# Transformer Failure Detection Using Dissolved Gas Analysis (DGA)

[1]Quist Mirable Jethro, [2]Amevi Acakpovi

1. Accra Institute of Technology – Open University of Malaysia (AIT – OUM), Graduate Studies-Mechanical Engineering, Email: quistmirablejethro@yahoo.com, Accra, Ghana;
2. Accra Technical University, Electrical/Electronic Engineering Department, Email: acakpovia@gmail.com, Accra, Ghana.

Abstract – A transformer is a device that transfers electrical energy from one circuit to another through inductively coupled conductors – the transformer coils.

Transformer is very important and plays a vital role when it comes to transmission and distribution of electrical energy. Transformer failures are disruptive to electrical grid and costly. Transformer insulation failures from various blasts like bombs or lightning strikes are major culprits. Methods and techniques used in data collection, analyses, computer simulation and results interpretation on transformer failures are presented.

Dissolved Gas Analysis (DGA) of the insulation was used to detect the health condition of transformers. Secondary data of burnt transformers were also considered.

A multi-layer casing for transformers was designed to protect transformers. The casing will protect transformers from detonation by bombs, rocket and other explosives, and lightning strikes.

A stabilized power supply will mean higher productivity in the country as a result of good energy supply**.**

*Keywords: distribution Transformer, void formation,transformer winding, transformer failure, box model, deflectors, Dissolved Gas Analysis Test.*

## 1. Introduction.

Transformer is very important and plays a vital role when it comes to transmission and distribution of electrical energy in armed forces in developing countries. The increasing failure rate of distribution transformers in power systems can be regarded as one of the dangerous phenomenon of distribution system management; especially in developing countries in the armed forces (military), like Ghana Armed Forces, Liberia Armed Forces, Nigerian Armed Forces, Sierra Leon Armed Forces, Guinea Armed Forces. A distribution failure transformer really worries the stability and reliability of power supply to clients.

DGA is a well established diagnostic method, where oil samples are taken routinely and the composition of the gases dissolved within the oil analyzed. There are mainly eight gases of interest for diagnostics. Namely, Hydrogen(H), Methane (CH4), Ethane (C2H6), Ethylene(C2H4), Acetylene(C2H2), Carbon Monoxide(CO), Carbon Dioxide(CO2), Oxygen (O2), Nitrogen(N2).

The amount and type of gases formed and dissolved into the oil can be determined with the help of Dissolved Gas Analysis (DGA). Similar to a blood test or scanner examination of human body, dissolved gas analysis can give an early diagnosis and increase the chances of finding the appropriate cure [ Ranjana and Amarjit Singh, 2017].

**2.** 0.Background on Transformer failures by age based on Dissolved Gas Analysis.

There are general failures of Distribution transformer systems but the aspect of bomb blast, explosives and mine detonations, bush burning and gun shot from hunters also exist.

This risks can be mitigated by examining the impact of Distribution Transformer failure on developing nations, mitigate causes and effects of Distribution Transformer system failure in developing nations, to promote efficiency and good energy supply in developing nations as a result of an inform policy on the mitigating the economic consequences of Distributing Transformer Failures in developing nation.

## Replacement estimate of distribution transformer

Table 1. below explain prices and the installation cost of some PMT transformers

| VOLTAGE LEVEL | TYPE PMT/GMT | KVA RATING | UNIT PRICE | INST. FEE | TOTAL GH¢ |
|---|---|---|---|---|---|
| 33/0.433KV | PMT | 50 | 9,922.53 | 606.39 | 10528.92 |
| 33/0.433KV | PMT | 100 | 18,226.06 | 618.28 | 18844.36 |
| 33/0.433kv | PMT | 200 | 25,005.84 | 642.06 | 25647.90 |
| 33/0.433KV | PMT | 315 | 29,591.48 | 713.40 | 30304.88 |
| 11/0.433KV | PMT | 50 | 8,998.91 | 606.39 | 9605.30 |
| 11/0.433KV | PMT | 100 | 16,469.48 | 618.28 | 17,087.76 |
| 11/0.433KV | PMT | 200 | 23,756.70 | 642.06 | 24,398.76 |
| 11/0.433KV | PMT | 315 | 26,945.02 | 713.40 | 27658.42 |
| TOTAL GHc | | | | | 164,078.28 |

In 2010 about 46 various distribution transformers were damaged in the western region electricity company of Ghana including those in military establishments, prison establishments, Customs Exercise and Preventive Service Institutions, police stations, hospitals, universities, colleges, schools and other governmental and Nongovernmental Organizations, which amounted to about GHc 865,328.54. It can be seen that losing so many Ghana cedis in a region because of transformer failure within a year is very unwelcoming. Transformer failures in general and Distribution Transformer systems failure have impact on the economy of developing nation. Distribution Transformer system failure in developing nations adds to insufficient energy production and power outage problems.

Insufficient energy production and power outage have effect on the efficiency and good energy supply in developing nations. As a result of that there is the need for an inform policy on the mitigating the consequences of Distributing Transformer Failures in developing nation.

Transformers are considered the most crucial and expensive piece of plant within a transmission system. Most transmission systems currently have large populations of aging transformers. With the growing demand for electricity, the loading of transformers is increasing. Current economic strategies call for reduced maintenance as well as capital expenditure. These challenges, which face utilities world-wide, necessitate improved management of transformers. The impact of a distribution transformer failure can be catastrophic. It would therefore be beneficial to know the risk status of the transformer population in order to facilitate better asset management, optimization of maintenance, refurbishment and replacement strategies which will ensure maximum asset utilization and minimize system risk. Statistical analysis on historical transformer failure data is therefore required, in order to obtain a model to determine the probability of failure of the transformers currently in service. Currently, ECG Transmission has over 550 Power transformers in service. It is of interest to see the age profile of the existing population, in terms of years since manufacture, since reliability is often related to age. From an analysis performed on insurance claims for transformers, it was found that the average

age of failure was approximately 15 years *[Bartley, W.H. (2003). "Analysis of transformer failures, imia wgp33(03). In 36th Annual Conference of the International Association of Engineer's in Insurers, Stokholm,(2003)"]*. The expected design life of a power transformer is 40 years. This figure is reduced, sometimes substantially, depending on the utilization of the transformer, i.e. its loading or the environment to which it is exposed.

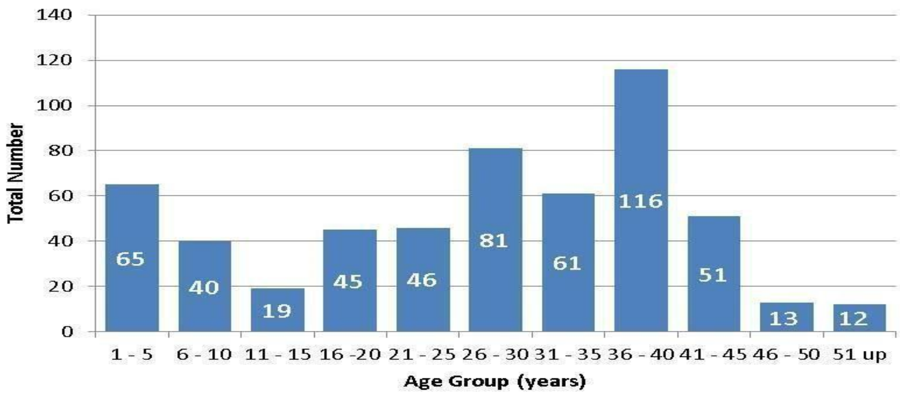



Figure 1: Breakdown of transformer age.

The breakdown of age within the population of Eskom Transmission transformers is shown in Figure 1. From this graph it is evident that the majority (61%) of the population is above the age where probability of failure would be expected to be high and 14% is beyond the expected design life. The average age of transformers in Transmission is currently 29 years, showing an aging, high risk fleet, according to traditional thinking.

With the increased age of the transformers and the following additional risk factors [*Bartley, W.H. (2003). "Analysis of transformer failures, imia wgp33(03). In 36th Annual Conference of the International Association of Engineer's in Insurers, Stokholm, 2003)"*]:

i.Increased utilization of equipment. ii. Deferred capital expenditure. iii. Reduced maintenance expenses. iii. Increased power consumption/load demand

It seems inevitable that the transformer failure rate within Transmission can be expected to increase rapidly in the near future. Therefore, it seems prudent to investigate the development of a model that will enable determination of probability of failure of these transformers, and hence improve asset management.

Eskom has adopted an asset management tool for optimizing asset replacement strategies. The Condition, Criticality and Risk Assessment (CCRA) model enables the phasing and structuring of project plans based on suitable cost-to benefit ratios. A complete CCRA is required to avoid replacement strategies based on age and condition alone, in an attempt to reduce failure rate. The purpose of this model is to incorporate the consequence of failure as well as the risk into the investment decision making process.

**Transformer Aging**

Bartley (2016) did not categorize "age" as a cause of failure. According to him, aging of the insulation system reduces both the mechanical and dielectric-withstand strength of the transformer. As the transformer ages, it is subjected to faults that result in high radial and compressive forces. As the load increases with system growth, the operating stresses increase. Bartley lamented that in an aging transformer, the conductor insulation is debilitated to the point where it can no longer withstand mechanical stress of a fault and these can cause transformer to failure. Turn to turn insulation then suffers a dielectric failure, or a fault causes a loss of winding clamping pressure, which reduces the transformer's ability to withstand future short circuit forces.

Table 2.3. Distribution of Losses by Age of Transformer

| **Age at fault** | **Number of Failures** | **Cost of Failure** |
|---|---|---|
| 0 to 5 years | 9 | $11,246,360 |
| 6 to 10 years | 6 | $22,465,881 |
| 11 to 15 years | 9 | $3,179,291 |
| 16 to 20 years | 9 | $10,518,283 |
| 21 to 25 years | 10 | $16,441,930 |
| Over 25 years | 16 | $15,042,761 |
| Age Unknown | 35 | $207,734,306 |

The table above displays the distribution of transformer failures by age.

The average age at failure was 18 years. The analysis of the ages of the transformers mentioned on the table 2.3, above shows that as transformers are growing they turned to fail. The age of transformers deserves special attention.

Looking into the analysis above, the author of the paper is right because it could be seen from the table and the explanation that transformer fails as they are catching up by age. But not withstanding, transformers under certain ages must not also be left out because they are not too old. Looking into the table above, transformers that failed between 0 to 5 year and 16 to 20 years are the same even though 16 to 20 years transformers are older. Even the cost for transformers between 0 and 5 years cost more than transformers between 16 to 20 years. Even the difference between in terms of cost of transformers between 0 and 5 years and over 25 years is not much.

This means that when transformers are installed equal care must be taken on all ranges of age. It also means that transformers failing because they are older are not always true. All transformers, irrespective of age must be given equal attention because they are very important device and also costly.

