Innovative Redesign of Adaptronic Transformer Casing

[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. Adaptronic transformer casing refined design development based on box model from computer science has a low failure rate and 99.7% reliable. Adaptronic is a design process that makes a product environmentally friendly, uses knowledge of many areas, complex product and conserves material through recycle, by use of high heat resistant, bullet resistance, anticorrosion and ballistic incorporated material, which are deflectors and bomb resistant, using box model from computer science. Statistical models were developed and tested for goodness of fit.

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

1.Introduction

Transformer is a device which uses the phenomenon of mutual induction to change the values of alternating currents and voltages. A distribution transformer is a static device constructed with two or more windings used to transfer alternating current. Distribution transformers provides the final voltage transformation in the electric power distribution system, stepping down the voltage used in the distribution lines to the level used by the customer. A Power transformer is also normally called transformer which is a device that transfers electrical energy from one circuit to another through inductively coupled conductors i.e the transformer's coils. They normally transfer high voltage.

Adaptronic transformer casing refined design with colour coding in figure 2.8, below shows a box model adopted from computer science with black being the external or outermost casing which developed from stealth. The next to this external casing is denoted with colour red which stands for explosive resistance bombproof materials blast bags and concrete walls. The next colour code yellow is the third outer bullet proof or bullet resistance casing nearer to the innermost green colour. The last green innermost casing nearer to the internal windings and the core materials of the transformer represents the actual transformer casing made up of high heat resistant elements like tungsten alloy used in the manufacturing of aircraft and space exhaust cones which uses jet propulsive engines like pure jet, turbofan, turbo-rocket, and gas turbine engines found on found on aircrafts like B52 stratofotress bomber, thunderbirds, firebirds, from United States Air force, K8 ighter jets from China and TU 95from Russia.

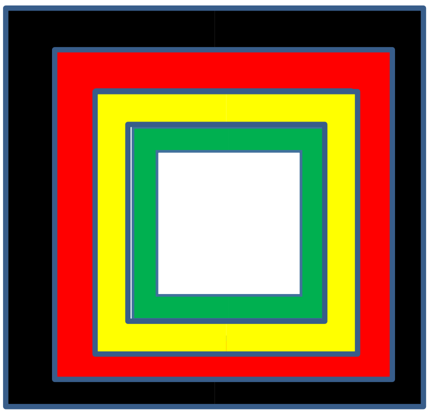


Fig:2.8. Adaptronic transformer casing refined design

Fig. 2.8. shows adaptronic transformer casing refined design using box model and colour coing from computer science.

Green ……………..Actual transformer casing

Yellow…………....Bullet resistance (Bulletproof material)

Red ……………….Explosive resistance bombproof materials blast bags and concrete walls

Black……………..Stealth casing.

Bullet proof materials; The ballistic panels deployed in bullet proof vests comprised of various materials, depending on the manufacturer. Body armor must past specific tests before it can be sold, so the variety of materials should all provide the same minimum levels of protection. Unfortunately this is not always the situation, as some types of body armor, such as Zylon, have been shown to fail after short periods of use. After its release Zylon was shown to degrade rapidly by independent tests, and this led to the recall of all Zylon based bullet proof vests in 2005.[1]. When purchasing body armor you should always unsure that it is made from confidential, high quality materials such as those listed below.

Para-Aramids; Aramids are manmade synthetic fibers that are heat impervious and extraordinarily strong, with exceptional strength-to-weight ratios. They were industrialized in the early 1960's, and mainly pioneered by DuPont™, who developed a meta-aramid called Nomax®. Further developments by DuPont™ led to the
introduction in 1973 of their first para-aramid,

Kevlar®.[2]. This led to a revolution in the bullet proof vest industry, as Para-aramid materials allowed body armor manufacturers to develop bullet proof vests that were lightweight and flexible, and which also offered a high level of protection to the wearer. Materials used for bullet proof vests such as Kevlar® and Twaron® are para-aramids, and have been used in body armor for nearly 30 years. Aramids are generally formed by a reaction between an amine group and a carboxylic group, which creates an AABB polymer. This liquid chemical blend is then made into its solid form by spinning it together with sulfuric acid, which when cooled can be made into a fiber, powder or pulp. Kevlar® has gone through several stages of development since first being introduced in the 1970's. The first version, Kevlar® 29, was a revolutionary step for body armor, as it allowed for the production of protective panels that were both flexible and also easily concealable. This meant DuPont™ could now produce a piece of body armor that was lightweight, and that could be worn comfortably by people on a day to day basis. DuPont™ continued to develop Kevlar® for use in body armor, and in 1988 they released their second version, known as Kevlar® 129. This improved version was even lighter than Kevlar® 29, and offered increased ballistic protection for the wearer, including protection against high energy rounds from weapons like a 9mm FMJ. The most recent version of Kevlar® was released in 1995 and named Kevlar® Correctional. This revolutionary version could stop attacks from knives and other such weapons, and led to the production of lightweight, multi-threat forms of body armor that were capable of stopping both ballistic and stab threats.

Ultra-high-molecular-weight polyethylene.Ultra-high-molecular-weight polyethylene, or UHMWPE for short, is additional popular choice for ballistic panels, and is used in many modern types of body armor. It has many similar characteristics to para-aramids, but is a type of polyolefin, made up of extremely long chains of polyethylene. UHMWPE can be formed in several different ways, including ram extrusion and compression molding. However when it is used for body armor it is created using a process called gelspinning. This involves drawing dissolved ethylene through a series of small holes, which creates a gel material. Two pieces of this gel are then sealed within polyethylene film, and this creates a composite that can then be made into both flexible, lightweight ballistic panels, or more rigid, hard armor plates. Research has shown that the strength-to-weight ratio for UHMWPE can be up to 40% higher than that of traditional para-aramid fibers Because of this it is becoming increasingly popular in the body armor market, with many companies releasing UHMWPE based types of body armor, such as Dyneema® and Spectra®, as their flagship models.

Hard Armor Materials. Early types of tough armor were made from metals such as steel, nevertheless they were heavy and often ineffective. Like soft body armor panels, modern hard armor can be made from a range of materials. The most popular form of hard armor plates are made from ceramics or ceramic composites, often with a paraaramid backing such as Kevlar®. These are generally the cheapest forms of hard armor, but are also the heaviest. Manufacturers have also developed Polyethylene and Monolithic based hard armor plates, which are lighter than ceramic types of hard armor, but are also more expensive.[3]. Ballistic helmets and other types of hard armor are normally made from a para-aramid base, such as Kevlar®, which is then coated in a tough thermoplastic resin. The resulting material is both thin and lightweight, making them comfortable when worn for long periods of time.

_Different Sections Of Body Armor. Body armor is made up of several different sections. At its most basic level a vest will have two protective panels, which are worn across the chest and back, contained within the vest carrier itself. These protective panels are what provides the protection for the wearer, without them inserted the vest is simply that, a normal vests.

Different Vest Covers. Body armor can come in a range of covers, these can be made from traditional material such as cotton, or manufacturer specific materials such as CoolMAX®, Cordura® or Gore-Tex®. Depending on the type of body armor the ballistic panels may be sewn into the vests carrier, however recently the industry has moved towards making the panels removable. This is preferable as it allows the vest carrier to be washed, and also allows for the user to change the protective panels if they wish, for example from an NIJ Level II panel to a NIJ Level IIIa panel.

CoolMAX®; CoolMAX® fabric features a moisture management system, and is used in many different types of performance clothing. CoolMAX® covers are lightweight and thin, and used mainly for covert vests. CoolMAX® is designed to be worn against the skin, and features technology that keeps the wearer cool by pulling sweat away from the body.

Cordura®; Cordura® is constructed using high tenacity fiber technologies, and is used to make a wide range of highly durable products. Used for overt vests, Cordura® is a heavy duty fabric which is designed to withstand the rigors of day to day wear.

Gore-Tex®; Breathable and waterproof, Gore-Tex® is used for thousands of different products in a variety industries. It is used to make the covers of the bullet proof vests worn by many police forces around the world.**[**www.bulletproof material 2016].[www.bullet proof materials 2017]

By Kelly David Burke, Alicia Acuna Published June 02, 2017 . Air Force cadet Hayley Weir had an idea that turned out to be a game changer. "It was just the concept of going out there and stopping a bullet with something that we had made in a chemistry lab."

