Midlands State University
Established 2000
Our Hands, Our Minds, Our Destiny

FACULTY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF SURVEYING AND GEOMATICS

TOPIC

Spatial Electric Load Forecasting; A case of Zimbabwe Electricity Transmission and Distribution Company (ZETDC), Mutare.

SUBMITTED BY: HELEN NGONIDZASHE SERERE

REGISTRATION NUMBER: R131162P

SUPERVISOR: DR. P.T. MAKANGA

This dissertation was submitted in partial fulfilment of the requirements of the Bachelor of Science Honours Degree in Surveying and Geomatics at the Midlands State University.

APPROVAL FORM

The undersigned certify that they have supervised the student Helen Ngonidzashe Serere's dissertation entitled "Spatial Electric Load Forecasting; A case of Zimbabwe Electricity Transmission and Distribution Company (ZETDC) Mutare", submitted in partial fulfilment of the requirements of the Bachelor of Science Honours degree in Surveying and Geomatics at Midlands State University.

Dr P.T. Makanga (Supervisor) Date signed

-------------------------------------- ---------------------------------------

Mr D Njike (Chairperson) Date signed

------------------------------------- ---------------------------------------

RELEASE FORM

NAME OF STUDENT:

Helen Ngonidzashe Serere

REGISTRATION NUMBER:

R131162P

DISSERTATION TITLE:

"Spatial Electric Load Forecasting; A case of Zimbabwe Electricity Transmission and Distribution Company (ZETDC) Mutare".

DEGREE TITLE:

Bachelor of Science Honours Degree in Surveying and Geomatics.

YEAR THIS DEGREE GRANTED: 2017

Permission is hereby granted to the Midlands State University Library to produce single copies of this dissertation and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights. Neither the dissertation nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

SIGNED: …………………………………………..

PERMAMENT ADDRESS:

P.O Box 1724
TRANQUILLAR ROAD
GREENVALE
GWERU

DATE: NOVEMBER 2017

DEDICATION

I dedicate this thesis to God Almighty my creator, my strong pillar, my source of inspiration, wisdom, knowledge and understanding. He has been the source of my strength throughout and on His wings only have I soared. Secondly I dedicate this thesis to my parents; Mr and Mrs Serere who have encouraged me all the way and whose encouragement have made sure that I give it all it takes to finish that which I start. Last but not least I dedicate this thesis to my siblings; Hillary, Hilton, and Hildah Serere who have supported me both socially and financially and in every way possible.

Thank you all. My love for you all can never be quantified. God bless you.

ACKNOWLEDGEMENTS

At the very beginning I would like to express my deepest gratitude to the Almighty God for giving me the strength and the composure to complete my honours degree in Surveying and geomatics and prepare this complete this thesis within the scheduled time.

During the course of my studies, I have received generous help from many quarters, which I would like to put on record here with deep gratitude and great pleasure.

First and foremost, I am grateful to my supervisors, Doctor P. T. Makanga and Engineer P. Mundenda. These people allowed me to encroach upon their precious time freely right from the very beginning till the completion of my thesis. Their guidance, encouragement and suggestions provided me with necessary insight into some research problems and paved the way for the meaningful ending of the honours degree.

I give also much credit to my university lecturers which I will however fail to mention by name because of their numbers, for their academic mentoring. I have no hesitation to say that, without their constant support and valuable advices from time-to-time, I would have probably failed to complete and achieve all that I did during this period. Last but not least my thanks and appreciations also goes to my colleagues and everyone else who gave their suggestions and offered their help in so many different ways.

TABLE OF CONTENTS

APPROVAL FORM I

RELEASE FORM II

DEDICATION III

ACKNOWLEDGEMENTS IV

TABLE OF CONTENTS V

LIST OF TABLES VIII

LIST OF ACRONYMS IX

CHAPTER ONE 1

1.1 INTRODUCTION 1
1.2 BACKGROUND 1
1.3 PROBLEM DEFINITION 2
1.4 AIM 2
1.4.1 Specific objectives 2
1.4.2 Research questions 3
1.5 JUSTIFICATION OF STUDY 3
1.6 STUDY AREA 3
1.7 PROJECT OUTLINE 4

CHAPTER TWO: LITERATURE REVIEW 6

2.1 INTRODUCTION 6
2.2 WHY LOAD FORECASTING 6
2.3 TYPES OF LOAD FORECASTING 8
2.4 LOAD FORECASTING METHODS AND FACTORS 9
2.4.1 Short term 9
2.4.2 Medium and long term 10
2.5 THE ENHANCEMENT OF SPATIAL DATA IN LOAD FORECASTING 14
2.6 SPATIAL LOAD FORECASTING CASE STUDIES 15
2.7 CONCLUSION 16

CHAPTER THREE: METHODS 17

3.1 INTRODUCTION 17
3.2 RESEARCH DESIGN 17

3.3 Study population 17
3.4 Data collection technique 17
3.5 Data description 18
3.5.1 Historical electric usage data 18
3.5.2 City of Mutare master plan 18
3.5.3 Mutare cadastral map 18
3.6 Data processing 18
3.7 Methods flow chart 20
3.8 Data modelling: Statistical analysis tools 21
3.8.1 Exploratory regression 21
3.8.2 Ordinary least squares regression 21
3.9 Conclusion 21

CHAPTER FOUR: RESULTS 22

4.1 Introduction 22
4.2 Exploratory regression 22
4.3 Ordinary least squares regression 23
4.4 Conclusion 29

CHAPTER FIVE: ANALYSIS OF RESULTS 30

CHAPTER SIX: LIMITATIONS, RECOMMENDATIONS AND CONCLUSIONS 31

6.1 Research limitations 31
6.2 Recommendations 31
6.3 Conclusion 31

REERENCE LIST 33

ANNEXURES 35

Annexure 1: Exploratory regression 35
Annexure 2: Ordinary least squares report 39
Variable Distributions and Relationships 40
Histogram of Standardized Residuals 41
Residual vs. Predicted Plot 42
Ordinary Least Squares Parameters 43

LIST OF FIGURES

Figure 1 Showing increase in population growth 1

Figure 2: Zimbabwe Map 3

Figure 3: Methods flow chart.. 20

Figure 4: Linear regression equation 21

Figure 5: Scatterplot of the standard residuals against the predicted variables 24

Figure 6: Moran I spatial autocorrelation report.. 25

Figure 7: Expected domestic load additions per suburb 26

Figure 8: Annual expected load additions 27

Figure 9: Mutare proposed map layout.. 28

LIST OF TABLES

Table 1: Short term load forecasting methods and techniques 10

Table 2: Medium to long term load forecasting methods and factors. 13

Table 3: Summary of exploratory regression results 22

Table 4: Exploratory regression results of the considered model 23

Table 5: Summary of OLS model variables. 23

LIST OF ACRONYMS

AC:	Air Conditioning systems.
EV:	Evaporative Colling systems.
GIS:	Geographic Information Systems
OLS	Ordinary Least Squares
S.E.L.F	Spatial Electric Load Forecasting
T&D:	Transmission and Distribution
ZETDC:	Zimbabwe Electricity Transmission and Distribution Company

CHAPTER ONE

1.1 Introduction

Electricity is a critical source of energy worldwide. To the day to day consumer; it is used in hospitals, households, industries, schools, commercial areas etc. This makes it a critical service which must constantly be managed well in advance to avoid straining resources in time. Unexpected electrical interruptions can cause serious repercussions on electrical utilities and even human life (Khadem et al., 2010). There is therefore a need for electric distribution companies to continuously reevaluate infrastructure in line with spatial developments so as to make certain of the availability and reliability of electricity to all users. This thesis hence intends to incorporate GIS into electrical transmission and distribution utilities in an effort to produce more accurate electrical forecasts which intern effects the availability of a reliable electrical network system.

1.2 Background

The period 1960 to 2015, Zimbabwe has seen 18% increase in population growth (Zimbabwe central Statistics, 2015). The World Bank recorded a population increase of 13% for city centres whilst recording only 5% population increase in rural areas owing mainly to the migration of citizens to cities. This phenomenal growth in urban population has seen the expansion of towns and cities all over Zimbabwe in a process of continuous urbanisation (Mugumbate et al., 2013).

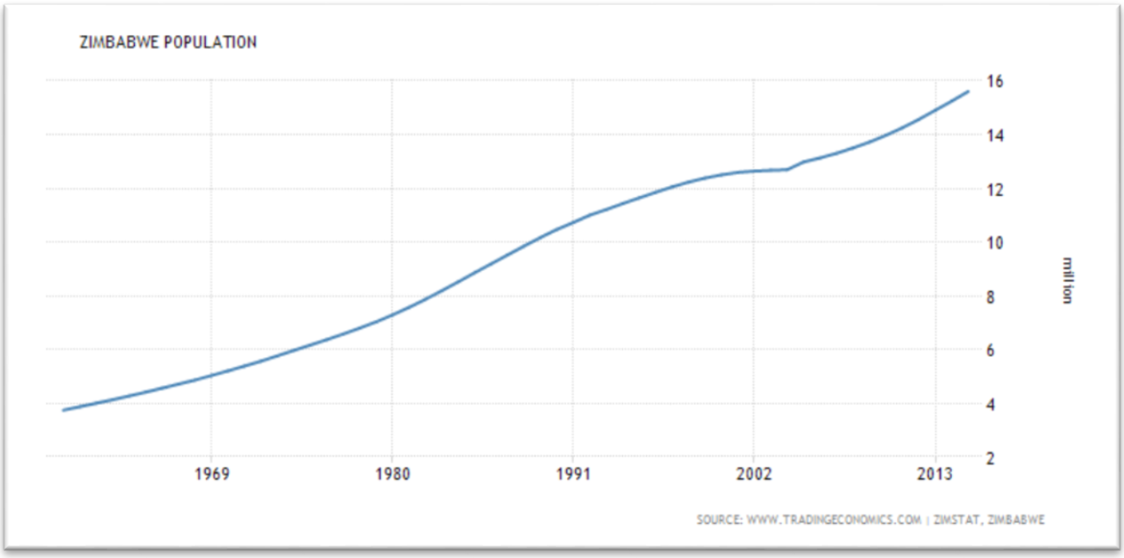


Figure 1 Showing increase in population growth

Source: www.tradingeconomics.com|ZIMSTAT.ZIMBABWE

With the constant rise in population growth, Zimbabwe has perceived a significantly high number of uprising settlements and town expansion. Mugumbate et al., (2013) states that; new residential, commercial and industrial stands are being proposed and developed every now and again. Owing to this current land development, there has been an increasingly high statistical demand of electricity on distribution utilities.

1.3 Problem definition

Electric distribution utilities such as the Zimbabwe Electricity Transmission and Distribution Company (ZETDC) are currently facing challenges when it comes to maintaining a reliable electric system mostly due to its inability to accurately forecast future electrical demand. ZETDC does not at the moment contain a detailed customer database with such information as average income and description of electrical appliance per household, which in most cases is perceived one of the primary datasets for spatial load forecast modelling. The development, population and up to date maintenance of such database is said to be tedious and expensive to the organization. Due to this reason, ZETDC has reverted to using various assumptions for future load predictions which in most cases tend to be inaccurate. Kaseke (2013) mentioned how these inaccurate load forecasts have resulted in Zimbabweans experiencing unexpected power outages affecting industries, farms, mines and households. Owing to that, there has been serious repercussions on the Zimbabwean government's efforts to successfully turnaround the economy and achieve sustainable economic and social growth (Zinyama and Tinarwo, 2015, p. 54).

1.4 Aim

- The main objective of this study is to spatially forecast future electric demand for the city of Mutare using readily available datasets.

1.4.1 Specific objectives

- To map out the Mutare residential areas and assign average load consumption per domestic household.
- Make use of spatial analyst tools to identify variables that are of significance to electric consumption and how they affect load.
- Model the relationship of load consumption to explanatory variables.
- Forecast future domestic electrical demand.

1.4.2 Research questions

- What are some of the key factors that affect electrical load forecasting?
- Which methods have been implemented in the event of unavailable datasets?
- How can we apply these methods to predict future load demand for uprising settlements?

1.5 Justification of study

The application of spatial electric load forecasting model is in an effort to reduce or eliminate unexpected electrical interruptions, thus providing a more reliable electrical flow (Carreno et al., 2010). This will help distribution utilities to predict not only the magnitude of new load expected but also where and when the new consumers are bound to occur (Willis, 2002). Electrical distribution utilities can hence be more planned for rising electrical demand and make prioritized decisions as to where and when to make infrastructure improvements.