The ideal strategy is a life assessment or life cycle management programme, that sets loading priorities, and provides direction for the identification of:

Transformer defects that can be corrected

Transformers that can be modified or refurbished

Transformers that should be relocated

Transformers that should be retired

**Significance of the study.**

There have been many researches done on the causes of transformer failures in developed countries and some developing countries in Asia. Very little has been done in Africa, especially Ghana. Even most of the research done in the developed countries is on power transformers. Since environmental, culture, attitude, terrain and other situations differ from developed countries to developing countries, knowing the actual causes of transformer failures in Ghana, will contribute immensely to the body of knowledge. It will also help to maintain quality and efficient power supply and to improve the economy. For example in western region alone during 2014 year about 46 distribution transformers failed and amounted to about GHC 865,328.54, when taken Ghana as a whole for one year it will be very disastrous. It will also benefit the government or utility companies when purchasing transformers for installation. The findings will be very useful to engineering companies and student because it will be a source of reference for electrical engineering students in the country. It will also help in designing alternative solutions to mitigate the problem, and will also add to the body of knowledge.

Unplanned electrical outages can cost millions of dollars for manufacturers, industrial process plants, commercial buildings and other organizations. A Lawrence Berkeley National Laboratory study estimated that the annual cost of power interruptions in US is about $80 billion - 72% allocated to commercial customers and 26% to industrial customers. And that's just reflective of failures on the supply side. Electricity is a key infrastructural element for economic growth.

Distribution transformer is employed to convert the electricity of higher voltage (11-33 kV) to a lower voltage (250-433V). Little research conducted in the country has revealed that many transformers got damaged very often which affects many customers in Ghana. The failure rate has been very high also in the developing countries. Most of the failures have been credited to lightning strikes, overloading of the transformer and insulation breakdown. Another vital area that researchers have not done much work is proper database which monitors transformer failures in the country. It will also help to address better control and protection devices for distribution and power transformers. Another reason being that failure of transformers impedes the growth of a nation because investors would not want to invest in nations with frequent power outages. Stabilized power supply means higher productivity in the country. Since the Government is seriously embarking on nationwide electrification project, transformers are the quintessential equipment for power transmission and distribution. Because of high failure rate the resultant reliability and safety implications particularly with distribution transformers, in-depth valuation is indispensable to ascertain a healthy transformer is very imperative. Experience has shown an increasing number of transformer explosions and fires in all types of power and distribution transformers worldwide (internationally). Establishing the failures will help reduce the failure rate.

All the findings from developed to developing countries vary from one place to the other, this also point to the fact that distribution transformer failures in Ghana needs to be given a serious attention.

In summary, the following contribution will be made;

Causes of transformer failures will be establish since their failure is costly and affect the quality of service.

2.1. Methods and technques.

Thuraijah Haigh and Amaratunga (2016), describes research methodology as the overall approach to the design process from the theoretical underpinning to the collection of data and analysis of data. These chapters as well present the research approach and the methods used in achieving the objectives for the research. The implications and concerns on the choice of the study approach, and framework validation approach are also justified. Methods and techniques used in sampling, data collection, analyses, computer gas simulation and interpretation are presented. It also discusses the concept formation through critical thinking, and is based on theories derived in literature, proposed study area. It also elaborates on the proposed study design, target population, sample size and sampling procedure, as well as the proposed methods of lab test, simulation, and data analysis methods. Methodology refers to the major approaches or paradigms, which guide the conduct of a research. Methods conversely refer to the specific research techniques or tools which a researcher employs to collect the relevant data for answering her/his research questions. The decision to use particular research method is generally determined by their appropriateness, the purpose of the study, the research problem, research questions, objectives and the researcher's preferences.

The research employed different methods; Combination of research methods. Lab testing were key factor using simulation software available to simulate some findings or results.

Secondary data of burnt transformers will be considered. Multiple methods will be employed.

## 2.2. Research Questions

The main aim of the research questions was to develop a framework for the categorization of transformers in developing countries. The research was motivated by the continuous light outs in Ghana, with the lack of appropriate means of determination of such causes and ways of solving them(Asamoah, 2015) . The central research question is therefore: i. What is the appropriate method of determining causes of light outs and in case of transformer failures that would meet the needs of power industries and clients in developing countries. To appropriately handle the main research question, further sub-questions were considered. The entire research methodology thus evolved around the research question and research objectives.

The framework for integrating risk during categorization of transformer failures in developing countries can be investigated and the various risk indicators identified. The research also viewed the investigation to be conducted as practical as possible rather than an abstract of ideas and place. A rigorous process of lab test was conducted by electrical engineers engaged directly on electrification transformer installation and replacement projects and existing procedures were also followed with regards to physical inspections, repair and replacements.

Epistemologically, this research was geared towards the positivist position, with the belief that intriguing process of categorizing transformer failures in developing countries could be explored further through a systematic but cognitive approach. The research could be carried without bias and that objective conclusions could be drawn from experimental lab tests, physical inspections and data collected through DGA, acidity, moisture content test and simulation of gasses.

Two research strategies are appreciated in traditional research: qualitative and quantitative research work. Though the philosophical stance underpinning the research has been discussed, Bryman (2014), holds that there is the need to clarify the orientation if the researcher has to conduct the research. Due to the difficulty in applying one method wholly, researchers have combined both approaches referred to as the mixed approach.

## 2.3. Strategy Adopted in this Research

The research adopted the mixed approach using both qualitative and quantitative strategies in collecting data. The approach to be used in doing this work is broken down into desktop study through literature, field experiment and observation through lab tests, probability and statistical analysis of results, generation of conceptual framework, development of models and the validation of the results of the model. The concepts developed are subject to critical laboratory tests, probability and subsequent statistical analysis. The result of the lab test constituted a basis for the model to be developed. It was further validated qualitatively through vetting and testing in subsequent stages in this work. An experimental approach was adopted in this research using both quantitative and qualitative methods. Quantitative research methods were chosen due to the engineering nature of the problem, to deduce the relationship between risk factors which has a measurable effect on the transformer failures in developing nations. The process involved lab test, DGA, simulation of gases, model development, validation, probability and statistical analysis of results obtained through tests.

According to Frazer and Lawley (2014), a research design is a plan of information required to answer research problems and how such information can be collected. Bryman (2004) and Yin (2013), hold that research design enables the researcher to connect empirical data to its conclusions, in a logical sequence to the initial research question of the study. Research design is the structure that guides the execution of the technique for collecting and subsequently analyzing data. It is, therefore, the framework within which the research method is employed. This sub-section outlines the various frameworks available for research data collection and analysis. It first discusses research design frameworks and then justifies the choice adopted by this research. Finally, it describes attitude measurement, which was adopted for the testing and validation of the designed framework.

### 2.4. Research Design Options

The research design option selected depends on: the type of research question, the control the investigator has over actual behavioral events and the focus on contemporary as opposed to historical phenomena (Yin, 2003). Bryman (2004) described five main research design options: i. Experimental, ii. Cross-sectional, iii. Longitudinal, iv. Case study and

v. Comparative research designs. It is important to realize the differences within these options and their relationship with research designs. Yin (2013), classified the options into experiment, survey, archival analysis, history and case study. Bryman (2014), pointed out that the selection of a particular design should be done to reflect the importance with which the researcher attaches to: i. expressing causal connections between variables. ii. generalizing to larger groups of individuals than those actually forming part of the investigation. iii. understanding behaviors and the meaning of that behavior in its specific social context. iv. having a temporal time appreciation of social phenomena and their interconnections.

Table:3.1.Research Strategy and Research Design.

| **Research Design** | **Research Strategy (Typical Form)** | |
|---|---|---|
| | **Quantitative Research** | **Qualitative Research** |
| **Experimental** | Typical Form: Most Researchers using an experimental design employ qualitative comparison Between Experimental and control groups with regards to the dependent variable | No typical form |
| **Cross-Sectional** | Typical Form: Survey Research or structured observation on a sample at a single point in time . Content analysis on a sample of documents. | Typical Form: Qualitative interviews or focus groups at a single point in time. Qualitative content analysis of documents relating to a single period. |

| | | |
|---|---|---|
| **Longitudinal** | Typical Form: Survey research on a sample on more than one occasion, as in panel and cohort studies. Content analysis of documents relating to different time periods. | Typical Form: Ethnographic research over a long period, qualitative interviewing on more than one occasion or qualitative content analysis of documents relating to different time periods. Such research warrants being dubbed longitudinal when there is a concern to map change. |
| **Case Study** | Typical Form: Survey research in which there is direct comparison between two or more cases as in cross cultural research. | Typical Form: The intensive study by ethnography or qualitative interviewing of single case which may be an organization on life, family or community. |
| **Comparative** | Typical Form: Survey research in which there is direct comparison between two or more cases, as in cross cultural research. | Typical Form: Ethnographical or qualitative interview research on two or more cases. |

Source: Bryman (2014).

Yin (2013), holds that in the selection of a relevant strategy, the listed factors in table 3.1.may be considered. Bryman (2014), summarizes the various research strategy and design options in their typical forms under the two main research strategies discussed earlier.

2.5.Research design

Two main methodological approaches to the conduct of research that have been identified in literature are quantitative and qualitative approaches (Cohen, et al., 2012,Bryman, 2015, 2015). Bryman (2015) argues that even though many writers on methodological issues attempt to distinguish between these two paradigms, the status fundamental, others regard it as no longer useful and still others as false. The author further contends that quantitative research as one which lays emphasis on quantification in the collection and analysis of data which entails a deductive approach to the relationship between theory and research, in which the accent is placed on testing of theories and embodies a review of social reality as an external and objective reality. Most of this distinction is unclear, because while some writers regard the distinction as Conversely, Bryman conceives qualitative research as a research strategy which lays emphasis on words rather than quantification in collection and analysis of data and which predominantly emphasizes an inductive approach to the relationship between theory and research, in which the emphasis is placed on the generation of theories and also embodies a view of social reality as a constantly shifting emergent property of individuals' creation (Bryman, 2004) .The researcher will employ mixed method (both qualitative and quantitative in data collection) research design for the study as that is deemed appropriate for the study.

The research design adopted in this thesis is a modelling and simulation work in the field of electrical engineering. Precisely, the design firstly consists of developing an adaptronic transformer casing resistance to explosives, bombs and bullets through the use of stealth technology, blast bags, concrete walls, explosive resistant glass and bomb proof materials as outlined in chapter one. Laboratory tests were performed including monitory DGA (Dissolved Gas Analysis). DGA gas including, $H_{2,}$ $CH_{4,}$ $^{C}2^{H}6$, $^{C}2^{H}4$, $^{C}2^{H}2$, CO, $CO_{2,}$ $O_{2,}$ $N_{2,}$ were analysed and simulated. It can therefore be classified under experimental research method. The research method is quantitative. The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena. Similarly, this research study uses analytical modelling which involves a lot of mathematical models and theory to prove a better way of analysing transformer failures in developing countries. It is also known that quantitative data is any data that is in numerical form such as statistics, percentages.