The 21-year-old Weir approached Air Force Academy Assistant Professor Ryan Burke with the idea. He was skeptical."I said, 'I'm not really sure this is going to work, the body armor industry is a billion-plus-dollar industry," he noted. Weir's idea was to combine anti-ballistic fabric with what's known as a shear thickening fluid to create a less heavy material to use in body armor. She demonstrated the principle to Burke by combining water and cornstarch in a container and asking the professor to jam his finger into the paste-like goo. "I jam my finger right into this bowl, and I almost broke my finger! Hayley's laughing because I've got this finger that I'm shaking and I'm saying, 'You know, that's pretty impressive stuff.'" Convinced, Ryan worked with Weir for several months in a small lab at the Air Force Academy in Colorado Springs. They were helped and advised by Dr. Jeff Owens, Senior Research Chemist at the Air Force Civil Engineer Center at Tyndall Air Force Base in Florida.They tried combining several different ingredients to come up with the exact formula for the shear thickening fluid, and the correct way to layer it with ballistic fibers. "The pieces are not new," Weir explains, "everything that we've used in there has been researched (before) in some capacity for ballistics protection." They tested their combinations on the firing range, failing time and again, until one day their quarter-inch thick design repeatedly stopped a round fired from a 9mm handgun.Weir and Ryan's excitement was tempered by the range safety officer who pulled his .44 Magnum and told them bluntly, "This will fail." Ryan says, "We loaded it in and it stopped it. And it stopped it a second time, and then a third time." They realized they had hit on something special, that could potentially lighten the average 26-pound body armor kit worn by servicemen in the field by as much as two thirds. "This is something that our competition doesn't have right now," Weir explained. "And with this advantage our soldiers, if they wear this body armor, will be able to move faster, run farther, jump higher." Body armor for the military and first responders may not be the only thing that can be improved by the new fabric. It could possibly be used to reduce or replace the thick metal plates that protect military aircraft, tanks and other vehicles. "And there' is some significant gravity and weight behind that," Ryan said. "And what it could mean for people like my friends who are still active duty in the military, that are going downrange, serving overseas." A patent for the as yet unnamed design is pending, and if money is ultimately made, the Air Force will share the profits with Weir, Ryan and Owens."It doesn't feel like it's that great of an achievement," Weir muses, "just because it's been something that we've enjoyed doing." The Air Force believes it is definitely a great achievement. They are providing the newly graduated 2d Lt Weir with a full-ride scholarship to Clemson University, where she will earn her Master of Materials Science and Engineering, before returning to the Air Force to continue her work. *Alicia Acuna joined Fox News Channel (FNC) in 1997 and currently serves as a general assignment reporter based in the network's Denver bureau.*

WWW.Bombproofmaterisls December16, 2015).There have been many incidents of airplanes being targeted with bombs that were exploded inflight, one doesn't have to look too far back in history to find an example of such a horrendous event, but an international team of scientists have come up with a revolutionary new material which may bombproof airplanes and protect against such incidents. The bombproof material would line the luggage hold so even if a device explodes in someone's checked baggage the plane would not be blown up while it's in the sky. [4]. The material is called Fly-Bag, it's a flexible bombproof material which will line the luggage hold. Fly-Bag is made from layers of heat resistant and very high strength materials like Aramid, which is a synthetic fiber that's used in body armor.

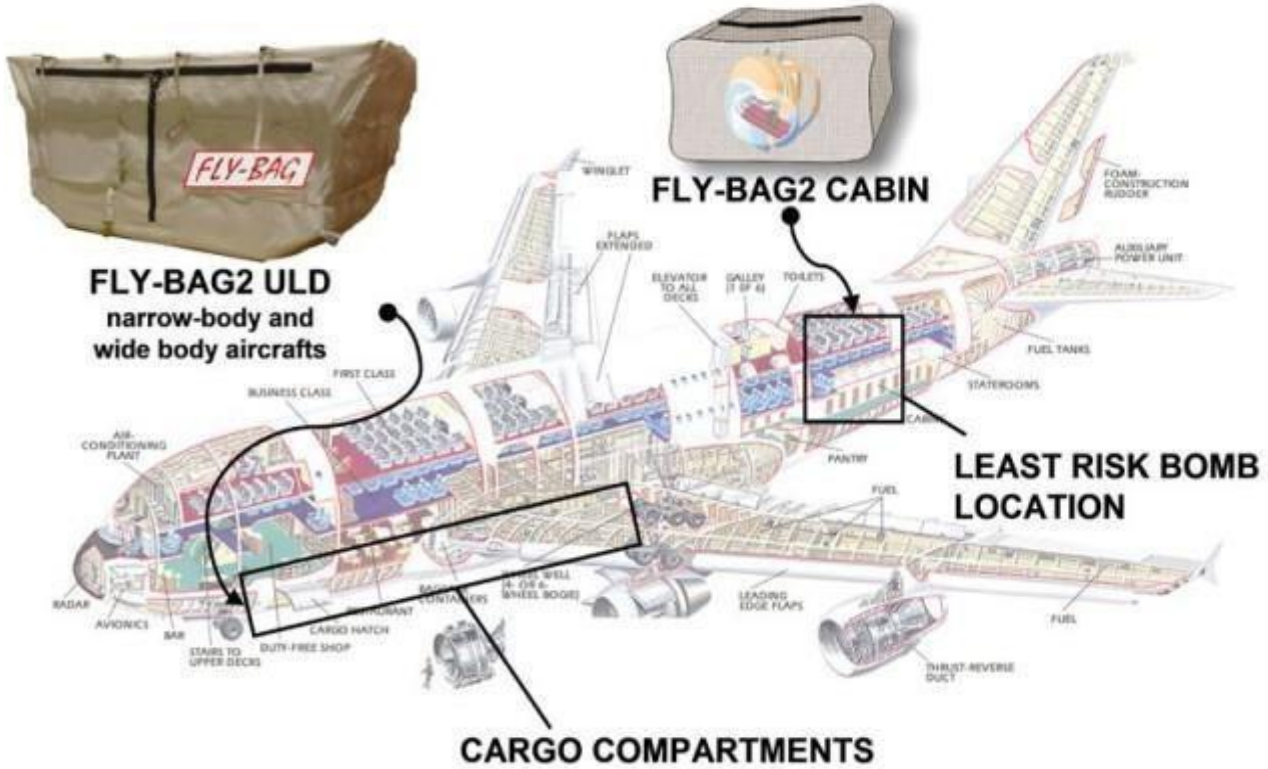


Figure.2.9. Bomb proof materials used in aircraft cargo compartments known as fly bags.

It's flexibility enables it to contain the power of a blast as well as the blast fragments that might fly as a result of it, the material is being developed by a group of European companies like Blastech with partners joining in from countries like Sweden, Italy, Germany, Spain and Greece.

Fly-Bag was initially tested at the blast laboratory of the University of Sheffield, the test proved that the prototype worked well, the team then conducted real-life controlled tests in luggage holds of an Airbus A321 and Boeing 747. The bombproof material works, giving us hope that perhaps in the future it may get much more harder for the unsavory elements to cause death and destruction by placing a bomb in an airplane's hold. [www.bombproof materials.com].[5]

Explosion-Proof Materials and Design; Date Published: November 18th, 2016. Categories: Construction Management Companies. Tags: Construction Management Companies, Construction Management Services, construction research laboratory, lab construction, technical construction. Imagine You've worked really hard to create a new material that is strong and yet, flexible. Now you get to test this material. To test this material, you and your coworkers head to the bomb range. This is often how many materials used in a variety of products from bomb-proof clothing to blast-resistant building materials have to be tested in order to truly test their effectiveness. As a construction management company for labs and their lab design needs, at LCS Constructors, we have to be on the cutting edge of technology and understand what options are available for the construction of blast resistant buildings. There have been many improvements in explosion-proof materials recently. Take a look at these neat options that are being developed by scientists around the world that can better help us provide your lab with top-quality, effective blast-resistant solutions.

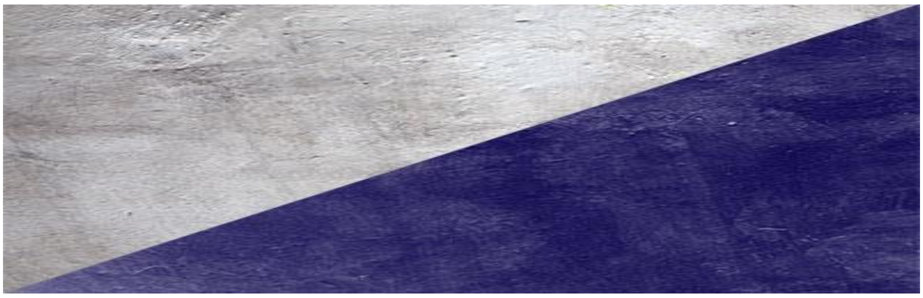


Figure:3. Explosion Resistant Concrete [Source: WWW.Explosion-Proof materials and Design (2016)].

Explosion-Resistant Concrete [Source ; WWW.Explosion-Proof Materials and Design 2016]. One of the most interesting developments in explosion-proof materials as in figure 3, is the development of a more explosionresistant concrete. Developed at the University of Liverpool, a new cement mixture provides a higher tensile strength than traditional concrete. This particular concoction utilizes less water than traditional concrete and also incorporates silica sand as the main aggregate over gravel or other crushed minerals. The real genius in the material is that it also incorporates steel fibers which have a high compression rating. Tests on this martial have shown that the greatest asset that this material can provide is that it does not crumble and create shrapnel when exposed to close-proximity explosions, like those that may happen from a chemical production facility. There is some debate whether this material could ever be more cost effective than steel, but for times when concrete is the best option during a lab build, it's great to know that there are more advanced explosion-proof materials in the works.[6]

Figure: 3.1. Shock and Explosion-Resistant Glass [Source; www.Explosion proof materials, 2016].