1.6 Study area

Figure 2: Zimbabwe Map

*Image s*ource: Google Zimbabwe City maps

Being the fourth largest city in Zimbabwe, Mutare makes up the provincial capital of the Manicaland Province. It is located on the eastern boundary of Zimbabwe, about 300 km from the Port of Beira in Mozambique and approximately 263 km from the Zimbabwean capital city of Harare. Like most cities, Mutare classifies its residential suburbs according to population density. The low density areas such as Murambi and the Fairbridge Park are located on the north end of the city along its foothills, while Palmerstone, Darlington, Greenside, Greenside extension and Bordervalle are east of the city centre, near the Mozambique border post. The west part comprise medium-density suburbs such as Yeovil, Westlea and Florida, as well as the high density suburb of Chikanga, with phases 1, 2 and 3. Further west of Chikanga lies Garikai and Hobhouse. South of the railway track lies the high-density suburb of Sakubva, which has the highest population density and considered the poorest of all Mutare’s suburbs. A few miles to the south, hidden behind a series of hills, is the high-density suburb of Dangamvura, the low-density areas of Weirmouth and Fern valley with lots exceeding an acre, and market gardening being the most practised economic activity. Further to the south along the Masvingo road is the high-density town of Zimunya. All the prior mentioned constitute most of Mutare suburbs as of now but with further developments being done, more residential, commercial and industrial stands are being expected.

1.7 Project Outline

Chapter one is a basic outline and discussion of the background of study and the problem definition which clearly outlines the problems being faced by electrical power utilities in as far as meeting electrical demand is concerned. Apart from defining the problem, the aim and corresponding objectives of the thesis is stated as a proposed solution to the highlighted challenges. A justification of research is also included together with a review of the area under study.

Chapter two deals with an in-depth literature review of the research. It gives a careful and structured exploration of the works done by other authors in similar fields. As such the chapter is a combination of definition of key concepts as proposed by different authors. It highlights the importance of load forecasts to distribution utilities, explores the different load forecasting types as well as the essence of GIS or spatial data in load forecasting. Different methods and factors considered by various scholars are stated and an analysis of the advantages and disadvantages of implementing any of the methods is given. A detailed review is also made on how other countries developed their load forecasting models from the data sets used, methods of creation as well as the limitations and challenges experienced.

Chapter three expounds on the methods used to carry out the research, the data gathering techniques, statistical formulas used as well as methods of modelling implemented. This chapter identifies; explains and justifies methods used by the author in coming up with a spatial electric load forecast in order to produce conclusive results in research.

Chapter four gives an overview of the major findings of this research. It clearly outlines the results that were obtained through the application of the various statistical regression tools stated in chapter three.

Chapter five gives a detailed analysis of the results produced in chapter four. It makes a quantitative and qualitative results analysis. The analysis of results forms a back born to further investigations on similar research studies.

Chapter six states the limitations encountered in the research. It also take account of recommendations passed by the researcher and gives a general conclusion reached following findings from prior chapters.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Electricity is one of the largest sources of renewable energy being used worldwide (Ernest and O'Connor, 2014; Khadem et al., 2010). It is defined by Hassan and Akhtar, (2012), as a general flow of electric current from its generating stations through transmission lines and finally to electrical appliances of various consumer classes. According to R.L.Sheilds associates, (2010) anyone who uses electricity is referred to as a consumer and the maximum amount of electrical energy consumed at a given time; demand. Electrical demand is not stagnant but varies according to a number of factors such as: time of the day, day of the week, season of the year, weather extremes, major public events etc. (Balantrapu, 2013; Feinberg and Genethliou, 2003a). Eventually, this means that balancing the relationship between demand and supply of electricity is not a once off exercise, but should be done proactively owing to the various factors considered.

Although electrical demand fluctuates with time, it alternates between its lowest and its peak electrical values. Peak electrical values refers to the highest electrical demand that has occurred over a specified amount of time and is usually altogether constant. (Balantrapu, 2013; Feinberg and Genethliou, 2003a). For this reason, electrical distribution utilities use peak load values to make long term investments on the infrastructure size needed for a particular area. Furthermore; because electricity is expensive and difficult to store (Gates, 2016), electrical utilities need to predict the demand needed on short, medium and long term basis (Khattak, 2015; Kocavelioğlu, 2012). This is referred to as load forecasting and is defined by Swaroop and Hussein (2012), as the processes of predicting future electric demand basing on available information.

2.2 Why load forecasting

One of the biggest challenges faced by electrical distribution utilities is the ever-increasing growth in demand for electric power (Amlabu et al., 2013; Willis and Northcote-Green, 1983). To be able to maintain a reliable network system, utilities thus have to be in a position to predict to the nearest possible accuracy this rising demand. According to Willis and Northcote-Green, (1983) improper forecasts lead to the saturation of electrical facilities and loss of power supply along with the consequent economic loss and social distress. In addition to this, Alfares and

Nazeeruddin (2002) together with Ranjana (2016) mention the financial consequences for forecast errors to be immensely significant such that even a small fraction reduction in the forecast error results in major financial benefits for the utility. On a similar note, Ranjana (2016) states that forecasting errors effects unnecessary electricity purchasing cost, in the event of an overestimate, or breaking contract penalty cost, in the event of demand underestimates, in order to keep the electricity supply and consumption balance. It is for this reason that Carreno et al., (2010) suggests the need for electric utilities to consider load forecasting as a business priority.

Overestimating electric power for instance causes the start-up of too many generating units and lead to unnecessary increase in the reserve and operating costs (Ranjana, 2016). Underestimating demand on the other hand results in difficulties to manage overload conditions, which then leads to the collapse of the power network system (Baliyan et al., 2015; Ranjana, 2016). Baliyan et al (2015) goes on to mention the negative consequences on both demand response and power installation as part of the challenges brought about by underestimating future electric demand. Correct forecasts can henceforth eliminate the financial, economic and social burdens brought about by inaccurate forecasts. Various scholars have thus come up with a number of advantages of increased forecasting accuracy. Some of the most prominent mentioned advantages being;

- Its ability to make important decisions such as purchasing and generating electric power, load switching, voltage control, network reconfiguration as well as infrastructure development (Badar, 2011; Baliyan et al., 2015; Feinberg and Genethliou, 2003; Ranjana, 2014; Singh et al., 2013; Swaroop and Hussein, 2012).
- Its ability to control operations and decisions like economic dispatch, unit commitment, fuel allocation, generator maintenance as well as online and offline network analysis.(Alfares and Nazeeruddin, 2002; Baliyan et al., 2015; Ranjana, 2016; Singh et al., 2013)
- It can hold as a great saving potential when accurately done (Alfares and Nazeeruddin, 2002; Baliyan et al., 2015; Singh et al., 2013).
- Load forecasting ensures a reliable and more economically efficient electric power system (Ranjana, 2016).
- It helps ISO, energy suppliers and financial institutions in making sure that they will be able to cope with future committed energy supply (Baliyan et al., 2015; Feinberg and Genethliou, 2003; Pijoan et al., 2013; Ranjana, 2014).

- Helps check the risky operation, fluctuating demand, demand for spinning reserve and vulnerability to failures (Baliyan et al., 2015).
- Load forecasting can also be used by ISO, energy suppliers and financial institutions for contract evaluations (Badar, 2011).

2.3 Types of load forecasting

Different types of load forecasting guarantees level of accuracy to a certain degree. Electric distribution utilities hence need to know the reason for their forecasts as well as the desired forecasting accuracy so as to decide on the type of forecasting to implement. Many scholars describe mainly three load forecasting types namely; short-term, medium-term and long-term forecasting. Most of these scholars however, contradict on the time duration for each of these load forecasting types. Alfares and Nazeeruddin (2002) classify load forecasting basing on the different time frames with short-term ranging from 1 hour to 1 day, medium-term ranging from 1 day to 1 year and 1-10 years for long-term load forecasts. Feinberg and Genethliou (2003), together with Swaroop and Hussein (2012), agree with Alfares and Nazeeruddin classifications but have a contrast on the range of short-term load forecasts by stating an hour to a week's interval instead.

Amlabu et al., (2013) however, totally contradicts with the above mentioned scholars' classes and ranges. Amlabu (2013) and his colleagues mention a forth load forecasting type being; very short-term load forecasting and allocates an hour to a week's duration for it. They then go on to state short-term forecasting as ranging from 1 week to about 4 years, 4 to 10 years for medium-term and at least 10 years for long term forecasting.

Although many scholars do not quiet agree on the time frames and classification of load classes. They however all seem to have a general agreement on the accuracy of the different load forecasting classes and hence the applications become a lot similar. According to Feinberg and Genethliou (2003), due to its higher accuracy, short-term load forecasting is used to estimate load flows and to make decisions that can prevent overloading. The timely implementations of such decisions lead to the improvement of network reliability and to the reduced occurrences of equipment failures and blackouts. In addition to that, Swaroop and Hussein (2012), state the use of short-term load forecasting in supplying information for the system management of day to day operations and unit commitment. Medium-term forecasts are solely used for scheduling fuel supplies as well as unit maintenance (Alfares and Nazeeruddin, 2002; Feinberg and Genethliou, 2003; Swaroop and Hussein, 2012). Planning infrastructure sizes, prediction of future needs for expansion, siting new transmission and distribution facilities etc. being common applications for long-term load forecasting (Badar, 2011; Hahn et al., 2009; Ranjana, 2014; Swaroop and Hussein, 2012).

2.4 load forecasting methods and factors

2.4.1 Short term

The accuracy of any load forecast depends not only on the method employed but largely on the accuracy of the considered factors (Feinberg and Genethliou, 2003; Ranjana, 2014). The most prominent short term load forecasting factors being;

- Time (Badar, 2011; Feinberg and Genethliou, 2003; Hahn et al., 2009; Ranjana, 2014)
- Weather data (Badar, 2011; Feinberg and Genethliou, 2003; Hahn et al., 2009; Ranjana, 2014)
- Possible customer classes (Badar, 2011; Feinberg and Genethliou, 2003; Ranjana, 2014)
- Electrical prices (Badar, 2011)
- Historical load data (Hahn et al., 2009).

Time factors are partitioned into time of the year, day of the week as well as the hour of the day. Ranjana, (2014) emphasizes temperature and humidity data sets as being the most influential of all meteorological information. The electric usage patterns are also said to be different for customers that belong to different classes but somewhat alike for customers within each class, thus a consideration of the various customer classes is valid. In addition to these factors Badar (2011), points out electrical prices as playing a major part in varying electrical demand with rising electrical prices being inversely proportional to electrical usage. On the other hand Hahn et al (2009) states historical electric usage data as a factor worth considering. Short term load forecasting methods hence rely on one or more of the stated factors.

Category	**Method**	**Technique**
Parametric	Similar day approach	Uses historical data having the same characteristics to the day of the forecast for instance, similar weather conditions, week day, season etc.
	Regression based approach	Measures degree of correlation between depended and the in depended variables. Most widely used
	Time series analysis	Models patterns in a time series plot and extrapolates the patterns into the future using historical data as input, it fits a model according to seasonality and trend

	Exponential smoothing	The approach is first to model the load based on previous data, then to use this model to predict the future load.
	Iterative reweighted least squares	Makes use of an operator that controls one variable at a time.
Artificial Intelligence	Artificial neural networks (ANN)	Forecasting is based on the pattern observed from the past event and estimates the values for the future. ANN are able to approximate numerically any continuous function
	Fuzzy logic	Based on the Boolean logic which is used for digital circuit design and is very strong for electric load forecasting when properly implemented.
	Statistical learning algorithms	Regression mathematical model that uses explanatory variables to model relationship between the depended and non-depended variables
	Support vector machines	More recent powerful technique for solving classification and regression problems by making use of simple linear functions to create linear decision boundaries
	Expect systems	Computer program, which has the ability to act as an expert. It can reason, explain, and have its knowledge base expanded as new information becomes available to it.

Table 1: Short term load forecasting methods and techniques

2.4.2 Medium and long term

Trend analysis, end use modelling, econometric and the statistical based approach are the most common methods discussed under the medium to long term load forecasting techniques. Badar (2011) together with Feinberg and Genethliou (2003) mention the end use modelling and econometric forecasting methods as being the most prominent of the four.

According to Campillo et al., (2012) and Ranjana, (2014) trend based analysis simply extends the growth of electrical demand into the future by mainly considering historical electric patterns. It is thus said to be simple, quick and easier to use especially in the case of inadequate data sets and when time and funding do not allow for more elaborate approaches. Badar (2011); Campillo et al., (2012) and Feinberg and Genethliou (2003) state that although the advantages of using this approach are alluring its disadvantages often cloud its advantages. The most outlined disadvantage of using trend analysis is its failure to provide explanations of what causes electricity fluctuations. In cases of technological advancements, economy fluctuations and change in electrical prices, trend analysis becomes greatly inaccurate. Some scholars thus believe that this method should only be applied in medium term forecasts and should not be extended to long term horizons.

