on page 15.)

[68] M. H. Variani, E. Liu, L. Dow, J. Dirkman, T. Martin, W. W. Cornish, P. Baker, and O. Trinko, "Smart meter data to identify customer voltage fluctuations caused by solar pv," in *Power & Energy Society General Meeting, 2017 IEEE*, pp. 1–5, IEEE, 2017. (Cited on page 15.)

[69] P. K. Wong, R. A. Barr, and A. Kalam, "Voltage rise impacts and generation modeling of residential roof-top photo-voltaic systems," in *Harmonics and Quality of Power (ICHQP), 2014 IEEE 16th International Conference on*, pp. 229–233, IEEE, 2014. (Cited on page 15.)

[70] S. Ronnberg and M. Bollen, "Solar power in northern scandinavia, its impact on the grid and methods to mitigate this impact," in *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, pp. 1–5, IEEE, 2013. (Cited on page 15.)

[71] Y. J. Dai, C. Y. Sun, X. Y. Wang, and W. H. Yang, "Research on china's first residential grid-connected photovoltaic system," in *Applied Mechanics and Materials*, vol. 392, pp. 563–567, Trans Tech Publ, 2013. (Cited on page 15.)

[72] H. Pezeshki, P. J. Wolfs, and M. Johnson, "Multi-agent systems for modeling high penetration photovoltaic system impacts in distribution networks," in *Innovative Smart Grid Technologies Asia (ISGT), 2011 IEEE PES*, pp. 1–8, IEEE, 2011. (Cited on page 15.)

[73] K. Chaudhari, N. K. Kandasamy, R. K. Kanamarlapudi, H. B. Gooi, and A. Ukil, "Modeling of charging profiles for stationary battery systems using curve fitting approach," in *Industrial Electronics Society, IECON 2017-43rd Annual Conference of the IEEE*, pp. 2777–2781, IEEE, 2017. (Cited on page 15.)

[74] B. Faessler, M. Schuler, M. Preißinger, and P. Kepplinger, "Battery storage systems as grid-balancing measure in low-voltage distribution grids with distributed generation," *Energies*, vol. 10, no. 12, p. 2161, 2017. (Cited on page 15.)

[75] Y. Arafat, L. B. Tjernberg, and P.-A. Gustafsson, "Field test on multiple smart meter switching to study the effect on power quality at customers level," in *PowerTech, 2015 IEEE Eindhoven*, pp. 1–6, IEEE, 2015. (Cited on page 15.)

[76] K. V. Suslov, N. Solonina, and A. Smirnov, "Smart meters for distributed filtering of high harmonics in smart grid," in *Power Engineering, Energy and Electrical Drives (POWERENG), 2011 International Conference on*, pp. 1–5, IEEE, 2011. (Cited on page 15.)

[77] K. Suslov, A. Smirnov, and N. Solonina, "Distributed filtering of high harmonics in microgrids," in *Electrical Power Quality and Utilisation (EPQU), 2011 11th International Conference on*, pp. 1–4, IEEE, 2011. (Cited on page 15.)

[78] H. Chang, Y. Huang, S. Ebrahimi, and J. Jatskevich, "Smart meter based selective harmonics compensation in buildings distribution systems with ac/dc microgrids," in *Power & Energy Society General Meeting, 2017 IEEE*, pp. 1–5, IEEE, 2017. (Cited on page 15.)

[79] R. D. T. Feeders, "Ieee distribution planning working group report," *IEEE Trans. Power Syst*, vol. 6, no. 3, pp. 975–985, 1991. (Cited on page 16.)

[80] T. G. Dietterich and R. S. Michalski, "Discovering patterns in sequences of events," *Artificial Intelligence*, vol. 25, no. 2, pp. 187–232, 1985. (Cited on page 18.)

[81] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, pp. 3–14, IEEE, 1995. (Cited on page 18.)

[82] J. Han and M. Kamber, "Mining frequent patterns, associations, and correlations," *Data Mining: Concepts and Techniques (2nd ed., pp. 227-283). San Francisco, USA: Morgan Kaufmann Publishers*, 2006. (Cited on page 18.)

[83] G. Zhanjun, P. Zhengliang, G. Nuo, and C. Bin, "A distribution network fault data analysis method based on association rule mining," in *Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific*, pp. 1–5, IEEE, 2014. (Cited on page 18.)

[84] S. Russell and P. Norvig, "Artificial intelligence: a modern approach," 1995. (Cited on page 18.)

[85] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006. (Cited on page 18.)

[86] D. Tian, A. Gledson, A. Antoniades, A. Aristodimou, N. Dimitrios, R. Sahay, J. Pan, S. Stivaros, G. Nenadic, X.-j. Zeng, *et al.*, "A bayesian association rule mining algorithm," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pp. 3258–3264, IEEE, 2013. (Cited on page 18.)

[87] S. Speaks, "Reliability and mtbf overview," *Vicor reliability engineering*, 2010. (Cited on page 19.)

[88] C. R. David, "Regression models and life tables (with discussion)," *Journal of the Royal Statistical Society*, vol. 34, pp. 187–220, 1972. (Cited on page 25.)

[89] A. Barabadi, J. Barabady, and T. Markeset, "Maintainability analysis considering time-dependent and time-independent covariates," *Reliability Engineering & System Safety*, vol. 96, no. 1, pp. 210–217, 2011. (Cited on page 25.)

[90] J. P. Klein and M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005. (Cited on page 26.)

[91] R. J. Cook and J. Lawless, *The statistical analysis of recurrent events*. Springer Science & Business Media, 2007. (Cited on page 26.)