Again, DGA generated gases, $H_2$, $CH_4$, $^C2^H6$, $^C2^H4$, $^C2^H2$, CO, $CO_2$, $O_2$, $N_2$, were tested using numerical data and statistics in order to assess the aging failure rate of distribution transformer systems failure to improve upon energy supply system . This method of modelling and simulating gasses is very relevant because it provides a platform to assess the expected result. Simulations have also helped finding tune the design in case the expected results were not met. The whole design process was then modified and the process was continuously repeated until satisfaction. This saves cost and reduces the risk of getting unsatisfactorily results.
The relevance of the design method is also proven by similarities with related work studied in the literature review. Transformer distribution systems failure has always been categorized through a process of modelling and simulation.
Based on the fore gone discussions, an experimental design adopted for this research to enable the author achieve his research objectives. The choice of this strategy is because the research design employ qualitative comparison Between Experimental and control groups with regards to the dependent variable ,computer simulation, official statistics, documentation and diaries.

**Dissolved Gas Analysis**

DGA is a well established diagnostic method, where oil samples are taken routinely and the composition of the gases dissolved within the oil analyzed. There are mainly eight gases of interest for diagnostics, outlined in 1.6 Table 1.6. Diagnostic gases.

| Gas | Chemical formula |
|---|---|
| Hydrogen | $H_2$ |
| Methane | $CH_4$ |
| Ethane | C2H6 |
| Ethylene | C2H4 |
| Acetylene | C2H2 |
| Carbon Monoxide | CO |
| Carbon Dioxide | $CO_2$ |
| Oxygen | $O_2$ |
| Nitrogen | $N_2$ |

The gas concentrations are measured and a trend of gas production is recorded. A developing fault within the transformer will lead to larger quantities of gases being generated and dissolved within the oil. The speed at which the fault gases are produced is an indication of the magnitude of the fault. For this reason, the rate of production is often considered to be of greater importance than the actual gas concentrations. The fault gases are also generated in small quantities during normal operation, due to natural aging. For this reason, relatively high concentrations of gas can also not be alarming if those concentrations are not increasing rapidly.

There are various methods of interpreting the gas results. These include well known methods such as: Total Dissolved Combustible Gases (TDCG), Duval's triangle, Roger's ratios, etc. [ IEC 60599,1999 and IEEE std, 2015]. An additional method developed by analyzing the DGA signatures of failed transformers and relating the identified patterns to gassing transformers in service to determine the fault cause and potential failure mode has been developed in [S Govender and A Singh, 2014].

The fault gases are produced in different quantities over a range of temperatures, due to the breakdown of the solid and oil insulation. An indication of how the gases are produced relative to temperature is shown in Figure 3.6. The temperature of the fault is an indication of the type of fault present. A partial discharge fault will generate relatively low temperatures and therefore, the main gases indicating this type of fault are Hydrogen and Methane. Arcing will generate extremely high temperatures and therefore, Acetylene is the prominent gas for this type of fault. Bare metal faults generate heat in the range 300-700 $CSo^-$ in this case, the gases that are expected to present would be Ethane and Ethylene. Since the paper is made of cellulose molecules, it is expected that any fault involving paper covered components would yield higher concentrations of Carbon Monoxide and Carbon Dioxide. $CO/CO_2$ is a ratio that can be used to determine whether the fault is thermal or dielectric[IEEE std. C57.104, 2014].

Since these gases are naturally occurring within the atmosphere, some caution must be exercised when interpreting these results.

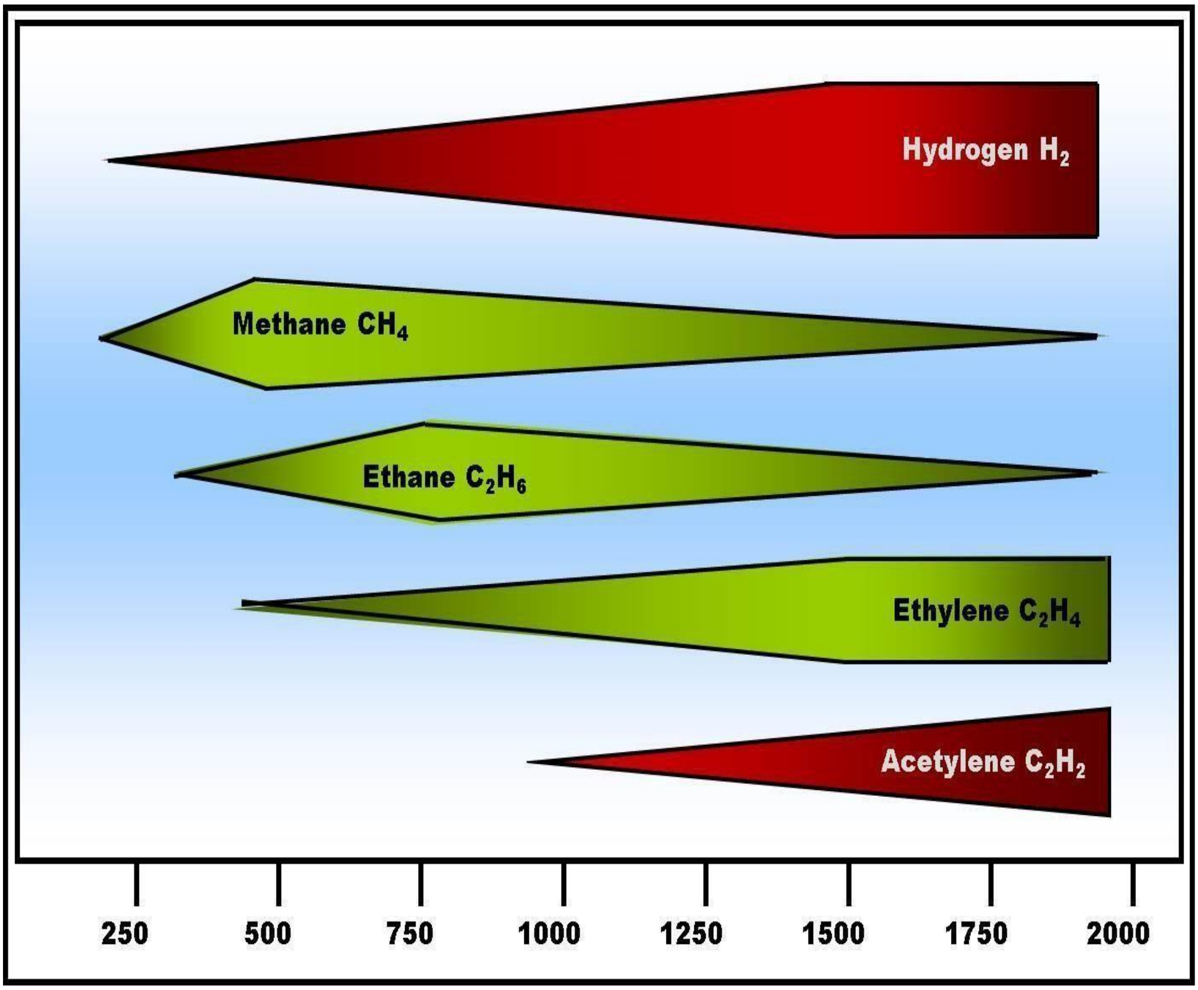
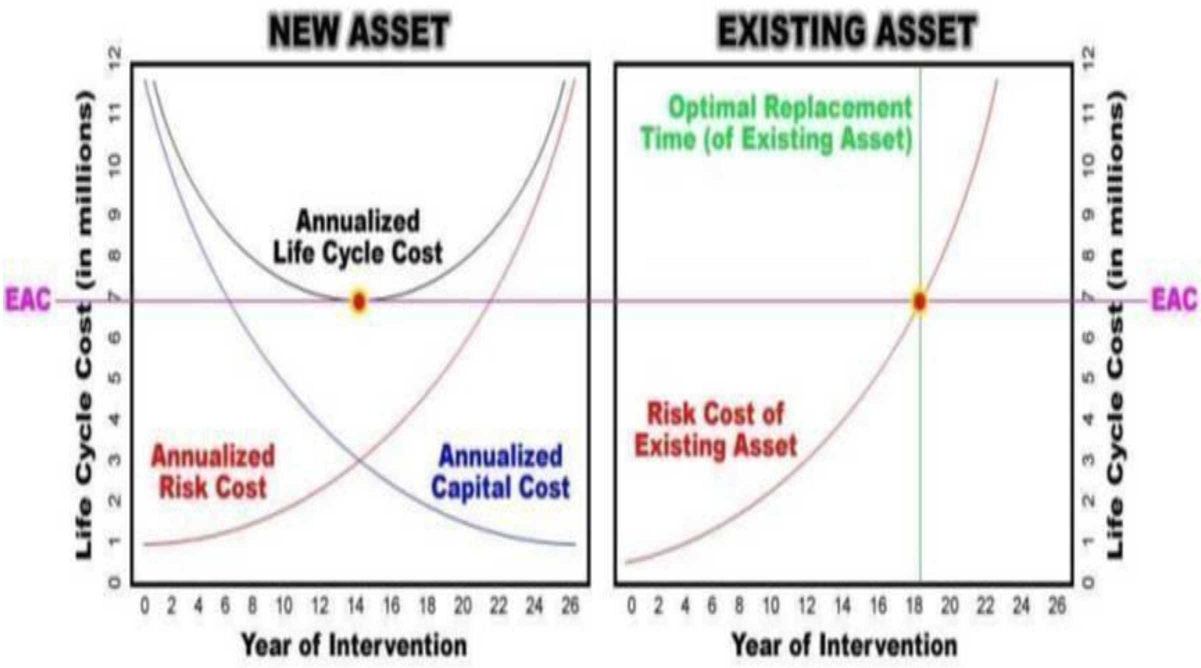



Figure 3.6: Gas concentrations at different temperatures *[P Gervais and J Aubin,21995]*.

In work done by Cigre working group A2-111 *[In Cigre, A2-111, 2012]*, it was found that reliable statistical data is often difficult to get since relatively few failures of distribution transformers are experienced over long periods of time. A method of life determination based on measurements of condition data and mechanisms of breakdown and failure, established from physics and chemistry tests, has been developed. The condition data is obtained from forensic analysis of failed units on site and a \Degradation Model" based on DGA and oil condition. A \Life Model" developed using the influence of temperature on the degradation of paper insulation and the model developed in [AIEE Transactions, 67, 1948] is then used. Failure rates based on condition and service time are then determined. Failure analysis using three tests: oil analysis, Furan derivatives analysis and HMM analysis is described in *[M.A.R Uzair, M Mohiuddin, and M.K Shujauddin, 2013]*. Oil condition and diagnostic gases (used in DGA, laboratory and online monitoring), as well as Furanic analysis (DP) are used to determine performance while HMM are used to determine probability of failure. In this method, similar to *[HICSS 2006]*, the state within the HMM is hidden and the outcome, which is dependent on state is visible. Hidden variables are co-related through a Markov process to determine the outcome, rather than independent of each other. In this way, the model is trained in a similar way to an ANN.