More advanced methods such as the end use and econometric approach try to provide some explanations as to why electricity demand behaves the way is does. According to Badar (2011), end use approach is based on the fact that the demand for electricity is depended on what it is used for thus its end use. This approach hence identifies exactly where electricity goes, the amount used for each purpose as well as the potential for additional conservation for each end use (Badar, 2011; Feinberg and Genethliou, 2003). End use models hence divide electrical demand into commercial, residential and industrial areas. A major disadvantage of this method however, is that it requires extensive information on customers and their appliances that in most cases might not be available (Badar, 2011; Campillo et al., 2012; Feinberg and Genethliou, 2003). In the event that such information is obtained, end use models produce accurate forecasts. These forecast are nonetheless only accurate before technological advancements such as energy saving kick in. Another setback of end use models is that they do not take into consideration the changes in the electricity prices which then leads to inaccurate predictions when the economy changes in long term periods (Ranjana, 2014).

Econometric models are employed to alleviate the setback of the end use models. They consider all of the end use model factors and add on electrical prices, employment levels and other economic factors such as per capita incomes (Badar, 2011; Feinberg and Genethliou, 2003; Ranjana, 2014). Econometric models therefore maintain forecasts accuracy for much longer periods than the end use and trend analysis methods. Limitations of the econometric methods are the unavailability of data and also that the projected demand factors might not always have the assumed proportionality.

In 2003; Feinberg and Genethliou developed the statistical model based learning method to alleviate the problem of unavailability of data. Although end use and econometric methods were considered accurate, essential data sets were in most cases unavailable. Feinberg and Genethliou, (2003) stated the complexity of applying end use and econometric models. They mentioned that in order to forecast future load demand, end use and econometric models would require the consideration of large number of factors as well as higher levels of human participation. The statistical model based learning was thus designed to simplify the medium to long term forecasting methods, avoid the use of unavailable data whilst increasing forecasting accuracy. Feinberg and Genethliou came up with a mathematical formal which would study historical data with factors such as the load of the day, day of the week, the weather conditions, time of the day etc. and use that for future load predictions.

Summary of medium and long term methods and factors

Method	Advantages	Disadvantages	Factors considered
Trend Analysis	• Simple (Badar, 2011; Campillo et al., 2012; Ranjana, 2014) • Quick (Badar, 2011; Campillo et al., 2012; Ranjana, 2014) • Inexpensive to perform. (Badar, 2011; Ranjana, 2014) • Useful when there is not enough data to use. (Badar, 2011)	• It produces only one result - future electricity demand. (Badar, 2011; Ranjana, 2014) • It does not help analyze why electricity demand behaves the way it does. (Badar, 2011; Ranjana, 2014) • Provides no means to measure the impact of changes in energy prices and government policies. (Badar, 2011; Ranjana, 2014) • Relies on past patterns of electricity to project future demand thus producing inaccurate forecasts in times of change (Badar, 2011) • Mostly suitable for short term forecasts (Campillo et al., 2012)	• Historical electricity usage (Ranjana, 2014)
End use modelling	• Accurate but sensitive to the amount and quality of end use data. (Badar, 2011; Feinberg and Genethliou, 2003) • Requires less historical data. (Badar, 2011; Feinberg and Genethliou, 2003) • Identifies exactly where electricity goes and how much is used for each purpose and the potential for additional conservation. (Badar, 2011) • It provides specific information on how energy requirements can be reduced with time.	• It assume a constant relationship between electricity and end use. This however will not hold over a ten to 20 year period due to energy saving technology. (Badar, 2011; Ranjana, 2014) • It requires extensive data for correct calculations. (Badar, 2011; Campillo et al., 2012; Feinberg and Genethliou, 2003) • It does not cater for electricity or fuel price changes (Badar, 2011)	• Descriptions of appliances used by customers (Campillo et al., 2012; Feinberg and Genethliou, 2003; Ranjana, 2014) • Sizes of the houses (Campillo et al., 2012; Feinberg and Genethliou, 2003; Ranjana, 2014) • Age of equipment (Campillo et al., 2012; Feinberg and Genethliou, 2003; Ranjana, 2014) • Technology changes (Feinberg and Genethliou, 2003; Ranjana, 2014)

	• It also breaks down electricity into residential commercial and industrial demand. (Badar, 2011; Feinberg and Genethliou, 2003)		• Customer behavior (Campillo et al., 2012; Feinberg and Genethliou, 2003; Ranjana, 2014) • Population dynamics (Campillo et al., 2012; Feinberg and Genethliou, 2003; Ranjana, 2014)
Econometric	• Unlike end-use models, econometrics can allow for variations in the relationship between electricity input and end-use. (Badar, 2011) • Provides detailed information on future levels of electricity demand, why future electricity demand increases or decreases, and how electricity demand is affected by various factors. (Badar, 2011; Ranjana, 2014) • Provides separate load forecasts for residential, commercial, and industrial sectors. (Badar, 2011; Feinberg and Genethliou, 2003) • It is flexible and useful for analyzing load growth under different scenarios. (Badar, 2011)	• Constant elasticity is often hard to justify, especially where very large electricity price changes. (Badar, 2011) • It is only as accurate as the forecasts of factors which influence demand. (Badar, 2011) • Projected depended factors may not also have the assumed proportionality. (Badar, 2011)	• Includes all end use factors (Badar, 2011; Feinberg and Genethliou, 2003; Ranjana, 2014) • economic factors such as per capita incomes (Badar, 2011; Feinberg and Genethliou, 2003; Ranjana, 2014) • employment levels (Badar, 2011; Feinberg and Genethliou, 2003; Ranjana, 2014) • electricity prices (Badar, 2011; Feinberg and Genethliou, 2003; Ranjana, 2014) • Policy factors (Badar, 2011)
Statistical model based learning	• Simple (Feinberg and Genethliou, 2003; Ranjana, 2014) • Accurate (Feinberg and Genethliou, 2003; Ranjana, 2014) • Avoids the use of unavailable data (Feinberg and Genethliou, 2003; Ranjana, 2014)	• Most effective for medium term forecasts rather than long term (Feinberg and Genethliou, 2003).	• Mainly historical data; (load, time of the day, day of the week, weather, (Feinberg and Genethliou, 2003; Ranjana, 2014) • Population & economic data for long term loads (Feinberg and Genethliou, 2003)

Table 2: Medium to long term load forecasting methods and factors.

2.5 The enhancement of spatial data in load forecasting

One of the most common applications of long-term load forecasting is in the prediction of future needs for expansion so as to plan for growth in electrical demand (Amlabu et al., 2013; Ranjana, 2014;). According to Carreno et al., (2010) and Pijoan et al., (2013), the growth in electrical demand inside the service area of an electric utility can be expected for mainly two reasons; first, natural growth because of the natural behaviour of existing consumers, and second, new loads because of new consumers. Carreno et al., (2010) then goes ahead to state that the natural behaviour of existing consumers is usually stationary, with low expected growth, hence attributes the main reason for load growth to be the new consumers inside and outside the actual service zone This being the case, electric distribution utilities have to consider the location of the new consumers when planning for network expansion (Pijoan et al., 2013).

The use of locational information to predict future electric demand is referred to as spatial load forecasting and is defined by Willis (2002) as a prediction of future electric demand that includes the location (where), as one of its chief elements, in addition to magnitude (how much) and temporal (when) characteristics. Noonan, (2005) in his article entitled, "Bringing GIS to transmission and distribution utilities", states how spatial electric load forecasting moves network planning engineers beyond trend based analysis that determines how much load is expected but not where, into the spatial analysis capabilities of GIS. Noonan goes on to say that through the enhancement of spatial data in electric utilities, planning departments are now able to identify changing end use patterns, predict future load centres, identify substation property requirements, magnitude and location and ensure the most defendable and cost effective capital expenditure for substation reinforcements. Distribution utilities are also capable of ensuring that money is spent on solving immediate problems while cooperatively creating long term benefits to a healthier distribution and transmission network system (Noonan, 2005; Willis et al., 1995).

Carreno et al. (2010) emphasis that electric distribution utilities should consider spatial load forecasting as a major business priority. This is because spatial load forecasting allows area engineers to predict large load additions as well as when and where they will appear. For this reasons it would assist distribution utilities to negotiate the acquisition of property from real estate agents, apply for requisite permits and acquire needed rights of way. Noonan (2005) also mentions the capability of planning engineers in using the power of GIS to visualise the distribution system's load, predict future load centres, identify substation property requirements, prioritize projects and obtain budgeting approval while minimising risk.

2.6 Spatial load forecasting case studies

In a conference paper entitled, "Spatial Load Forecasting: Bringing GIS to T&D Asset Management", Noonan (2005) talks about the spatial load forecasting model created for the Utah region. The area consisted of 56 cities and had an average population growth of approximately 2.6% per year over the prior 15 years. Historical electric usage data was used to determine the general electric usage patterns, effects of increase in new consumers, as well as effects of technological advancements. It was from this historical data that Noonan determined that the convention made from evaporative cooling methods to central air cooling systems had played a significant role in increased electrical demand.

In developing the spatial load forecasting model, Noonan's first step was the determination of the consumer load classes. The load classes distinguished customers basing on their distinctive load behaviour, usage and patterns that led to typical daily load curve. That is; a household in say a medium density residential was classified as either medium density residential with evaporative cooling or medium density residential with central air conditioning. A total of 8 load bearing classes used included; Medium density residential with AC, Medium density residential load with EV, low and high density residential, commercial retail, commercial office and institutional, light industrial with major industrials being excluded and commercial business district.

The second stage was to create a land use model showing current and future land uses. Master plans and zoning regulations were used for future land use determination. The major drawback on this stage was the unavailability of current land use information for most of the cities considered. Each land use type obtained was translated into a corresponding spatial forecast load class in order to define the land use's typical load and patterns of usage. Aerial photographs were used to help define a land use proper class.

To make long term load forecasts, Noonan used the end use modelling approach. Most houses had to be re-evaluated to determine the appliances used as well as the ages of the appliances. For houses which did not contain either EV or AC information and could not be accessed, other factors were considered such as square footage of the building, year built as well as the house's total value. Electrical data for anticipated growth was predicted through application of different scenarios and zoning rules for anticipated appliances expected for the different load classes. A load curve was developed for each particular load class and calibration process was done to match the load's calculated load to the actual load on the system.

In 2003, Feinberg and Genethliou in a book titled, " Applied mathematics for power systems", discuss a statistical based method which they employed to spatially forecast future electric growth in the North-eastern US. In an effort to avoid the use of unavailable information Feinberg et_al developed a statistical model that learns the load parameters from historical data. Several data models were tried out and the following multiplicative model was chosen as being the most accurate;

$$L\ (t) = F\ (d\ (t),\ h\ (t)) \cdot f\ (w\ (t)) + R\ (t),$$

Where L (t) represents the actual load at time t, d (t) is the day of the week, h (t) being the hour of the day, F (d, h) is the daily and hourly components, w (t) is weather data, f (w) is the weather factors (temperature and humidity) and R (t) being a random error.

Feinberg and his colleagues justified their model by stating that electric load depends not only on the weather conditions at the current time but also on the weather information of the preceding hours and days. As an example Feinberg and Genethliou (2003) emphasised that the use of air conditioners increases when high heat temperatures continue for several days. To estimate the various weather variables, Feinberg et_al made use of regression based models.

2.7 Conclusion

Literature has shown variations in the methods, factors and models employed in forecasting future electric demand over the years. Scholars have attributed these variation to the difference in political and economic situations for the different parts of the world. Additionally these variations are said to exist due to the differences in consumer behaviour, unavailability of data sets, limited time as well as funding. Literature has thus proved that no spatial electric load forecasting model could be used globally holding the same load forecasting accuracy. There hence exist the need to evaluate factors that affect demand for a desired area and come up with a load forecasting model that best predicts future electric demand basing on the evaluated factors and available data sets.

CHAPTER THREE: METHODS

3.1 Introduction

This chapter highlights the procedures, activities and techniques employed in forecasting future domestic electrical demand for the city of Mutare. It outlines such information as how the data was acquired, cleaned, manipulated and processed to produce the desired results.