While some lab designs may call for the exclusion of windows, we know it's definitely a better working environment for employees to not work in a concrete bunker. However, when you need a facility to be strong and blast-resistant due to its proximity to other locations that may work with hazardous materials, you need to have the discussion about shock and explosion-resistant glass as shown in figure 3.1. There are options that involve many inches thick layers of glass, and there are other options available using Plexiglass, which is a type of thermoplastic; but what if there was an option available that felt and looked just like glass with the blast-resistant properties of Plexiglass? Researchers at the University of Missouri and the University of Sydney have come together to create a new glass material that may be able to stand up to hurricane force winds and even be truly blast-resistant. This new glass material is more like a fabric in some respects. It's a weave of glass fibers that are then coated in liquid plastic and bound together with an adhesive. While not yet available on the market, it's testing well in small scale studies, even at close range to a blast. Blast Doors; Even blast doors have changed in design over time from just steal plating to a much more complicated design intended to stand up to substantial, nearby blasts better. Think of blast-resistant doors a bit like a sandwich. The bread, or exterior, is typically two steel plates. Between these two plates, there is a steel rebar grid that is then surrounded by concrete to create a solid center cavity. Blast doors have to be perfectly sized, and when they are, this design is able to protect people from chemical explosions and other high-temperature threats.

What is Hardening? While many of the materials we've talked about are easy to use during the lab construction process, going back and making a previously constructed lab resistant to any type of blast can seem like a monumental task. It's actually much more common than you think, however, and LCS Constructors has plenty of experience hardening labs to meet safety requirements in blast zones. Hardening is a process where the main "objective is to prevent collapse, minimize hazards associated with flying debris, and facilitate evacuation of a facility-all with the primary goal of minimizing human injury and loss of life," according to Joseph A. Demkin, part of the American Institute of Architects. This process takes an already existing building and adds specific safeguards to help protect individuals in case there is a need. There are four main points that go into hardening a building against the possibility for an explosion or blast based on a paper by Robert Smilowitz (2016): i. Establishing a secure perimeter ii. Mitigating debris hazards resulting from the damaged façade. iii**.** Preventing progressive collapse. iv**.** Isolating internal threats from occupied spaces. These four points all come down to how a blast, or the blast loading, affects a building. Blasts that occur close, will be strong but targeted, and blast further away reduce in strength, but they will also affect more of the building. By addressing the four points listed, the effect of the blast load can be greatly reduced in order to protect individuals in a building. The hardening process involves not just evaluating these important points that help reduce risk, but also finding ways to mitigate them.[7]. For example, preventing a progressive collapse can be improved upon in a standard construction building by reinforcing the main structure's beams and flooring through a variety of different processes. Materials and construction designs that are currently available can help to provide labs with the technical construction they need in order to meet safety standards and protect their employees.[8]. Additionally, some types of facilities can't even be insured without the proper blast-resistant design or hardening processes. Making sure you work with the right construction management company that not only has plenty of experience executing blast-resistant constructions but also knows how to renovate your existing building can help save you time, stress, and money. LCS Constructors specializes in lab designs and even blast-resistant buildings to help meet all of our clients' needs. [www.bulletproof material 2016]

Stealth Materials; The stealth casing deflects the radar waves emanating from ground to air, or air to ground, to locate friendly or enemy aircrafts in the territorial airspace during flight or on the ground under realistic battle conditions. The ground radar and aircraft radar detect aircrafts or other flying air vehicles such as UAVS (Unmanned Arial Vehicles like drones, fire scouts, predators and JDAM bombs (Joint Direct Attack Munitions). The same radar detects bad weather and prevents fighter pilots from stalling during flight as a result bad weather such as high turbulences red clouds denoted on the flying instruments panel of aircrafts during flights by red colour. This stall warning is taken care of by the stall warning management system which is interconnected with the radar electronically on the instrument panel on the CASA C-295 M designed and built by EADS (European aeronautics Defense and Space) in 2017.[9]. This stealth technology was adopted from military aircraft design especially fighters like F-117 night hawk, developed based on project risk management in the United states of America and bombers like B-2 Spirit also designed and built and being used by the United states Air force or the American Air force.

2. Background on Adaptronic transformer casing design .

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 failure 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.[10]. 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 (250433V).[11]. 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.[12]. 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.[13]. 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.[14]. Causes of transformer failures will be established since their failure is costly and affect the quality of service. All the findings from developed to developing countries differ 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. The contribution will help achieve the following; Collection of data on transformer failures in the country. Analyzing the failures into different categories. Suggesting best practice for operation of transformers. Design alternative solutions to mitigate the problem.

Since most transformer failures have been attributed to lighting strike, insulation failure due to old age, overloading, establishing the fact that Void formation and other unidentified factors which contribute to transformer failure will help to accomplish better control and protection devices or methods for transformers. Excellent supply of electricity to customers. Promote business growth without collapsing. Add new knowledge into the body of knowledge (void formation, environmental issues, culture issues etc.)

In 2014 alone 46 different distribution transformers were damaged which amounted to GH¢ 865,328.54 in western region alone. This implies that minimizing the failure will result in sound economy for the utility companies including the military and the country at large.Stabilized power supply which will mean that higher productivity in the country. Another vital area of this work is serving as appropriate database which monitors transformer failures in the country. Reference material for students and academics and ontribute to the general body of knowledge.

2.1.Research goals.

Adaptronic transformer casing refined design development based on box model from computer science has a low failure rate and 99.7% reliable. Adaptronic is a design process that makes a product environmentally friendly, uses knowledge of many areas, complex product and conserves material through recycle, by use of high heat resistant, bullet resistance, anticorrosion and ballistic incorporated material, which are deflectors and bomb resistant, using box model from computer science. The expected result will be better control and protection devices or methods for transformers. Excellent supply of electricity as transformers are now going to be protected from bomb detonation and other environmental influences. Prevent detonation of transformers by bombs, rocket and other explosives as a result of better protection of transformers. Stabilized power supply which will mean that higher productivity in the country as a result of good energy supply. It will serve as appropriate database which will monitors transformer failures in the country. It will serve as reference material for students and academics . It will also contribute to the general body of knowledge.

2.2. Research Questions.

The following are some research questions past and present researchers have asked.

i.Which type of distribution transformer frequently fails? The answers got in this are not specific, some researchers outline that, smaller sizes of transformers normally burn. Although this may be true in some ways, not all smaller transformers burn early. ii. What are some causes that lead to the transformer failure? A lot has been said on this because others mentioned insulation breakdown, transformer protections, lighting, oil contamination and many more. iii. Which rating (kVA) of distribution transformers frequently get damaged? iv. How does void formation contribute to transformer failures? In this particular question no researcher has asked before and is a new research area the researcher wants to investigate.

iv. How does the environment contribute to transformer failures? v.How does the maintenance culture of the utility companies contribute to transformer failures? iv. How does the primary and secondary substation contribute to transformer failure? V.In all many has been said on the above mentioned question and the responses have been in different approached. In this content, the researcher wants to investigate it deeper than before.

2.3. Methods and techniques.

The approach used in doing this work is broken down into desktop study through literature, observation and statistical analysis of results, generation of conceptual framework and validation of the framework, development of a working model and the validation of the same. The concepts developed were subject to experimental laboratory test, observation, probability and subsequent statistical analysis. The result of the experimental laboratory test, observation and subsequent statistical analysis constitute the basis for a framework to be developed. It was validated qualitatively through vetting and testing in subsequent stages of this work. Simulation was adopted in this research, life estimation based on physical age by probability of density function (PDF), Degree of Polymerization (DF), using Artificial Neural Network(ANN), to correlate insulation degradation with measurement taken during maintenance activities. Oil quality test using methods of moisture content, IFT (Interfacial tension), dielectric strength, acidity, tan δ, colour/appearance and sludge. Both quantitative and qualitative methods were used. Quantitative research methods were chosen, due to the engineering nature of the problem, to deduce the relationship between cause which has a measurable effect on the transformer failure in developing countries. The process involved experimentation through laboratory testing by DGA, electrical testing and visual inspection. With these methods, incipient faults and defects can be identified. [IEC 60599: (2011). "Mineral oil impregnated electrical equipment in service Guide to interpretation of dissolved and free gasses analysis, 2011]". [IEEE std. C57.104: IEEE Guide for the interpretation of gasses generated in oil-immersed transformers, (2014)].[Govender S,and Singh. A,(2015) "Toolkit for Transformer Life Management Report on: Condition Assessment Tools and Methodologies". Trans-Africa Projects, 2015]. 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).[16].

Electrical testing are tests performed to evaluate the components of the active part and identify defects. The electrical testing includes: Sweep Frequency Response Analysis (SFRA) is a method used to evaluate the mechanical integrity of core, windings and clamping structures of power transformers. DC resistance tests are performed in the factory, to determine the I2R losses and end temperature in a temperature rise test, and in the field for assessing possible damage.

[Y-Land, (2017) "Theory, Design, Maintenance and Life Management of Power Transformers, Power Series Volume 5"].

Dissipation factor and winding capacitance test is performed to provide information about movement and leakage losses within the distribution transformer. [Y-Land, (2017) "Theory, Design, Maintenance and Life Management of Distribution Transformers, Power Series Volume 5"].

Infrared scanning provides information about external connections, internal connections, bushing oil levels, cooling system blockages and hot spots.

Magnetizing current test provides information about the presence of core faults, inter-turn faults or unintentional loops in the earthing structure.

Visual inspections are performed routinely by the operator on site and as part of maintenance activities. These include checks and tests on the auxiliary components of the transformer.

All bushings routinely have tan δ and capacitance measurements taken. The values of these measurements should be within acceptable limits as stipulated by the OEM. The bushings are inspected for damage to the bushing bodies and insulator sheds. They are checked to be free of chips/tears, radial cracks, flashover burns, copper splash and copper wash. The cementing and fasteners of the bushings are checked to be secure. The bushings are checked for evidence of oil leaks and correct oil levels. Should any defect be identified, corrective action is taken.