2.7 Condition, Criticality and Risk Assessment Model Asset management model An asset management model (CCRA) has been adopted by Eskom for optimizing investment decisions. The outcome of this model is dependent on the determination of probability of failure based on a HI that is developed using the life assessment parameters . It is there-fore necessary to evaluate the basis of this model, as well as the use of the HI parameters.




**2.8. Overview of methodology**

Eskom has adopted an asset management tool for optimizing asset refurbishment/replacement/retirement strategies. The Condition, Criticality and Risk Assessment (CCRA) model enables the phasing and structuring of project plans based on suitable cost-to-benefit ratios.

The inputs to this model include plant age, condition/HI, probability of failure and consequences of failure. The output of the model is the overall risk allowing analysis of cost/benefits of replacement/refurbishment. The timing of replacement is optimized according to the strategy shown in Figure:3.6.

Figure 3.7: Optimization of asset replacements [WERF, 2009].

Bathtub curves, as shown in Figure 3.7, developed from Weibull distributions depict the probability of failure of the plant as a function of age (P(age)). There are three distinct regions in a bathtub curve:

1.Infant mortality: indicated by a high failure rate in the first few years after manufacture, that decreases over time. These failures are generally attributed to inherent design defects or manufacturing errors.
2.Random failures: indicated by a at region (constant failure rate) over the age range in the middle of the expected design life, where random failures are expected throughout the population.
3.Wear out: indicated by an increasing failure rate with an increase in age. This region has the highest failure rate. This area is related to wear out and is the area of most interest in this study.
A HI for each plant type is determined, based on the condition of the plant, relative to end-of-life. The HI is a quantification of condition measurements that are taken on the transformer and an overall score is obtained. This is then used to determine a probability of failure as a function of HI (P(HI)). Bathtub curves, developed from Weibull distributions depict the probability of failure of the plant as a function of age (P(age)). If P(HI) > P(age), then P(age) is modified according to P(HI), i.e. the bathtub curve becomes steeper, indicating an increased rate of aging, and a new probability of failure is determined as a function of both age and HI
(P(age, HI)). This final probability of failure is then used for further analysis. This then comprises the \Condition" portion of the CCRA model

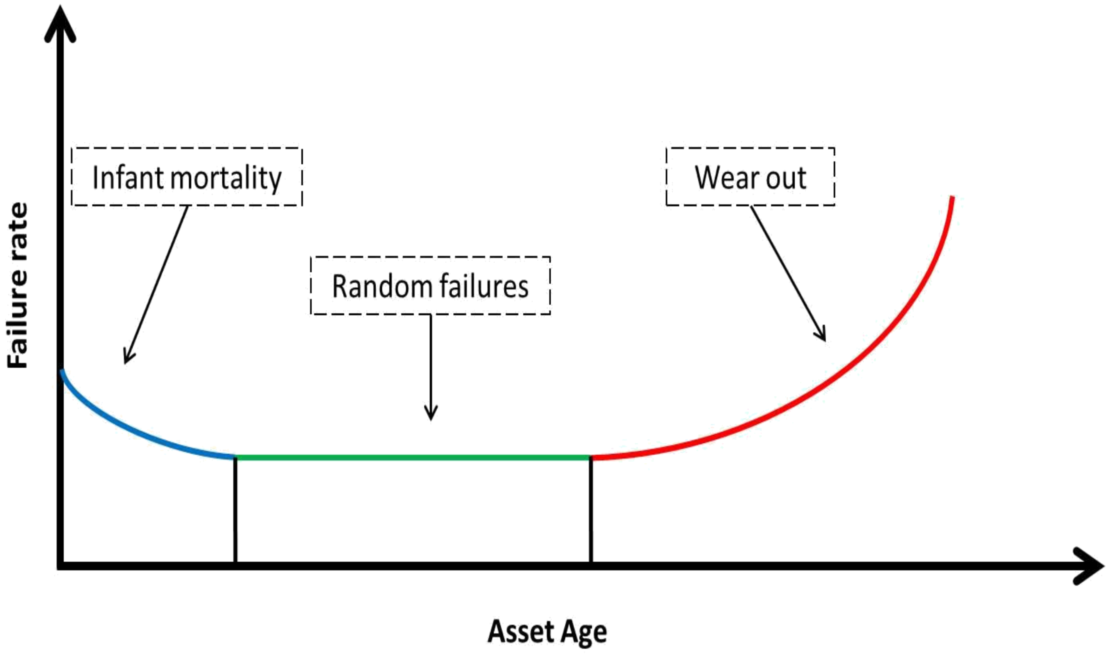



Figure 3.8: Example of equipment bathtub failure curve *[WERF, 2009]*.

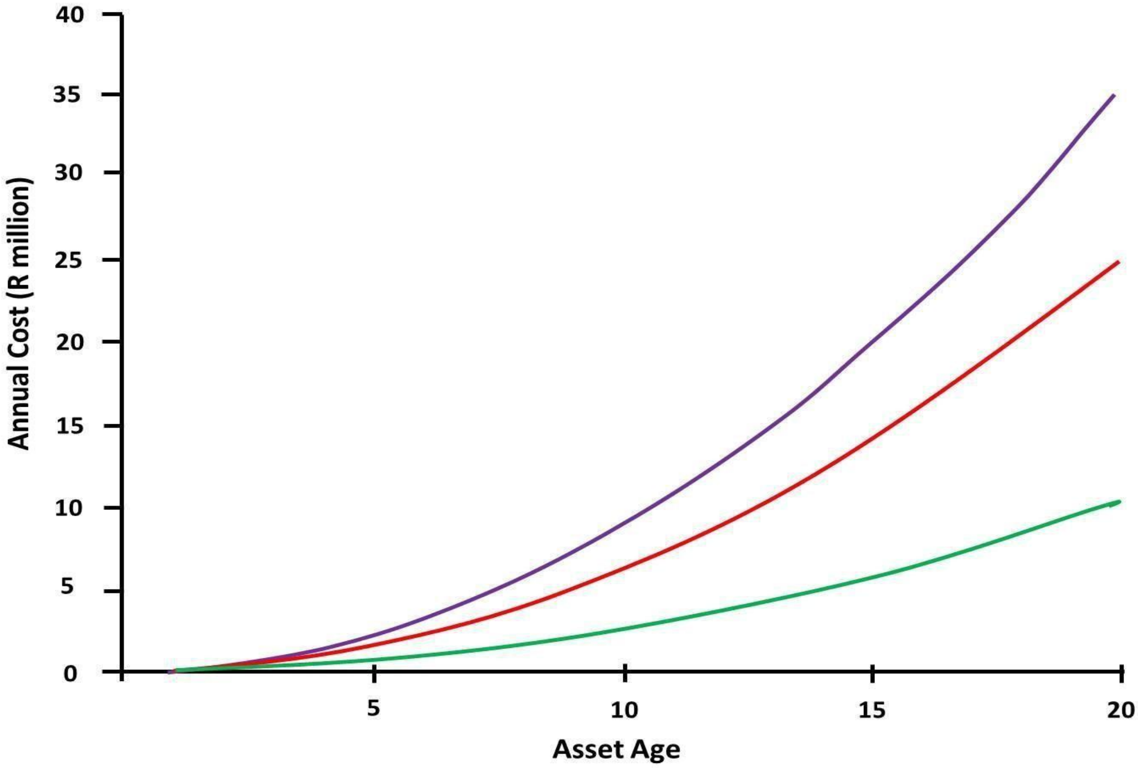



Figure : 3.9. Effect of Health Index modifier on probability of failure curve *[CCRA Economic Life Manual, 2009]*.

The consequence (or impact) of failure of the plant is then determined, to ascertain the impact such a failure would have on the system, should it occur. This is done by analyzing:

- the position of the plant in the network
- whether it has (N - 1) redundancy
- whether or not contingencies are available
- what consequential damage could occur or is expected should the failure of the plant be catastrophic the impact of adjacent plant failing at the same time, or during the outage of the initial failure.

Different scenarios are drawn up for each failure event. These scenarios are developed from the Emergency Preparedness Plans (EPP) of each substation in which the transformers are installed. EPPs outline the processes to follow in the case of an emergency or loss of plant, to recover load. The consequences include the following:

- Replacement cost of failed plant
- Cost of emergency repairs, if applicable Cost of damage to adjacent plant
- Cost of unserved energy or customer interruption, should it occur Safety or environmental costs • Each scenario is then analyzed, including its probability of occurrence. This then comprises the \Criticality" portion of the CCRA model and is a monetary value.

Risk is defined as the product of the probability of an event occurring and the consequences associated with that incident or, the frequency and severity of the losses *[Routledge, 2009]*.

A method similar to that employed in the CCRA model is that of a time-dependent failure probability, based on available condition data *[Z Zhang, Y Jiang, and J.D McCalley Iowa State University]*. The primary objective of this method is for the optimization of required maintenance interventions. A probability of failure based on the commonly used Weibull distribution is used to determine equipment time-to-failure. A Bayesian approach is used with this model due to limitations in empirical data available.

By using a Bayesian approach, uncertainties due to lack of information are expressed via probability distributions. Unknown parameters, such as shape and scale in the Weibull distribution are considered as random variables. In this way, Baye's theorem is used to determine the posterior probabilities based on condition data. This acts as a modification of the expected prior probability of failure that is assumed, based on age and failure rate, as done in *[Jagers J.N, Khosa J, de Klerk P.J, and Gaunt C.T, 2013]*.

$$\text{Risk} = \text{Probability x Consequences} \qquad (3.1)$$

Risk is defined as the product of probability and consequence or, the frequency and severity of the losses *[Routledge, 2013]*. The risk of each item of plant is calculated from the probability of failure that was calculated as a function of age and HI and the consequence cost that was calculated from the impact of the failure. This comprises the \Risk" portion of the CCRA model and is also a monetary value.

Due to limitations in financial resources, it is necessary to perform some prioritization in justifiable projects. This is done in order to justify spending alternatives. The benefit versus the risk of implementing a project this year rather than delaying it by a year or more is determined. In order to do this, the benefit to cost ratio is evaluated. The risk value is compared to the cost of refurbishment/replacement of the asset. If this ratio is very high, the refurbishment/replacement strategy is prioritized. If it is very low, the refurbishment/replacement strategy will more than likely be deferred to a later time. This is useful in determining the impact on the business of a decision to delay projects.

An accurately defined probability of failure, relevant to the population of transformers being analyzed is critical, since it forms the basis of this calculation. With a probability of failure that is not representative of the population being analyzed, the risk analysis is inaccurate and decision-making will be awed and unoptimized. This leads to an increased risk of wasteful expenditure and an attempt to decrease the asset failure rate, rather than the real risk associated with the impact of the failure, which is the ultimate goal.