3.2 Research design

The research design gives a comprehensive and systematic set of procedures linked together to solve research problems whilst at the same time providing answers to corresponding research questions. Provision is made for a detailed framework of the research plan of action. The research questions mentioned in the first chapter are going to be employed to provide guidance to the research by ensuring that the researcher does not diverse. Exploratory regression was used in this study to identify variables that explain electric usage patterns. The researcher advocates for this particular research design because it explores each individual variable and all possible combinations and points out the passing models. Ordinary least squares (OLS) is then used to identify the weights that each of the passing variable has on the depended variable.

3.3 Study population

The study was concentrated on the high, medium and low density domestic households which constitute the city of Mutare. However not all settlements were included. Residential suburbs such as the Murambi area and plots such as Fern Valley, Odzi and Cecil Kopje plots were deliberately excluded from the study population. Electrical usage patterns in these areas were producing outliers and hence leading to a lager standard deviation and model bias. A total of 33 suburbs was thus used in this study.

3.4 Data collection technique

For the purposes of answering the research questions stated in the first chapter of this research, the researcher made use of secondary datasets. Historical electric usage data, Mutare city cadastral map and its corresponding master plan was collected from ZETDC Mutare. Google earth images were used to cater for areas that were not represented in the city’s cadastral map.

3.5 Data description

3.5.1 Historical electric usage data

This dataset was obtained in excel format. It contained such information as the total number of customers per suburb, land use or connection type as well as the total individual electric consumption for August 2017. It also included the corresponding electric totals for each suburb.

3.5.2 City of Mutare master plan

The obtained city of Mutare master plan was made in 2013. It contained in itself information on both the developed and proposed settlements. Maps were given in pdf format showing the locations of both current and proposed settlements as well as the considered boundary for development. The master plan also contained a phasing development plan book in excel format which showed the start of the development year for proposed locations as well as the year through which development is expected to be complete. Information pertaining the proposed suburb area, class, average expected stand size and consumers was given in excel format. In addition to this, the master plan also contained the expected populations and their corresponding population densities for proposed settlements in accordance to the World Bank population statistics as well as Mutare population trends according to Census results.

3.5.3 Mutare cadastral map

The Mutare cadastral map was acquired as a shapefile in WGS 84 projection system. It contained about 80% of Mutare developed locations. Suburbs lying in the outer skates of the city like Zimunya, Penhalonga and Odzi Township were not included and hence Google Earth imagery was used to map out these locations.

3.6 Data processing

Electric load data for each of the used 33 suburbs comprised of different electric usages such as domestic, commercial, industrial, agriculture etc. The research was however only interested in domestic usage data. Query functions were thus used to depict domestic load from all other electrical usages. The total number of domestic consumers per suburb as well as their corresponding electrical usage data was obtained. From these two datasets, average domestic load data was derived.

In order to map out the city of Mutare developed and proposed settlements, the map had to be first converted into a tiff image using the pdf to tiff conversion tool. The resulting image file did not however contain projection data and hence georeferencing was done with the cadastral map being used as a referencing dataset. The georeferenced image map was then digitized into suburbs in WGS84 projection system. Suburb name, class, development status as well as the year of development were amongst the attributes constituting the digitized layer.

Google earth imageries were used to identify areas that were missing from the cadastral map. The missing suburb locations were digitized and the area constituting the households was derived. The kml to layer conversion tool was used to project the google earth data into ArcMap. Average stand sizes were then calculated by dividing the total area occupied by households in a particular suburb with the total number of domestic consumers in that suburb.

3.7 Methods flow chart

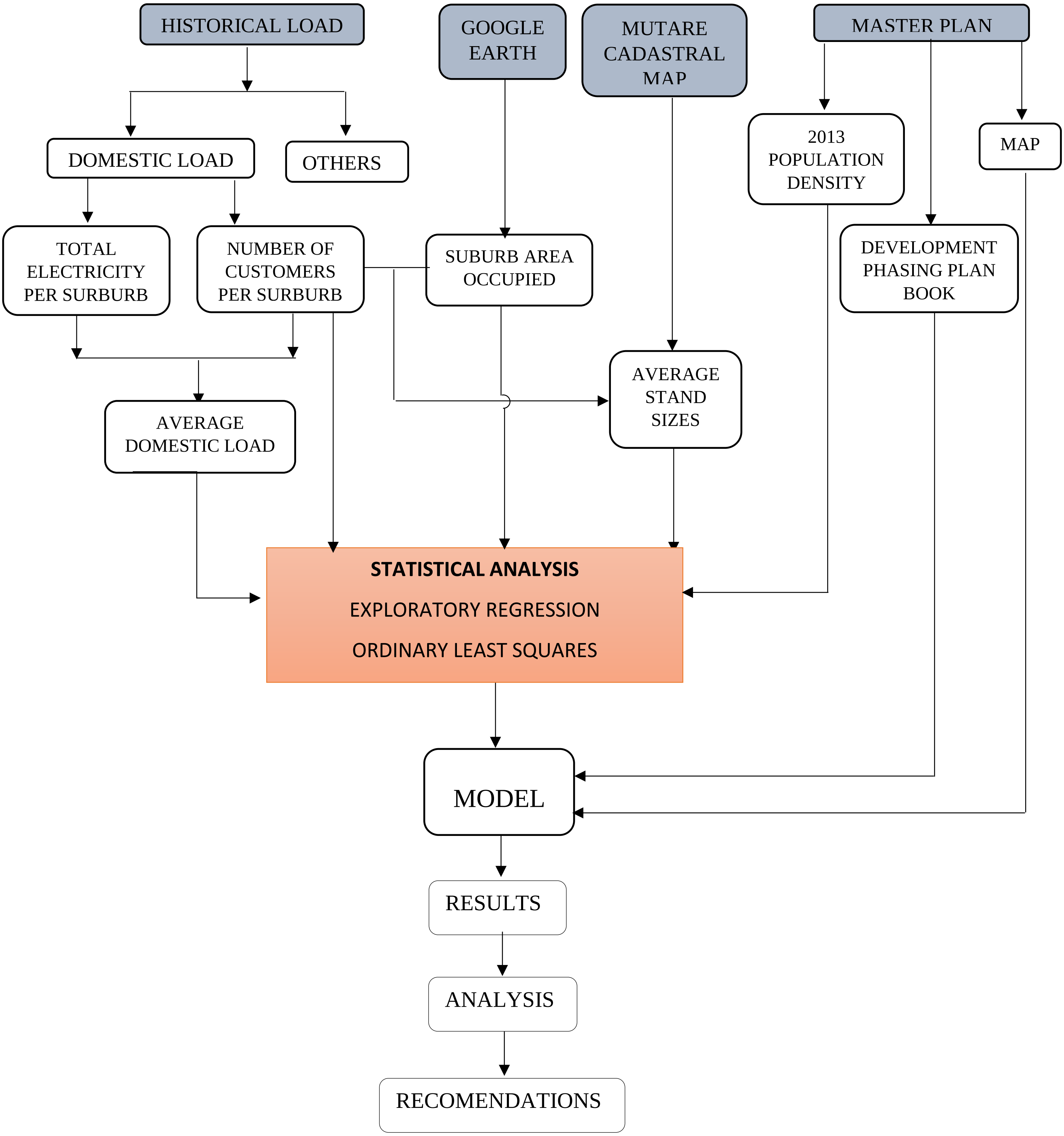


Figure 3: Methods flow chart

3.8 Data modelling: Statistical analysis tools

3.8.1 Exploratory regression

This statistical analysis tool was used to evaluate all given candidate variables to find the one or a combination of variables that best explained the depended variable (average load). Four candidate variables were evaluated and these were total domestic consumers in an area, population density, area occupied per suburb as well as the average household sizes. The minimum adjusted R^2 was set at 50% meaning that all passing models had to be able to accurately predict the depended variable at least 50% of the time. The 'p' value defined as the acceptable confidence interval was set at 95% percent. Jarque-Bera coefficient which shows whether the model residuals are normally distributed was set at a 10% allowance factor. Minimum accepted spatial autocorrelation (S.A) was given at 10%. Any variable with less than a 10% S.A was missing key exploratory variables and hence regarded as an unstable model thus could not be used to explain the electric load usage.

3.8.2 Ordinary least squares regression

OLS modelled the depended variable in terms of its relationship to a set of explanatory variables. The coefficients and standard error for the participating variables were obtained from this statistical analysis tool. The linear regression equation given below was then used to predict the future load demand using OLS coefficients and explanatory variable values from the master plan.

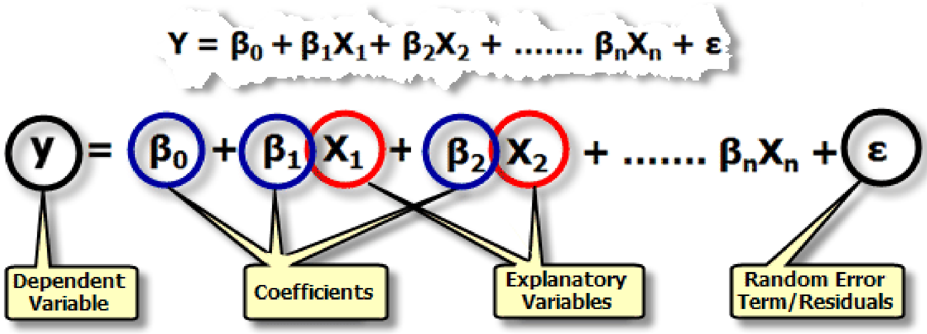


Figure 4: Linear regression equation

3.9 Conclusion

This chapter highlighted the data collection, processing and modelling techniques employed by the researcher. It also gave a brief description of the data used including its format, coverage and limitations. The following chapter is going to present the results which were obtained from the methods that have been discussed.

CHAPTER FOUR: RESULTS

4.1 Introduction

This chapter gives an overview of the major findings of this research. It clearly outlines the results that were obtained through the application of the various statistical regression tools stated in the previous chapter. The results are presented in quantitative form with maps, graphs and tables being the main presentation formats employed.

4.2 Exploratory regression

This statistical analysis tool evaluated a total of four variables with average electric load being the depended variable. Models were considered passing if they satisfied the defined search parameters. The table below gives a summary of the candidate variables considered and their capabilities in explaining the depended variable.

Candidate variable	*Min Adjusted R^2 (>0.5)*	*'p' value <0.05*	*Jarque-Bera p value >0.1*	*Min accepted spatial autocorrelation >0.1*
Area occupied per suburb	0.04	0.95	0.74	0.00
Population density	0.77	0.03	0.29	0.23
Total domestic consumers	0.18	0.37	0.86	0.02
Average stand sizes	0.58	0.05	0.71	0.79

Table 3: Summary of exploratory regression results

The exploratory regression returned a total of three models listed below as passing models:

- Negative population density
- Positive average stand sizes
- + average stand sizes – population density

The results show that population densities for the city of Mutare is capable of explaining average load consumption data accurately 77% of the time. The population density and the average load consumption data have an inverse relationship meaning that an increase in population density results in a decrease in the average household electric usage data. On the other hand the results also indicated a directly proportional relationship between the average stand sizes and the average electrical usage data. Average stand size was in this case capable of explaining the depended variable accurately 58% of the time.

The third passing model produced was a combination of the population density and the average stand sizes. The proportional relationships of these two individual variables were maintained. A combination of the average stand sizes and population densities produced an even higher adjusted R^2 as compared to the use of each of the individual variables. The table below summarizes this model results.

Min Adjusted R^2 >0.5	'p' value <0.05	Jarque-Bera p value > 0.1	Variance Inflection factor < 7.5	Min acceptable spatial autocorrelation >0.1
0.79	0.04	0.47	2.27	0.899

Table 4: Exploratory regression results of the considered model

The results show that this model could explain the average electric usages accurately 79% of the time at 96% confidence interval as evidenced from the R^2 and 'p' value respectively. The variance inflection factor (V.I.F) reflects the amount of redundancy or multicollineariaty among the exploratory variables that can be tolerated. A higher V.I.F that is greater than 7.5 would show model instability. For the purposes of producing better forecasts in this research, models one and two were neglected and this third model was used throughout the research.

4.3 Ordinary least squares regression

This statistical analysis tool performed an OLS linear regression to model a depended variable in terms of its relationship to a set of exploratory variables. In order to evaluate the magnitude of effect each of the passing model variables had on the depended variable, OLS was performed. The OLS produced such results as the variable coefficients, individual standard error, linear plots of the depended variable against explanatory variable as well as a scatterplot of the standard residuals. The table below summaries the OLS model variables.

Exploratory variable	*Coefficient*	*Standard error*
Average stand size	0.029985	0.014567
Population density	-0.405270	0.071217

Table 5: Summary of OLS model variables.