All tap changers have the following checks performed: i. Oil tests: moisture and dielectric strength are checked to be within limits stipulated by the OEM. ii.Speed test: results are checked to be within the limits stipulated by the OEM.

Transition resistance: results are checked to be within the limits stipulated by the OEM. Iii. Number of operations: number of operations determine the maintenance intervals for the tap changer.

General inspections of bushing-metal interfaces, gaskets, weld seals, flanges, valve fittings, gauges and monitors are checked for oil leaks and moisture ingress. Should any of these be identified, corrective action is taken.The transformer tank, marshaling kiosk and tap changer mechanism box are checked for rust or corrosion. Cabinets are checked for evidence of condensation, moisture or insect/rodent ingress. Weld seals, flanges, valve fittings, gauges and monitors are checked for rust or corrosion. Seals, condensation heaters and locking mechanisms are checked for damage. The conservator/Oil Preservation System is inspected for rust, corrosion and paint damage on the tank body. The weld seals, flanges, valve fittings, gauges and monitors are checked for rust, corrosion and evidence of moisture ingress. The cooling system is inspected for rust, corrosion or oil leaks on the body of the radiators or pipework. Fan and pump enclosures are checked to ensure they are free of rust, corrosion and oil leaks and securely mounted in position with no signs of vibration. Fan and pump bearings are inspected to ensure they are in good condition and fan controls are operating as per design. The overall physical condition of the transformer is inspected to ensure that it is externally clean and corrosion free. The condition of all primary and secondary connections is checked. The condition of all monitoring, protection and control, pressure relief, gas accumulation and silica gel devices and auxiliary systems (including online DGA monitoring and drying systems) that are mounted on the power transformer, is checked. External evidence of overheating or internal overpressure are inspected. Maintenance and service records are checked. Information modeling, review of various research philosophies and methods that suggest that an experimental scientific engineering methodology is the only one that could adequately address both causes and effects of transformer failure in developing nation. Qualitative methods were also adopted due to the exploratory nature of the research and the subjective impact of the opinion of people during data gathering.[17]. Secondary data would be gathered through systematic review of literature through text books, institutional and statutory publication, periodicals, trade/academic journals, seminar and conference papers, and browsing engineering websites. Primary data was gathered by observation and through laboratory tests and physical, visual or general inspections carried out during maintenance, servicing or replacement of various transformers. This was aimed at identifying, investigating and analyzing the causes and effects of transformer failures in developing countries.

The target population will be the personnel in the military in electrical engineering department and those of ECG (Electricity company of Ghana), working on transformers because they have the actual figure.

The research will employ different methods; Combination of research methods, Personal observation, Lab testing will be a key factor and using simulation software available to simulate some findings or results. There will also be physical inspection of burnt transformers.

Secondary data of burnt transformers will be considered. Multiple methods will be employed. Bryman (2014) and Yin (2013), holds 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. Bryman

(2014) described five main research design options: experimental, cross sectional, longitudinal, case study and comparative research designs. An Experimental and comparative was adopted for the research. The choice of this strategy was because, the research comprises a wide variety of designs including model, laboratory tests, computer simulations, content analysis official documentation and diaries.

Table 1: Number of distribution transformers destroyed as a result of bush burning, lightning strike, gun shot from hunting.

REGION	**NUMBER OF DISTRIBUTION TRANSFORMERS DESTROYED IN GHANA**
Eastern Region	15
Western Region	13
Central Region	5
Southern Region	10
Northern Region	11
Total	54

SOURCE: Electricity Company of Ghana; number of distribution transformers destroyed from 2000 to 2019.

Eastern region had fifteen (15) distribution transformers destroyed as a result of fire detonation, activities from hunters like gun shot, explosives from mining industries and bush burning. That was from the year, 2000 to 2017.

Western region has recorded thirteen (13) distribution transformers destroyed as a result of lightning Strike, fire destination and gun shot from hunting activities, including bush burning from the year, 2001 to 2017.

Central region has on record five (5) distribution transformers destroyed as a result of lightning strike, fire destination and gun shot from hunting activities, including bush burning from the year, 2001 to 2017.

Southern region has on record ten (10) distribution transformers destroyed which are out of service as a result of hunting activities such as gun shot and bush burning resulting to fire detonation, from the year, 2001 to 2018.

Northern region had all distribution transformers destroyed from gunshot from hunting activities, fire destination from bush burning and lightning strikes. That was from the year, 2001 to 2018.

This has brought down to a total number of fifty – four (54) transformers destroyed from the year, 2000 to 2019, in the country as a result of bush burning, lightning strike and fire detonation. Those s data were obtained from the statistical analyses of the Electricity Company of Ghana (ECG) in the year, 2019.

Replacement estimate of distribution transformer

Table 2. 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.

Source of data

The source data to be used in this study was obtained from the Eskom databases used for storing reliability and condition data. Since this study involves the statistical analysis of distribution transformer failures, it was necessary to obtain historical data for both failed transformers, as well as \healthy" ones.

Distribution transformers do not have a high failure rate, with a high average failure rate being approximately only ten per annum. For this reason, failure data is scarce and in order to obtain a reasonable data sample, it was necessary to use data over an 18 year period (2000 - 2018).

Failure data was obtained from failure records and reports. Maintenance reports and factory records were also interrogated to obtain maximum information. Only failures satisfying the definition of failure level categorized as severe, intermediate and minor as per Section 3.3 were considered.

The definition of a healthy transformer follows the assumption: if the transformer has been in uninterrupted service since 2000, then in 2000 it was healthy. All data for \healthy" transformers was therefore collected for transformers within the existing population from 2000. This presents some limitations, since complete historical condition data is not available for all transformers

A total of 512 transformers were obtained through this process comprising 193 failed and 319 healthy transformers.

A number of challenges arise when handling data for analysis. These include: missing data, detection of outliers, transformation of data as well as sample bias. Methods of overcoming these challenges are outlined below.

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

A missing variable is predicted using a regression model which is determined from the other complete 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.

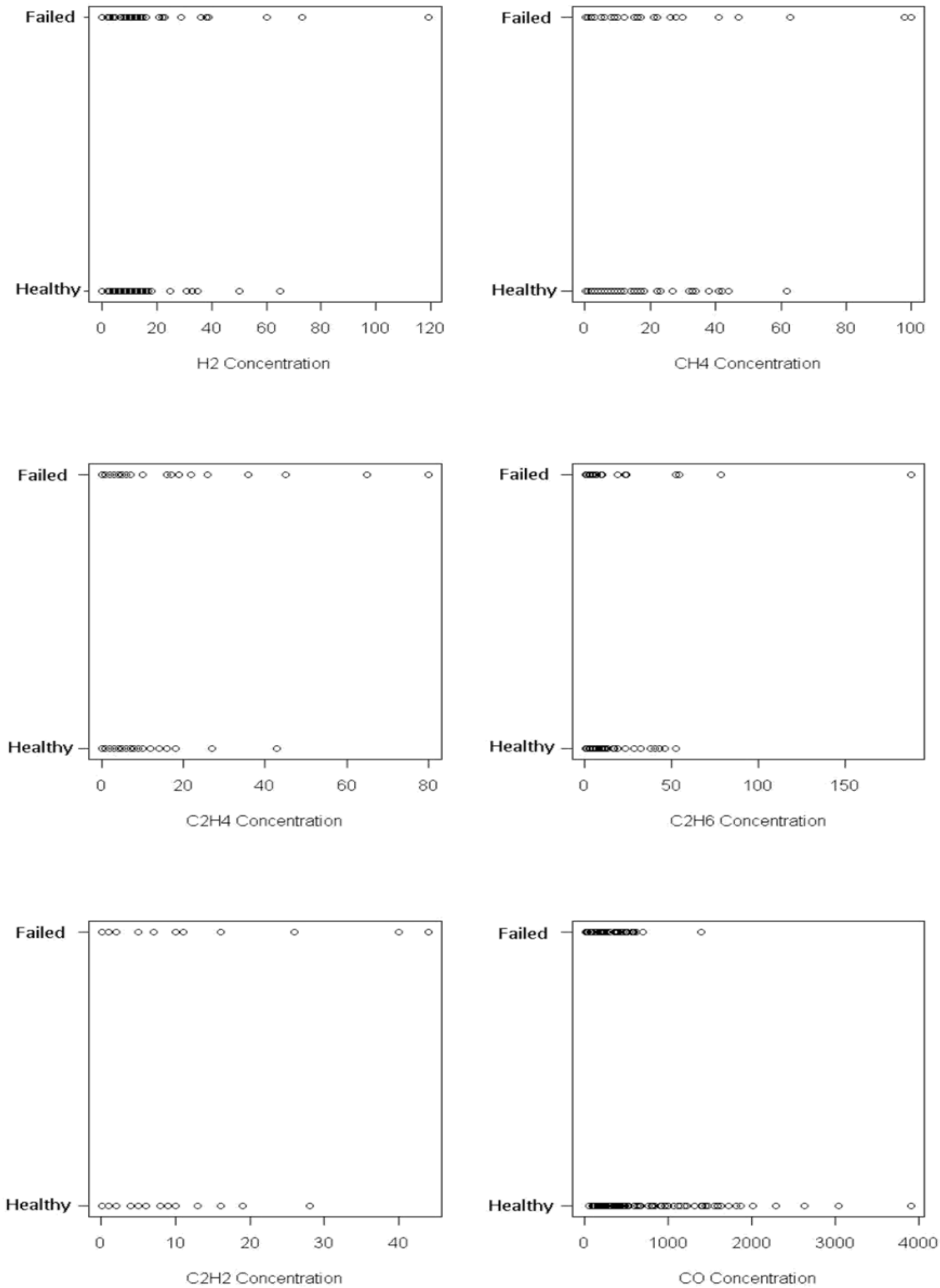


Figure .4.5: Plot of gas concentrations vs failures From the results shown in Figure:4.5, it can be seen that, with the exception of Carbon Monoxide, the expected trend in gas concentrations is evident, with higher gas concentrations present in the failed transformers and lower concentrations in the 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.