**. Failure definition**

Asset management tools are refined based on probability and subsequent consequences of failure. Failure can range from anything between a minor defect which can be repaired on site, to a catastrophic event, necessitating the replacement of the failed transformer, as well as adjacent plant and possibly interruption of supply. Failure of a power transformer is therefore defined as per the definition of failure found in *[Taleb N., 2016]*.

“Failure: the termination of the ability of a circuit, bay or item to perform a required function” Failures

are then separated into three levels which are defined as follows:

• Severe: The transformer requires replacement or removal from site to facilitate repair within a factory. In both instances, a new transformer will be installed to return the circuit to service.

• Intermediate: The transformer requires repair, but this can be implemented on site. This is usually intrusive work to restore the transformer to working condition and return it to service.

• Minor: these are trip events that remove the transformer from service temporarily. No work is required in order to return to the plant to service, since the transformer’s major components have not been affected.

Failures of components that are critical to the operation of the transformer are also considered failures of the transformer. For example, the failure of an HV bushing, or an OLTC will be considered failures of the transformer since the transformer cannot operate in the absence of those components.

For the purpose of this study, only failures classified as severe or intermediate are considered since the minor failures have no direct impact on the end-of-life of the transformer, or the decision to refurbish/replace/retire it.

**Modes of failure**

[Land Y, 2017].As outlined in Cigre working group 12.05 report *[Bossi A, 203]*.
, international surveys are performed on failures of large power transformers and failure statistics are reported with 10 year intervals. These statistics are based on standard modes of failure with the following root causes: Core

Windings

Bushing

Tapchangers

Main tank and oil system Auxiliaries

Other

The breakdown of failures, by cause, experienced within Eskom Transmission during the period 1996 - 2014 is as shown in Figure 4. From this graph, it can be seen that the vast majority (56%) of all failures experienced during this period is related to external components such as bushings and tapchangers.

**Model parameters**

The existing HI model for distribution transformers consists of a number of measurable condition parameters. These are separated into four categories, each with different weightings. The high level composition of the HI is shown in Tables 1.7, 1.8, 1.9 and 2.

Table 1.7 provides the parameters related to the most influential component of the distribution transformer impacting the end-of-life and expected remaining life. These parameters provide information about both the solid and liquid insulation.

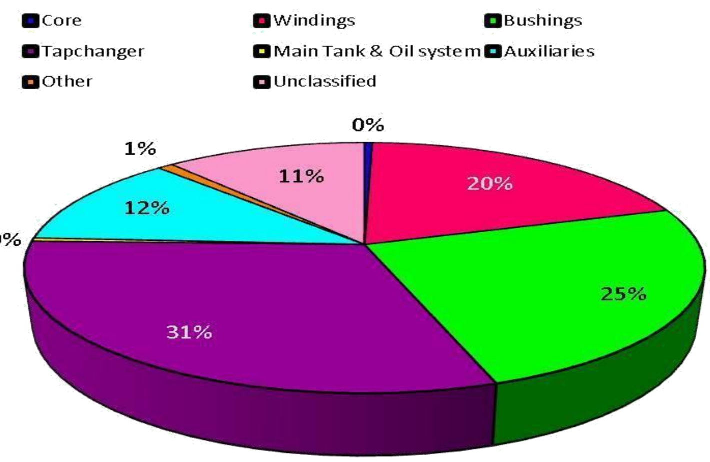



Figure 4: Breakdown of failures in Eskom Transmission as per Cigre reporting structure [*Bossi A Electra, No 88, 1983*].

| | |
|---|---|
| Insulation | Oil test |
| Furanic analysis | DP |
| Oil quality | Moisture |
| | IFT |

| | |
|---|---|
| | Dielectric Strength |
| | Acidity |
| | Colour/appearance |
| | Tan |
| | Sludge Table: 1.7 Overview of distribution transformer HI insulation parameters. |

Table 1.8., provides the electrical tests used to identify potential faults in various components of the active part of the transformer. Faults in any of these components can lead to catastrophic failure of the transformer.

Table: 1.8. Overview of distribution transformer HI electrical test parameters

| Electrical test | Sub – test |
|---|---|
| SFRA | N/A |
| DC resistance | N/A |
| tan  and winding capacitance | N/A |
| Core insulation resistance | N/A |
| Infrared scanning | N/A |
| Magnetizing current | N/A |

Table 1.9. outlines the various visual inspections, checks and tests that are performed during routine inspections and maintenance activities. The purpose of these checks is to identify deterioration that has the potential of developing into a transformer fault. Any defects identified during these inspections can be corrected with maintenance activities.

Table: 1.9. Overview of distribution transformer HI visual inspection parameters

| Component | Visual inspection |
|---|---|

| | |
|---|---|
| Bushings | tan and capacitance |
| | Visual inspection |
| Tapchanger | Oil test |
| | Speed test |
| | Contact thickness |
| | Transition resistance |
| | Number of operations |
| Visual inspections | Oil leak |
| | Conservator condition |
| | Cooling system condition |
| | Tank and overall physical condition |

Table 2. shows the DGA results that are used in condition assessment evaluations.

| Parameter | Gas |
|---|---|
| Concentration | $H_2$ (Hydrogen) |
| | $CH_4$ (Methane) |
| | $C_2H_6$ (Ethane) |
| | $C_2H_4$ (Ethylene) |
| | $C_2H_2$ (Acetylene) |
| | CO (Carbon Monoxide) |
| | $CO_2$ (Carbon Dioxide) |
| Daily rate of production | $H_2$ |
| | $CH_4$ |
| | $C_2H_6$ |
| | $C_2H_4$ |
| | $C_2H_2$ |
| | CO |
| | $CO_2$ |

The combustible gas concent rations as well as the rates of production are of interest and are included. Of the four HI cate gories, only two are used in this study, namely: Insulation and DGA. The reason is that this study is concerned with the determination of probability of failure relative to end-of-life of the transformer and not all of these components are related to end-of-life determination.

The primary purpose of the electrical tests is to diagnose faults once the transformer has been removed from service. A deviation in the test results is indicative of an immediate threat to the transformer and would be repaired as necessary and retested, prior to re-energization. After such repair, the test results would again be satisfactory. For this reason, the test results are not considered useful in terms of condition monitoring or indicative of long term plant health deterioration. The requirement/reason for repair is more useful for identification of potential faults and investigation of root cause of failure.

Table: 2. Overview of distribution transformer HI DGA parameters

The visual inspections are related to maintenance activities and identified deficiencies can be addressed relatively easily. Again, should any fault be identified within the tapchanger or bushings, these will be repaired/replaced. Although these components have a large impact on the failure of power transformers, the routine condition monitoring/checks are not useful from a modeling perspective since they have binary condition values rather than the continuous values which are indicative of slow deterioration.

Also the following were carried out;

Lab testing was key factor .simulation software available were used to simulate some findings or results and secondary data of burnt transformers were considered.

**Data types**

Data can be classified into four main types as proposed by Stevens [Afifi . A; Clark .V; and May S;(2016) “Computer-Aided Multivariate Analysis, 4th Edition”. Chapman and Hall/CRC.], namely:

nominal, ordinal, interval and ratio. These are explained in Table 3.4 below.

Table 3.4: Steven's measurement system

| Type | Explanation | Example |
|---|---|---|
| Nominal | There are a number of distinct categories that the variable can be classified into. | Names Religion |
| Ordinal | There are a number of distinct categories that the variable can be classified into, and the categories have a known order. | Service ranking Factory ratings |
| Interval (discrete /continuous) | This is an ordinal variable with an equal distance between successive values. | Temperature Calendar dates |
| Ratio | Interval variables with fixed zero measurement points, hence preserving ratios independent of the unit of measurement. | Height Difference in time |

The data used in this study are comprised of three of these data types. These include: nominal, interval and ratio. Each data type is handled differently since different information is available from each variable

Missing data

When performing statistical analyses, the issue of missing data is always a concern. Due to the fact that this study is based on historical data (18 years old), it is apparent that some of the data required in this study will be missing. This may be due to a number of factors including: operating practices changing, carelessness of data storage and inability to recover certain portions of transformer records, etc.

There are three categories of missing data [Rubin D.B. (1978). "Multiple imputations in sample surveys a phenomenological bayesian approach to nonresponse". In The Proceedings of the Survey Research Methods Section of American Statistical Association, pages 20–34, ]: Missing Completely At Random (MCAR) where the data are missing independently of the DV or IV, Missing At Random (MAR) where the data are missing dependent on one or more IVs but independently of the DV, and non-ignorable where the missing data is dependent on both the IVs and DV. There are a number of ways of handling the missing data. These are outlined below.

Listwise/Casewise deletion

This is the simplest method of dealing with missing data. Any record that contains a missing variable is deleted from the sample. This can cause substantial reduction in sample size and lead to large biases [Carter. R.L. (2016) "Solutions for missing data in structural equation modeling". Research and Practice in Assessment, 1( Oommen T.V., (2015).

Pairwise data deletion

With this method, data records with missing data variables are used in the analysis only when the analysis does not involve the missing variable. Again this method can produce large biases and unequal sample sizes [Carter. R.L. (2016) "Solutions for missing data in structural equation modeling". Research and Practice in Assessment, 1([Carter. R.L. (2016) "Solutions for missing data in structural equation modeling". Research and Practice in Assessment, 1( Oommen T.V., (2015).

Mean substitution

Mean substitution involves substituting each missing variable with the mean of all corresponding variables within the entire data set. This is problematic since it reduces the variance of the variables substituted in this way, which can lead to underestimating the spread of the data [Burke. S(2011). "Missing values, outliers, robust statistics and non-parametric methods". Statistics and Data Analysis, 1:19–24].

Hot deck imputation

In this instance, the record that is most similar to that with a missing variable is found and the value of the variable in this record that corresponds to the missing variable is substituted for the missing value [Allison. P.D.(2012) "Missing Data". Sage Publications]. The difficulty that arises is in defining similarity, since this is contextual and is therefore not a simple task [Allison. P.D. (2012) "Missing Data". Sage Publications]. This method also does not account for uncertainty in the approximation.

Regression methods

variables, i.e. the missing variable becomes the response variable and the other variables become the predictor variables [Aiken .L and West. S.(2013) "Multiple regression: Testing and Interpreting Interactions". Sage Publications.]. This leads to a complete data set with a reduced standard error

Expectation maximization and Raw maximum likelihood

These methods can be used to handle data that is MAR. In these methods estimates are found of the most likely value that the missing variable might have. A vector of means and a covariance matrix are developed that are superior to those that are developed from the previous methods of approximating missing data that are mentioned above [Rubin. D.B.(2013) "Multiple imputations in sample surveys a phenomenological bayesian approach to nonresponse". In The Proceedings of the Survey Research Methods Section of American Statistical Association, pages 20–34], [Shafri .H; Suhaili. A; and Mansor. S (2015) "The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis". Journal of Computer Science, 3(6):419– 423.]. The disadvantage of these methods is that large sample sets are required.