The coefficient variables from the OLS results summarized above shows that population density has a much greater influence on electrical usage as compared to that of average stand sizes.

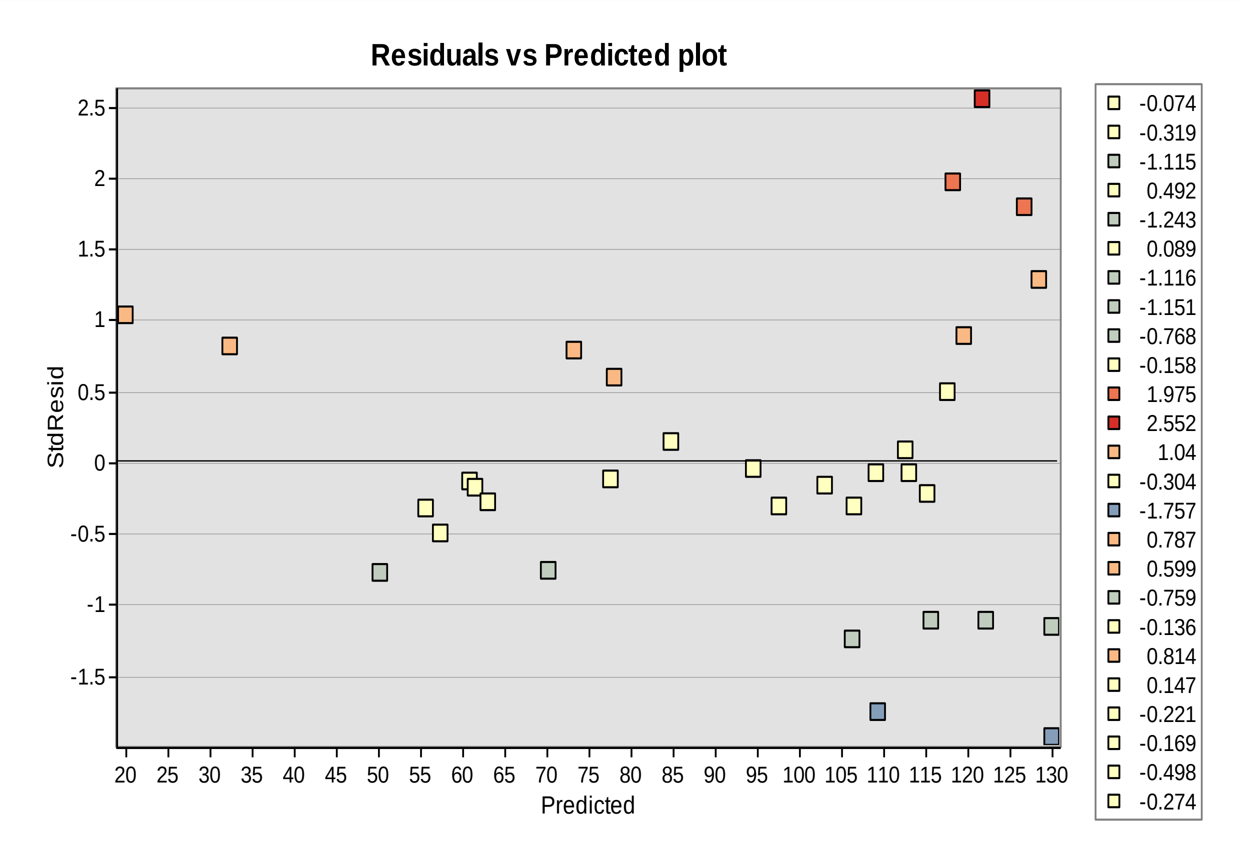


Figure 5: Scatterplot of the standard residuals against the predicted variables

The residual scatterplot shows model's over and under predictions.. The scatterplot shows no structure with the standard errors being randomly distributed around zero. The mean of the standard residuals approximates zero showing that the prediction line cuts well between the over and under estimated values and hence proving model stability. The Moran I spatial auto correlation report given below shows the resultant distribution of the standard errors. Since the distribution is random, the model produced shows stability in predicting future load demand.

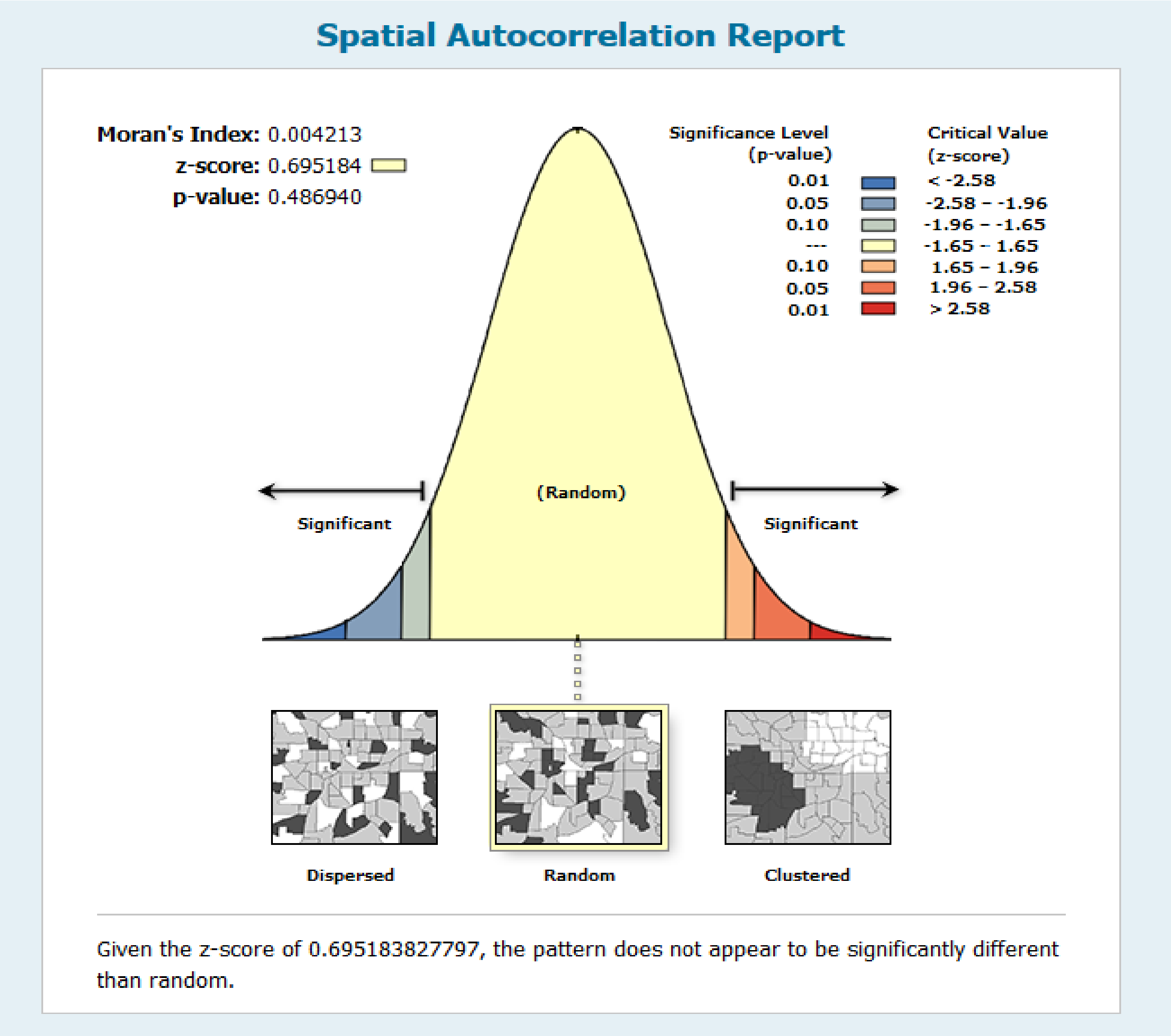


Figure 6: Moran I spatial autocorrelation report

After modelling and evaluating the model's relationship with the depended variable, the linear regression equation was used in predicting future electrical demand with the use of the variable coefficients from the OLS report. The graphs below show the predicted electrical usages for each of the proposed settlements as well as the yearly predicted load additions.

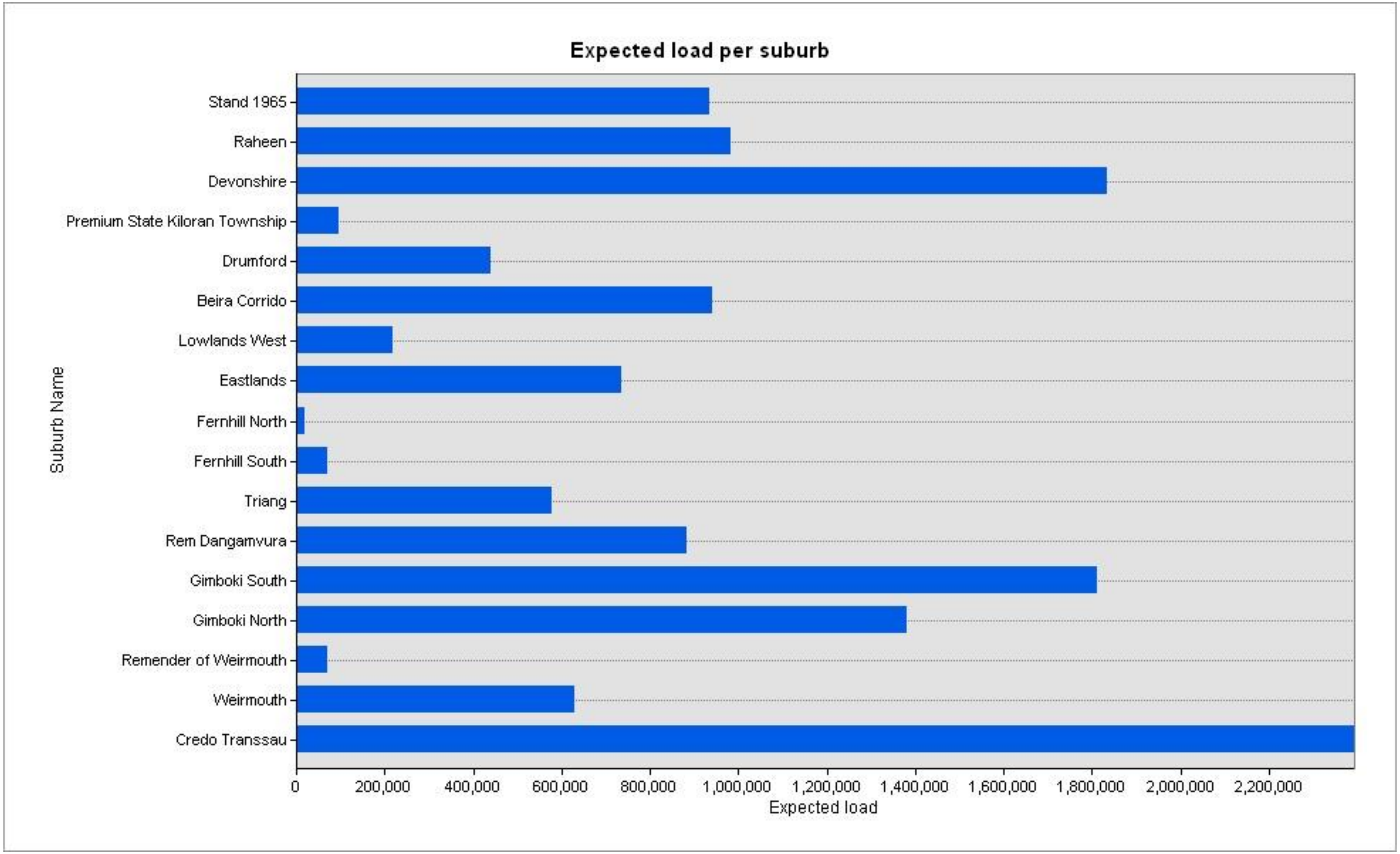


Figure 7: Expected domestic load additions per suburb

The expected load per suburb table shows predictions for 17 proposed settlements from 2018 till 2027. The highest predicted electric demand is recorded for Credo Transsau. Devonshire and Gimboki South follow up with closely similar predicted electrical demand. Fernhill North records the lowest expected load demand of all proposed settlements.