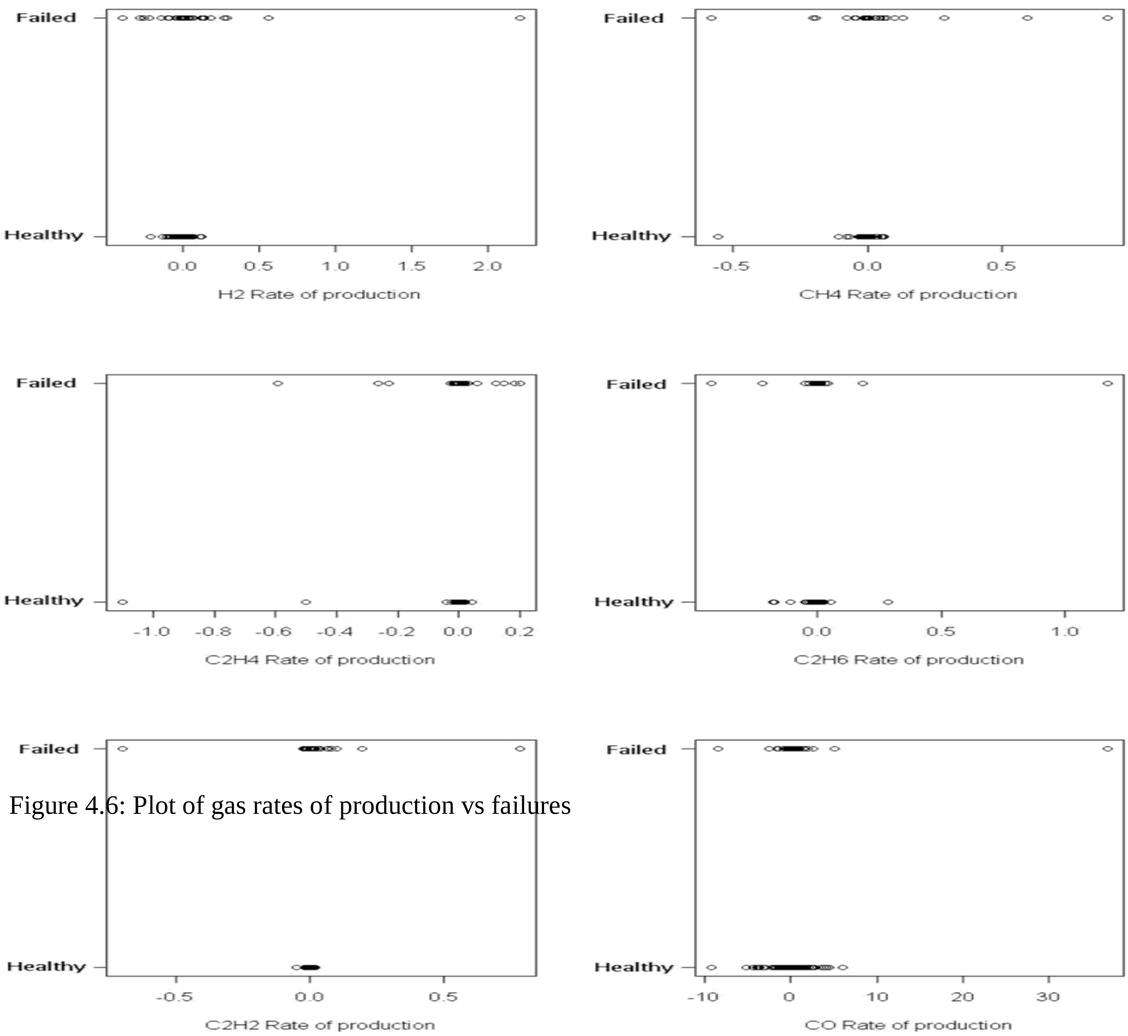


Figure 4.6: Plot of gas rates of production vs failures

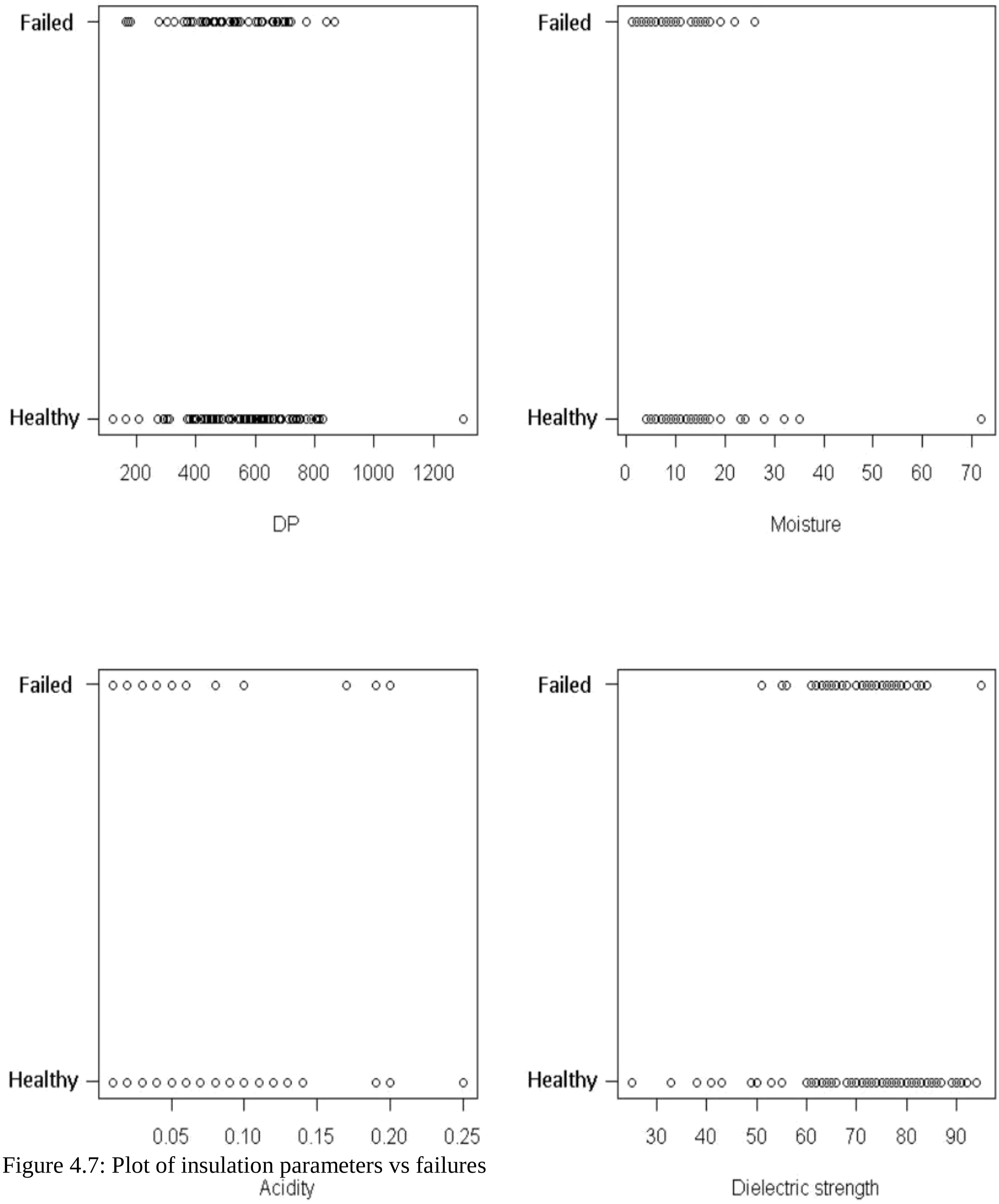


Figure 4.7: Plot of insulation parameters vs failures

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.