Multiple imputation

Multiple imputation has a number of advantages over other methods of missing data approximation. This method involves having more than one estimate for a given missing variable, computed using other values within the data set. In this way, the variance between the estimates gives information about the uncertainty of the imputation. In this way, biases in the data are also reduced. Since there are a number of estimates for each missing variable, after multiple imputation, there are a number of complete data sets instead of just one. Each data set is analyzed individually and the results are then compressed to form only one final solution [ Rubin. D.B.(2013) "Multiple imputations in sample surveys a phenomenological bayesian approach to nonresponse". In The Proceedings of the Survey Research Methods Section of American Statistical Association, pages 20–34]. This method is also disadvantageous since large data sets are required.

Since the data set available for this study exhibits non-ignorable missing data and is of limited size, the preferred methods of handling the missing data are not practical. For this reason, the listwise deletion method was utilized. This leads to significant reduction in sample size and the data needs to be evaluated for biasing.

Data visualization

Visualization of the raw data set is a simple method for determining the integrity of the data. By plotting the raw data, it is easy to visualize any discrepancies within the data, for example, if you have a dichotomous variable, but have values lying at points between 0 and 1, then it is obvious that the data set contains bad data. It is prudent to perform this quick check, prior to any data processing or analysis to confirm that the data is in fact as expected.

Gas concentrations and production rates

The concentrations of the different gases was plotted against the failures for the data set to be used in this study. Since the data was collected for only failed and healthy transformers, the failed variable can only have a value of either \0" (indicating a healthy transformer) or \1" (indicating a failed transformer)

The data is examined to determine conformity to an expected pattern. In this case, the data is expected to form two distinct groups, one at low concentrations of gas in healthy transformers and another at high concentrations in failed units.

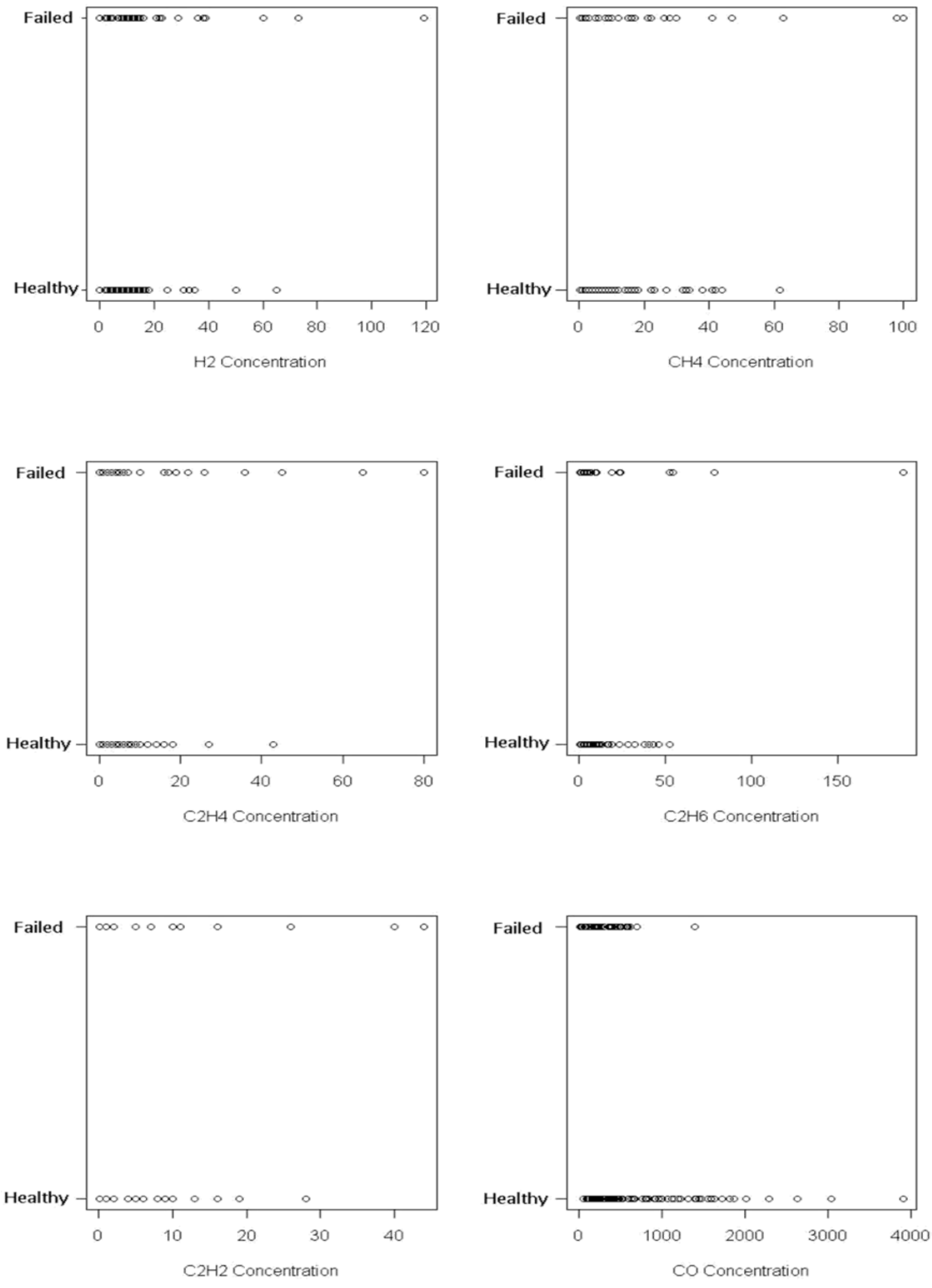



Figure .4.5: Plot of gas concentrations vs failures From the results shown in Figure:4.5, it can be seen that, with the exception of Carbon Monoxide, the expected trend in gas concentrations is evident, with higher gas concentrations present in the failed transformers and lower concentrations in the healthy transformers. The opposite trend is present in the CO data, which is not expected.

The daily rates of production of the different gases was plotted against the failures for the data set to be used in this study. From the results shown in Figure 4.6 it can be seen that, with the exceptions of Carbon Monoxide and Ethylene, the expected production rates are evident, with higher production rates in the failed transformers and lower production rates in the healthy ones.

Similarly to the gas concentration data, the daily rate of production data is examined to determine conformity to an expected pattern. In this case, the data is expected to form two distinct groups, one at low rates of gas production in healthy transformers and another at high rates of gas production in failed units.

The trends of CO and $C_2H_4$ show higher rates of decrease in concentrations for failed transformers. This decrease in production rate of CO is congruent with the concentrations of the gas in failed transformers.

DP and oil quality

The values of the DP and oil quality tests were plotted against the failures for the data set to be used in this study.

The results shown in Figure 4.6 indicate that there are higher moisture and lower dielectric strength measurements in the sample of failed transformers as expected.

The measurements of acidity and DP indicate values that are fairly evenly distributed across their ranges for failed and healthy transformers alike. Lower DP values and higher acidity values in failed units would have been expected.