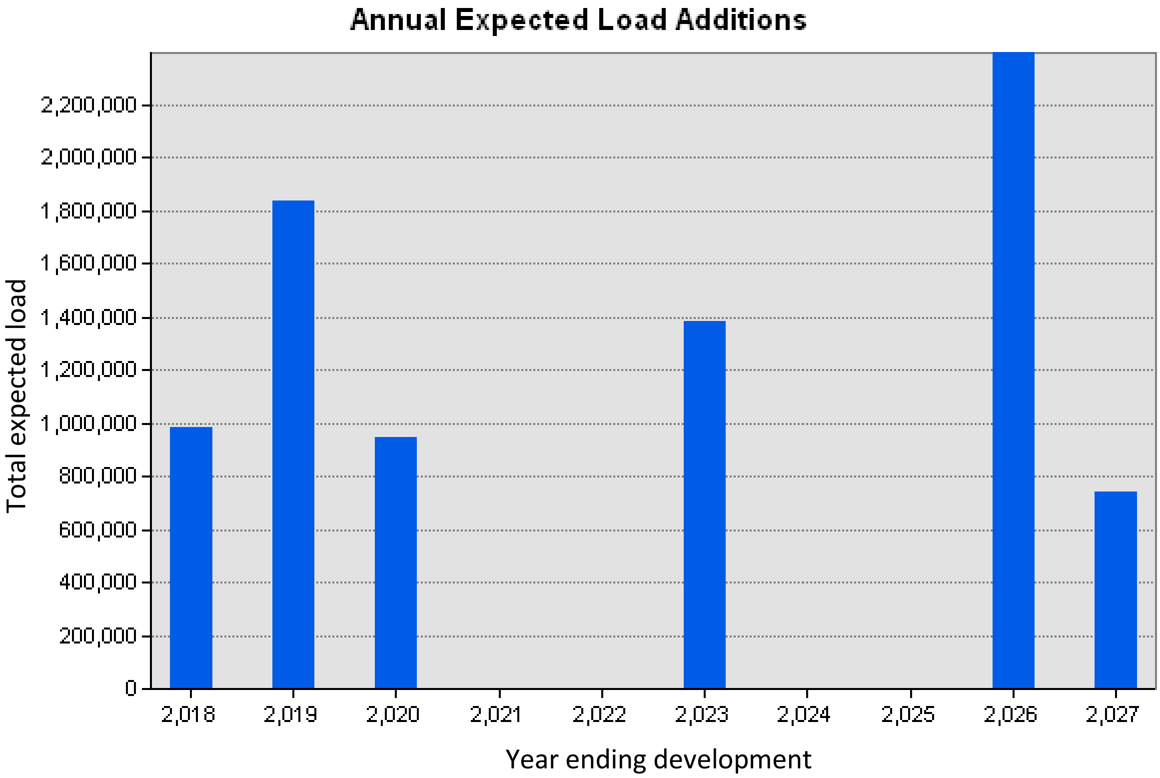


Figure 8: Annual expected load additions

For annual expected load additions, 2026 is expected to have the highest domestic load addition as compared to all other years with predicted load additions surpassing 2, 4 million units. The second highest predicted load demand is for the year 2019 which has an expected load demand of just over 1,8 million units. 2021, 2022, 2024 and 2025 shows years in which settlements are still under development phasing. ZETDC is to expect load additions during these years although magnitude of the load could not be calculated. It should thus be emphasised at this point that the predicted load additions shown in the results section are only after full development has been attained.

The principles of spatial electric load forecasting is the ability to depict temporal, spatial as well as the magnitude of the expected load. The map below shows the locations and years in which development for each of the individual settlements is expected to be completed.

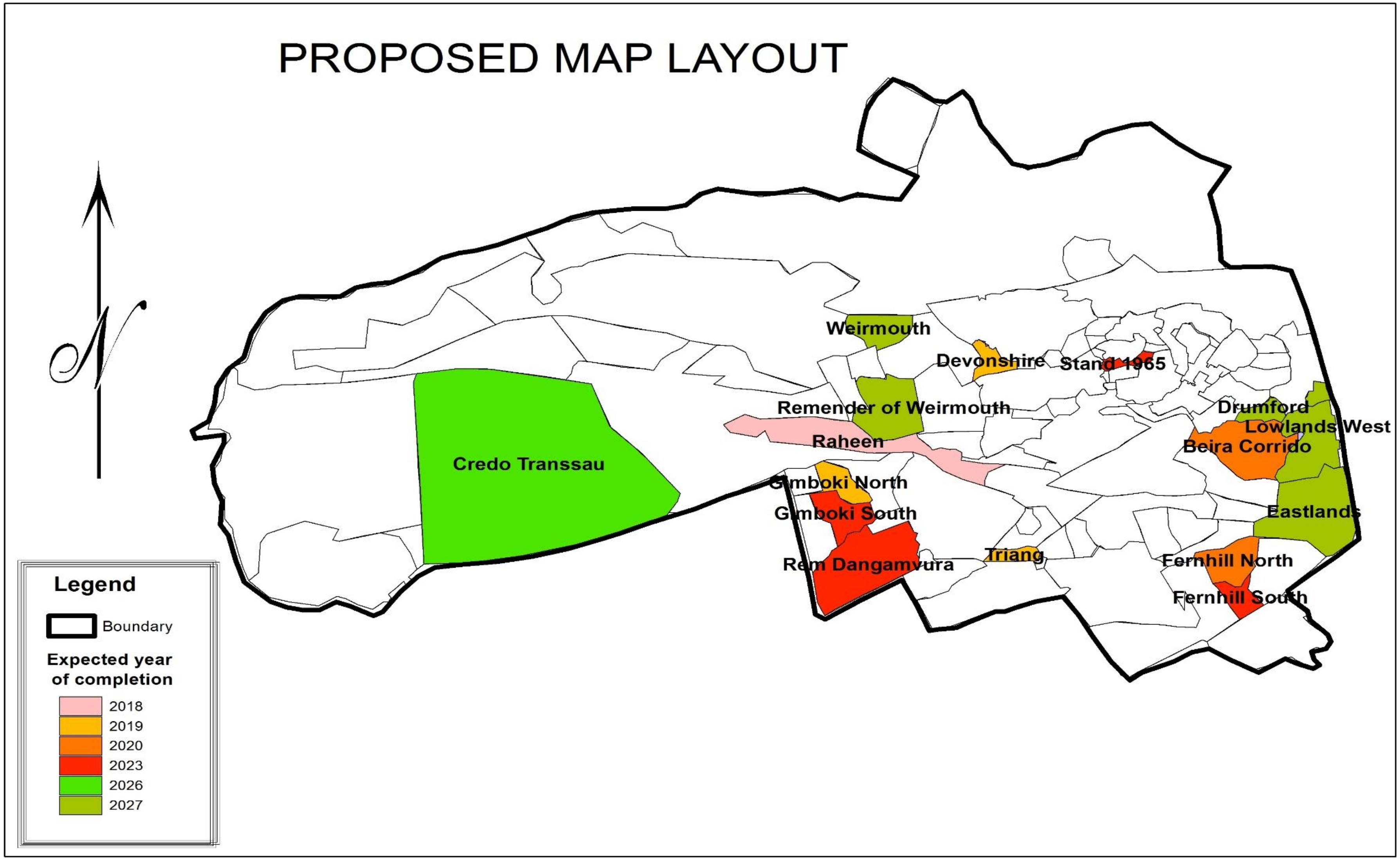


Figure 9: Mutare proposed map layout

4.4 Conclusion

The results presented in this chapter show how electrical load forecasting is possible even in cases of inadequate data sources. It shows the relative accuracy of each variable and combined accuracy of multiple variables. The results section also highlighted the location of the new consumers, time at which development is expected to be completed as well as the magnitude of electrical additions once the development has been completed. It can thus be said that this chapter managed to provide answers to the research questions stated in the first chapter.

CHAPTER FIVE: ANALYSIS OF RESULTS

In cases of unavailable datasets, as evidenced from the results chapter, population density and average stand sizes are seen to be useful variables in spatial electric load forecasting. With population density maintaining an inverse relationship and average stand size a direct relationship to average electric usages the two variables combine to give a 79% forecasting accuracy which is relatively much larger than that produced individually. As evidenced from literature, accuracy of load forecasting can be further increased by including other factors which can provide explanations to electrical usage patterns. Information on appliances used per neighbourhood and general average income would be expected to better model variations in electrical usages per neighbourhood. This intern will help produce much detailed forecasts as compared to using smaller number of variables.

Although the produced S.E.L.F model had a reasonably high prediction accuracy, there is still need to evaluate its generalizability before it can be implemented on different areas. This is because, the behaviour of consumers changes from place to place and therefore influencing factors differ. Spatial electric load forecasting models al*so may* not hold constant even when modelling the same area. Consumer behaviour in the same area can change due to political or economic changes. In addition to this it is also important to note that changes in electrical prices, employment rate, and enhancement of new energy saving technologies as well as technological advancements which result in the invention of more electrical appliances will affect consumer behaviour and intern the predicted load values. This being the case, forecasting of future electric demand seizes to be a one step process but should thus be an ongoing process as changes are bound to occur anytime and anywhere.

5.1 Conclusion

This chapter made an analysis of the model results obtained in chapter four with the use of both qualitative and quantitative data. The discussion can hence be used by ZETDC authorities in planning future infrastructure maintenance as well as future load forecasts.

CHAPTER SIX: LIMITATIONS, RECOMMENDATIONS AND CONCLUSIONS

6.1 Research limitations

One of the limitations faced by the researcher was the unavailability of efficient historical electricity data. Historical load data available was only for the month of august. This hence made it impossible for the researcher to evaluate different load usage patterns for the different seasons and in the process possibly discover other factors that could help better explain electric usage patterns and thus provide much better electrical forecasts.

The other limitation faced was the un-aggregated historical load data. The historical electric load data combined electrical usages for households in different locations. Chikanga phase one, two and three for example was given as one neighbourhood. This proved to be a challenge especially when it came to places like Dangamvura where the neighbourhood has both high and medium density classes. The separation of the datasets was impossible and hence an average stand size and load usage had to be attributed to all areas included. A detailed aggregation of the neighbourhoods would have provided the researcher with more data sets and would have produced a much more accurate forecasting model.

6.2 Recommendations

Future research should explore various datasets and possibly come up with additional factors that can forecast electric load. Information such as daily peak load values can be studied to get a general knowledge of possible appliances used in particular neighbourhoods or if possible at a household level. These results can then be used to derive some factors affecting load that might not be easily available. ZETDC should also try to develop an organic database containing such information as the electrical appliances available per household, the number of buildings each stand contains, the average house sizes etc. Such information will help marginally increase forecast accuracy for future electrical generations.

6.3 Conclusion

It is evidenced from previous chapters that the unavailability of detailed consumer information does not prohibit the forecasting of electricity. General factors such as average stand sizes and population density can be immensely useful in predicting electrical demand. However the spatial load forecasting model e will have to be first tested before it can be implemented on different areas as different areas tend to have different consumer behaviour. The load forecasting model will also have to be revaluated and areas re-forecasted as a number of social, economic and political variations can cause changes in consumption behaviour. There is also a great need for more factors concerning electrical usage patterns to be implemented to bring about more accurate spatial load forecasts.

REERENCE LIST

Alfares, H.K., Nazeeruddin, M., 2002. Electric load forecasting: literature survey and classification of methods. Taylor Fransis, International Journal of Systems Science 33, 23–34.

Amlabu, C., Agber, J., Onah, C., Mohammed, S., 2013. Electric Load Forecasting: A Case Study of the Nigerian Power Sector. Int. J. Eng. Innov. Technol. IJEIT 2.

Badar, I., 2011. Comparison of Convectional and Modern Load Forecasting Techniques Based on Artificial Intelligence and Expert Systems. Int. J. Comput. Sci., 3 8, 504–604.

Balantrapu, S., 2013. Factors influencing weather forecast.htm.

Baliyan, A., Gaurav, K., Mishra, 2015. A Review of Short Term Load Forecasting using Artificial Neural network Models, in: Procedia Computer Science. Presented at the International Conference on Intelligent Computing, Communication & Convergence, Odisha, India, pp. 121–125.

Campillo, J., Wallin, F., Torstensson, D., Vassileva, I., 2012. Energy demand model design for forecating electricity consumption and simulating demand response scenarios in Sweden. Presented at the International Conference on Applied Energy.

Carreno, E.M., Padilha-Feltrin, A., Leal, A.G., 2010. Spatial electric load forecasting using an evolutionary heuristic. Sba Controle Autom. Soc. Bras. Autom. 21, 379–388.

ERNST, D., O'CONNOR, D., 2014. COMPETING IN THE ELECTRONICS INDUSTRY. THE EXPERIENCE OF NEWLY INDUSTRIALISING ECONOMIES [WWW Document]. URL http://146.83.150.183/handle/10533/90134 (accessed 8.22.17).