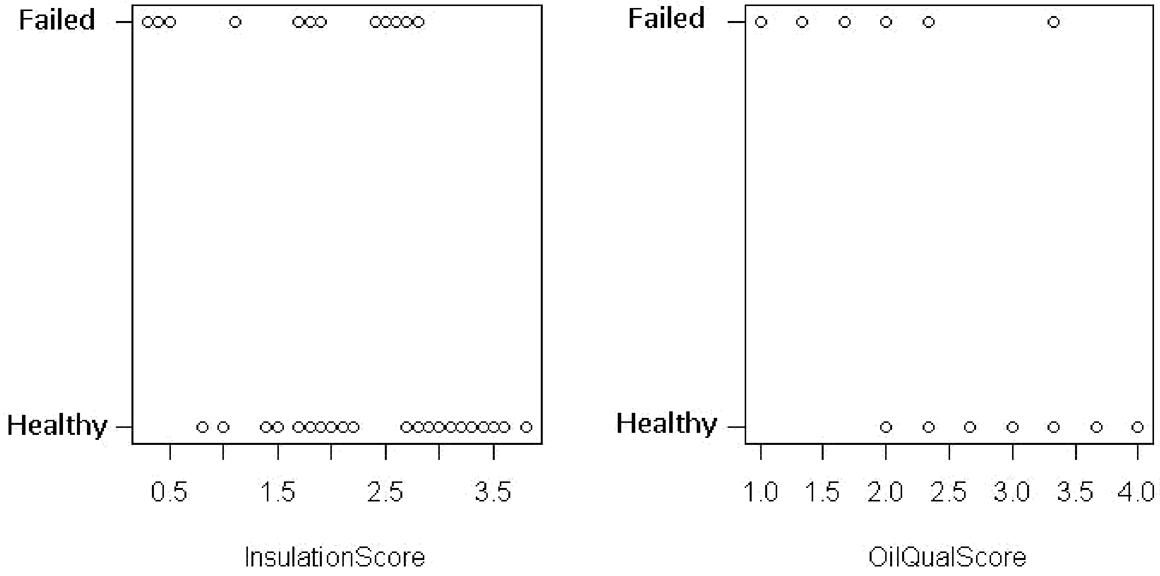


Figure 4.8: Plot of HI scores vs failures

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

$$x_{norm} = \frac{x \quad x_{min}}{x_{max} \quad x_{min}} \tag{5.1}$$

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.

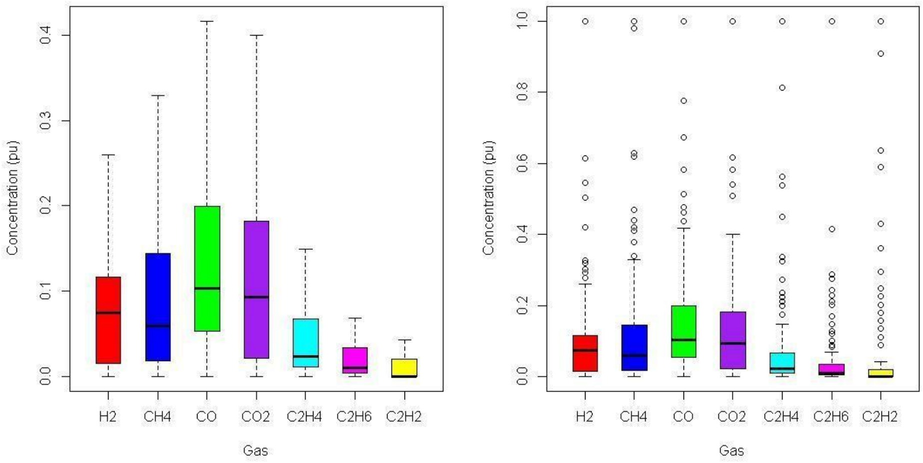
The median is represented by the horizontal line within the box. A line that is not perfectly centered is an indication of skewness of the data which is a measure of asymmetry about the sample mean. Skewness in the data is an indication that the data are not normally distributed. This is not of concern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

Gas concentrations and production rates



	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

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

1st Quad	2	2	215.5	244	1	1	0.1
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 .

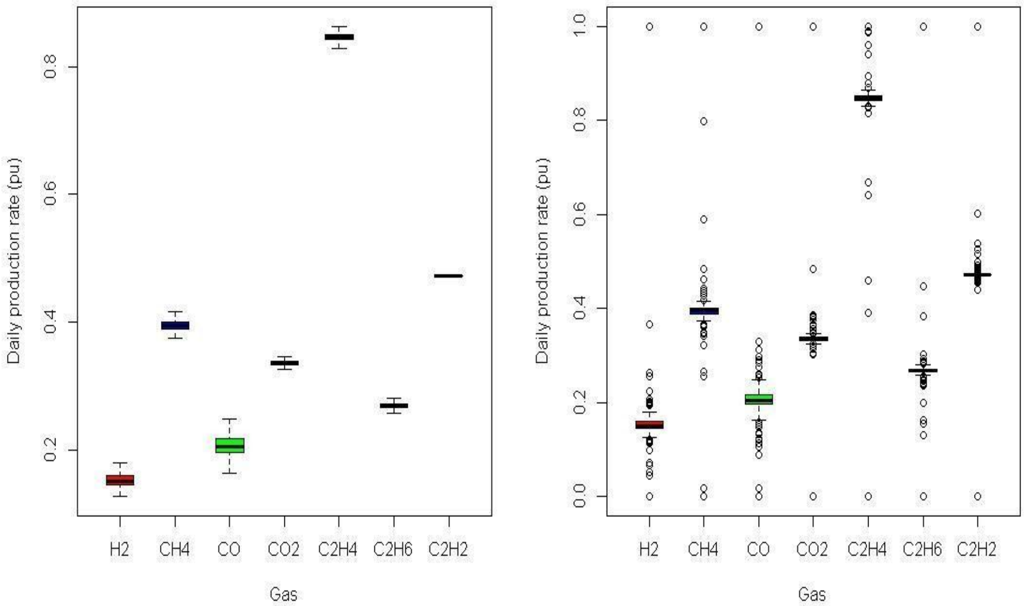


Figure 5: DGA rate of production(left), including potential outliers(right)

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

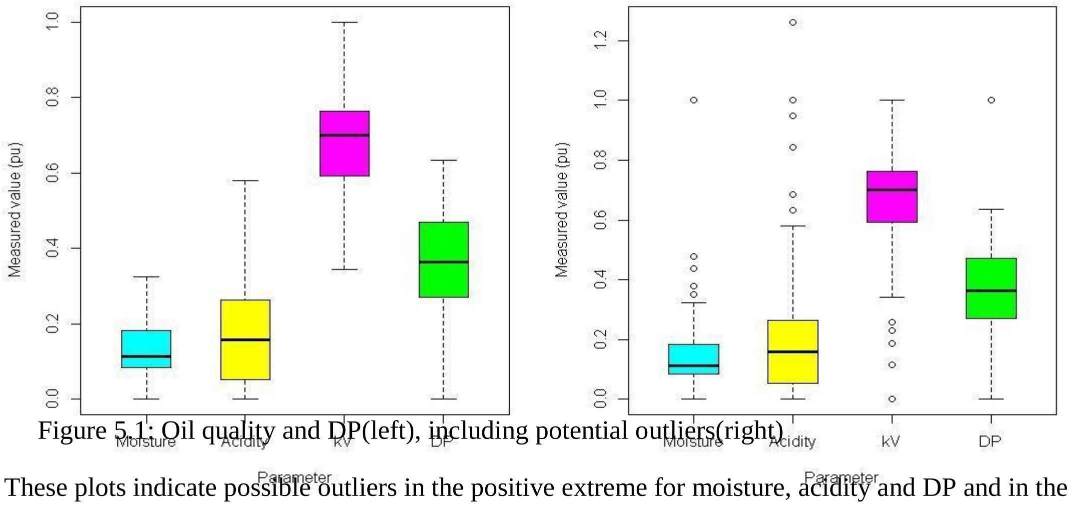


Figure 5.1: Oil quality and DP(left), including potential outliers(right)

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.

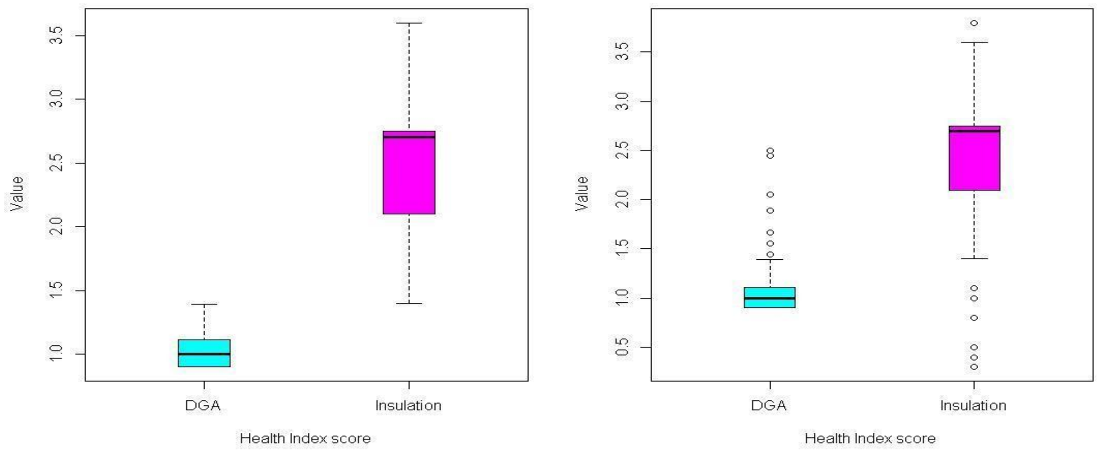


Figure 5.2: HI parameters(left), including potential outliers(right)

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.

Methods of statistical analysis

Statistical analysis will be performed on a portion of the condition data set only, in order to determine the weightings, and statistical significance of each condition parameter. Once the significance of each parameter has been determined from historical failure data, analysis will be carried out on the data set containing relevant data about both failed transformers, as well as healthy ones. A sensitivity analysis will be performed to determine whether or not the model is better at predicting failure than simply predicting by chance.