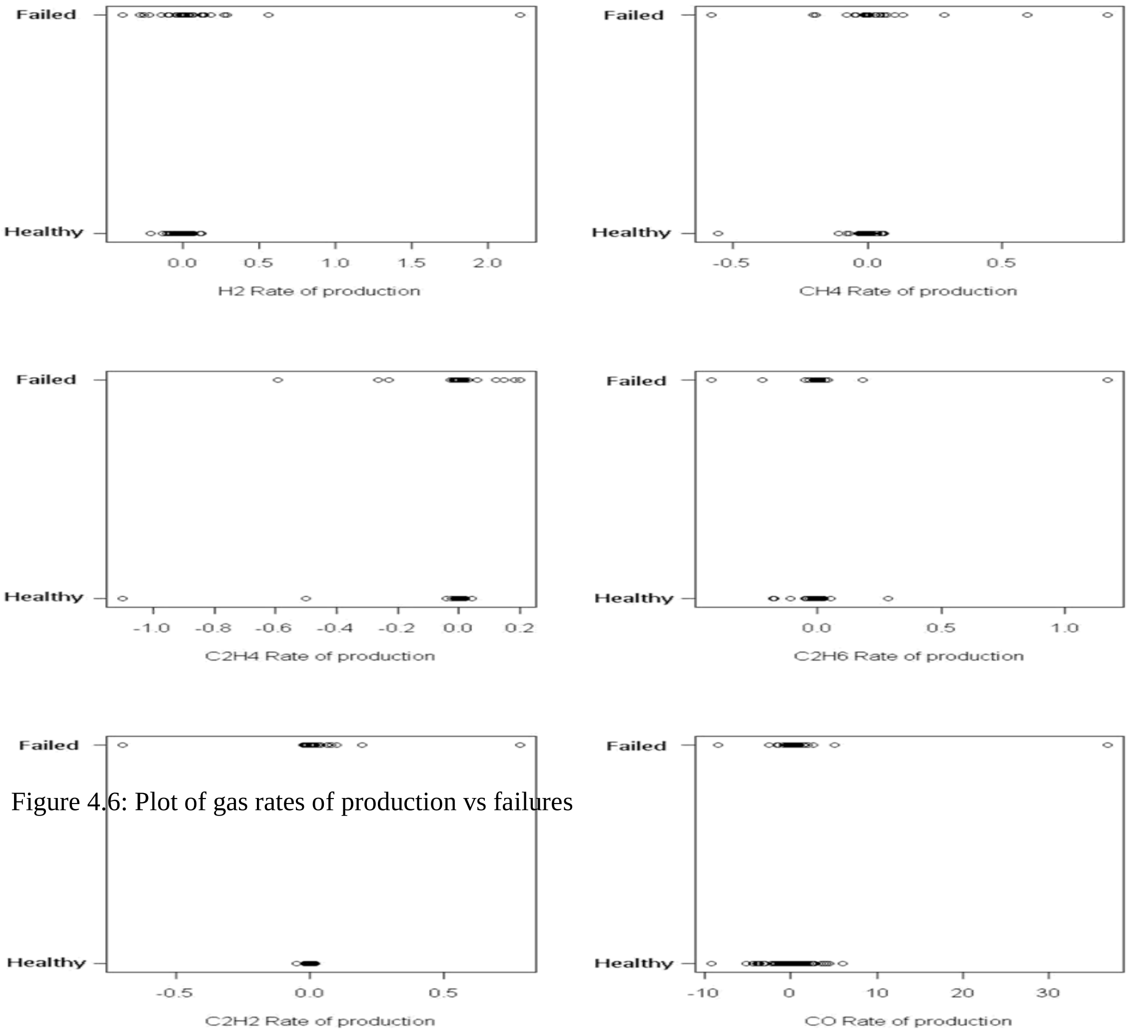



Figure 4.6: Plot of gas rates of production vs failures

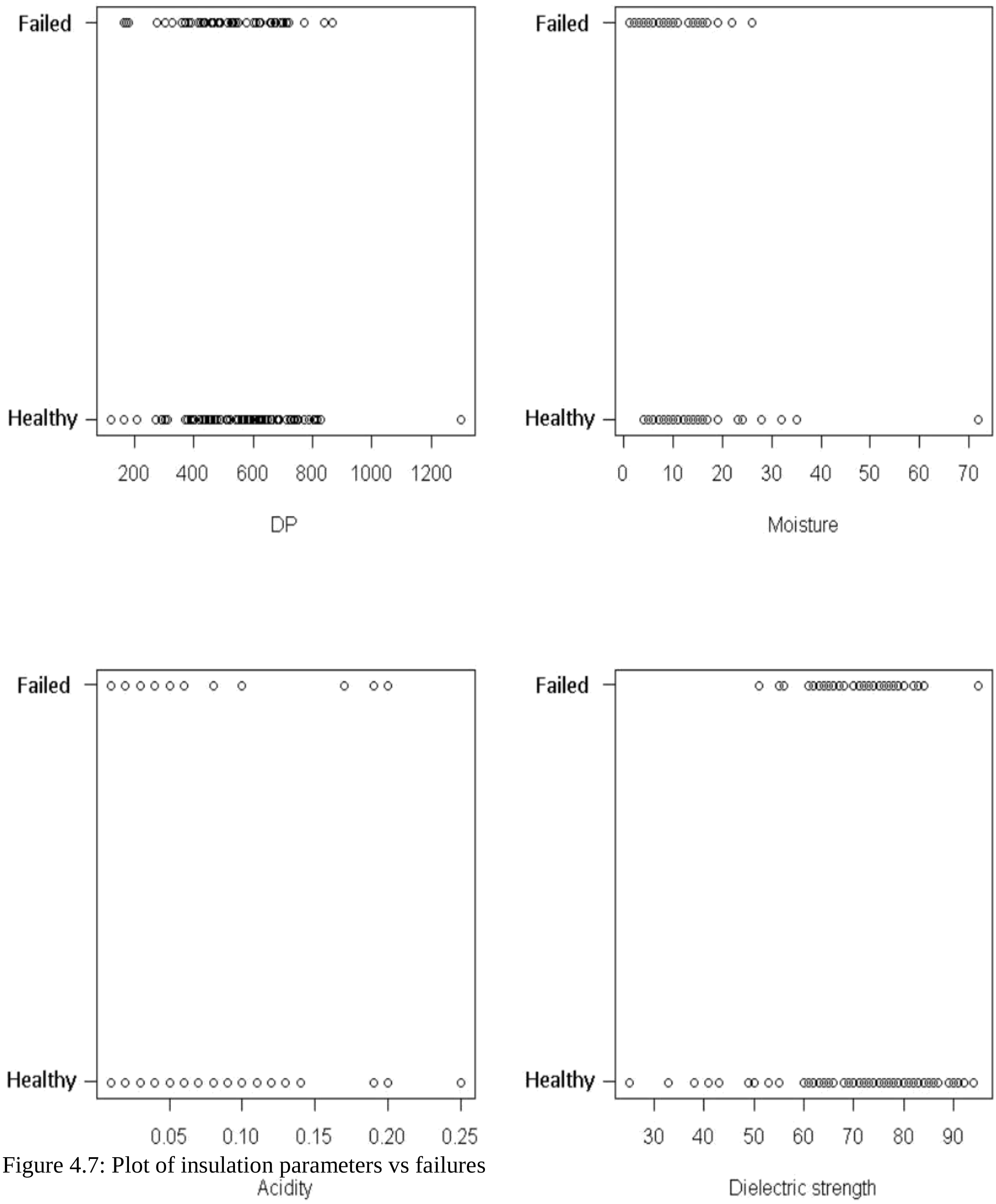


Figure 4.7: Plot of insulation parameters vs failures

Health Index

The values of the DGA and Insulation HI scores were plotted against the failures for the data set to be used in this study.

The results shown in Figure 4.8, indicate the expected trend in scores for failed transformers is lower than for healthy ones in both insulation and oil quality.

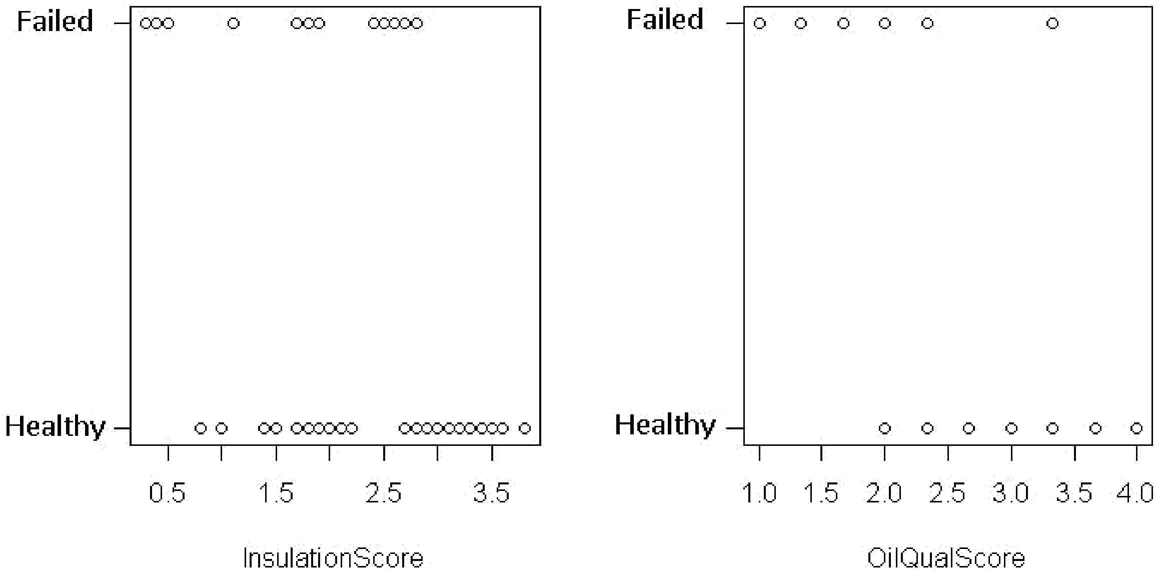



Figure 4.8: Plot of HI scores vs failures

Transformation of data

Data are usually normalized according to the min-max minimization algorithm into a range [Eskom and ABB Powertech (2018). “Theory, Design, Maintenance and Life Management of Power Transformers, Power Series Volume 5. Y-Land”.] according to equation. Equation 5.1 below

$$x_{norm} = \frac{x \;\; x_{min}}{x_{max} \;\; x_{min}} \qquad (5.1)$$

where:

$x_{norm}$ is the normalized value in range [0 1] x

is the value to be normalized

$x_{max}$ is the maximum possible value for a particular observation x $x_{min}$ is the minimum possible value for a particular observation x

Various other transformation techniques can be employed to normalize/linearize the data as required by the specific statistical analysis being performed. Normalization is however not a requirement for logistic regression and such data processing is therefore not required for this data set. The logit function used in the Logistic Regression linearizes the DV.

Outliers

Box plots are then used to determine the skewness of the data, and identify possible statistical outliers [ Schwertman .N; Owens M; and Adnan R (2014) “A simple more general boxplot method for identifying outliers”. Computational Statistics and Data Analysis, 47:165–174]. All identified outliers are included in the sample analyses, unless they have been confirmed to be erroneous values.

This method uses examination of the statistical percentiles of the data set. The statistics of interest in this analysis are: maximum, minimum, median, mean, 1st quartile (25th percentile) and 3rd quartile (75th percentile). These statistics are then plotted as follows:

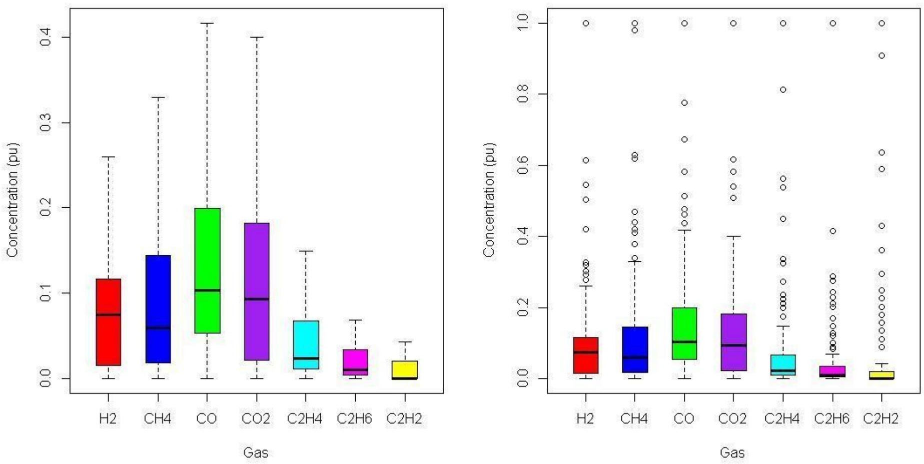
1st and 3rd quadrant form the top and bottom of each box for each variable. 50% of the data is represented within this box, with the length of the box being the interquartile range.

The median is represented by the horizontal line within the box. A line that is not perfectly centered is an indication of skewness of the data which is a measure of asymmetry about the sample mean. Skewness in the data is an indication that the data are not normally distributed. This is not of concern in this study since normality is not a requirement of logistic regression.

The maximum and minimum of the sample are represented by lines (sometimes referred to as \whiskers") extending from the top and the bottom of the box.

A general assumption is that an outlier is a value that falls more than 1.5 times the interquartile range away from either the top or the bottom of the box [ Schwertman .N; Owens. M; and Adnan.R.(2014). "A simple more general boxplot method for identifying outliers".Computational Statistics and Data Analysis, 47:165–174.]. Therefore, if no outliers are present, the maximum value within the sample falls on the top of the upper \whisker" and the minimum falls on the bottom of the lower \whisker". Potential outliers are indicated by the presence of data points either above or below the upper and lower \whiskers" respectively.

While outliers may be identified statistically, the data samples need to be evaluated critically since the \outliers" may in fact just be extreme values that have a critical impact on the analysis and should not be removed without just cause.




| | $H_2$ | $CH_4$ | CO | $CO_2$ | $C_2H_4$ | $C_2H_6$ | $C_2H_2$ |
|---|---|---|---|---|---|---|---|
| Min | 0.1 | 0.1 | 6 | 0.1 | 0.1 | 0.1 | 0.1 |
| $1^{st}$ Quad | 2 | 2 | 215.5 | 244 | 1 | 1 | 0.1 |

Figure 4.9: DGA concentrations(left), including potential outliers(right)

Both Carbon Monoxide and Carbon Dioxide ,have higher concentrations and daily rates of production than thehydrocarbon gases. For this reason, the values of all gases were scaled to put values for the purpose of graphical comparison. The boxplots in Figure 4.9, show the gas concentrations of the DGA data both with and without identified outliers.