Feinberg, E.A., Genethliou, D., 2005. Load forecasting, in: Applied Mathematics for Restructured Electric Power Systems. Springer, pp. 269–285.

Feinberg, E.A., Genethliou, D., 2003a. Load forecasting.

Feinberg, E.A., Genethliou, D., 2003b. Load forecasting, in: Appied Mathematics for Power Sysytems. State university of New York, Stony Brook, pp. 269–284.

Gates, B., 2016. It Is Surprisingly Hard to Store Energy [WWW Document]. gatesnotes.com. URL https://www.gatesnotes.com/Energy/It-Is-Surprisingly-Hard-to-Store-Energy (accessed 9.4.17).

Hahn, H., Meyer-Nieberg, S., Pickl, S., 2009. Electric load forecasting methods: Tools for decision making. Eur. J. Oper. Res. 902–907. doi:10.1016/j.ejor.2009.01.062

Hassan, H.., Akhtar, M.F., 2012. Mapping of Power Distribution Network using Geographical Information System (GIS). Int. J. Emerg. Technol. Adv. Eng. 2, 343–347.

Kaseke, N., 2013. Emergence of electricity crisis in Zimbabwe reform response and cost implications. J. Bus. Manag. Soc. Sci. Res. 2.

Khadem, S.K., Basu, M., Conlon, M., 2010. Power Quality in Grid Connected Renewable Energy Systems: Role of Custom Power Devices. Conf. Pap.

Khattak, M., 2015. Short-Term Forecasting of Electricity Demand [WWW Document]. Experfy Insights. URL http://www.experfy.com/blog/short-term-forecasting-of-electricity-demand (accessed 9.4.17).

Kocavelioğlu, M., 2012. Short and long term forecasting of electricity demand in Turkey. Instanbul Technical University, Turkey.

Mugumbate, J., Maushe, F., Nyoni, C., 2013. Ruralisation of Urban Areas: Reversing Development in Zimbabwe. Int. J. Adv. Res. Manag. Soc. Sci. 2, 13–30.

Noonan, J., 2005. Spatial load forecasting: bringing GIS to T&D asset management. Presented at the 25th Annual ESRI International User Conference. San Diego.

Pijoan, A., Esteban, O., Borges, C.E., Penya, Y., 2013. Multi-Agent GIS system for Improved Spatial Load Forecasting.

Ranjana, W., 2016. Medium Term Load Forecasting using Adaptive Neuro Fuzzy Inference System. Int. J. Innov. Emerg. Reasearch Eng. 3, 610–614.

Ranjana, W., 2014. An Approach to Electricity Load Forecasting And Techniques: A Review. J. Inf. Knowlege Res. Electron. Commun. 3, 1080–1089.

R.L.Sheilds associates, 2010. Explanation of Electrical Demand.htm.

Singh, A., Khatoon, I., Muazzam, M., 2013. An overview of Electricity Demand Forecasting Techniques 3, 38–49.

Spatial load forecasting, n.d.

Swaroop, R., Hussein, A., 2012. Load forecasting for power system planning and operation using artificial neural network at Al Batinah region Oman. J. Eng. Sci. Technol. 7, 498–504.

Willis, H.L., 2002a. Spatial electric load forecasting. CRC Press.

Willis, H.L., 2002b. Spatial Electric Load Forecasting. CRC Press.

Willis, H.L., Engel, M.V., Buri, M.J., 1995. Spatial load forecasting. IEEE Comput. Appl. Power 8, 40–43.

Willis, H.L., Northcote-Green, J.E., 1983. Spatial electric load forecasting: a tutorial review. Proc. IEEE 71, 232–253.

Zinyama, T., Tinarwo, J., 2015. How Do Institutions Get Back To Life After a Crisis? The Case of Zimbabwe Electricity Supply Holdings. Glob. J. Res. Bus. Manag. 2, 54–63.

ANNEXURES

Annexure 1: Exploratory regression

Messages

Executing: Exploratory Regression Export_Output; Av_Lod_Dom; Area_occ; Av_stand_siz; Domestic; Ppln_den # # # 4 1 0.5 0.1 7.5 0.1 0.1

Start Time: Thu Oct 12 12:33:44 2017

Running script ExploratoryRegression...

WARNING 001605: Distances for Geographic Coordinates (degrees, minutes, seconds) are analyzed using Chordal Distances in meters.

****

Choose 1 of 4 Summary

Highest Adjusted R-Squared Results

AdjR2 AICc JB K(BP) VIF SA Model

0.77 282.32 0.29 0.03 1.00 0.23 -PPLN_DEN***

0.58 302.12 0.71 0.05 1.00 0.79 + AV_STAND_SIZ ***

0.18 324.16 0.86 0.37 1.00 0.02 -DOMESTIC***

Passing Models

AdjR2 AICc JB K(BP) VIF SA Model

0.769475 282.320947 0.285531 0.031152 1.000000 0.230753 -PPLN_DEN***

0.579961 302.120581 0.714312 0.045660 1.000000 0.789951 + AV_STAND_SIZ ***

****

Choose 2 of 4 Summary

Highest Adjusted R-Squared Results

AdjR2 AICc JB K(BP) VIF SA Model

0.79 280.56 0.47 0.04 2.28 0.90 + AV_STAND_SIZ ** -PPLN_DEN***

0.77 284.32 0.21 0.11 1.06 0.31 -AREA_OCC -PPLN_DEN***

0.76 284.92 0.28 0.08 1.36 0.23 -DOMESTIC -PPLN_DEN***

Passing Models

AdjR2 AICc JB K(BP) VIF SA Model

0.791270 280.562333 0.469083 0.035920 2.276610 0.899170 + AV_STAND_SIZ ** -PPLN_DEN***

****

Choose 3 of 4 Summary

Highest Adjusted R-Squared Results

AdjR2 AICc JB K(BP) VIF SA Model

0.79 283.01 0.37 0.10 2.31 0.88 -AREA_OCC + AV_STAND_SIZ ** -PPLN_DEN***

0.78 283.30 0.49 0.07 2.50 0.93 +DENSITY** +DOMESTIC -PPLN_DEN***

0.76 286.19 0.24 0.13 3.45 0.14 -AREA_OCC +DOMESTIC -PPLN_DEN***

Passing Models

AdjR2 AICc JB K(BP) VIF SA Model

****

Choose 4 of 4 Summary

Highest Adjusted R-Squared Results

AdjR2 AICc JB K(BP) VIF SA Model

0.79 284.79 0.29 0.18 3.45 0.86 -AREA_OCC + AV_STAND_SIZ * +DOMESTIC -PPLN_DEN***

Passing Models

AdjR2 AICc JB K(BP) VIF SA Model

****

************* Exploratory Regression Global Summary (AV_LOD_DOM) *************

Percentage of Search Criteria Passed

Search Criterion	Cutoff	Trials	# Passed	% Passed
Min Adjusted R-Squared	> 0.50	15	12	80.00
Max Coefficient p-value	< 0.10	15	4	26.67
Max VIF Value	< 7.50	15	15	100.00
Min Jarque-Bera p-value	> 0.10	15	15	100.00
Min Spatial Autocorrelation p-value	> 0.10	12	9	75.00

Summary of Variable Significance

Variable % Significant % Negative % Positive

AV_STAND_SIZ 100.00 0.00 100.00

PPLN_DEN 100.00 100.00 0.00

DOMESTIC 25.00 62.50 37.50

AREA_OCC 0.00 75.00 25.00

Summary of Multicollinearity

Variable VIF Violations Covariates

AREA_OCC 2.68 0 --------

AV_STAND_SIZ 2.31 0 --------

DOMESTIC 3.45 0 --------

PPLN_DEN 2.67 0 --------

Summary of Residual Normality (JB)

JB AdjR2 AICc K(BP) VIF SA Model

0.857066 0.180882 324.160605 0.371419 1.000000 0.018113 -DOMESTIC***

0.832043 0.172312 326.023003 0.561062 2.430391 0.016952 +AREA_OCC -DOMESTIC**

0.741317 0.039464 329.416287 0.953096 1.000000 0.000721 -AREA_OCC

Summary of Residual Spatial Autocorrelation (SA)

SA AdjR2 AICc JB K(BP) VIF Model

0.932108 0.784447 283.298765 0.494958 0.072389 2.500256 + AV_STAND_SIZ**+DOMESTIC -PPLN_DEN***

0.899170 0.791270 280.562333 0.469083 0.035920 2.276610 + AV_STAND_SIZ ** -PPLN_DEN***

0.883875 0.786296 283.014353 0.370147 0.095155 2.307247 -AREA_OCC+ AV_STAND_SIZ**-PPLN_DEN***

Table Abbreviations

AdjR2 Adjusted R-Squared

AICc Akaike's Information Criterion

JB Jarque-Bera p-value

K(BP) Koenker (BP) Statistic p-value

VIF Max Variance Inflation Factor

SA Global Moran's I p-value

Model Variable sign (+/-)

Model Variable significance (* = 0.10, ** = 0.05, *** = 0.01)

Completed script ExploratoryRegression...

Succeeded at Thu Oct 12 12:33:48 2017 (Elapsed Time: 4.23 seconds)

Annexure 2: Ordinary least squares regression report

		Summary of OLS Results - Model Variables						
Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	102.674193	11.151991	9.206804	0.000000*	9.278312	11.066043	0.000000*	--------
PPLN_DEN	-0.405270	0.071217	-5.690620	0.000003*	0.057035	-7.105597	0.000000*	2.276610
AV_STAND_S	0.029985	0.014567	2.058393	0.048329*	0.013928	2.152939	0.039488*	2.276610

		OLS Diagnostics	
Input Features:	Export_Output	Dependent Variable:	AV_LOD_DOM
Number of Observations:	33	Akaike's Information Criterion (AICc) [d]:	280.562333
Multiple R-Squared [d]:	0.804316	Adjusted R-Squared [d]:	0.791270
Joint F-Statistic [e]:	61.654128	Prob(>F), (2,30) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	133.412899	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	6.652923	Prob(>chi-squared), (2) degrees of freedom:	0.035920*
Jarque-Bera Statistic [g]:	1.513949	Prob(>chi-squared), (2) degrees of freedom:	0.469083

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable. [b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

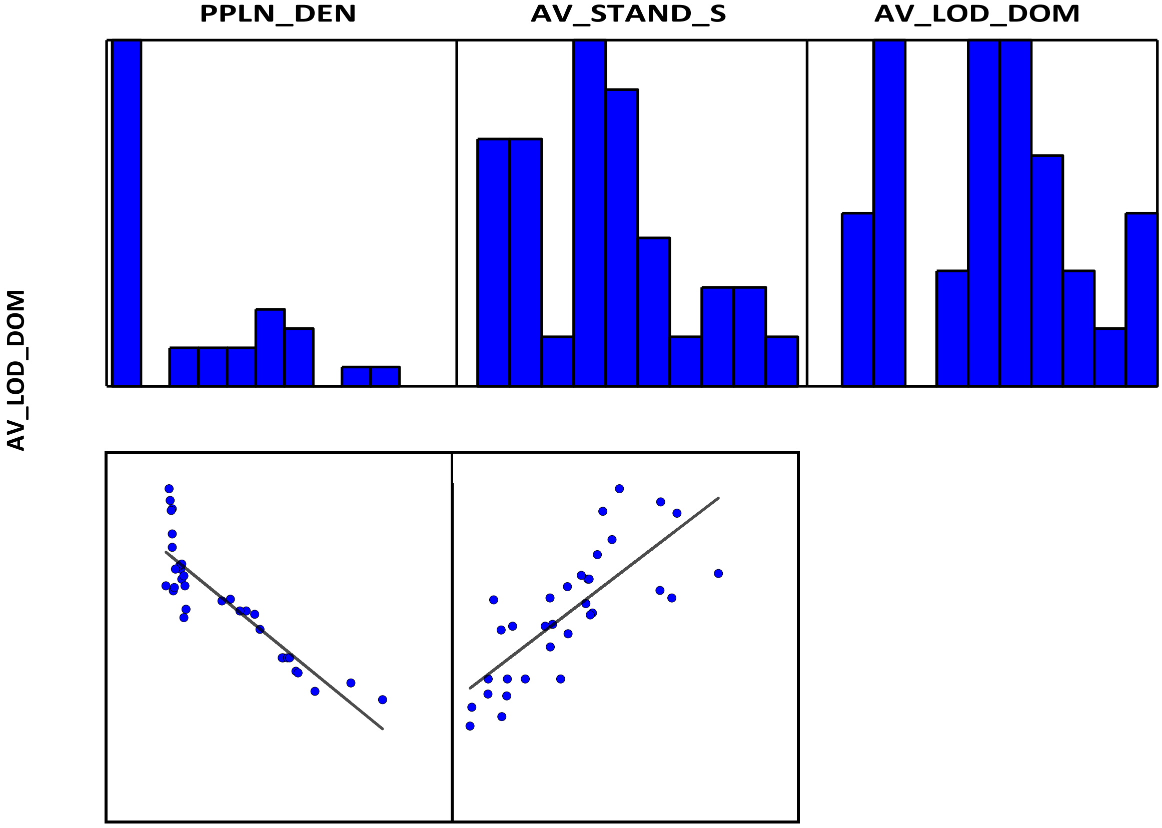
[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the

Wald Statistic to determine overall model significance.