Care must be exercised in determining the correct statistical test to use for a specific analysis. Depending on the type and number of both the dependent and independent variables in question, different tests and methods are used.

In this case, there is one dependent variable: transformer failure, which is a categorical, dichotomous variable since the transformer can only be in one state at a time. It can either be failed or healthy, not both simultaneously. There are multiple dependent variables and these variables, as outlined and the condition data are all continuous variables.

According to the summary of statistical tests that can be performed for any given analysis given in [Leepe J.D, 2016], there are two types of statistical models that can be developed that t the data to be analyzed in this study. These are: multiple logistic regression and discriminant analysis.

Multiple logistic regression

Multiple logistic regression is much the same as regular logistic regression, but with more than one independent variable used to determine the dependent variable.

This method assumes that there are only two groups to which an individual sample can belong. Each sample can only belong to one group at a time .(In McLachlan G.J,2014). this study, there are only two groups, namely: failed transformers and healthy transformers. A specific transformer can be classified as failed or healthy, not both simultaneously.

The purpose of logistic regression is to determine the linear relationship be-tween y and x. Where y is the natural logarithm of the odds ratio of the probability of failure and x is the input vector x = (x_1; x_2; x_3; :::; x_n) which contains different features of the individual.(In McLachlan G.J, 2014). The output y is not itself a useful quantity, but is useful in determining p, which is the probability of failure.

The logistic regression model equation is shown in Equation 4.4:

$$\text{Ln}\left(\frac{p}{1-p}\right) = {}_0 + {}_1x_1 + {}_2x_2 + ::: + {}_nx_n \qquad (4.4)$$

and denotes the relationship between y and x. Where ($_1$; :::; $_n$) are weighted coefficients that determine the magnitude of the influence each variable in x has on the populations of each group (In McLachlan G.J,2013), and $_0$ is the intercept.

Two probabilities are calculated:

Prior probability: This is the probability that an individual is more likely to belong to one group than another (Chapman and Hall/CRC, 2016). This is useful in reducing the probability of misclassifying an individual. For example, at any point in time there is a much higher percentage of the transformer population that is healthy than has failed. Therefore, the probability of picking a healthy transformer at random from the current population is much greater than picking one that is about to fail.

Post probability: This is the probability that an individual belongs to a specific group [Afifi A, Clark V, and May S, 2016]. This is useful in determining the probability of a transformer failing, since for the transformer to belong to the group of failed transformers, the transformer would have had to have already failed, in which case prediction is not necessary as the outcome is already known.

The cost of misclassification can also be determined using this model. This is important since the model is not assumed to be ideal and misclassifications can be expected. The cost of misclassification will be different based on the individual that is misclassified. For example, the cost of classifying a high risk transformer as healthy will be far greater than the cost of classifying a healthy transformer as high risk. Care must also be taken to not overstate risk, since, while the cost of misclassifying a healthy transformer as high risk may be higher than the opposite, if too many healthy transformers are misclassified, the cost of any action taken to mitigate this risk could outweigh any potential bent.

The model should also be tested in order to determine whether or not the classification based on p is more accurate/reliable than a classification that was made based Statistical

Assumptions.

The following assumptions that are required for normal Linear Regression and General Linear models, based on ordinary least squares algorithms are not required for Logistic

Regression models:

Linearity: no linearity between IVs and DV is required

Normality: there is no requirement for IVs to be multivariate normal

Homoscedasticity: random variables need not have homogeneity of variance

Variable type: the logistic regression can handle nominal and ordinal data as independent variables, not only interval and ratio on chance alone.

For binary Logistic Regression, based on maximum likelihood estimates, the following assumptions are required (Aiken L and West S, 2007):

The dependent variable must be binary.

Independence of error terms is required. There should be independence between data points

No multicollinearity must be present, i.e. independent variables must be independent of each other.

Linearity between the independent variables and log odds is required, although it is not required between the independent variables and the dependent variable .

Statistical analysis

Decision theory

All improvements and developments are made as a result of intelligent decision making. Decision making is the process of arriving at a conclusion or resolution based on the consideration of various alternatives. Decision making can only be improved with better understanding of the problem.

Decision theory is defined as the study of principles and algorithms for making correct decisions. While simple decisions can be made without a theory, often complex decisions, ones involving high risk, levels of uncertainty and time dependency, will require mathematical or statistical models to produce optimized outcomes. This is a probabilistic approach which moves away from heuristic decision making. The deficiency with this method is that uncertainty related to known unknowns is taken into account and the extreme influences of the unknown unknowns are ignored.

Although many robust statistical methods are available for mathematical approximation of real situations, the results should always be analyzed critically. Limitations within these models are ever-present and should not be relied upon blindly or unquestioningly. This is unfortunately, common practice and is referred to as the Ludic Fallacy (Freeling .A,2016) and is advised against. Decisions should therefore be made with the best possible information at hand, while minimizing risk (risk cannot be completely eliminated).

Decision making

In Bayesian decision theory or Bayesianism (Freeling A.,2015), the aim is to reduce a Decision Maker's incoherence, and to make the Decision Maker approximate the behaviour of the hypothetical Rational Agent, so that after aiding he should satisfy Maximizing Expected Utility"

Bayesian decision making is summarized by the following four principles (Savage L.J,2014).

The Bayesian subject has a coherent set of probabilistic beliefs and these beliefs are in compliance with mathematical laws of probability

The Bayesian subject has a complete set of probabilistic beliefs, all alternatives have a degree of belief.

The Bayesian subject changes their beliefs in accordance with their conditional probabilities, on presentation of new evidence.

The Rational Agent chooses the option with the highest expected utility.

A statistical model is developed by gathering evidence and testing the model's effectiveness against a defined hypothesis. As new evidence is presented, belief in the hypothesis changes accordingly, as per point 3 above and the decision is potentially changed.

Decision classification

The purpose of the statistical model in this study is to assign various transformers to two different classes: healthy and failed, and to define their degree of belonging to each class. Each assignment is viewed as a decision and needs to be evaluated accordingly. This is done by means of a confusion matrix, as shown in Table:3.6. The risk associated with errors and misclassification can be more severe in some instances.

In this case:

a True Positive (TP) is a failed transformer correctly classified

a False Positive (FP) is a healthy transformer classified as a failed one a False Negative (FN) is a failed transformer classified as a healthy one a True Negative (TN) is a healthy transformer classified correctly.

Table: 3.6. Confusion matrix for decision making.

Actual	Predicted	
	True	False
True	TP	FP Type I error
False	FN Type II error	TN

The accuracy of the model is evaluated with Equations 4.1, 4.2 and 4.3 respectively [Fawcett2 T,2015]. Equation 4.1 gives an indication of the accuracy of the model in general.

$$\text{Accuracy} = \frac{TP + FP}{TP + FP + TN + FN} \qquad (4.1)$$

The Fall-out or False Positive Rate is given by Equation 4.2. The implication of a high Fall-out is a large number of healthy transformers being replaced unnecessarily. With the high capital cost of a power transformer, this can have a large negative financial impact on the power utility.

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \quad (4.2)$$

The Miss Rate or False Negative Rate is given by Equation 4.3. The implication of a high Miss Rate, is not being able to pre-emptively replace the failed transformers and avoid the potential damage to adjacent plant, safety and environmental risks should the transformer fail catastrophically. Loss of income and reputation should power outages result from the failure are also a consequence.

$$\text{False Negative Rate} = \frac{FP}{FP + FN} \quad (4.3)$$

Both Fall-out and Miss Rate have negative implications that cannot be ignored. Analysis of the individual accuracies is however not conclusive in the final decision making process. These error rates need to be evaluated in terms of risk and not solely probability.

Methods of statistical analysis

Statistical analysis will be performed on a portion of the condition data set only, in order to determine the weightings, and statistical significance of each condition parameter. Once the significance of each parameter has been determined from historical failure data, analysis will be carried out on the data set containing relevant data about both failed transformers, as well as healthy ones. A sensitivity analysis will be performed to determine whether or not the model is better at predicting failure than simply predicting by chance.

Care must be exercised in determining the correct statistical test to use for a specific analysis. Depending on the type and number of both the dependent and independent variables in question, different tests and methods are used.

In this case, there is one dependent variable: transformer failure, which is a categorical, dichotomous variable since the transformer can only be in one state at a time. It can either be failed or healthy, not both simultaneously. There are multiple dependent variables and these variables, as outlined and the condition data are all continuous variables.

According to the summary of statistical tests that can be performed for any given analysis given in [Leepe J.D, 2016], there are two types of statistical models that can be developed that t the data to be analyzed in this study. These are: multiple logistic regression and discriminant analysis.

Multiple logistic regression

Multiple logistic regression is much the same as regular logistic regression, but with more than one independent variable used to determine the dependent variable.

This method assumes that there are only two groups to which an individual sample can belong. Each sample can only belong to one group at a time .(In McLachlan G.J,2014). this study, there are only two groups, namely: failed transformers and healthy transformers. A specific transformer can be classified as failed or healthy, not both simultaneously.

The purpose of logistic regression is to determine the linear relationship be-tween y and x. Where y is the natural logarithm of the odds ratio of the probability of failure and x is the input vector $x = (x_1; x_2; x_3; :::; x_n)$ which contains different features of the individual.(In McLachlan G.J, 2014). The output y is not itself a useful quantity, but is useful in determining p, which is the probability of failure.