These plots indicate a number of potential outliers in the data by the points above the top whisker. Table 3.5, shows the actual values of the percentile statistics for each gas. From these values, it can be seen that none of the gases have excessively high maximum values and in fact the maximums of each gas are still considered low. Generally these values would be discarded as outliers, however, the outliers that have been identified result from failed transformers' data and can therefore not be excluded. In this case, valuable information would be lost if these outliers were excluded from analysis. No obvious, real outliers are evident.

The gas production rates were also scaled to per unit values for graphical comparison. The boxplots in Figure 5, show the gas concentrations of the DGA data both with and without identified outliers.

These plots indicate the presence of outliers in both the upper and lower regions. On examination of the actual values of the

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Median | 9 | 6 | 409 | 1041 | 2 | 2 | 0.1 |
| Mean | 11.78 | 11.79 | 611 | 1384.6 | 6.2 | 8.3 | 2.7 |
| 3rd Quad | 14 | 14.5 | 783.5 | 2040.5 | 5.5 | 6.5 | 1 |
| Max | 119 | 100 | 3912 | 11200 | 80 | 188 | 44 |

Table 3.5: Percentile summary of DGA concentration percentile **statistics show**n .

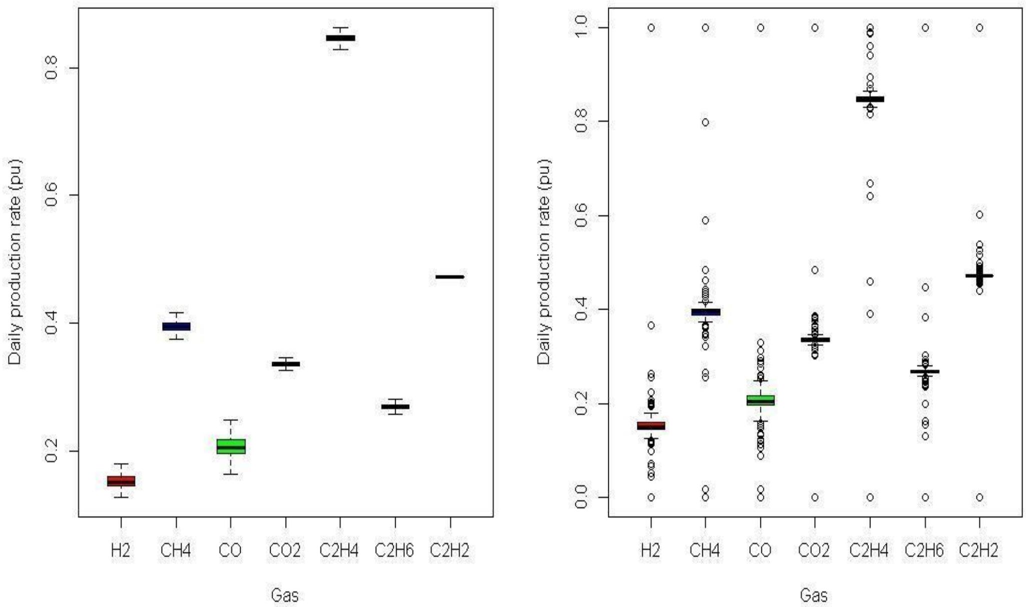



Figure 5: DGA rate of production(left), including potential outliers(right)

in Table 3.5, none of the values are unrealistic and consequently cannot be eliminated as outliers.

Some uncertainty in interpretation of the production rate values is present due to the method of calculation employed. Manual oil samples do not lend themselves to reliable production rate calculations.

DP and oil quality.

| | $H_2$ | $CH_4$ | CO | $CO_2$ | $C_2H_4$ | $C_2H_6$ | $C_2H_2$ |
|---|---|---|---|---|---|---|---|
| Min | -0.394 | -0.583 | -9.298 | -137.4 | -1.101 | -0.432 | -0.703 |
| 1st Quad | -0.018 | -0.009 | -0.271 | -0.520 | -0.006 | -0.006 | 0 |
| Median | 0 | 0 | 0.135 | 0.125 | 0 | 0 | 0 |
| Mean | 0.019 | 0.002 | 0.266 | 2.255 | -0.014 | 0.001 | 0.004 |
| 3rd Quad | 0.021 | 0.007 | 0.742 | 1.708 | 0.006 | 0.004 | 0 |
| Max | 2.212 | 0.892 | 36.838 | 273.9 | 0.199 | 1.174 | 0.788 |

The measurements recorded for DP and oil quality were reduced to per unit values for the purpose of graphical comparison. The boxplots indicating distribution of the DP and oil quality data both with and without possible outliers are shown in Figure 5.1.

Table: 3.6. Percentile summary of DGA production rates

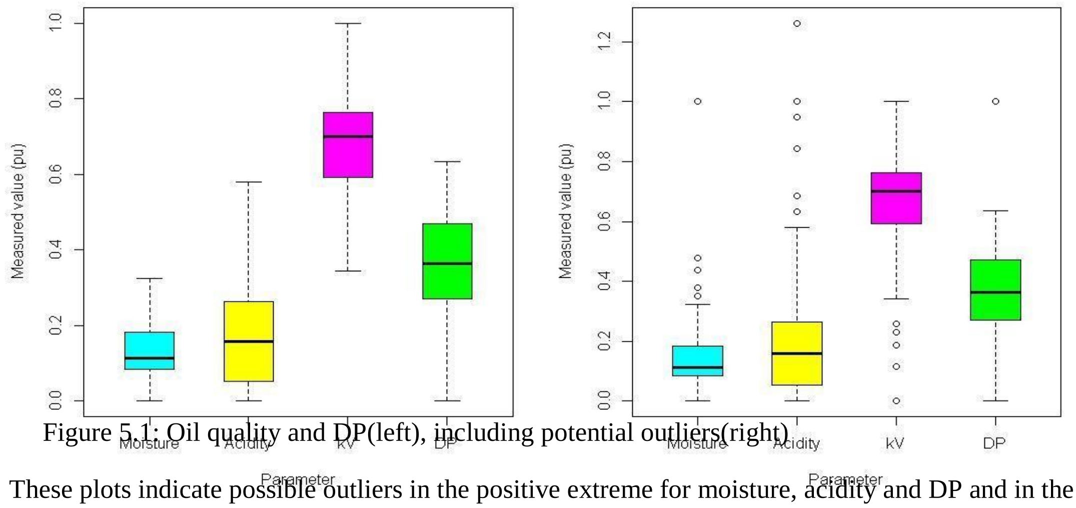



Figure 5.1: Oil quality and DP(left), including potential outliers(right)

These plots indicate possible outliers in the positive extreme for moisture, acidity and DP and in the negative extreme for dielectric strength.

Table 3.7: Percentile summary of Oil Quality and DP measurements

| | DP | Acid | kV | $H_2O$ |
|---|---|---|---|---|
| Min | 121 | 0.01 | 25 | 1 |
| 1st Quad | 440 | 0.02 | 67 | 7 |
| Median | 550 | 0.04 | 74 | 9 |
| Mean | 578 | 0.05 | 72 | 11 |
| 3rd Quad | 676 | 0.06 | 78 | 14 |
| Max | 1300 | 0.25 | 95 | 72 |

On examination of the actual values of the statistical percentile analysis shown in Table 3.7, it can be seen that the range of data values is reasonable, with the exception of the maximum value for DP.

This value appears to be unrealistically high since the expected value of DP for brand new paper is 1200 and once it has been processed during the manufacture of a transformer it has reduced to approximately 900. This value is therefore considered an outlier and removed from the dataset.

Health Index

The values of HI score are uniform for all components, in the range (0-5). For this reason, it was not necessary to reduce the values to per unit equivalents. The boxplots of the DGA and Insulation HI scores, both with and without potential outliers are shown in Figure 5.2. These plots indicate a number of potential outliers.

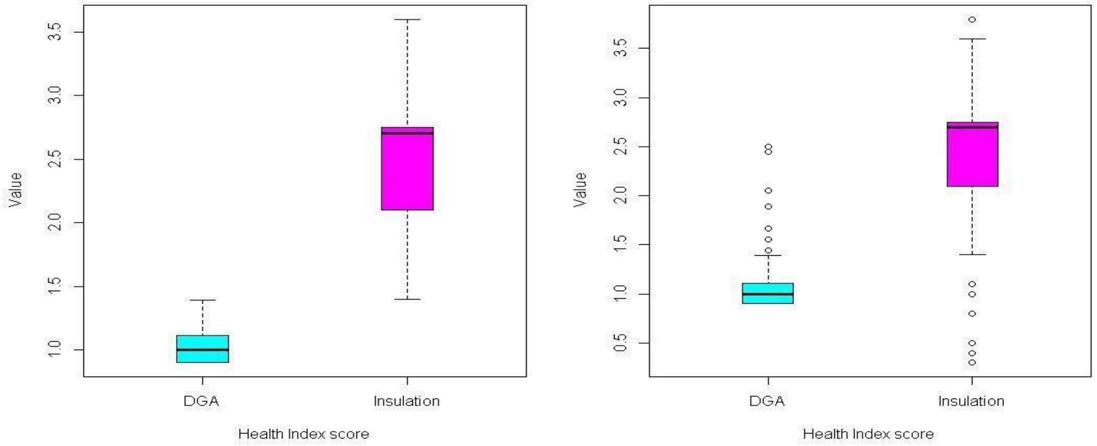



Figure 5.2: HI parameters(left), including potential outliers(right)

Table 3.8. Percentile summary of HI parameters.

| | DGA | Insulation |
|---|---|---|
| Min | 0.9 | 0 |
| 1st Quad | 0.9 | 2 |
| Median | 1 | 3 |
| Mean | 1.1 | 2.6 |
| 3rd Quad | 1.1 | 3 |
| Max | 2.5 | 4 |

Examination of the statistical percentiles in Table 3.8, shows values that are reasonable and cannot be eliminated as outliers. This is congruent with the decision not to remove outliers in the base data used to calculate the HI scores.

Conclusion

The data set to be used in this study was defined. The data was evaluated visually and found to display the expected trend. The data set was found to contain records with missing data and these records were then deleted, since the data set is too small to utilize methods of missing data approximation. The data was examined graphically, using boxplots to identify potential outliers. In this case, only the outliers identified in the DP measurements were removed since there was no evidence supporting the removal of the outliers identified in the other variables. The data set processed, using the methods outlined in this chapter, will be used in the development of the statistical model for probability of failure determination.

*REFERENCES*


1. Ranjana and Amarjit Singh, 2017

2. K**umar** at el 2017

3.Yin (2013),

3. Eduful & Mensah, 2016

4. Cohen, et al., 2012

5. Berg, Fritze, 2017

6. Frazer and Lawley (2014)

7. Bryman, 2004

8. *[Bartley, W.H. (2003). "Analysis of transformer failures, imia wgp33(03). In 36th Annual Conference of the International Association of Engineer's in Insurers, Stokholm,(2003)"]*.