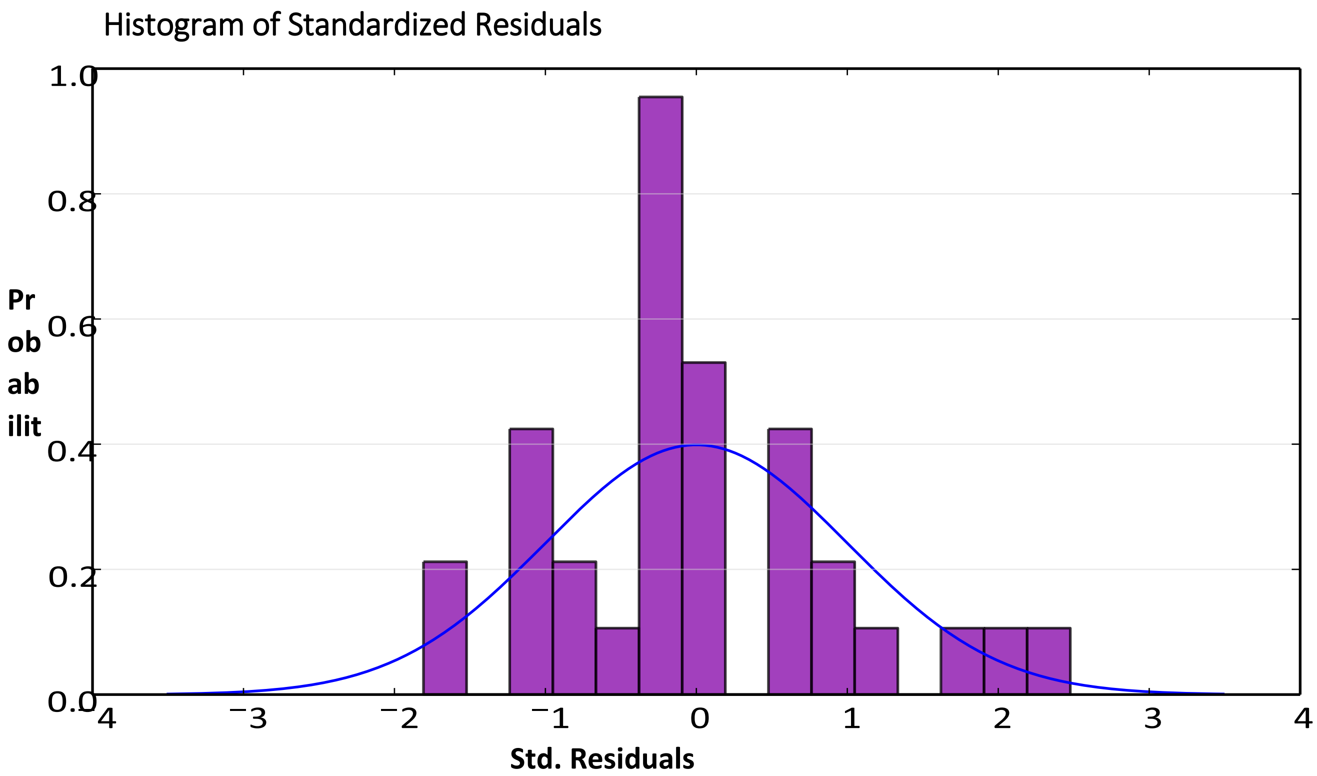
[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

Variable Distributions and Relationships



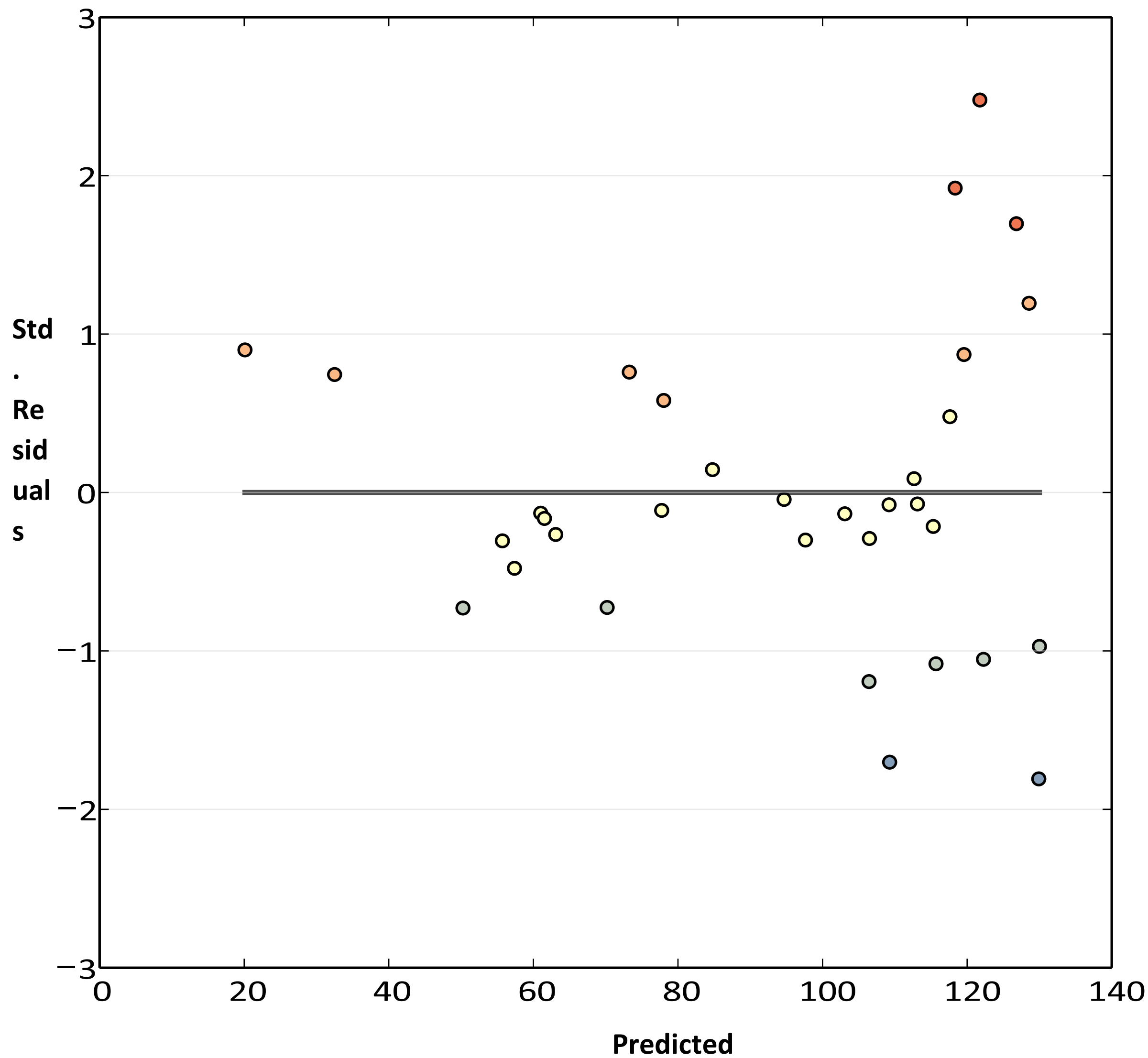
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

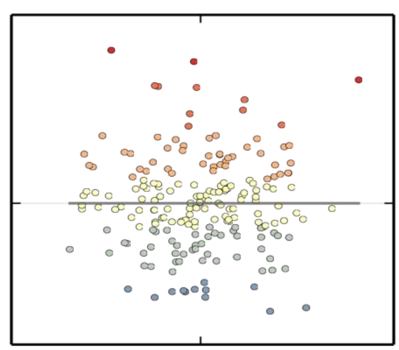


Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).

Residual vs. Predicted Plot



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Random Residuals

Ordinary Least Squares Parameters

Parameter Name	Input Value
Input Features	Export_Output
Unique ID Field	id
Output Feature Class	None
Dependent Variable	AV_LOD_DOM
Explanatory Variables	PPLN_DEN AV_STAND_S
Selection Set	False

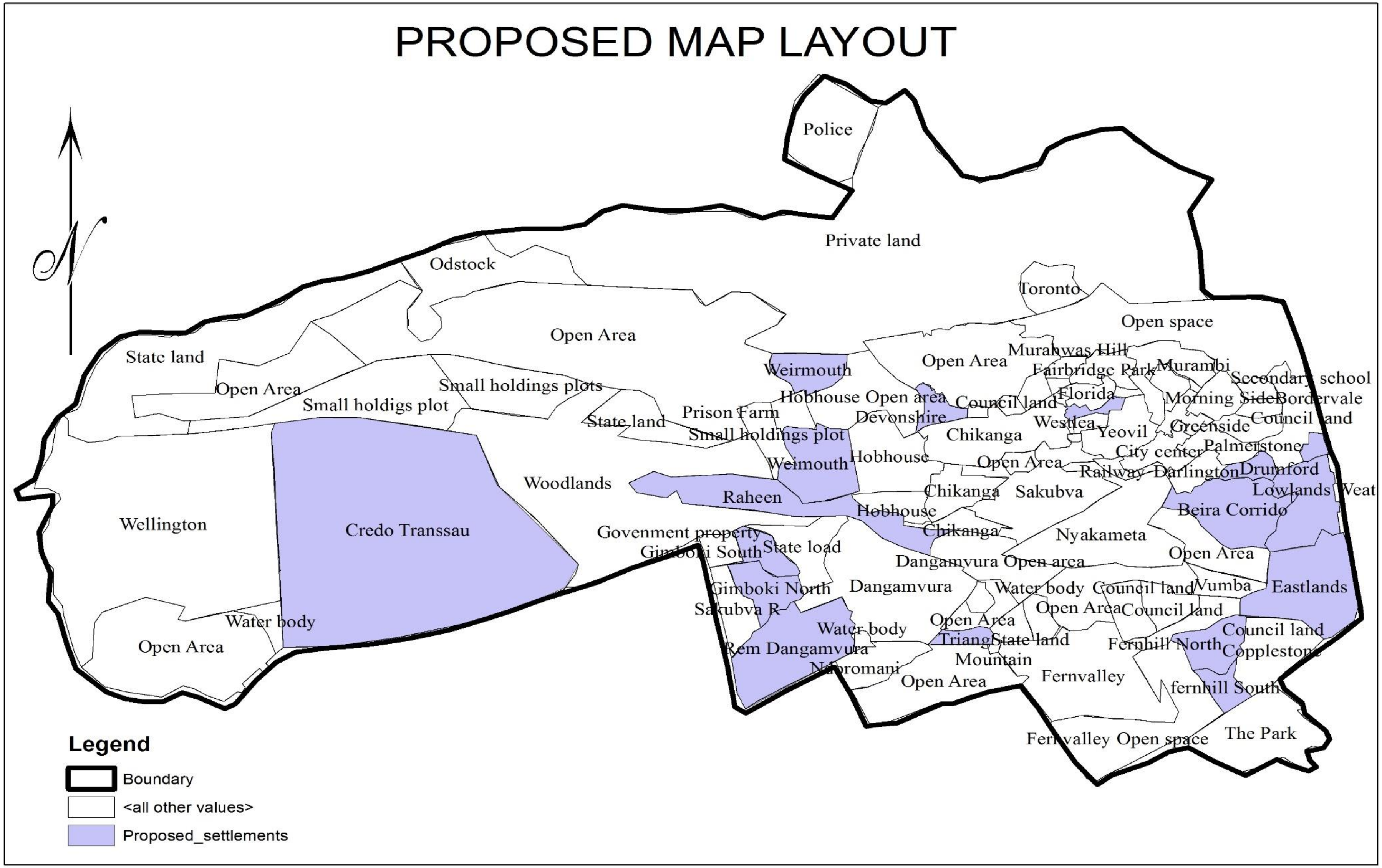


Figure 10: Mutare locational map