The logistic regression model equation is shown in Equation 4.4:

$$Ln \left(\frac{p}{1 \quad p} \right) = {}_0 + {}_1x_1 + {}_2x_2 + ::: + {}_nx_n \quad (4.4)$$

and denotes the relationship between y and x. Where ($_1$; :::; $_n$) are weighted coefficients that determine the magnitude of the influence each variable in x has on the populations of each group (In McLachlan G.J,2013), and $_0$ is the intercept.

Two probabilities are calculated:

Prior probability: This is the probability that an individual is more likely to belong to one group than another (Chapman and Hall/CRC, 2016). This is useful in reducing the probability of misclassifying an individual. For example, at any point in time there is a much higher percentage of the transformer population that is healthy than has failed. Therefore, the probability of picking a healthy transformer at random from the current population is much greater than picking one that is about to fail.

Post probability: This is the probability that an individual belongs to a specific group [Afifi A, Clark V, and May S, 2016]. This is useful in determining the probability of a transformer failing, since for the transformer to belong to the group of failed transformers, the transformer would have had to have already failed, in which case prediction is not necessary as the outcome is already known.

The cost of misclassification can also be determined using this model. This is important since the model is not assumed to be ideal and misclassifications can be expected. The cost of misclassification will be different based on the individual that is misclassified. For example, the cost of classifying a high risk transformer as healthy will be far greater than the cost of classifying a healthy transformer as high risk. Care must also be taken to not overstate risk, since, while the cost of misclassifying a healthy transformer as high risk may be higher than the opposite, if too many healthy transformers are misclassified, the cost of any action taken to mitigate this risk could outweigh any potential bent.

The model should also be tested in order to determine whether or not the classification based on p is more accurate/reliable than a classification that was made based Statistical

Assumptions.

The following assumptions that are required for normal Linear Regression and General Linear models, based on ordinary least squares algorithms are not required for Logistic

Regression models:

Linearity: no linearity between IVs and DV is required

Normality: there is no requirement for IVs to be multivariate normal

Homoscedasticity: random variables need not have homogeneity of variance

Variable type: the logistic regression can handle nominal and ordinal data as independent variables, not only interval and ratio on chance alone.

For binary Logistic Regression, based on maximum likelihood estimates, the following assumptions are required (Aiken L and West S, 2007):

The dependent variable must be binary.

Independence of error terms is required. There should be independence between data points

No multicollinearity must be present, i.e. independent variables must be independent of each other.

Linearity between the independent variables and log odds is required, although it is not required between the independent variables and the dependent variable.

Model evaluation/Goodness of t

Various methods exist for evaluating the goodness of t of a statistical model. The tests used for ordinary linear regression are different to those used for logistic regression. Interpretation of the various test statistics is different for both models. Goodness of t of the models developed in this study is evaluated by examining both likelihood ratio tests and pseudo-R^2.

Data separation

The optimal evaluation of goodness of t of a statistical model requires three data sets, namely: training data, validation data and test data (Hosmer D, and Lemeshow S, 2016).The training and validation data sets comprise the data set used in developing the final statistical model. With these data sets, both the model inputs and output are known. The test data set is comprised of new data where only the inputs are known and the outputs are determined using the statistical model. These results are then analyzed.

The training data is used with statistical methods to develop the prediction model. The model is then tested using the validation test set, where the IV data are input into the prediction model and the output of the model is compared with the known outputs in order to determine the degree of accuracy of the model. Once the model has been found sufficiently accurate, the test data is input into the model and the results are then analyzed. This process is outlined in Figure: 6.2,below.

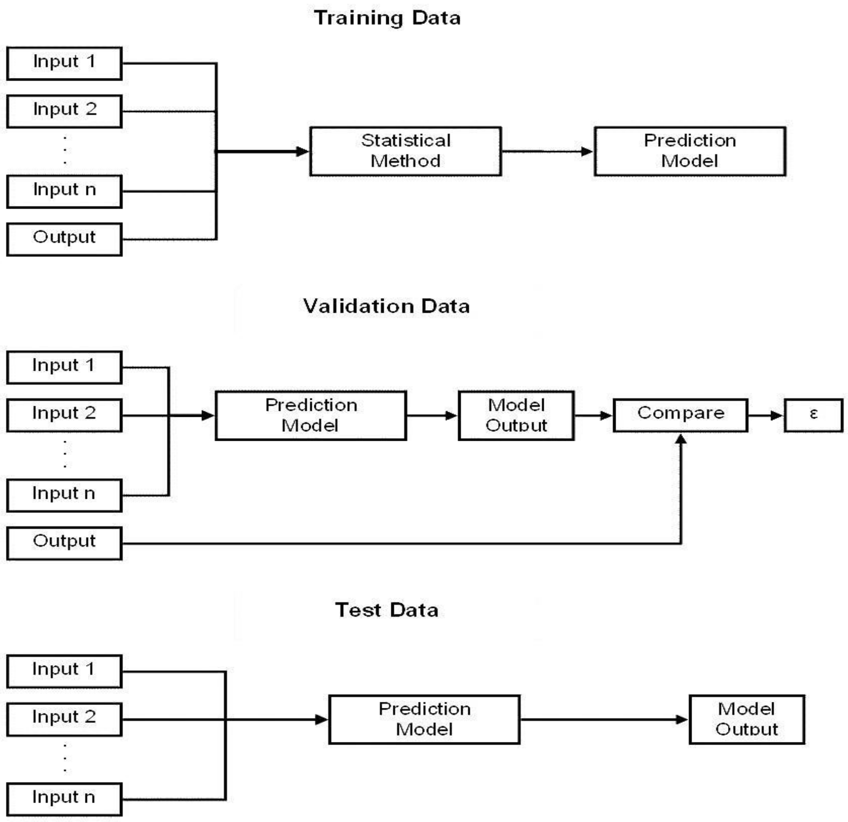


Figure: 6.2. Relationship between different datasets and prediction model

This method is however used in the ideal case where a large data set is available. In cases such as this study, the limited size of the data set this methodology and other model goodness-of- t methods will be utilized.

Deviance and likelihood ratio tests

The likelihood ratio test is shown in Equation 4.5 (Aiken L and West S, 2015). The deviance of the model is the difference between the fitted values and the expected values. For this reason, the smaller the deviance of the fitted model compared to the deviance of the null model, the better the model is at predicting the outcome.

$$D_{fitted} \quad D_{null} = 2 \ln \frac{\text{likelihood of fitted model}}{\text{likelihood of null model}} \qquad (4.5)$$

Where :D_{fitted} is the deviance of the fitted model, D_{null} is the deviance of the null model.

Pseudo-R^2

In a linear model, R^2 is used to evaluate goodness of t of a model by the proportion of variance in the dependent variable explained by the independent variables. Since Logistic Regression is heteroscedastic, the proportionate reduction in error is not constant across the range of predicted outcomes. The R^2 statistic is therefore not interpreted in the same was as for linear regression where very low values are expected(Hu B, Shao J, and Palta M, 2016).

The pseudo-R^2 statistic is the proportion of variance of the latent variable, inferred from other variables but not directly observed which is explained by the covariate, the variable affecting the relationship between the IVs and DV, and is shown in Equation 4.6.

$$R_L^2 = \frac{D_{null} \quad D_{fitted}}{D_{null}} \qquad (4.6)$$

where D_{fitted} and D_{null} are as defined for Equation 4.5.

Hosmer-Lemeshow test

The Hosmer-Lemeshow test (Hosmer D and Lemeshow S, 2015), is a^2 statistic that is calculated on data that are grouped into groups with approximately the same number of observations per group (usually 10 groups). This test statistic is given by Equation 4.7.

$$H = \sum_{g=1}^{G} \frac{(O_g \quad E_g)^2}{N_g \; _g (1 \quad _g)} \qquad (4.7)$$

where:

O_g are the observed events

E_g are the expected outcomes

N_g are the observations

$_g$ is the expected risk

G is the number of groups

The disadvantages of this test are that is has a large dependence on the number of observations grouped, as well as the number of groups. This test also has reduced accuracy in predicting certain types of lack of model t. For this reason, this test is not implemented in this study.

Conclusion

The need for statistical analysis in the decision making process required to implement the CCRA model, that was introduced in the previous chapter, was discussed. An outline of the statistical method of Multiple Logistic Regression which was used for this study was outlined, along with the statistical assumptions made. All models require testing and various goodness-of- t tests are available for evaluating logistic regressions. These methods are outlined in this chapter. The use of the deviance and likelihood ratio tests, as well as pseudo-R^2 tests are used in this study.

Statistical analysis is often an iterative process that is highly dependent on the integrity of the data used in the analysis. Methods of data processing are therefore introduced and applied to the data used in this study.

REFERENCES

1. *Kelly David Burke, Alicia Acuna Published June 02, 2017*
2. Y-Land, (2017) "Theory, Design, Maintenance and Life Management of Power Transformers, Power Series Volume 5"].
3. *www.bulletproof material 2017*
4. *IEEE std. C57.104: IEEE Guide for the interpretation of gasses generated in oil-immersed transformers, (2014)*
5. *Govender S,and Singh. A,(2015) "Toolkit for Transformer Life Management Report on: Condition Assessment Tools and Methodologies". Trans-Africa Projects, 2015.*
6. *Dong in Allan 2012*
7. *LOKH 2012*
8. *Alicia Acuna joined Fox News Channel (FNC) in 1997 and currently serves as a general assignment reporter based in the network's Denver bureau.*